

Simpozij OBDOBJA 42

Lucija Mandić

**Kako meščanski je v resnici meščanski roman? Kvantitativna vsebinska analiza
daljše slovenske pripovedne proze v dolgem 19. stoletju**

objavljeno v:

Andraž Jež (ur.): *Slovenska literatura in umetnost v družbenih kontekstih.*

Obdobja 42. Ljubljana: Založba Univerze v Ljubljani, 2023.

<https://centerslo.si/simpozij-obdobja/zborniki/obdobja-42/>

© Univerza v Ljubljani, Filozofska fakulteta, 2023.

Obdobja (e-ISSN 2784-7152)

Univerza v Ljubljani
Filozofska fakulteta



KAKO MEŠČANSKI JE V RESNICI MEŠČANSKI ROMAN? KVANTITATIVNA VSEBINSKA ANALIZA DALJŠE SLOVENSKE PRIPOVEDNE PROZE V DOLGEM 19. STOLETJU

Lucija Mandić

Inštitut za slovensko literaturo in literarne vede, ZRC SAZU, Ljubljana
lucija.mandic@zrc-sazu.si

DOI:10.4312/Obdobja.42.49-58

V prispevku smo uporabili metodo modeliranja tem za raziskovanje tem, povezanih z industrializacijo, meščanstvom, političnim udejstvovanjem in nacionalnimi idejami v daljši slovenski pripovedni prozi v dolgem 19. stoletju. Analizo smo izvedli na korpusu ELTeC-slv, ki vsebuje 100 daljših prozskih del v slovenščini, ki so izšla med letoma 1836 in 1921. Pridobljene rezultate smo predstavili v odnosu do literarnih zvrsti in žanrov.

modeliranje tem, digitalna literarna veda, daljša pripovedna proza, slovenski roman

This article uses the topic modeling method to investigate topics related to industrialization, the bourgeoisie, political activity, and national ideas in long Slovenian narrative prose in the long nineteenth century. The analysis was carried out on the ELTeC-slv corpus, which contains one hundred long prose works in Slovenian published between 1836 and 1921. The results obtained are presented in relation to literary forms and genres.

topic modeling, computational literary studies, long narrative prose, Slovenian novel

1 Uvod

Vzpon meščanstva v dolgem 19. stoletju, ki se začne s francosko revolucijo in zaključuje s koncem 1. svetovne vojne (Hobsbawm 1989: 6), s seboj ni prinesel le razmaha industrializacije, razcveta periodike in oblikovanja nacionalne ideologije, temveč tudi novo literarno zvrst – roman (Kmecl 1981: 20–22). Prvi slovenski roman je izšel le leto pred prvimi volitvami v deželne stanove in je bil sprva neločljivo povezan z novim narodnozavednim meščanskim slojem. Roman se je skozi drugo polovico 19. stoletja žanrsko diferenciral, večje število časopisov in revij ter številčnejša slovensko beroča javnost pa sta prispevala k rasti števila objav. Dela, ki so nastajala, so se razlikovala po obsegu, ciljnem bralstvu, tematiki, kakovosti in žanrih (Perenič 2021: 62); pojavili so se sentimentalni roman, zgodovinski roman, znanstvena fantastika, ženski roman, takoj po prelomu stoletja pa tudi prvi socialni roman – Finžgarjev *Iz moderne sveta*. Proti koncu 19. stoletja se ni večala le produkcija romanov, v velikih nakladah so izhajale tudi povesti Mohorjeve družbe, pogosto s kmečko in družinsko tematiko – tudi Ciglerjev tip bidermajerske vzgojne povesti je vztrajal vse do konca 19. stoletja (Hladnik 1982). Avtorji v drugi polovici 19. stoletja niso strogo razlikovali med

oznakami roman, povest in včasih celo novela, ampak so jih pogosto uporabljali zamenljivo, zato so poleg romanov v korpus ELTeC-slv vključene tudi povesti in novele. Povest naj bi bila »krajša in večidel tudi poljudnejša pripovedna zvrst« (Kos 1985: 28), a je predstavljala tudi »splošnejši pojem, ki je označeval nekatere tipe romana« (prav tam). V tem prispevku nas je zanimala vsebina literarnih del, ki se nanaša na družbeno-zgodovinsko realnost časa, v katerem je nastala, poleg tega pa smo preverili, kakšna je interakcija vsebine z različnimi tipi daljše pripovedne proze.

2 Modeliranje tem

Za analize, ki se posvečajo vsebini literarnih del, je metoda modeliranja tem ena najpogostejših izbir (Jockers, Mimno 2013; Uglanova, Gius 2020; Shadrova 2021; Schröter, Du 2022). Gre za nenadzorovano tehniko strojnega učenja, ki pregleda nabor dokumentov, zazna vzorce besed in besednih zvez v njih ter samodejno združi skupine besed, ki naj bi najboljše povzemale vsebino dokumentov (Brett 2012). Tematski modeli so uporabni za združevanje dokumentov v skupine, organiziranje velikih korpusov besedilnih podatkov in pridobivanje informacij iz nestrukturiranega besedila. Obstaja več modelov, ki jih lahko uporabimo za modeliranje tem. Pri tej analizi smo uporabili model LDA (latentna Dirichletova alokacija), ki izhaja iz predpostavke, da vsa besedila v korpusu vsebujejo določeno število tem, pri čemer vsaka vsebuje nabor verjetnostno razporejenih besed, vsak dokument pa je sestavljen iz več tem.

3 Korpus

Za analizo smo uporabili že obstoječi korpus ELTeC-slv (Erjavec idr. 2020), ki je nastal v sklopu COST Akcije *Distant Reading for European Literary History*. Vsebuje 100 daljših proznih besedil v slovenščini, ki so bila objavljena med letoma 1836 in 1921. Korpus ELTeC-slv je del obsežnega korpusa ELTeC (European Literary Text Collection), ki združuje korpus v 17 evropskih jezikih, ti pa so sestavljenih po skupnih parametrih. Ti parametri določajo obseg korpusa, obseg posameznih besedil (nad 10.000 besed), zastopanost avtoric, zastopanost posameznih avtorjev, enakomerno razporejenost števila besedil po posameznih obdobjih itn. (gl. Schöch idr. 2021). Korpus ELTeC-slv naj bi bil reprezentativen vzorec objavljenih besedil v slovenščini v obdobju med letoma 1836 in 1921, a v praksi ne predstavlja povsem realnega stanja v slovenski daljši prozi v dolgem 19. stoletju.

4 Kakšne tematske modele proizvede LDA in na kakšen način ti reflektirajo vsebino literarnih del

Čeprav je uporaba modeliranja tem v digitalni literarni vedi splošno razširjena, se mnogim zdi vprašljiva. Shadrova (2021) izpostavlja, da se koncept »tém«, ki jih pridobimo z LDA, konceptualno razlikuje od lingvističnih konceptov, ki se nanašajo na vsebino besedila (in ki tudi sami niso enoznačno definirani), saj jih ne ekstrahiramo iz besedila, temveč jih konstruiramo. Izraz »tema« ne predstavlja sinonima za

npr. »literarno temo«, saj gre z vidika strojnega učenja le za objekte s podobnimi strukturnimi, funkcionalnimi in drugimi značilnostmi, zbranimi v bolj ali manj koherentno gručo (Uglanova, Gius 2020). To ne pomeni, da z modeliranjem tem ne dobimo zanimivih informacij o vsebini (literarnih) besedil. Schröter in Du (2022) ugotavljata, da so informacije, pridobljene z modeliranjem tem, na primerih literarnih besedil najbližje temu, kar v literarni vedi po ruski formalistični šoli razumemo kot siže. Schröter in Du ga po Tomaševskem definirata kot način predstavitve fabule na ravni diskurza, ki ne vključuje le narativnih struktur literarnega dela, temveč tudi dogajalni kraj, čas in opise junakov. S tega vidika je modeliranje tem kljub skepticizmu Shadrove uporabno orodje za vsebinske analize literarnih korpusov.

5 Metodologija

S postopkom modeliranja tem smo želeli v korpusu ELTeC-slv zaslediti vzorce pojavljanja tem s področja industrializacije, meščanstva, političnega udejstvovanja in nacionalnih idej. Za modeliranje tem na primeru korpusa ELTeC-slv smo uporabili paket Gensim (Řehůřek, Sojka 2010) v programskem jeziku Python. Gensim deluje po načelu modela LDA; ta dovoljuje izbiro parametrov (segmentiranje besedil, določitev števila tem in števila besed na temo), ki omogočijo najboljše rezultate glede na njihovo interpretabilnost.

5.1 Predpriprava besedil

Iz dosedanjih študij na področju modeliranja tem je razvidno, da je kakovost pridobljenih tem v veliki meri odvisna od predpriprave korpusa (Jockers 2013: 131; Uglanova, Gius 2020; Sherstinova idr. 2022). Korpus ELTeC-slv je kodiran v jeziku XML skladno s smernicami TEI (TEI Consortium 2020), avtomatično tokeniziran in stavčno segmentiran z orodjem ReLDA tokeniser (Ljubešić, Erjavec 2016) ter lematiziran in oblikoskladenjsko označen z modelom CLASSLA-StanfordNLP (Qi idr. 2018). Imenske entitete so avtomatično označene z orodjem Janes-NER (Fišer idr. 2020).

Običajna praksa je, da se pred izdelavo tematskih modelov iz korpusa izločijo t. i. nepomembne besede (angl. *stop words*). Seznam nepomembnih besed zajema npr. nepolnopomenske besede, imena mesecev, dni v tednu, literarnih junakov ipd. Po preizkušanju različnih možnosti smo ugotovili, da z izločanjem besed na podlagi vnaprej sestavljenih seznamov nepomembnih besed ne dobimo tako dobrih rezultatov, kot če besede izločamo na podlagi njihovih oblikoskladenjskih oznak.

Najuporabnejše modele smo dobili z upoštevanjem lem ter filtriranjem nepolnopomenskih besed in imenskih entitet. Upoštevanje lem je pomembno pri vseh pregibnih jezikih, saj se besede v različnih variantah pogosto pojavljajo v bližini. Rezultat tega so teme, ki združujejo enake besede v različnih pregibnih oblikah, ne pa semantično sorodne, a različne besede. Ker je bil korpus že predhodno tokeniziran, lematiziran, oblikoskladenjsko označen in ima označene tudi imenske entitete, je bil postopek predpriprave zelo enostaven. Z nastavitvijo parametrov v knjižnici Gensim smo pri modeliranju tem upoštevali le leme samostalnikov, izločili pa smo tudi vse

imenske entitete, saj so se te izkazale za moteče zaradi svoje polisemije (ime Janez v korpusu npr. referira na mnogo različnih oseb).

Ker so nas zanimala aktualna tema med letoma 1836 in 1921, smo za natančnejše rezultate izločili zgodovinske romane in zelo dolg pustolovski kolportajni roman *Tolovajski glavar Črni Jurij in njegovi divji tovariši*. Izključili smo tudi delo *Moja duša vasuje* Frana Saleškega Finžgarja, ki se je izkazalo za zbirko novel in kot tako ne sodi med daljša prozna dela. Pri filtriranju besedil smo si pomagali z metapodatki, pridobljenimi za Korpus daljše slovenske pripovedne proze KDSP (Mandić, Erjavec 2023). Končne analize smo tako izvedli na 62 besedilih.

5.2 Izgradnja modela LDA

Izračunali smo več različnih modelov LDA in na koncu prišli do optimalnih parametrov, s katerimi smo pri segmentih po 500 besed dobili model s 25 temami po 10 besed.

6 Rezultati analize

Pridobljene teme smo nato interpretirali in jih naslovili – interpretacija tematskih modelov je še vedno problematična točka modeliranja tem, saj je v veliki meri odvisna od raziskovalke ali raziskovalčeve subjektivne presoje, kar velja tudi v tem primeru. Z upoštevanjem raziskav na tem področju (gl. Gillings, Hardie 2022) je bila interpretacija v primeru te analize podprta tudi z natančnim branjem besedil. Izmed 25 tem smo izluščili tiste, ki se glede na našo interpretacijo lahko nanašajo na meščanstvo, industrializacijo, politično udejstvovanje in nacionalne ideje. Za relevantne so se izkazale tudi teme s področja plemstva, sentimentalnosti, družine in podeželskega okolja. Prvi dve lahko povežemo z razvojem romana, predvsem ženskega (Mihurko Poniž 2010; Hladnik 1981; Perenič 2021), slednji pa z vzgojno-poučnimi in kmečkimi povestmi (Žejn 2021; Hladnik 1982). Med temami, ki so bile interpretabilne, so se pojavile tudi vojaške, popotniške in pomorske teme, ki pa jih tu ne bomo obravnavali. Nekaj pridobljenih tem je bilo po pričakovanjih tudi težko interpretabilnih, zato smo jih označili kot »nejasne«.

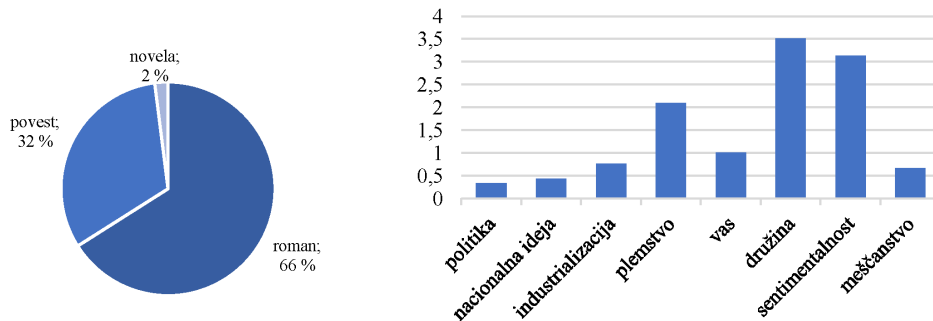
Za potrebe analize smo izluščili osem tem:

- tema 6 (politika): volitev, komisar, glavar, kandidat, shod, stranka, narod, volilec, poslanec, večer;
- tema 8 (nacionalna ideja): narod, država, stoletje, jezik, pesnik, zgodovina, korist, nega, rod, duh;
- tema 15 (industrializacija): zdravnik, gospod, seme, nadzornik, delavec, ravnatelj, tovarna, vrata, delo, glava;
- tema 18 (plemstvo): baron, grofica, gospod, dan, misel, beseda, družba, čas, srce, gospa;
- tema 20 (podeželje): pot, deklica, glava, večer, beseda, vas, mladenič, oče, mlin, raz;
- tema 21 (družina): mati, oče, dan, otrok, leto, sin, pismo, hiša, denar, mož;

- tema 24 (sentimentalnost): oči, glava, pogled, obraz, ljubezen, gospod, duša, ženska, lice, miza;
- tema 25 (meščanstvo): gospod, beseda, doktor, duh, društvo, prijatelj, sluga, gospa, oči, odvetnik.

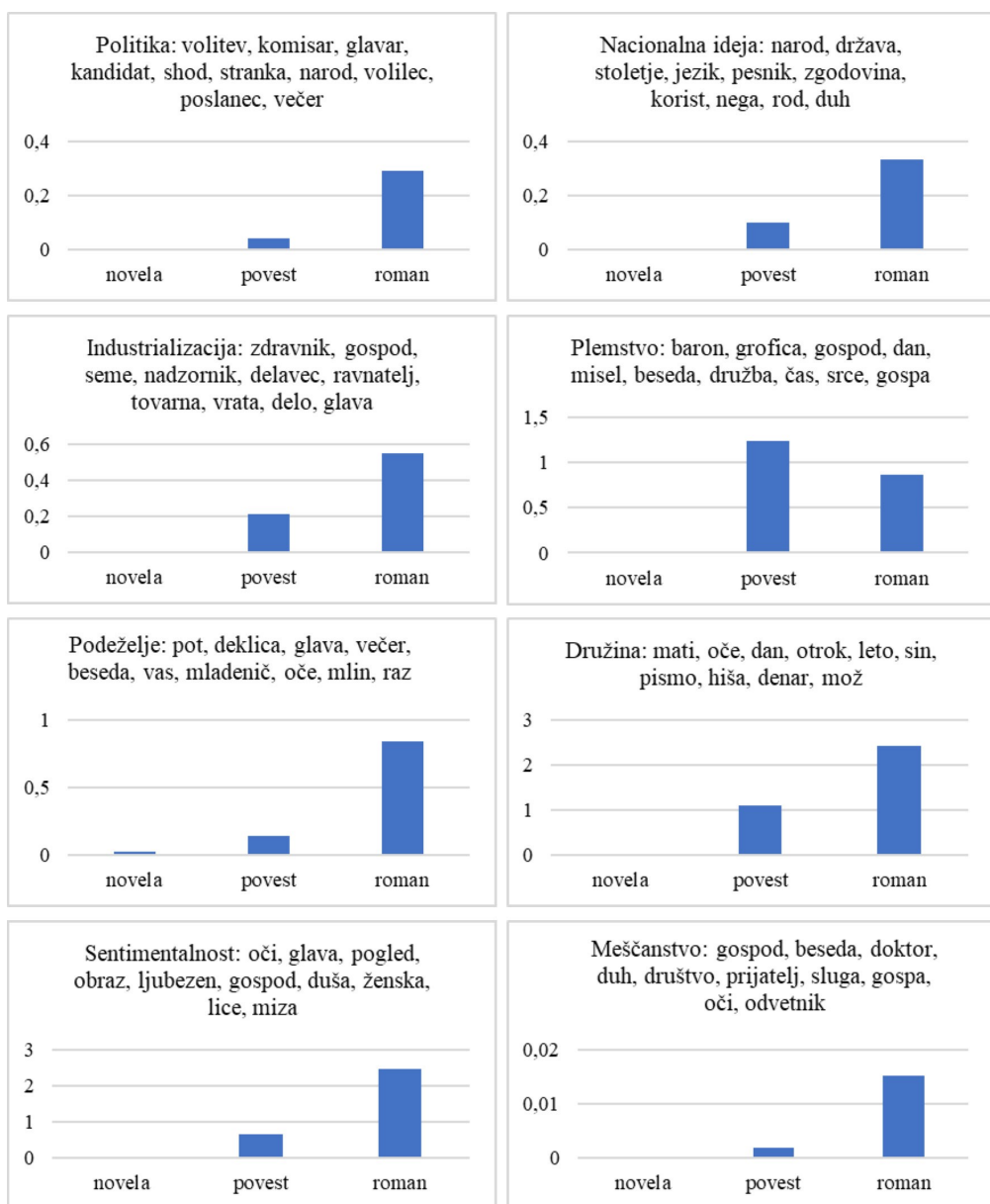
6.1 Teme in literarne zvrsti

Literarna zgodovina povezuje razvoj romana z vzponom meščanstva, kljub prevladi romanov v korpusu (Graf 1) pa so osrednje teme, povezane z meščanstvom (poleg teme »meščanstvo« še temi »nacionalna ideja« in »politika«), zastopane v manjšini (Graf 2). Glede na pridobljene podatke bi lahko sklepali, da je slovenski roman v 19. stoletju predvsem sentimentalni in družinski. Ker smo želeli dodatno preveriti prisotnost posamezne teme v različnih literarnih zvrsteh, smo izračunali verjetnost pojavljanja posameznih tem v različnih literarnih zvrsteh. Pri tem smo uporabili povprečne vrednosti, da večja zastopanost romanov v korpusu ne bi vplivala na pridobljene rezultate.



Graf 1: Razmerje med romani, povestmi in novelami v korpusu. Graf 2: Zastopanost izbranih tem v korpusu.

Analiza distribucije tem glede na literarno zvrst (Graf 3–10) je prikazala bolj pričakovane rezultate; teme »politika«, »nacionalna ideja«, »meščanstvo« in »sentimentalnost« so v večji meri prisotne v delih z oznako roman kot v delih z oznako povest ali novela. Zanimivo je, da je tema »industrializacija« v primerljivi meri prisotna tako v povestih kot v romanih, prav tako tema »družina«. Izstopa tema »plemstvo«, ki je v večji meri prisotna v povestih. Edina tema, ki je v največji meri prisotna v edini noveli v korpusu, je »podeželje«. Gre za novelo *O, ta testament!* Frana Jakliča, ki je res postavljena v vaško okolje, presenetljivo pa je, da je tema »podeželje« v večji meri zastopana v romanih kot pa v povestih, med katerimi jih je devet opredeljenih kot »kmečka povest« (v primerjavi s štirimi romani s kmečko tematiko).

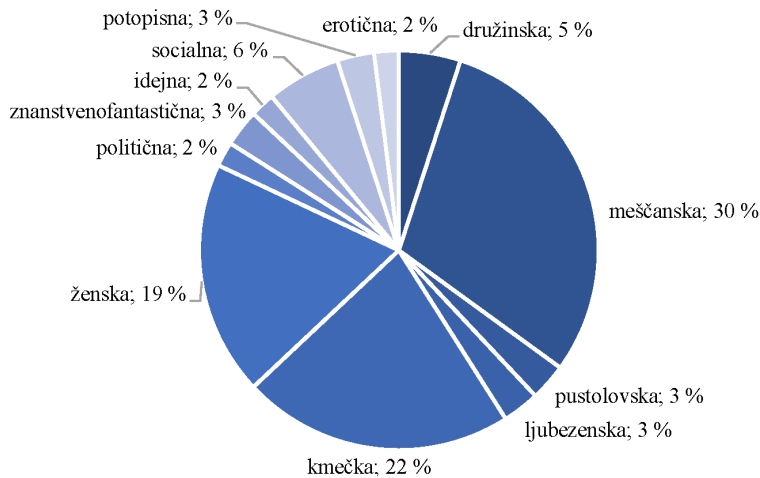


Grafi 3–10: Prisotnost posamezne teme v novelah, povestih in romanih.

6.2 Modeliranje tem in žanri

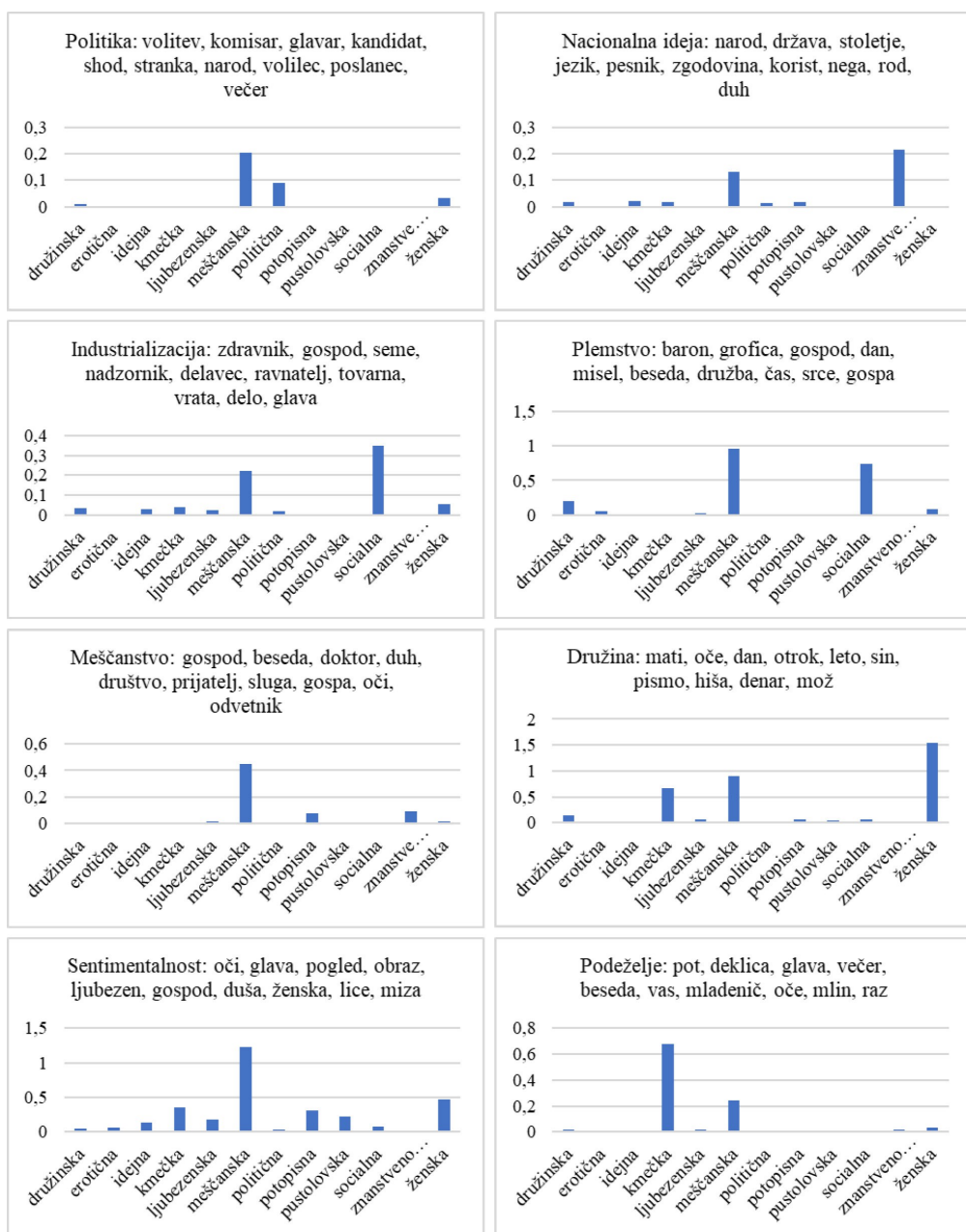
V zadnjem delu prispevka smo se posvetili odnosu med tematskimi modeli in metapodatki, ki se nanašajo na tematiko literarnih del in ki so bili pridobljeni s pomočjo žanrskih oznak, literarnozgodovinskih opredelitev in natačnega branja. Zbrani metapodatki se ne navezujejo strogo na žanrske oznake, kljub temu pa nam

lahko podajo okvirno informacijo o korelaciji žanra in strojno pridobljenih tematskih modelov.



Graf 11: Sestava korpusa glede na tematiko (podatki zbrani na podlagi žanrskih oznak in natančnega branja).

Kot vidimo na Grafu 11, v korpusu prevladujejo dela z meščansko, kmečko in žensko tematiko. Model LDA je v večini primerov potrdil informacije, pridobljene na podlagi žanrskih oznak in natančnega branja (Graf 12–19); tema »politika« je pričakovano v večji meri prisotna v meščanskih in političnih delih, tema »industrializacija« pa v meščanskih in socialnih povestih in romanih. Podobno velja za teme »družina«, »podeželje« in »meščanstvo«. Tema »nacionalna ideja« je v večji meri prisotna v meščanskih in znanstvenofantastičnih delih. Pri slednjih gre za Mencingerjevega *Abadona* in Tavčarjev roman *4000*; ker gre za antiutopični deli, ki komentirata aktualno slovensko politično situacijo s konca 19. stoletja, tak rezultat ni presenetljiv. Zanimivo je, da se v delih s socialno tematiko v večji meri kot drugod pojavlja tema »plemstvo«. Tema »sentimentalnost« je sicer največkrat prisotna v delih z meščansko tematiko, vendar je glede na pridobljene podatke tudi najbolj univerzalna za vse žanre. Izstopa tudi tema »družina«, ki se največkrat pojavlja v ženskih romanih, presenetljivo malo pa v delih, ki naj bi obravnavala družinsko tematiko. Čeprav na tem mestu ni prostora za poglobljene kvalitativne analize posameznih literarnih del, na katera se rezultati modeliranja tem nanašajo, ti predstavljajo obetavno izhodišče za nadaljnje raziskave slovenske daljše proze v dolgem 19. stoletju.



Grafi 12–19: Prisotnost posameznih tem glede na metapodatkovne oznake tematik.

7 Zaključek

V prispevku smo z metodo modeliranja tem preučevali teme, povezane z industrializacijo, meščanstvom, politiko in nacionalnimi idejami, na primeru daljše slovenske pripovedne proze v dolgem 19. stoletju. Opisali smo postopek modeliranja

tem z uporabo modela LDA na primeru korpusa ELTeCslv. Pridobljene podatke smo predstavili v dveh perspektivah: literarnozvrstni in žanrski. V prvem delu analize smo primerjali povprečne vrednosti prisotnosti posameznih tem med povestmi, romani in novelo. Izkazalo se je, da so vse teme, razen tem »plemstvo« in »podeželje«, v večji meri zastopane v romanih, medtem ko je razlika v zastopanosti tem »industrializacija« in »družina« v romanih in povestih majhna. Tema »plemstvo« je prevladovala v povestih, edina tema, ki je prevladovala v noveli, pa je »podeželje«. V zadnjem delu smo primerjali podatke o tematiki literarnih del, ki so bili pridobljeni z upoštevanjem žanrskih oznak in natančnega branja, s strojno pridobljenimi tematskimi modeli. Primerjava je pokazala, da večina tematskih modelov podpira »analogno« pridobljene informacije o tematiki literarnih del; odstopanja so se pokazala le v primeru teme »plemstvo«, ki je bila v nepričakovano veliki meri zastopana v delih s socialno tematiko, in teme »družina«, ki je bila manj zastopana v delih z družinsko tematiko, zato pa toliko bolj v delih s kmečko in žensko tematiko. Čeprav je korpus po izključitvi nerelevantnih del štel le 62 besedil, je modeliranje tem pokazalo obetavna izhodišča za nadaljnje študije slovenke književnosti na tem področju, kvantitativne rezultate strojno podprtih analiz pa bi bilo v prihodnosti dobro nadgraditi z dodatnimi kvalitativnimi študijami del, ki so bila vključena v korpus.

Literatura

- BAJT, Jan, ROBNIK-ŠIKONJA, Marko, 2022: Strojna analiza tematik in sentimenta. *Uporabna informatika* XXX/1. 19–36. Na spletu.
- BRETT, Megan R., 2012: Topic Modeling: A Basic Introduction. *Journal of Digital Humanities* II/1. Ni paginacije. Na spletu.
- ERJAVEC, Tomaž, HLADNIK, Miran, JUVAN, Marko, MIHURKO PONIŽ, Katja, 2021: *Slovenian Novel Corpus (ELTeC-slv): April 2021 release (v2.0.0)*. Zenodo. <https://doi.org/10.5281/zenodo.4662600> (dostop 16. 8. 2023).
- FIŠER, Darja, LJUBEŠIČ, Nikola, ERJAVEC, Tomaž, 2020: The Janes Project: Language Resources and Tools for Slovene User Generated Content. *Language Resources & Evaluation* 54. 223–246. Na spletu.
- GILLINGS, Matthew, HARDIE, Andrew, 2022: The Interpretation of Topic Models for Scholarly Analysis: An Evaluation and Critique of Current Practice. *Digital Scholarship in the Humanities* XXXVIII/2. 530–543. Na spletu.
- HLADNIK, Miran, 1981: Slovenski ženski roman v 19. stoletju. *Slavistična revija* XXIX/3. 259–293. Na spletu.
- HLADNIK, Miran, 1982: Mohorjanska pripovedna proza. *Slavistična revija* XXX/4. 389–414. Na spletu.
- HOBSBAWM, Eric J., 1989: *The Age of The Empire 1875–1914*. New York: Vintage Books.
- JOCKERS, Matthew, 2013: *Macroanalysis*. Urbana, Chicago, Springfield: University of Illinois Press.
- JOCKERS, Matthew, MIMNO, David, 2013: Significant Themes in 19th-century Literature. *Poetics* XLI/6. 750–769. Na spletu.
- KMECL, Matjaž, 1981: *Rojstvo slovenskega romana*. Ljubljana: Mladinska knjiga.
- KOS, Janko, 1985: Evropski izviri slovenskega romana v 19. stoletju. *Slavistična revija* XXXIII/1. 27–50. Na spletu.
- LJUBEŠIČ, Nikola, ERJAVEC, Tomaž, 2016: Corpus vs. Lexicon Supervision in Morphosyntactic Tagging: The Case of Slovene. Nicoletta Calzolari idr. (ur.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. 1527–1531. Na spletu.

- MANDIĆ, Lucija, ERJAVEC, Tomaž, 2023: *Corpus of Longer Narrative Slovenian Prose KDSP 1.0*. CLARIN.SI. <http://hdl.handle.net/11356/1823> (dostop 16. 8. 2023)
- MIHURKO PONIŽ, Katja, 2010: Prispevki slovenskih pripovednic k žanrski podobi proze 19. stoletja. *Jezik in slovstvo* LV/1–2. 47–60. Na spletu.
- PERENIČ, Urška, 2021: Prispevek Luize Pesjak h konstituiranju slovenske kratke pripovedne proze. Alenka Žbogar (ur.): *Ustvarjalke v slovenskem jeziku, literaturi in kulturi. 57. seminar slovenskega jezika, literature in kulture*: Ljubljana: Znanstvena založba Filozofske fakultete. 63–70.
- QI, Peng, DOZAT, Timothy, ZHANG, Yuhao, MANNING, Christopher D., 2018: Universal Dependency Parsing from Scratch. Daniel Zeman (ur.): *Proceedings of the CoNLL 2018*. 160–70. Na spletu.
- ŘEHŮŘEK, Radim, SOJKA, Petr, 2010: Software Framework for Topic Modelling with Large Corpora. *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks*. Valletta: University of Malta. 46–50. Na spletu.
- SCHÖCH, Christof, PATRAS, Roxana, ERJAVEC, Tomaž, SANTOS, Diana, 2021: Creating the European Literary Text Collection (ELTeC): Challenges and Perspectives. *Modern Languages Open* XXV/1. 1–19. Na spletu.
- SCHRÖTER, Julian, DU, Keli, 2022: Validating Topic Modeling as a Method of Analyzing Sujet and Theme. *Journal of Computational Literary Studies* I/1. 1–20. Na spletu.
- SHADROVA, Anna, 2021: Topic Models do not Model Topics: Epistemological Remarks and Steps Towards Best Practices. *Journal of Data Mining and Digital Humanities* XVIII/1. Na spletu.
- SHERSTINOVA, Tatiana, MOSKVINA, Anna, KIRINA, Margarita, ZAVYALOVA, Irina, KARYSHEVA, Asya, KOLPASHCHIKOVA, Evgenia, MAKSIMENKO, Polina, MOSKALENKO, Alena, 2022: Topic Modeling of Literary Texts Using LDA: On the Influence of Linguistic Preprocessing on Model Interpretability. Sergey Balandin, Tatiana Shatalova (ur.): *31st Conference of Open Innovations Association (FRUCT)*. Helsinki. 305–312. Na spletu.
- TEI Consortium, 2020: TEI P5: *Guidelines for Electronic Text Encoding and Interchange*. <https://tei-c.org/guidelines/P5/> (dostop 16. 8. 2023).
- UGLANOVA, Inna, GIUS, Evelyn, 2020: The Order of Things. A Study on Topic Modelling of Literary Texts. Folgert Karsdorp idr. (ur.): *Proceedings of the Workshop on Computational Humanities Research, Amsterdam, Netherlands*. 57–76. Na spletu.
- ŽEJN, Andrejka, 2021: *Izhodišča slovenske pripovedne proze*. Ljubljana: Založba ZRC, ZRC SAZU. Na spletu.