

This work is licensed under a Creative Commons Attribution 4.0 International License.

Ovaj rad dostupan je za upotrebu pod međunarodnom licencom Creative Commons Attribution 4.0.



<https://doi.org/10.31820/f.32.2.5>

Andrejka Žejn

RAČUNALNIŠKO PODPRTA STILOMETRIČNA ANALIZA PRIPOVEDNE LITERATURE JANEZA CIGLERJA IN CHRISTOPHA SCHMIDA V SLOVENŠČINI

*dr. sc. Andrejka Žejn, ZRC SAZU Inštitut za slovensko literaturo in literarne vede,
Ljubljana, Slovenija, andrejka.zejn@zrc-sazu.si*

izvorni znanstveni članak

UDK 821.163.6.09Cigler, J.-32

rukopis primljen: 14. srpnja 2020; prihvačen za tisk: 9. prosinca 2020.

V tridesetih letih 19. stoletja je nastala prva (izvirna) slovenska pripoved Sreča v nesreči (1836) avtorja Janeza Ciglerja. Raziskovalci slovenske literature so velik vpliv na nastanek tega dela pripisovali bavarskemu pisatelju Christophu Schmidu, tako v vsebini kot strukturi, ideji in pripovednemu slogu, pa tudi sledenju duhu časa in upoštevanju tradicije. Podrobneje so se posvečali zlasti vsebinskim primerjavam, ne pa tudi analizam podobnosti stila. S paketom Stylo v statističnem programskem okolju R za napredne stilistične analize zbirk besedil smo zato izvedli stilometrično analizo Ciglerjevih in Schmidovih prevedenih del ter nekaterih prevodov pripovedne literature obravnavanega obdobja. Ob upoštevanju določenih omejitev, kot so omejena količina besedil, prevedena in izvirna literatura ter žanrska raznolikost, je analiza potrdila izviren Ciglerjev slog in odprla nova vprašanja glede avtorstva nekaterih del.

Ključne besede: *slovenska pripovedna literatura; Janez Cigler; Christoph von Schmid; stilometrija; stil*

1. Uvod: iskanje motivnih povezav med Schmidom in Ciglerjem

Janez Cigler (1792–1869), ki je kot pisatelj in duhovnik deloval od dvajsetih do šestdesetih let 19. stoletja, ima v zgodovini slovenske pripovedne literature posebno mesto; njegovo delo *Sreča v nesreči* iz leta 1836 je v večini pregledov slovenske literature navedena kot prva slovenska izvirna povest ali pripoved. Zgodba *Sreče v nesreči* poteka v času po Napoleonovih vojnah v Evropi. Prikazuje slovensko družino Svetin, mati Nežo, očeta Franceta ter njuna otroka, dvojčka Janeza in Pavleta. Verna družina se mora po spletu nesrečnih okoliščin raziti, člani družine preko vseh preprek dosežejo uspešno in umirjeno življenje ter se na koncu ponovno združijo. Sama zgodba je pospremljena z obilico moralnih zgledov in nauk. Fran Levstik, eden pomembnejših pisateljev slovenskega realizma, je že leta 1858 v svojem literarnem programu Ciglerjevo delo postavil za zgled ene od smeri slovenskega pripovedništva in dopustil možnost, da Ciglerjeva pripoved ni v celoti izvirna, vendar se v vprašanje o morebitnih tujejezičnih predlogah ni poglobljajal (Levstik 1858: 54). Prvi, ki je Ciglerjevi *Sreči v nesreči* iskal konkretne vzornike, je bil verjetno češki literarni zgodovinar Vaclav Burian. V pregledu začetkov slovenske »razvedrilne« proze je zapisal, da se Janez Cigler tako v »pripovednem izrazu« kot tudi v snovi zgledoval po pripovedih Christopa Schmida (Burian 1930: 248). Schmid je bil nadvse uspešen bavarski otroški in mladinski bidermajerski pisatelj, njegova dela so bila prevedena v številne evropske jezike. Zaradi njegove priljubljenosti so tudi v nemščini še za časa njegovega življenja izhajala dela, ki so poskušala posnemati Schmidovo pripovedno literaturo (Wild 1990: 143) in ga celo – seveda napačno – navajala kot avtorja. V slovenščino so bile najprej leta 1815 prevedene njegove *Zgodbe Svetega pisma za mlade ljudi*, od začetka tridesetih let 19. stoletja dalje pa še številne povesti, ki so izšle v več izdajah, zaradi česar so nekateri starejši literarni zgodovinarji iz konca 19. in začetka 20. stoletja Schmida celo pojmovali kot začetnika slovenske pripovedne literature. Vendar se Burian pri podrobnejši obravnavi vplivov izogone podobnosti v t. i. pripovednem izrazu in išče le motivna prekrivanja med različnimi Schmidovimi deli v nemščini in v slovenskih prevodih ter Ciglerjevo *Srečo v nesreči*.¹ Burianovo pretresanje snovnih virov, iz katerih bi lahko

¹ Iskanje motivnih virov nadalje razširi še na dela Leopolda Chimanija in Josefa S. Ebersberga, ob Schmidu dveh vidnejših avstrijskih bidermajerskih mladinskih pisateljev, ter kratka poročila iz *Lublanskih novic*, časopisa, ki ga je urejal in pisal slovenski razsvetljenec, jezikoslovec in pesnik Valentin Vodnik.

zajemal Cigler, je popravljaj in dopolnjeval Rudolf Kolarič v izdaji *Sreče v nesreči* ob stoletnici prvega izida. Kolarič najprej opozarja na nelogičnost določenih Burianovih ugotovitev, saj so nekatera Schmidova dela, ki jih Burian navaja kot možne motivne vire za Ciglerja, izšla šele po letu 1836, torej po izidu Ciglerjeve *Sreče v nesreči*, odkriva pa motivne podobnosti s Schmidovim delom *Evstahij*, ki jih Burian ni prepoznal. Kolarič v sklepu ugotavlja, da neposrednega zglada za *Srečo v nesreči* ni mogoče določiti in da so vsi »vplivi Schmidovih in drugih podobnih vzgojnih povestic tedanje dobe [...] splošnega, bolj slogovnega ko snovnega značaja« (Kolarič 1936: 11–15). Martin Jevnikar, natančen poznavalec Schmidovega opusa v slovenščini, na eni strani pritrjuje Kolariču in prepozna Schmidov vpliv prvenstveno v obliki in ideji, vendar se vseeno ne more izogniti iskanju motivnih drobcev, ki bi jih Cigler lahko povzel po Schmidu (Jevnikar 1968: 17–19).

Kmecl (1974: 98) vzroke za sorodnost med Schmidovimi in podobnimi tujimi deli ter Ciglerjevo *Srečo v nesreči* pripisuje skupni slovstveni tradiciji, temelječi na sorodnih idejah, vrednotah, vlogi in vsebini. Janko Kos (1981: 245) pa spet izhaja iz spoznanj raziskovalcev pred Kmeclom; motivne podobnosti med Ciglerjem in Schmidom išče le v prevodih do leta 1836 in dotedanje raziskave motivov dopolni z analizo »temeljnih strukturnih vzorcev« in znotraj teh vzorcev posebne pisateljske tehnike. Preko prevzetega pripovednega modela ali strukturnih vzorcev ugotavlja navezave Ciglerjevega pripovedništva – preko Schmidovega *Evstahija* – na pozno-antični krščanski roman (Kos 1981: 250).

2. Stilni vplivi

Podrobnejše obravnave Ciglerjeve *Sreče v nesreči* se torej ne morejo izogniti primerjavam s prevedenimi in tudi neprevedenimi pripovednimi deli Christopha Schmidu. Prepoznavajo motivne, torej vsebinske, in idejne podobnosti, potencialna prevzemanja, sorodne izvore in posnemanje pripovednega modela. Skoraj povsod zasledimo tudi omembe posnemanja pripovednega sloga ali pripovedne tehnike, kar pa nikjer ni nadalje podrobneje analizirano.

V prispevku bomo poskušali z analizo sloga z orodjem za računalniško podprto stilometrično analizo, paketom Stylo, določiti slogovne povezave med Schmidom in Ciglerjem v širšem kontekstu pripovedne literature, ki je nastala v prvih desetletjih 19. stoletja.

3. (Računalniško podprta) stilometrija

Stilometrija, statistična ali kvantitativna analiza stila, raziskuje podobnosti in razlike med besedili na različnih jezikovnih ravneh (Eder et al. 2017: 1), v prvi vrsti z namenom določanja avtorstva. Izhodišče stilometrije je, da je avtorjev stil nekaj nezavednega, česar ni mogoče zavestno usmerjati, obenem pa ima kvantitativne razlikovalne lastnosti. Z merjenjem in preštevanjem teh lastnosti želi stilometrija določiti »karakteristike« avtorja, njegov »stilistični prstni odtis« (Holmes, Kardos 2003: 5). Za enega prvih, ki je uporabil izraz stilometrija in pravilno napovedal njen prihodnji razvoj še pred računalniško podprto stilometrijo, velja poljski filozof Wincenty Lutosławski, ki je leta 1897 z računskimi metodami določal kronologijo Platonovih spisov (Binongo 2003: 10–11), prve ideje o povezavi med kvantitativno določljivimi lastnostmi besedila, konkretno dolžino besed, in določanjem avtorstva pa so se pojavile že vsaj leta 1851 v Angliji. Pot za računalniško stilometrijo sta odprla Frederick Mosteller in David Wallace v začetku šestdesetih let 20. stoletja. Ameriška statistika sta na podlagi najpogostejši funkcijskih besed določala avtorstvo dela *Federalističnih spisov*. Pomembna prelomnica so bile stilometrične analize angleške literature na podlagi pojavljanja nepolnopolnomenških besed, s katerimi je John Burrows v začetku devetdesetih let potrdil stilometrijo kot preverjeno metodo za ugotavljanje avtorstva (Holmes, Kardos 2003: 6).

Ko govorimo o stilometriji v okvirih digitalne humanistike, seveda govorimo o računalniško podprti stilometriji; kot taka je definirana kot del računalniških analiz besedil v okviru oddaljenega branja ali makroanalize, in sicer z upoštevanjem velikih (večjih) zbirk besedil, v katerih skuša najti razmerja in vzorce podobnosti in razlik, ki so »očesu človeškega bralca« skriti (Eder et al. 2016: 108). Običajno se stilometrija opira na enostavne jezikovne značilnosti, ki jih lahko v besedilnem dokumentu določamo avtomatsko, kot na primer relativna frekventnost (najpogostejših) besed, raba in razporeditev ločil, povprečna dolžina povedi ali besed (Eder et al. 2017: 1–2). Še vedno močno prednjači analiza na ravni leksike, vedno večja zmogljivost računalnikov in dostopnost besedil v elektronski obliki pa omogoča ta tudi skladišne in pomenske stilometrične analize (Holmes, Kardos 2003: 6). S skokovitim razvojem računalništva je računalniško podprta stilometrija postala skoraj nepregledno metodološko področje raziskav (prim. Schöch et al. 2017).

Možnosti aplikacije stilometrije so zelo široke, uporablja se na primer v forenziki, natančneje v forenzičnem jezikoslovju (ugotavljanje plagiatorstva,

določanje avtorjev anonimnih grozilnih pisem, napovedi terorizma), v zadnjem času pa njen pomen narašča zlasti na področju analize javnega mnenja, trženja in kadrovanja (Eder et al. 2016: 107; Zwitter Vitez in Pollak 2019: 168). Stilometrija je uporabljena in preverjena metoda tudi v literarnih vedah, filozofiji, zgodovini in korpusnem jezikoslovju (Evert et al. 2017: ii4).

3.1. Računalniško podprte stilometrične analize v raziskavah slovenskega jezika in literature

Uporaba stilometričnih metod v širšem družbenem diskurzu za slovenski jezik je značilna zlasti za raziskave, ki skušajo s pomočjo obsežnejšega referenčnega korpusa določiti avtorja besedila in profil anonimnega avtorja (prim. Zwitter Vitez 2011), konkretne raziskave ugotavljanja avtorstva pa so najpogosteje usmerjene na področje kaznivih dejanj (prim. Zwitter Vitez 2014) in ugotavljanje plagiatorstva.

Sporadične analize slovenske literature s stilometričnimi računalniškimi metodami se ne ukvarjajo z razmerjem posameznih besedil do referenčnega korpusa, ampak raziskujejo pomenljive podobnosti in razlike med posameznimi besedili. Avtorjem raziskav je uspelo z različnimi pristopi pokazati na določeno razmerje med kvantitativno merljivimi značilnostmi besedila in avtorstvom del, bodisi na podlagi dolžine besed in povedi (Dovič 2002), pogostnosti n-terčkov (Jakopin 2003), skupin najpogostejših 50 nepolnopomenskih ali funkcijskih besed (Limbek 2008) ali pogostosti posameznih ločil (Panker 2012).

3.2. Interdisciplinarnost stilometrije

Pri širokem področju raziskav (literarnega) stila je nespregljivo dejstvo, da je vsako preučevanje jezikovnega ali literarnega stila nujno zastavljeno interdisciplinarno, kar je bilo prepoznano tudi zunaj digitalno-humanističnih raziskav, z računalniško stilometrijo pa je interdisciplinarnost še izrazitejša, saj sega zunaj humanistike in družboslovja v naravoslovne in tehnične znanosti, zlasti računalništvo, matematiko in statistiko. Posamezne stilometrične raziskave kmalu razkrijejo, iz katerega področja izhaja raziskovalec: medtem ko so pri raziskovalcih, ki raziskujejo na področju literarnih ved, v ospredju literarna dela in je stilometrija metodološko orodje, s katerim je določena hipoteza, ki jo je postavilo že »klasično« literarnovedno raziskovanje, ovržena ali potrjena, je pri računalničarjih v ospredju razvijanje najustreznejše metode in so literarni teksti le gradivo za

preverjanje metod. Po Eder et al. (2016: 107) moramo stilometrijo v izhodišču razumeti kot večdisciplinarno področje raziskav, ki združuje prizadevanja tako računalničarjev oz. izkušenih programerjev, torej strokovnjakov s tehničnim ozadjem, kot tudi raziskovalcev v humanistiki, ki v računalništvu niso podkovani. Zato so bila razvita računalniška orodja, ki naj bi premostila razkol med obema disciplinama in so za uporabo dostopna raziskovalcem v humanistiki (Eder et al. 2017: 1–2), eno od teh orodij je paket Stylo v programskem okolju R, ki ga razvija skupina raziskovalcev s področja jezikoslovja, literarnih ved in računalništva z različnih raziskovalnih ustanov na Poljskem² in v Belgiji, *Computational Stylistics Group*.³

3.2.1. Paket Stylo

Stylo⁴ je paket v statističnem programskem okolju R⁵ (prim. R Core Team 2019) za napredne stilistične analize zbirk besedil. Poglavitne prednosti R za analizo podatkov so, da je brezplačen, ima širok spekter uporabnikov, deluje v različnih operacijskih sistemih in se sproti nadgrajuje. Sistem deluje v obliki paketov, ki predstavljajo razširitve osnovnega programa (Gregorič, Mole 2016: 5). V paketu Stylo so implementirane poglavitne metode, ki so trenutno najbolj v rabi pri stilometričnih analizah literature (Eder et al. 2016: 107–108).

Paket Stylo je bil preverjan tudi na poljščini, ki je tako kot slovenščina visoko fleksijski jezik, zato se zdi potencialna primernost paketa Stylo za analize slovenskih literarnih besedil toliko večja. Računalniško podprta stilometrija se je namreč uveljavila z raziskavami literature v angleškem jeziku, ki je, kar se tiče razpoložljivosti elektronskega korpusa in orodij za procesiranje naravnih jezikov, tudi najbolj razvit jezik, njena prednost je še, da praktično ne pozna slovnične pregibnosti.

Stilometrične analize s paketom Stylo temeljijo na merjenju pogostosti nepolnopolnomenških ali funkcijskih besed. Da so frekvence najpogostejših besed najučinkovitejši indikator stila, je bilo potrjeno že v več raziskavah,

² Avtorica se je v okviru *Short Term Scientific Mission (STSM)* pri *COST Akciji Distant Reading for European Literary History (CA16204)* udeležila 11-dnevnega izobraževanja na Inštitutu za poljski jezik pri Poljski akademiji znanosti v Krakovu na Poljskem, kjer deluje večina članov raziskovalne skupine s Poljske.

³ Prim. <https://computationalstylistics.github.io/>.

⁴ <https://cran.r-project.org/web/packages/stylo/index.html>.

⁵ <https://cran.r-project.org/>.

med drugim že v začetkih stilometričnih raziskav oz. v zgoraj omenjeni študiji o avtorstvu *Federalističnih spisov* sredi 19. stoletja ali v Burrowsovih pionirskih analizah. Nepochopomske besede so obenem najpogostejše pojavnice v korpusu – kar naj bi veljalo ne glede na jezik – zato na splošno govorimo o merjenju uporabe najpogostejših leksemov. Ne nazadnje je prednost tega pristopa tudi, da je podatke o relativni pogostosti pojavnice mogoče relativno enostavno izluščiti iz besedila, še zlasti računalniško (Eder et al. 2017: 9). Na ta način je analiza stila neodvisna od vsebine besedila.

4. Korpus analiziranih besedil

Stilometrija v literarni vedi ne pomeni obrata tradicionalnega literarnega raziskovanja literature, ampak njegovo dopolnitev z alternativnimi metodami (Holmes, Kardos 2003: 5), zato izbira besedil temelji na ugotovitvah dosedanjih literarnih raziskav, ki smo jih na kratko povzeli v uvodu. Prispevek je nastal v okviru širšega projekta, ki na novo premišlja izhodišča slovenske pripovedne literature od baročnega pridigarja Janeza Svetokriškega do Janeza Ciglerja;⁶ tu smo se osredotočili na stilometrično analizo pripovedne literature avtorjev, katerih osrednja (pripovedna) dela so v slovenskem jeziku izšla v prvih treh desetletjih 19. stoletja. Iz celotnega korpusa smo izbrali sledeča besedila za analizo:

– *Srečo v nesreči* Janeza Ciglerja iz leta 1836, prvo izvirno slovensko pripoved.

– Večji del pripovedne literature Christopha Schmida v slovenščini do leta 1836, ki je v slovenski literarni zgodovini najpogosteje označena kot poglobitni motivni, idejni, strukturni in stilni vpliv na Ciglerjevo *Srečo v nesreči*. Prevodi Schmida v slovenski jezik so nastajali v času, ko avtorstvo literarnega dela ni imelo take teže in vloge kot danes, zato v prevodih avtor izvirnika ni vedno naveden. V takih primerih smo kot Schmidova upoštevali dela, za katere v slovenski literarni zgodovini najdemo podatek, da so Schmidova in jih kot Schmidova zabeleži tudi slovenski knjižnični informacijski sistem *Cobiss*,⁷ čeprav je bilo kasneje Schmidovo avtorstvo problematizirano ali celo ovrženo.

⁶ Projekt *Izhodišča slovenske pripovedne proze* (1. avgust 2018–31. julij 2020) financira Agencija za raziskovalno dejavnost Republike Slovenije, šifra projekta Z6-9385 (<https://www.zrc-sazu.si/sl/programi-in-projekti/izhodišca-slovenske-pripovedne-proze>).

⁷ Gl. <https://www.cobiss.si/>.

Korpus literature, ki je bil v središču našega zanimanja, smo razširili z deli, ki so glede na ugotovitve slovenske literarne zgodovine pomembno prispevala k začetkom slovenske pripovedne literature, in v manjši meri tudi deli obeh osrednjih avtorjev, ki so izšla po letu 1836, in sicer:

– Leta 1828 je duhovnik Franc Veriti, Italijan, ki se je priučil slovenskega jezika, izdal delo *Popotnik široke inu vozke poti* in istega leta tudi prvo knjigo svetniških življenjepisov *Življenja svetnikov in prestavni godovi*. *Popotnik* je od Buriana (1930) dalje označen kot neposredni predhodnik ali spremljevalec začetkov slovenske pripovedne literature, vendar literarna zgodovina pripovednost v njem prepozna le v »pripovednem okviru« (Burian 1930: 244), v osnovni zgodbi v najbolj temeljnih prvinah (Kmecl 1974: 91) ali drobcih pripovednih sestavin (Kos 1981: 238). Čeprav sta glavna očitka Veritejevi pripovedi alegoričnost in didaktičnost, sta verjetno opredelitvam, da ne gre za pravo pripovedno literaturo, botrovala tudi dilema, ali je *Popotnik* izviren, preveden ali nekakšna samostojna obdelava že znane literarne tematike, in dejstvo, da je bil avtor Italijan, čeprav odličen v pisanju v slovenskem jeziku. *Svetniški življenjepisi*, ki jih je zbral in zapisal slovenski duhovnik in nabožni pisatelj Mihael Hofman in po njegovi smrti izdal Franc Veriti, pa veljajo za enega od zgledov začetkov slovenske pripovedne literature (Kmecl 1974; Kos 1981).

– Kasnejša pripovedna dela Janeza Ciglerja, svetniški življenjepis *Življenje s. Heme, brumne koroške grafine* (1839) ter pripovedi *Deteljica ali življenje treh kranjskih bratov francoskih vojakov* (1863) in *Kortonica, koroška dekllica* (1866). Obe Ciglerjevi deli iz šestdesetih let 19. stoletja naj bi nastali zaradi pohvalne navedbe v Levstikovem literarnem programu, kar potrjuje tudi dejstvo, da je *Deteljica* pisana po zgledu *Sreče v nesreči*, medtem ko je *Kortonici* opazen večji vsebinski in strukturni odmik od Ciglerjeve prve pripovedi.

– Dva prevoda Schmidove pripovedne literature, ki sta izšla po letu 1836: *Genovefa* (1851) in *Sveti večer* (1866).

– Pripovedni deli, ki sta izšli v začetku 19. stoletja, in sicer prevod dela *Prava pot k dobriemu stanu* (1812) Benjamina Franklina, prevedel Janez Nepomuk Primic, sicer izrazito didaktično besedilo, in *Izidor, brumni kmet* (1835), ki velja za delo Josefa Huberja, prevedel Luka Dolinar.

4.1. Priprava korpusa za stilometrične analize

Za večino besedil, zajetih v analizo, sta bila v slovenski digitalni knjižnici dLib⁸ dostopna sken in .txt datoteka. Skene nekaterih besedil, ki še niso bila dostopna na dLibu, je bilo mogoče pridobiti preko storitve e-knjige po naročilu (e-Books on Demand, EOD), posamezna besedila so bila dostopna preko Google knjige in Wikivira,⁹ nekatera dela je bilo treba najprej skenirati na optičnem bralniku s programsko opremo za optično prepoznavanje znakov (OCR). V vseh besedilih, tudi že prej dostopnih, je bilo treba najprej odpraviti napake, ki so nastale ob avtomatski prepoznavi znakov. Besedila so bila zapisana večinoma v bohoričici, manjši del v metelčici ali dajncici,¹⁰ le nekaj besedil pa je že bilo natisnjenih v današnji gajici, ki se je dokončno uveljavila po letu 1848. Nazadnje so bila poenoteno prečrkovana v gajico, odstranjeni so bili posebni znaki, kot npr. znak ' med predlogi s/z in k/g ter sledečo besedo ali naglasi. Sledile so tokenizacija, posodabljanje besed, oblikoskladenjsko označevanje in lematizacija s programom ToTrTaLe (Erjavec 2011).

Po izvedenem posodabljanju in lematizaciji besedil smo imeli na voljo tri skupine istih besedil:

1. Izvirna besedila, tj. besedila, ki so poenoteno zapisana v gajici.

2. Posodobljena besedila, v katerih so besede poenoteno posodobljene. Poenoten posodobljeni zapis zagotavlja, da so na primer variante zapisov veznika *in*, kot npr. *inu*, *ino*, ali zapisov poglasnika *z e* ob zlogotvornem *r*, npr. *vrt* in *vert*, zvedene na isto pojavnico.

3. Posodobljena lematizirana besedila. Vprašanje lematizacije postane aktualno, ko orodja za stilometrično analizo apliciramo na jezike z visoko pregibnostjo in s tem z velikim številom možnih oblik, kar lahko pomeni razpršenost podatkov. Lastnosti, ki temeljijo na pogostosti besed, so

⁸ Gl. <http://www.dlib.si/>.

⁹ Gl. https://sl.wikisource.org/wiki/Glavna_stran. Čeprav na Wikiviru izrazito prevladujejo besedila iz 19. stoletja, sta za našo raziskavo prišli iz tega vira v poštev le dve besedili. Deloma je to tudi posledica kriterija, da smo vedno upoštevali prvo izdajo.

¹⁰ Bohoričica, poimenovana po protestantu Adamu Bohoriču, je bila v rabi za zapisovanje v slovenskem jeziku (po zgledu nemškega črkopisa) vse od 16. stoletja. Od gajice se razlikuje v prvi vrsti po zapisu sičnikov in šumevcev, veliko nedoslednosti je pri zapisu *i in j* ter *u in v*. Metelčica in dajncica predstavljata poskus vpeljave nove pisave v 19. stoletju, s katero bi se oddaljili od nemškega vpliva in se približali slovanskim načinom zapisovanja, vendar se poskusa nista uveljavila. Nazadnje je vse tri pisave nadomestila gajica.

namreč lahko pri visokem številu besednih oblik zavajajoče, saj je vsaka oblika samostalniške besede, pridevniške besede, glagola ipd. šteta kot posamezna beseda. Ta učinek je manj opazen pri upoštevanju manjšega števila najpogostejših besed, na primer 100 do 500 najpogostejših, saj v tem primeru prevladujejo nepregibne slovnične vrste, kot so vezniki in predlogi, vseeno naj bi bila lematizacija besed za bolj relevantne rezultate nujna (Eder et al. 2017: 9). V tej zvezi pa ostaja dilema, ali raba določenih oblik pri pregibnih besednih vrstah nosi določeno informacijo o avtorjevih nezavednih izbirah. Zato smo izvedli tri ločene analize, vsakič z drugo skupino besedil, s čimer smo preverili razlike med različno procesiranimi besedili.

Pri poimenovanju datotek smo sledili priporočilom razvijalcev paketa Stylo, zato imena vključujejo osnovne metapodatke: priimek avtorja, prvo besedo ali smiselno prvih nekaj besed naslova dela, letnico prvega izida in pri prevodih za podčrtajem še priimek prevajalca.

5. Stilometrična analiza s paketom Stylo

Za t. i. laičnega uporabnika paketa Stylo, lahko bi zapisali tudi za raziskovalca, ki izhaja iz literarne vede, je najbolj priporočljiva in praktična uporaba grafičnega uporabniškega vmesnika (GUI). S paketom Stylo so naloženi štirje grafični uporabniški vmesniki. Stylo v osnovni nastavitvi upošteva 100 najpogostejših besed ali enograme najpogostejših besed, z neupoštevanjem velike/male začetnice. Izvedeno je razvrščanje podatkov v skupine po podobnosti in različnosti (cluster analiza oz. analiza razvrščanja v skupine) glede na relativne pogostosti 100 najpogostejših besed in z uporabo meril distance z namenom, da se kvantificira stilistična podobnost (Evert et al. 2017: ii5). Rezultati se izrišejo v dendrogramu (drevesnem izrisu), pridobljenem po metodi za merjenje razdalje delta, ki jo je v stilometrijo uvedel John Burrows (2001), ustreznost metode pa je bila preverjena tudi za druge jezike, ne le za angleščino, in za različne zvrsti in obdobja (Evert et al. 2017: ii5).¹¹ Vmesnik »*Input and Language*« omogoča nastavitve jezika besedil za 11 jezikov,¹² za ostale, tudi za slovenščino, je na

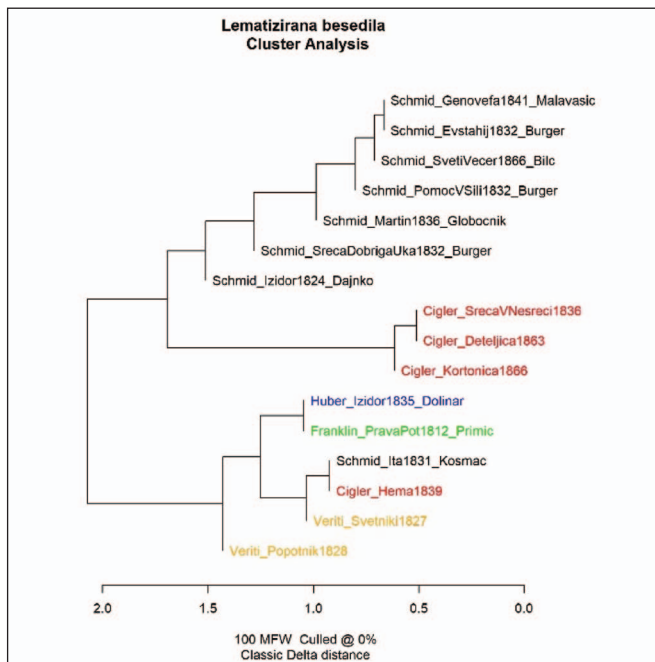
¹¹ Burrows je delo razvil za merjenje razdalje med posameznim besedilom in referenčnim korpusom, vseeno so jo raziskovalci kasneje uporabljali zlasti za merjenje razdalje med posameznimi besedili (Evert et al. 2017: ii6).

¹² Na voljo so izbire angleščina, latinščina, poljščina, madžarščina, francoščina, italijanščina, španščina, nizozemščina, nemščina in kitajščina/japonščina/korejščina.

voljo opcija »Other«. Za format besedil smo izbrali možnost golo besedilo.¹³ Preko vsakega od štirih uporabniških vmesnikov je mogoče osnovne nastavitve za analize poljubno spreminjati.

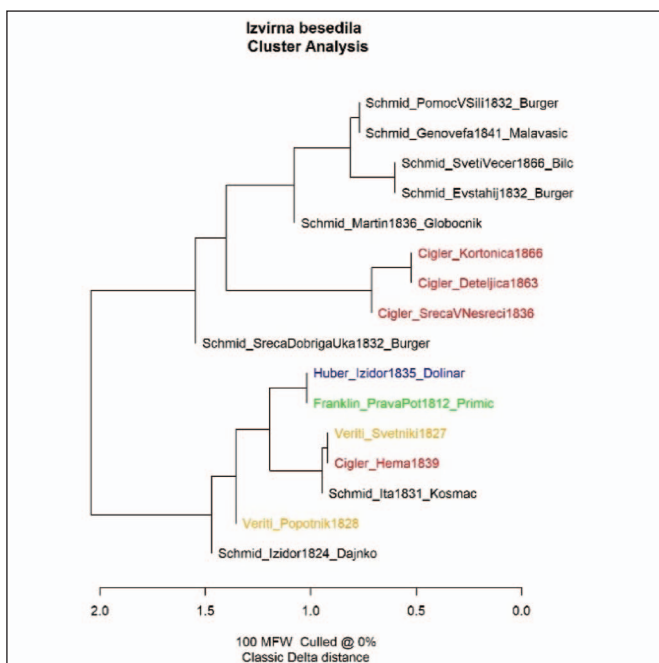
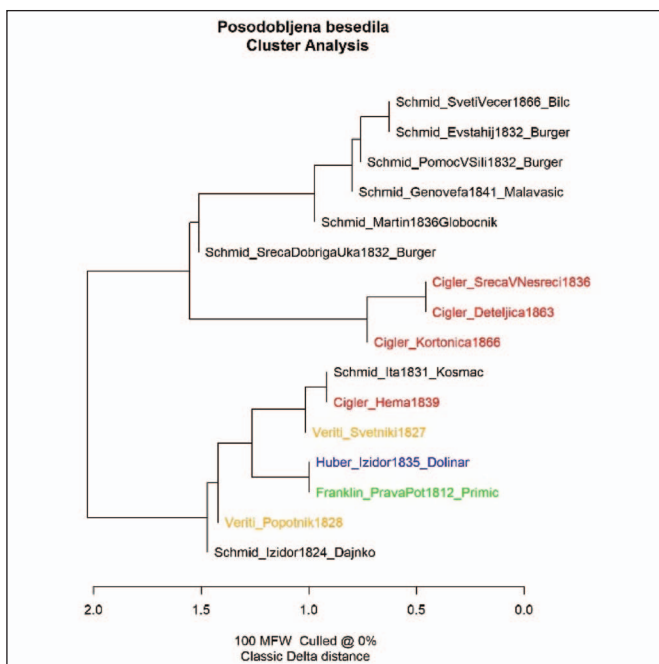
5.1. Rezultati analize

Rezultat združevanja skupin oz. hierarhija med skupinami je v osnovnih nastavitvah, ki smo jih uporabili, prikazan na dendrogramu, na katerem je vsak avtor¹⁴ izpisan s svojo barvo. V skupine so združena besedila, ki so si glede na relativno pogostnost 100 najpogostejših besed najbolj podobna. Število vej predstavlja skupine, v katere so razporejena ali razdeljena besedila, iz vozlišč/stičišč razberemo, na kateri razdalji se začnejo enote med seboj ločevati glede na postavljene kriterije. Najnižje vrednosti vozlišč (glede na os y) nam povedo, da so te enote (skupine enot) med seboj zelo podobne in obratno pri višjih vrednostih.



¹³ Možnosti sta še .xml in .html.

¹⁴ Avtor je prepoznan glede na poimenovanje datoteke, zato mora biti izpisan na prvem mestu, pred podčrtajem.



Na vseh treh dendrogramih razrez loči dve skupini na najvišji hierarhični ravni. Ena skupina se deli na večino prevodov Christopha Schmida ter tri Ciglerjeva dela, *Sreča v nesreči*, *Kortonica* in *Deteljica*. V drugi osnovni skupini so preostala dela, avtorjev Veritija, Huberja, Franklina, dve deli, ki sta bili pripisani Schmidu, in Ciglerjev življenjepis sv. Heme, ki se grupirajo v tri oz. štiri skupine. Iz podrobnejše primerjave treh dendrogramov razberemo, da so pri analizi t. i. izvirnih in posodobljenih besedil v rezultatih le manjše razlike: na izvirnih besedilih se pokaže večja oddaljenost Schmidove *Sreče dobriga uka* od njegovih preostalih del, ki pa je vseeno razvidna tudi iz stilometrične analize posodobljenih besedil, vendar z manjšo razdaljo do ostalih Schmidovih del. Pri Ciglerju je analiza t. i. izvirnih besedil pokazala večjo bližino med *Deteljico* in *Kortonico* in njuno skupno oddaljenost od *Sreče v nesreči*, pri posodobljenih besedilih pa večjo bližino med *Srečo v nesreči* in *Deteljico*. Poglavitna razlika rezultatov posodobljenih lematiziranih besedil v primerjavi z obema drugima tipoma besedil je, da je delo *Kmet Izidor s svojimi otroki ino ludmi*, ki je bilo sprva pripisano Schmidu, razvrščeno v skupino s Schmidom in Ciglerjem. Analiza izvirnih in posodobljenih besedil pokaže večje ujemanje med Ciglerjevim svetniškim življenjepisom sv. Heme in delom *Ita, togenburska grafinja* v razmerju do nekoliko bolj oddaljenih Veritijevih *Svetnikov*, medtem ko je pri lematiziranih besedilih večje ujemanje med *Ito* in Ciglerjevo *Sv. Hemo*. Kljub tem razlikam lahko sklenemo, da dendrogrami prikažejo podobna razmerja, in sicer razdelitev na dve osnovni skupini ter nekaj razlik znotraj manjših skupin.

Pregled seznama 100 najpogostejših besed po posameznih skupinah, ki ga generira paket Stylo,¹⁵ pokaže, da na nekoliko večjo razliko med dendrogramom lematiziranih besedil v primerjavi s preostalima dvema vpliva lematizacija glagola *biti*, ki je lahko v polnopomenski ali nepolnopomenski vlogi. Pri izvirnih in posodobljenih besedilih je najpogostejša pojavnica je (3. os. gl. *biti*),¹⁶ medtem ko je pri lematiziranih besedilih na prvem mestu seznama pojavnica *biti*; med prvimi 100 najpogostejšimi pojavnici prvih dveh skupin je pojavnica *biti* na 94. oz. 96. mestu, na seznamu pa so še ostale oblike glagola *biti* v sedanjiku, in sicer (od najpogostejše do oblike z

¹⁵ Sezname najpogostejših besed in podatki o njihovi relativni pogostosti v posameznem besedilu so za vsako izvedeno analizo dostopni v .txt datotekah, ki jih generira Stylo.

¹⁶ Kar ne nazadnje kaže na tretjeosebna pripovedovalca.

najmanj pojavnicami): *so*,¹⁷ *si*,¹⁸ *sem*,¹⁹ *sta*,²⁰ *ste*,²¹ pretekli deležnik na -l (po pogosti si sledijo *bil*, *bila*, *bilo*, *bili*) in prihodnjik (relativno najpogostejša pojavnica je spet oblika za 3. osebo ednine, *bo*, med pojavnicami sta še *bom* in *boš*). Pri tem iz seznama najpogostejših pojavnic ni mogoče razbrati, kdaj je glagol *biti* rabljen polnopomensko in kdaj slovnično, kot nepolnopomenski glagol.²² Oblike glagola *biti* skupno predstavljajo 12 odstotkov vseh prvih sto najpogostejših pojavnic, kar posledično pomeni, da je pri lematiziranih besedilih višji odstotek polnopomenskih besed, ki se jim skuša osnovna stilometrična analiza izogniti. Na primeru pojavljanja oblik glagola *biti* sklepamo, da v našem primeru najbolj relevantne rezultate pokaže stilometrična analiza posodobljenih besedil.

5.1.1. Vizualizacija rezultatov v programu Gephi

Za vsako posamezno analizo z različnimi nastavitvami je v paketu Stylo ustvarjena excelova datoteka, ki jo lahko uvozimo v odprtokodno programsko opremo za analizo omrežij Gephi (Bastian et al. 2009). V vizualizaciji v Gephiju je podobnost med besedili, ki so na sliki prikazana s točkami (*nodes*) in napisi – vsak avtor je zapisan s svojo barvo –, prikazana z veznimi črtami (*edges*) med besedili: debelejša črta pomeni večjo podobnost/povezanost glede na relativno pogostnost 100 najpogostejših besed, tanjša pa manjšo. V programu Gephi smo kot primer vizualizirali rezultate analize posodobljenih besedil, in sicer zaradi ugotovitve o relevantnosti teh rezultatov.

Vizualizacija v Gephiju pokaže iste povezave nekoliko drugače: vidimo na primer lahko, da so besedila, ki so na dendrogramu uvrščena v vejo s Schmidovimi deli, medsebojno povezana bodisi močneje bodisi šibkeje, v takih primerih je za uvrstitev v isto skupino odločilna večkratna medsebojna povezanost. Nasprotno pa je na vizualizaciji v Gephiju izrisana tudi povezava med deli iz osrednje skupine Schmidovih del in določenimi deli, ki

¹⁷ Pri izvornih besedilih na 6. mestu in pri posodobljenih na 7.

¹⁸ Pri izvornih besedilih na 27. mestu in pri posodobljenih na 25.

¹⁹ Pri izvornih besedilih na 15. mestu in pri posodobljenih besedilih, kjer sta v isto pojavnico združeni obliki *sim* in *sem*, na 51.

²⁰ Pri izvornih besedilih na 56. mestu in pri posodobljenih na 54.

²¹ Le pri izvornih besedilih, in sicer na 95. mestu.

²² Raziskave ali ocene, v kateri vlogi se pojavlja pogosteje, niso na voljo, le ocenimo lahko, da se pogosteje pojavlja v nepolnopomenski vlogi.

so na dendrogramu uvrščena na drugo vejo, vendar s to razliko, da je povezava izkazana le z enim delom (npr. *Evstahij* in *Ita*), močnejše pa je besedilo povezano z deli preostalih avtorjev.

6. Interpretacija rezultatov

Za sklep se bomo posvetili interpretaciji rezultatov stilometrične analize, pri kateri bomo upoštevali v prvi vrsti rezultate na podlagi posodobljenih besedil. Pozorni bomo zlasti na odstopanja od razporeditev v skupine, kakršne bi pričakovali iz izsledkov dosedanjih literarnozgodovinskih raziskav.

Pri stilometrični analizi je pričakovano, da bodo dela istega avtorja uvrščena znotraj istih skupin, močne povezave naj bi se pokazale med Schmidom in Ciglerjem. Izpostaviti moramo, da je stilometrija v osnovi primerjalna metoda, zato so rezultati razmerij med besedili v veliki meri odvisni tudi od izbora besedil, vključenih v analizo. V tukaj predstavljeni analizi smo bili omejeni z majhnim številom pripovednih besedil v obravnavanem obdobju. Nekateri avtorji so zato prisotni z le enim besedilom. Čeprav so rezultati relativni, torej glede na sorazmerni količinski delež avtorja v korpusu, posamezna dela, npr. Franklinova *Prava pot* ali oba prevoda *Izidorja*, kažejo povezave z drugimi avtorji, kar hkrati pomeni, da so od vseh razlikujejo približno enako in da najbolj podobno delo v korpus ni zajeto.

Določene omejitve predstavlja tudi primerjava prevedenih in izvirnih besedil. Pri presojanju razmerij znotraj Schmidovih prevodov je treba poudariti, da so Schmida prevajali različni prevajalci, brez kakega zaokroženega izbora ali cilja in do vključno leta 1836 znotraj dveh središč, ljubljanskega in celovškega.²³ Pri prevodih Schmida so tako lahko prisotni različne stopnje zvestobe prevoda izvirniku, vpliv sloga prevajalca in ne nazadnje žanrska uvrstitev izvirnega besedila. Kljub temu je pri prevodih Schmida pri vizualizaciji rezultatov dovolj prepoznavno osrednje jedro, ki pa ni zaokroženo, ampak odprto. Najmočnejše medsebojne povezave so se pokazale med deli *Evstahij*, *Sveti večer ter Pomoč v sili* in *Genovefa* in *Martin, mladi puščavnik*, čeprav so ta dela prevajali trije prevajalci in so izšla z več kot tridesetletno razliko.

²³ Po letu 1836 se je prevajanje razširilo na Štajersko (Radgona in Gradec) in Novo mesto.

Iz stilometrične analize lahko podamo še sledeče ugotovitve:

– Pri prevodu *Sreča dobriga uka*, ki jo je leta 1832 prevedel Jože Burger, glavni predstavnik ljubljanskega kroga prevajalcev (Jevnikar 1939: 196), lahko nekoliko bolj rahlo povezanost z ostalimi Schmidovimi deli pripišemo žanrskim razlikam, saj se zgodba odvija v obliki pisem med bratom in sestro.

– V isti knjigi kot *Sreča dobriga uka* je bila objavljena pripoved *Leseni križ*. Jevnikar (1939: 190) je pri analizi prevoda ocenil, da čeprav je glede na izvirnik dodan opis Blede in jezera in je dogajanje predstavljeno v slovensko okolje, ne gre za samostojno predelavo dela, medtem ko je Plavšak (1963, po Kos 1981: 239) na podlagi podrobne primerjave postavil tezo, da gre skoraj za Burgerjevo izvirno delo. Stilometrična analiza bolj pritrjuje Jevnikarjevi kot Plavšakovi tezi.

– V analizo sta bili vključeni tudi dve deli s sorodnim naslovom, ki sta sprva veljala za prevod Schmidu, in sicer *Kmet Izidor s svojimi otroki ino ludmi, ali pripodobni nauki dobrih staršev za svoje otroke in podložne*, prevedel Peter Dajnko in je izšla leta 1824 v Gradcu, ter *Izidor, brumni kmet*, prevedel Luka Dolinar leta 1835. V nobenem od prevodov Schmid ni naveden kot avtor izvirnika, vendar je Dajnkovemu prevodu Schmidu kot avtorja pripisal Franc Simonič (1903: 456) v *Slovenski bibliografiji*, po Burianu (1930: 245–246) in Jevnikarju (1968: 16) sta Dajnkov in Dolinarjev prevod starejši in novejši prevod istega dela, in sicer *Isidor, Bauer zu Ried* Josepha Huberja, kasneje pa se je uveljavila teza, da je le Dolinarjev *Izidor* prevod Huberjevega dela. Stilometrična analiza je pokazala, da nobenega od teh del ni mogoče pripisati Schmidu in da sta si tudi oba prevoda medsebojno slogovno zelo oddaljena. Tudi iz vsebinske primerjave Huberjevega nemškega »izvirnika« in obeh slovenskih prevodov lahko ugotovimo, da gre za vsebinsko in slogovno tri zelo različna dela.

– *Ita, togenburska grafinja*, je delo, ki ga je leta 1831 prevedel Jurij Kosmač. Dolgo je veljalo za Schmidovo, tudi v nemško govorečem prostoru, saj je bilo še v času Schmidovega življenja, ravno leta 1831, ko je izšel slovenski prevod, objavljeno v izboru Schmidovih del (Grießer, Heine 1968: 173–174). Kot prevod Schmidovega dela sta ga navajala še Simonič v *Slovenski bibliografiji* (1903: 455) in Burian (1930: 245), prvi je na napačno pripisano avtorstvo opozoril Jevnikar (1939: 209). V nemških bibliografijah je delo pripisano Franzu J. Waizeneggerju. Vizualizacija v Gephiju pa pokaže, da se *Ita* glede na kriterije za stilometrično analizo povezuje z le enim

Schmidovim delom, zato je na dendrogramu razporejena v skupino, ki je oddaljena od osrednjih Schmidovih del.

Najpomembnejši segment naše analize pa so dela Janeza Ciglerja. Na dendrogramih in v prikazu v Gephiju se jasno izriše medsebojno močno povezana skupina treh Ciglerjevih del: *Sreče v nesreči*, *Deteljice* in *Kortonica, koroške dekllice*. Vsako od besedil posamično izkazuje s prevodi nekaterih Schmidovih del in Dolinarjevim prevodom *Izidorja* zelo rahle povezave, zaradi katerih ne moremo govoriti o slogovnih vplivih ali slogovni podobnosti, še zlasti v primerjavi z intenzivnostjo in količino povezav znotraj Ciglerjevih ali Schmidovih osrednjih del. Glede na rezultate stilometrične analize lahko sklenemo, da so osrednja Ciglerjeva pripovedna dela slogovno dokaj enotna in oddaljena od prevodov Schmidove literature, torej izkazujejo samosvoj pripovedni slog. V zvezi s *Kortonico* velja izpostaviti Vrhovnikovo (1892: 37) trditev, da le-ta ni izvirna, ampak »posneta« po neki daljši povesti, objavljeni v časniku *Pustertaler Bote*,²⁴ in le predstavljena na Koroško. Vendar v izdajah časnika in v dostopnih prilogah²⁵ od začetka izhajanja do vključno leta 1866, ko je *Kortonica* izšla, morebitne predloge ali izvirnika ni bilo mogoče najti. Poleg tega je časopis izhajal na štirih straneh, tako da možnosti za daljšo povest, kot je *Kortonica*, praktično ni bilo. Seveda možnosti, da je *Kortonica* »posneta«, prevedena, prirejena ipd. po predlogi ali izvirniku iz tujejezičnega, najverjetneje nemškega vira, s tem ne moremo izključiti, domneva, da je *Kortonica* Ciglerjevo izvirno delo, pa je zelo verjetna.

Pri Ciglerjevih delih v dendrogramu in prikazu v Gephiju izrazito izstopa delo *Življenje s. Heme, brumne koroške grafine*, ki glede na stilometrično analizo ne izkazuje nobenih povezav s preostalimi njegovimi deli, ampak s tremi deli različnih avtorjev, kar torej kaže, da je ne moremo uvrstiti v nobeno močnejše povezano skupino del. V primeru *Heme* gre sicer za svetniški življenjepis, vendar zgodba kot celota opisuje življenje družine z dvema sinovoma in usode njenih posameznih članov, torej bi jo lahko uvrstili v družinske povesti. V njej je prepoznan tipični potek svetnikovega življenja, značilen za baročne svetniške legende, v določenih pasusih pa tudi ideje, tipične za Ciglerjeva pripovedna dela, zlasti *Srečo v nesreči*, kot na primer zgledno zakonsko življenje glavnih junakov, med katerima ni prepira, vsako nedeljo oz. vsak

²⁴ Lokalni časopis *Pustertaler Bote* je izhajal v nemščini v Brunecku/Bruneco, danes Južna Tirolska, dvakrat (1850–1852) oz. enkrat tedensko (1853–1927).

²⁵ Digitalizirani časopis je dostopen na <https://digital.tessmann.it/tessmannDigital/Zeitung/sarchiv/Jahresuebersicht/Zeitung/8>.

dan sta pri maši in sta zgled pobožnosti, svoja sinova vzgajata v globoki veri in usmerjenosti v večno življenje, hudobneži so prej ali slej pravično kaznovani, dobri in verni ljudi so zaničevani, a na koncu dobro premaga zlo, Bog pošilja ljudem nesrečo, zato da bolj skrbijo za svojo večno srečo. Slovenska literarna zgodovina Ciglerjevo *Hemo* običajno zgolj našteva med preostalimi Ciglerjevimi deli, nekoliko več o njej pišeta Burian in Jevnikar; Burian je menil, da se je Cigler s *Hemo* želel približati svojemu vzoru Schmidu oz. njegovemu delu *Genovefa* in da je uporabil zgodovinsko-legendarno snov iz *Carinthie* (leto 1921, št. 45, 46). Gradivo naj bi »zgolj razčlenil v deset [...] poglavij« (Burian 1930: 251), kar se zdi nelogično, saj legenda v *Carinthii* obsega vsega skupaj dobre tri strani. Burianove domneve o posnemanju Schmida je zavrnil že Jevnikar (1968: 19), pritrnil pa je Burianovi ugotovitvi, da je Cigler snov črpal iz *Carinthie*, in dodal, da je ob tem tudi »kritično izbiral« iz številnih legend o sveti Hemi. Vendar pa stilometrična analiza veliko bolj pritrjuje skopi navedbi, da je bila »Sveta Hema, brumna koroška grofinja« [sic!] prevedena iz nemščine (Schumi 1882: 163). Pri tem ni nujno, da jo je prevedel ravno Cigler, čeprav je v pribesedilju zapisano, da je delo »spisal in na svitlobo dal Janez Ciegler, Višnagorski fajmašter«. Iz leta izida *Heme* in iz dejstva, da gre za svetniški življenjepiš, lahko sklepamo na enega od možnih virov, in sicer na zapuščino Mihaela Hofmana. Hofman je namreč »zbiral gradivo za sistematično slovensko življenje svetnikov« (Kidrič 2013), ki ga je sicer uporabil oz. objavil Franc Veriti, vendar je dostop do Hofmanove zapuščine imel tudi Cigler, saj je istega leta kot *Hemo* objavil *Eksercije* iz zapuščine Mihaela Hofmana. V tem kontekstu so pomenljive stilometrične povezave med in Ciglerjevo *Hemo* in Veritijevimi svetniškimi življenjepisi ter velika stilometrična oddaljenost med Veritijevimi svetniškimi življenjepisi in njegovim *Popotnikom široke inu vozke poti*.

Z računalniško podprto stilometrično analizo pripovedne literature prvih desetletij 19. stoletja so se potrdile nekatere dosedanje ugotovitve in odprla nova vprašanja in pogledi na začetke slovenske pripovedne literature. Glede na naravo obravnavane literature smo soočeni z določenimi omejitvami, zlasti z obsegom razpoložljivega korpusa besedil in vprašanjem primerjave prevedene in izvirne literature. Stilometrična analiza je vseeno pokazala odmik Ciglerjeve pripovedne literature od sočasne prevedene proze in s tem potrdila njegov primat v začetkih izvirne slovenske pripovedne literature. V nadaljnjih raziskavah bo treba presojati razmerja Ciglerjeve izvirne pripovedne literature do izvirne slovenske pripovedne literature sledečih desetletij 19. stoletja.

Literatura

- Bastian, Mathieu, Sebastien Heymann, Mathieu Jacomy (2009) „Gephi: an open source software for exploring and manipulating networks”, *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009*, San Jose, California, ZDA, 361–362.
- Binongo, José Nilo G. (2003) „Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution”, *Chance*, 16, 2, 9–13.
- Burian, Václav (1930) „Počátky slovinské zábavné prósy [+Úvodní kapitola z chystané knihy “O slovinské novelistice”]”, *Časopis pro moderní filologii*, 16, 239–258.
- Burrows, John (2002) „Delta’ – A measure of stylistic difference and a guide to likely authorship”, *Literary and Linguistic Computing*, 17, 3, 267–287.
- Dovič, Marijan (2002) „Podbevšek in Cvelbar: poskus empirične preverbe namigov o plagiatorstvu”, *Slavistična revija*, 50, 2, [233]–249.
- Eder, Maciej, Maciej Piasecki, Tomasz Walkowiak (2017) „An open stylometric system based on multilevel text analysis”, *Cognitive Studies / Études Cognitives*, 17, 1–26. doi:10.11649/cs.1430.
- Eder, Maciej, Jan Rybicki, Mike Kestemont (2016) „Stylometry with R: a package for computational text analysis”, *R Journal*, 8, 1, 107–121.
- Erjavec, Tomaž (2011) „Automatic linguistic annotation of historical language: To1TrTaLe and XIX century Slovene”, *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, Portland, Oregon, ZDA, Association for Computational Linguistics, 33–38.
- Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, Thorsten Vitt (2017) „Understanding and explaining Delta measures for authorship attribution”, *Digital Scholarship in the Humanities*, 32, 2, ii4–ii16.
- Gregorič, Asta, Maruška Mole (2016) *R in RStudio: programsko okolje za statistično obdelavo in grafični prikaz podatkov: uvodni tečaj v analizo okoljskih podatkov na Fakulteti za znanosti o okolju*, Nova Gorica.
- Grieffler, Monika, Josef Heine (1968) „Bibliographie”, *Christoph von Schmid und seine Zeit*, ur. Hans Pörnbacher, A. H. Konrad, Weissenhorn, 170–198.

- Holmes, David, Judit Kardos (2003) „Who Was the Author? An Introduction to Stylometry” *Chance*, 16, 2, 5–8.
- Jakopin, Primož (2003) „Nizkoentropijski jezikovni model na besedilih Cirila Kosmača in Ivana Cankarja” *Slovenski roman*, ur. Miran Hladnik in Gregor Kocijan, Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete, Ljubljana, 421–428.
- Jevnikar, Martin (1939) „Krištof Schmid v slovenskih prevodih”, *Slovenski jezik* 2, 188–212.
- Jevnikar, Martin (1968) *Krištof Schmid in začetki slovenskega pripovedništva. Schmidov vpliv na začetke slovenskega pripovedništva*, posebni odtis iz *Izvestja za slovenske srednje šole v Trstu*, 1966/67, Trst, 15–22.
- Kmecl, Matjaž (1974) „Ciglerjeva Sreča v nesreči” Janez Cigler, *Sreča v nesreči*, Mladinska knjiga, Ljubljana, 91–114.
- Kidrič, Francè (2013) „Hofman, Mihael (1755–1826)”, *Slovenska biografija*, Slovenska akademija znanosti in umetnosti, Znanstvenoraziskovalni center SAZU. <http://www.slovenska-biografija.si/oseba/sbi234809/#slovenski-biografski-leksikon> (dostop 25. maj 2020).
- Kolarič, Rudolf (1936) „Uvod”, Janez Cigler, *Sreča v nesreči*, Družba sv. Mohorja, Celje, [5]–35.
- Kos, Janko (1981) „Začetki slovenske pripovedne proze in evropska tradicija”, *Slavistična revija*, 29, 3, 233–258.
- Levstik, Fran (1858) „Glasnik literarni”, *Glasnik slovenski*, 2, 3, 54–56.
- Limbeck, Marko (2008) „Usage of Multivariate Analysis in Authorship Attribution: Did Janez Mencinger Write the Story »Poštena Bohinčeka«?” *Metodološki zvezki*, 5, 1, 81–93.
- Panker, Ines (2012) *Avtomatsko določanje avtorstva slovenskih leposlovnih besedil*, diplomsko delo, Fakulteta za računalništvo in informatiko Univerze v Ljubljani.
- R Core Team (2019) *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. <https://www.R-project.org/>.
- Schöch, Christof, Joseph Rudman, Hugh Craig, Patrick Juola, Mike Kestemont, Maciej Eder (2017) „Stylometry Bibliography”, ADHO-SIG Digital Literary Stylistics, https://www.zotero.org/groups/643516/stylometry_bibliography/items.

- Schumi, Franz (1883) *Archiv für Heimatkunde: Geschichtsforschungen, Quellen, Urkunden und Regesten 1*, Verlag des Herausgebers, Laibach.
- Simonič, Franc (1903) *Slovenska bibliografija*, Slovenska Matica, Ljubljana.
- Vrhovnik, Ivan (1892) *Janez Cigler, slovenski pisatelj*, Ljubljana.
- Wild, Reiner (1990) *Geschichte der Deutschen Kinder- und Jugendliteratur*, J. B. Metzlersche Verlagsbuchhandlung, Stuttgart.
- Zwitter Vitez, Ana (2011) „Povej mi karkoli in povem ti, kdo si: ugotavljanje avtorstva besedil”, *Meddisciplinarnost v slovenistiki*, ur. Simona Kranjc, Znanstvena založba Filozofske fakultete, Ljubljana, 565–570.
- Zwitter Vitez, Ana (2014) „Ugotavljanje avtorstva besedil: primer »Trenirkarjev«”, *Jezikovne tehnologije [Elektronski vir]: zbornik 17. mednarodne multikonference Informacijska družba – IS 2014, 9.–10. oktober 2014, [Ljubljana, Slovenia]: zvezek G = Language technologies: proceedings of the 17th International Multiconference Information Society - IS 2014, October 9th–10th, 2014*, ur. Tomaž Erjavec, Jerneja Ž. Gros, Ljubljana, Slovenia, 131–134.
- Zwitter Vitez, Ana, Senja Pollak (2019) „Profiliranje avtorjev besedil: viri, metode in uporaba na področju humanistike in družboslovja”, *Znanost in družbe prihodnosti*, ur. Miroљjub Ignjatović, Aleksandra Kanjuo Mrčela, Roman Kuhar, Slovensko sociološko društvo, Ljubljana, 168–171.

SUMMARY

Andrejka Žejn

COMPUTATIONAL STYLOMETRIC ANALYSIS OF NARRATIVE
PROSE BY JANEZ CIGLER AND CHRISTOPH VON SCHMID IN
SLOVENIAN

The first original Slovenian story, *Sreča v nesreči* (Fortune in Misfortune), was published by Janez Cigler in 1836. Slovenian literature specialists have ascribed a major influence on its creation to the works of the Bavarian writer Christoph von Schmid, not only in terms of its content, structure, concept, and narrative style, but also in following the spirit of the time and its regard for tradition. These researchers primarily focused on comparisons of content, but not on analyses of similarities in style. The Stylo package in the statistical software environment R for advanced stylistic analyses of text collections was therefore used to perform a stylometric analysis of Cigler's works and Schmid's translated works, as well as some translations of narrative prose from the period studied. Given certain limitations—such as the limited volume of texts, translated and original literature, and genre diversity—the analysis confirms Cigler's original style rather than an imitation of Schmid, raising new questions regarding the authorship of certain works.

Key words: *Slovenian narrative prose; Janez Cigler; Christoph von Schmid; stylometry; style*