
Marko Juvan

ZRC SAZU

*Inštitut za slovensko literaturo in literarne vede
Univerza v Ljubljani, Filozofska fakulteta*

Mojca Šorli

ZRC SAZU

*Inštitut za slovensko literaturo in literarne vede
Univerza v Ljubljani, Filozofska fakulteta*

Andrejka Žejn

ZRC SAZU

Inštitut za slovensko literaturo in literarne vede

UDK 821.163.6.09:81'322'1968''

INTERPRETIRANJE LITERATURE V ZMANJŠANEM MERILU: »ODDALJENO BRANJE« KORPUSA »DOLGEGA LETA 1968«

Članek obravnava vlogo interpretacije v digitalnohumanističnem pristopu k literaturi, posebej v korpusni analizi. Pokaže, kakšen je v literarni vedi interpretativni pomen računalniških statistično-preštevni analiz pojavnic, oznak in metaoznaki v majhnem korpusu literarnih besedil, zajetem iz določenega časovnega obdobja. Na primeru korpusa izvornih literarnih besedil, natisnjenih med mednarodnim študentskim gibanjem »dolgega leta 1968« v periodiki pretežno študentske generacije (*Tribuna in Problemi*), kvantitativna analiza oz. »oddaljeno branje« raziskuje uveljavljanje modernizma, spreminjanje vrstnega repertoarja in literarnega jezika, besedilne sledi revolucionarnega ozračja in vlogo žensk v literarnem sistemu.¹

Ključne besede: digitalna humanistika, literarni korpus, oddaljeno branje, korpusna analiza, slovenska literatura, študentsko gibanje, 1960–1975

1 Digitalna humanistika in interpretacija literature

»Digitalna humanistika« (DH) je izraz, ki je očaral naše tisočletje, potem ko je odmevni zbornik *A Companion to Digital Humanities* (Schreibman idr. 2004) z njim zamenjal bolj tehnični termin »humanistično računalništvo« (*humanities computing*).² To je bilo sprva obrobno področje, ki se je kot mostišče med »dvema

¹ Članek je rezultat projekta *Maj 68 v literaturi in teoriji: zadnja sezona modernizma v Franciji, Sloveniji in svetu* (J6-9384), ki ga financira ARRS.

² Že leto pozneje izraz posloveni Matija Ogrin v uvodu k zborniku o znanstvenih izdajah (Ogrin 2005: 12–15).

kulturama« (Snow 2013)³ trudilo za priznanje v humanističnem *mainstreamu* vse od petdesetih let 20. stoletja. Pod posodobljeno blagovno znamko »digitalna humanistika« je v našem tisočletju končno doživelo kanonizacijo in ekspanzijo.⁴ V financiranju znanosti, zagledanem v nove tehnologije in meddisciplinarnost kot vrednoti sami na sebi, je DH danes uvrščena med indikatorje t. i. raziskovalne odličnosti in družbene aplikativnosti. Vodilnega toka humanistike in njegovih elitnih ustanov DH sicer v glavnem še ni osvojila, a trka na njihova vrata kot nova, neoliberalnemu duhu časa primerna znanstvena paradigma (prim. Allington idr. 2016; Schreibman idr. 2004: xxvi; Siemens idr. 2007: 17; Terras idr. 2013: 166, 175–176). »Analogna« humanistika, če ji tako rečemo, do mlajše digitalne sestre ostaja zadržana in kritična. Med drugim ji očita teoretsko podhranjenost, opiranje na zastarela pojmovanja teksta, naivni realizem ter trivialnost, nezadostnost ali celo zgrešene računalniški rezultati, na prvi pogled neovrgljivih.⁵ Kljub pomislekom se DH v obliki socialnega omrežja, ki daje vtis demokratskega sodelovanja, od devdesetih let 20. stoletja nezadržno širi, v veliki meri prek digitalnih medijev. Tudi računalniško slabo pismenim humanističnim strokovnjakom so njene metode in orodja približali osebni računalniki, dostopni arhivi digitaliziranih tekstov ter prijazni programi za hrambo in obdelavo podatkov.

Zdi se, da je jezikoslovje po svojem predmetu raziskovanja in strukturalni metodi najbližje formalizmu računalniških jezikov. Morda se je zaradi te homologije humanistično računalništvo najprej zasidrilo v analizi jezika in tekstov. Računalniških orodij, ki so s svojimi kapacitetami pomnjenja in obdelave nepreglednih količin podatkov presegle omejitve posameznikovega uma, so se nato prijemale še druge humanistične stroke, med prvimi literarna veda. Na začetku se je po zgledu jezikoslovja ukvarjala s stilno in semantično-tematsko analizo, izpeljano iz najpogostejših besed in konkordanc (prim. Scherber 1977). Veda o književnosti sicer ima (tudi pri nas) tradicijo preštevnih metod, denimo v verzologiji (Pretnar 1997), teoriji žanrov, stilistiki in študijah avtorstva (prim. Dović 2004; Hladnik 1993, 1995; Perenič 2010). Toda pred dobo DH literarna veda zvečine še ni segala po algoritmičnih in zahtevnejši statistiki, kar je temeljni metodološki pogoj v današnji računalniški stilistiki (prim. Allison idr. 2013; Rybicki idr. 2016), računalniškem prepoznavanju žanrov in drugih raziskavah »kvantitativnega formalizma« (Allison idr. 2011). Odkar se literarna stroka prepušča čarom DH, vedno raje uporablja računalniška orodja, zlasti pri elektronskih znanstvenokritičnih izdajah (gl. Ogrin 2005), spletnih knjižnicah, prepoznavanju avtorstva, raziskavi žanrov ali stilov (gl. Žejn 2021), v zadnjem času tudi georeferenciranju literarnozgodovinskih podatkov (gl. Juvan in Dokler 2015).

³ Snow razpravlja o prepadu med »kulturama« matematično-naravoslovnih znanosti in družboslovno-humanističnih ved, kot se odražata v izobrazbi, družbenem napredku in ugledu.

⁴ O zgodovini humanističnega računalništva in digitalne humanistike po svetu gl. Jannidis idr. 2017; Schreibman idr. 2004; Siemens 2016; Siemens in Schreibman 2007; Terras idr. 2013. O DH na Slovenskem gl. Hladnik 2012.

⁵ Podobna kritika prihaja tudi iz insajderskih vrst, gl. npr. Da 2019.

Bržčas največji premik, ki ga je v literarni stroki izzvala DH, so spodbudili korpusi, sestavljeni iz tisočev, celo sto tisočev besedilnih enot (prim. Hoover idr. 2014). Graditev obsežnih digitalnih repozitorijev, strukturiranih in označenih po notnih merilih, so omogočili množična digitalizacija tiskane dediščine, njena dosegljivost na internetu, pa tudi razvoj standardov za računalniško označevanje besedil (Hoover idr. 2014: 2; Terras idr. 2013: 133–135).

Besedilni korpusi so zahtevali bistveno spremembo literarnovedne metode, njeno zblizanje z družboslovno in računsko znanstveno domeno.⁶ Gre za premik od t. i. branja od blizu k branju na daljavo. Bližnje branje (angl. *close reading*) je metonimija za natančno interpretacijo posamične besedne umetnine (najpogosteje izbrane iz kanona), oddaljeno branje pa je metafora za analizo arhiva,⁷ tj. množice tekstov, v kateri nas ne zanimata ne pomen ne struktura enega samega besedila, temveč distribucija oziroma pogostnost določene lastnosti, strukture ali elementa v izbrani besedilni množici. Izraz *distant reading* je v teorijo svetovne literature vpeljal Franco Moretti (2011). Z njim je poimenoval metodo posploševanja podatkov, ki jih dobimo iz druge roke, iz metabesedil izvedencev, ki so preučili relevantna dela v jezikih svoje filološke specializacije. Branje na daljavo je po Morettiju nujen pogoj za kakršno koli sintezo o svetovnem literarnem pojavu, prepoznavnem nad ali pod ravnijo besedila (denimo za posplošitev o medliterarnosti posamičnega žanra, forme). Samo tako se namreč svetovnozgodovinska sinteza izogne prehitrim sklepom, izpeljanim iz zoženega kanona vodilnih književnosti sveta. Z ustanoviteljskim delovanjem v Literarnem laboratoriju Univerze Stanford je Moretti zaslužen tudi za prenos izraza »oddaljenega branja« iz tradicionalne humanistike v digitalno (Moretti 2018). Tu se nanaša na računalniško podprto analizo besedilnih korpusov; pot si utira med sorodnimi termini, na primer »makroanalizo« (Jockers 2013) ali »algoritemskim kritištvom« (Ramsay 2011).

Prehod od branja z bližine (mikroanalize) k oddaljenemu branju (makroanalizi) predpostavlja spremembo merila (prim. Clarke in Wittenberg 2017). Po analogiji s kartografijo nas večje merilo približa posamičnemu, manjše merilo, ki prikaže razsežnejše ozemlje, pa splošnemu. Pomanjšava merila je ne le količinska sprememba, temveč tudi kakovostna. Z razširitvijo se pogledu odprejo povezave, strukture in lastnosti, ki jih posnetek od blizu, narejen v velikem merilu, ne zazna. Z oddaljenim računalniškim pogledom je mogoče iz »velikega podatkovja« (*big data*) formulirati posplošitve, ki se po svoji nomotetičnosti približajo naravoslovnim zakonom. Takšen argument ima za metaopazovalca znanstvenih domen (Rosenbloom 2013) večjo dokazno moč kot humanistična interpretacija. Slednja je odvisna od tega, kako prepričljiva je v partikularnem kontekstu za skupnost ostalih interpretov.

⁶ O strukturi znanosti, ki se deli na štiri »velike domene« (fizično, življenjsko, družbeno in računsko; *physical, life, social, computing*), ki imajo sebi primerne, a različno »močne« metode in stopajo v razna razmerja, med njimi zlasti interaktivna, gl. Rosenbloom 2013.

⁷ Za opozicijo kanon/arhiv gl. Moretti 2018.

Pri prehodu od preučevanja posameznih književnosti k svetovnemu literarnemu sistemu kot sovisni celoti odpovedo humanistične metode, ki jih lahko uporablja um posameznika. Isto velja za analizo enojezičnega besedilnega korpusa. »Analogna« literarna veda se je pri preučevanju besedil utemeljevala na metodi interpretacije.⁸ V nasprotju z interpretacijo, ki zaradi omejitev spoznavne perspektive razmerje med delom in celoto neizogibno presoja nedoločno in subjektivno, korpusnemu pristopu posamično besedilo ne pomeni nič, če ni postavljeno v *merljiva* razmerja do besedilne množice, kakršne človeški um ne more niti spominsko obvladati. Prav masovni podatki besedilnega korpusa so predmet oddaljenega branja, možnega samo z uporabo računalniške statistike, algoritmov in strojnega učenja.

Interpretacija, katere cilj je razumevanje drugega, je v humanističnih znanostih izvorno mišljena kot intersubjektivno razmerje. Intersubjektivnost se prenese tudi v razmerje med soobstoječimi interpretacijami, ki so pravzaprav izjave različnih spoznavajočih subjektov o istem tekstu. V nasprotju s trdo znanostjo, ki napreduje prek matematičnih, logičnih in/ali empirično-eksperimentalnih ovrb nezadostnih razlag, se humanistika zaradi intersubjektivnosti razvija prek soočanja množice interpretacij. Te so podvržene vsakokratni preizkušnji argumentacijske prepričljivosti in kontekstualne relevantnosti svoje spoznavne perspektive.

V DH se v razmerje med subjekti vplete računalniški medij. V spoznavni proces poseže z algoritmom, ki so ga informatiki predhodno izšolali na drugih besedilih. Literarni strokovnjak nad spoznavnim intervalom, ki je v domeni stroja, nima nadzora, čeprav lahko spreminja nastavitve in vhodne podatke. Statistična analiza korpusa je resda strojna, a spoznavni proces ostaja okvirjen z metodo interpretacije (prim. Rommel 2004). Interpretacija prototipskih besedil je eno od izhodišč za hipoteze računalniške raziskave. Interpretacija, ki spominja na Barthesovo sledenje prepletu kodov v toku pripovedi, je raztresena v tok označevanja digitaliziranih enot v metajeziku računalniškega kodiranja. Oznake za kodiranje tekstov, npr. TEI, predpostavljajo teorijo besedila, ki so jo razvili interpretativni pristopi.⁹ Nekaterih oznak se računalniški program še ni naučil samodejno dodeljevati tekstu (npr. naratološki pojmi, retorične figure, modalnost), zato jih mora v korpus po svojem razumevanju vnašati človek. Tudi pri nenadzorovanem strojnem kategoriziranju – tu ni vnaprej kodiranih podatkov in jih s t. i. rojenjem (*clustering*) sortira računalniški program – se na koncu spet vključi spoznavni subjekt. Računalniško proizvedene rezultate, tabele in grafe je končno treba interpretirati, se pravi razumeti njihov smisel v referenčnih okvirih literarne vede. To je tudi cilj sledečih analiz.

⁸ Interpretacija kot proces intersubjektivnega razumevanja izjavo drugega prevaja v metakoncepte. Tako spoznavni subjekt prek metabesedilnega modeliranja oblikovno-pomenske zgradbe interpretiranega teksta izpiše, kako razume »del in celoto«, kakšen je zanj smisel besedila na ozadju nadredne enote (npr. avtorskega opusa, žanra, obdobja, filozofije, nezavednega).

⁹ Po mnenju Johna Cummingsa (v: Siemens idr. 2008: 458–460) se opira na razmeroma stare predstave novega kritištva in strukturalizma.

2 Dolgo leto 1968, transformacije družbe in literature: hipoteze

Sodelavke in sodelavci raziskovalnega projekta o modernistično-avantgardistični literaturi »dolgega leta 1968« (tj. obdobja svetovnega študentskega gibanja med ok. 1965 in 1975) smo si zastavili vprašanje, kako lahko na novo interpretiramo literarnozgodovinske probleme z računalniško podprto lingvistično-statistično analizo korpusa literarnih besedil, objavljenih v tedanji študentski periodiki (prim. Juvan idr. 2018).¹⁰ Na izhodiščih novejših strokovne literature o obravnavanem obdobju in z nedigitalnimi metodami humanistike, predvsem s strukturalno-semiotičnim in sistemskim pristopom, smo postavili hipotezo, da je bilo študentsko gibanje, marsikje povezano z delavskim, po svojem značaju svetovna revolucija. Kot takšno naj bi težilo k radikalni transformaciji družbe na vseh področjih, od ekonomskega in političnega do jezikovnega, kulturnega, literarnega ali erotičnega. Bližina revolucije je po naših domnevah prek neoavantgard prenovila tudi modernizem, čeprav se je ta prej že institucionaliziral in akademiziral. Z analizo bibliografskih metapodatkov in besedilnega korpusa smo želeli preveriti, ali in v kolikšni meri so se v »dolgem letu 1968« na literarnem polju zgodile naslednje transformacije:

- a) *Študentke in študenti se kot nov politični subjekt uveljavijo v javni sferi:* Ali tedanja študentska generacija prevladuje tudi kot proizvajalec literarnih besedil v obravnavani periodiki? Ali so mladi nosilci prenovne modernizma?
- b) *Emancipiranje žensk:* Je maj '68 kot »totalna revolucija« omajal patriarhalno strukturo slovenskega literarnega polja? Kolikšen je delež avtoric? Kakšno je vrednotenje žensk?
- c) *Prenova zvrstnega repertoarja:* Sta se glede na tedanje teorije o revolucionarnosti teksta prenovila značaj literarnih zvrsti in njihov repertoar? Je poezija, ki tvori jedro slovenskega literarnega kanona, obdržala vodilno vlogo?
- d) *Modernizem se prek neoavantgard radikalizira:* Ali v analizirani slovenski periodiki, ki velja za osrednje gibalno modernizacije v tistem času, modernizem v literaturi res prevladuje? Je zmeren ali radikalen?
- e) *Uveljavljanje vizualnega koda v tisku:* Če šestdeseta leta zaznamuje začetek prevlade vizualnih medijev nad tiskom, bi se to moralo v korpusu nakazati z okrepljenim pomenom vizualnega v strukturi tekstov.
- f) *Dekanonizacija knjižnega jezika:* Šestdeseta leta so kot obdobje »splošnega sekularizma« (During 2021) v modernih književnostih načele primat knjižnega jezika, utemeljenega na narodnem kanonu. Je s tem v literarnih besedilih prišlo do porasta prvin tujih jezikov in neknjižne slovenščine?

¹⁰ Projekt *Maj 68 v literaturi in teoriji: zadnja sezona modernizma v Franciji, Sloveniji in svetu* (2018–2022) je predstavljen na spletni strani <https://maj68.zrc-sazu.si/>.

3 Priprava korpusa

Besedila za korpus smo pridobili iz javno dostopne digitalizirane periodike, in sicer *Tribune* in *Problemov* v Digitalni knjižnici Slovenije (dLib.si). Ta izbira je bila sicer pragmatična (za digitalizacijo bi sami potrebovali dodatna sredstva in čas), a tudi vsebinska. Ta dva časopisa sta namreč tesno vpeta v študentsko gibanje dolgega leta 1968, obenem pa sta bila odprta teoretskim in literarnometriškim inovacijam. Izbrali smo letnike 1968–1972, ko je bilo gibanje na vrhuncu, in za primerjavo dodali leto 1964, ko je oblast ukinila revijo *Perspektive*, predhodnico kritične teorije in modernistične literature *Problemov*.

Ob pomoči zunanjih sodelavcev, študentk in študentov smo v Excelovih tabelah pripravili anotirano bibliografijo vseh izvornih in prevedenih teoretskih ter literarnih besedil. Tabelarni popis poleg standardnih bibliografskih informacij vsebuje še biografske in analitične metapodatke, kot so rojstne letnice, nacionalnost ter družbeni spol avtorjev in avtoric, zvrst besedila, stopnja modernizma in vizualnosti ter odstopanje od slovenskega knjižnega jezika. Za korpus so bila nato iz celotne bibliografije izbrana samo izvirna literarna besedila slovenskih avtorjev. Vsako od 874 izbranih besedil smo po optični črkovni prepoznavi (OCR) s popravki pretvorili v format .docx. Korpus je bil avtomatsko lematiziran in oblikoskladenjsko označen,¹¹ pri označevanju po smernicah TEI so bile upoštevane prilagoditve CLARIN.SI TEI (<https://github.com/clarinsi/TEI-schema>). Vsaki besedilni enoti smo dodali metapodatke, prenesene iz Excelovih tabel. Sodelavec Tomaž Erjavec je ožji korpus iz letnikov 1964, 1968 in 1972 pod oznako Maj68 1.0 (<https://www.clarin.si/repository/xmlui/handle/11356/1430>) postavil v repozitorij Clarin.si, za analize pa je dostopen prek spletnega konkordančnika NoSketch Engine (<http://www.clarin.si/noske/>). V sodelovanju z Andrejem Pančurjem smo besedila korpusa naložili v SIDIH, zbirko digitalne humanistike pod okriljem DARIAH (*The Digital Research Infrastructure for the Arts and Humanities*).

4 Analiza metapodatkov iz anotirane bibliografije

V prvi fazi raziskave smo preverili interpretativno vrednost preštevanja bibliografskih metapodatkov o 874 slovenskih literarnih besedilih korpusa *Maj68*. Ti podatki so bibliografski in interpretativno-analitični. Prvi so pridobljeni iz primarnih virov, naloženih na dLibu, in iz biografskih virov (npr. *Slovenska biografija*), drugi pa so nastali kot moderirana kvalitativna ocena, odvisna od interpretacij posameznega besedila: vnašalci podatkov so se ravnali po enotnih kriterijih določanja oznak, vnosi pa so bili preverjeni v projektni skupini. Izkazalo se je, da je na podlagi preštevanja bibliografskih metapodatkov mogoče razmeroma zanesljivo interpretativno preveriti hipoteze od a) do f).

¹¹ Podrobneje gl. <https://clarin.si/repository/xmlui/handle/11356/1430>.

4.1 Generacijska in družbenospolna struktura pišočių

45 % avtoric in avtorjev literarnih prispevkov v obeh revijah je bilo rojenih v štiridesetih letih 20. stoletja, 11 % pa v petdesetih letih. Močna je generacija, rojena 1930–39 (17 %), k njej sodijo tudi rojeni 1920–29 (5 %). Sklepamo, da gre predvsem za sodelavce *Problemov* iz perspektivovske generacije. V letu 1968 je bila najštevilčnejša generacija (1940–49) študentska, stara med 19 in 28 let; k njej sodi še dekada 1950–59 (samo en avtor je rojen po 1954), tedaj stara med 16 in 18 let (zadnji razredi gimnazije in bruci). *Tribuna* je generacijsko homogena, saj je kar 75 % pišočių rojenih 1940–49, *Probleme* pa si razvidno delita dve generaciji: rojenih 1940–49 je 41 %, tistih 1930–39 in 1920–29 pa skupaj 39 %. V *Problemih-Literatura* se tehtnica prevesi k mladim: 48 % tekstov so napisali rojeni 1940–49, 32 % pa rojeni 1950–54; le 19 % pripada generacijama med 1920 in 1939. Najmočnejša med pišoćimi v obe reviji je torej »baby boom« študentska generacija, rojena v letih 1946–48 (skupno 23 %). Preštevni podatki torej potrjujejo domnevo, da se je študentska generacija uveljavila v javnosti, a to gotovo velja le za analizirani periodični publikaciji. Za posplošitev bi bilo treba analizirati ves tisk tega obdobja.

V analizirani periodiki po številu objavljenih besedil prevladujejo avtorji (81 %), žensk je le 7 %, za preostanek (12 %) pa podatka ni. Podzastopanost avtoric je osupljiva. Razmerje med spoloma se v letih pred dogodki maja '68 in po njih ni znatno spremenilo. Leta 1964 je bilo v *Problemih* in *Tribuni* avtorstvo žensko v 8 % (ob kar 29 % neugotovljenih avtorstev), leta 1972 pa je avtoric malenkost več, 10 % (podatek manjka le za 1 % vnosov). Potrjuje se stališče, da je bilo leto '68 tudi na Slovenskem pretežno »moška zadeva« (Balžalorsky Antić 2021: 317). Po metapodatkih sodeč, študentska revolucija v nasprotju s pričakovanji na poljih kulture in literature (kolikor sta zastopana v analiziranih publikacijah) ni omajala izrazito patriarhalne strukture slovenske družbe.

4.2 Zvrstni repertoar

Daleč najvećji delež glede na število objavljenih besedil ima poezija (68 %), sledita proza (21 %) ter vizualna in konkretna poezija (7 %). Dramatika je tudi za revijalni tisk slabo zastopana (ok. 1 %). Dram je celo manj od nove avantgardne zvrsti, tj. hibridov med strukturalistično teorijo in literaturo (2 %). Delež besedil v zvrsteh, ki so značilno neoavantgardne (vizualna poezija, hibridi), je zaznaven, čeprav je v manjšini (skupaj ok. 10 %).¹²

V letu 1964 je bilo natisnjenih 242 literarnih besedil, razmerje med številoma proznih (114) in pesniških (124) je razmeroma izenačeno, je pa pesniških besedil več (51 %); dramskih je pričakovano le za vzorec (2 %). V letih študentskega

¹² Slika zastopanosti zvrsti se seveda spremeni, če namesto števila besedil upoštevamo število pojavnic v korpusu: s tega vidika predstavlja poezija 13,56 % korpusa, proza 66,09 %, dramatika 18,56 % ter hibrid 1,77 %.

gibanja (1968, 1972) se je zvrstna struktura drastično spremenila, le pri dramatikii delež ostaja tradicionalno nizek (1 %). Pri 630 tekstih je šokanten upad števila proznih besedil (68) v primerjavi s pesniškimi (471). Proznih besedil je le malo več od nove zvrsti, tj. vizualne in konkretne poezije (64). Če to eksperimentalno zvrst zaradi njenega poimenovanja prištejemo k poeziji, se njena prevlada še poveča. Že tako je delež poezije kar 75 %. Pojavita se, v skromnem obsegu (skupaj 3 %), še dve novi hibridni zvrsti, ki mešata poezijo in prozo oziroma literaturo in teorijo. Iz tega lahko zgolj za naš korpus hipotetično sklepamo, da obdobje, ki je težilo k radikalni transformaciji obstoječega, tradicionalnega prvenstva poezije med zvrstmi slovenskega literarnega kanona vsaj po številu besedil ni prizadelo. Morda se je za leta študentskega gibanja, nabita z voljo in afekti, usmerjena v prelomne trenutke sodobnosti in osvobajanje posameznika, zdela primernejša (lirska) poezija od ekstenzivnejše in objektivnejše proze.

Zvrstni repertoar starejše, »perspektivovske« generacije (rojene 1920–40) se po uvajanju novih zvrsti razlikuje od tistega pri mlajši, »tribunaški« generaciji (rojeni 1941–54). Samo za mlade je značilna eksperimentalna zvrst hibridov (skupaj 5 %), tudi delež vizualne in konkretne poezije med njihovimi objavami je precej večji od prejšnje generacije (9 % proti 3 %). Neoavantgardne zvrsti pri mladih skupaj štejejo 14 %, pri starejših pa 6 %. Očitno so po številu besedil glavni nosilci neoavantgardnih zvrsti v našem korpusu mladi, torej generacija, vpeta v študentsko gibanje.

Upoštevač dejstvo, da ima znano žensko avtorstvo le 62 besedil, moško pa več kot desetkrat toliko (707), ne preseneča, da je zvrstni repertoar moških bolj diferenciran. Zanimivo je, da avtorice kažejo večjo naklonjenost prozi kakor avtorji (njen delež je 28 % pri ženskah in 15 % pri moških). Delež enot neoavantgardne vizualne in konkretne poezije je pri ženskah 11 %, pri moških pa 6 %. To kljub neizenačenemu vzorcu nakazuje, da so avtorice, zastopane v našem korpusu, po zvrstnem repertoarju nekoliko bolj nagnjene k eksperimentu od moških.

4.3 Modernizem

Z ugotovitvijo o razmeroma šibki zastopanosti izrazito neoavantgardnih zvrsti (ok. 10 %) se skladajo ocene izrazitosti modernizma v posameznih tekstih. Ker odločitev o tem, kateri literarnoumetniški smeri pripada besedilo, zahteva interpretativno branje od blizu in zamudno preverjanje kriterijev za uvrščanje, smo pri presoji modernizma v korpusu sklenili kompromis med branjem od blizu in daleč. Oblikovali smo kontrolni seznam modernističnih lastnosti (sintaktičnih in semantičnih), ki so ugotovljive tako rekoč na prvi pogled.¹³ Po teh merilih v korpusu prevladuje zmerni modernizem, prepoznaven kot srednja stopnja izrazitosti (47 %). Sledijo teksti, v katerih je stopnja opaznosti modernizma visoka (22 %) ali nizka (22 %). Besedil, v

¹³ To so opustitev interpunkcije, svobodni verz, nestandardna členitev, opaznost vizualnega koda ter specifična tema in forma naslova. Za vsak izpolnjeni kriterij smo besedilu pripisali točko, vsota točk pa je dala oceno njegove modernističnosti (>3 = visoka stopnja modernizma).

katerih modernizma ni bilo opaziti, je le 10 %. Modernizem, čeprav ne radikalen, torej zaznamuje literaturo v *Tribuni* in *Problemih*. Kar 90 % vseh besedil je bilo ocenjenih kot nizko-, srednje- in visokomodernističnih.

Visoko stopnjo izrazitosti modernizma smo prisodili 14 % literarnih besedil generacije, rojene 1920–40, in 27 % besedil generacije, rojene 1941–54. Pri obeh generacijah prevladuje srednja stopnja modernizma (49 % pri starejših in 56 % pri mlajših). Pričakovano je ocenjena izrazitost modernizma večja v mlajši generaciji kakor v starejši.

4.4 Vizualnost

Vizualni kod dinamizira tradicionalni videz tiska v 22 % besedil, kar se ujema z vrednostmi za visoko stopnjo izrazitosti modernizma. V 39 besedilih (dobri 4 % enot celotnega korpusa) je vizualnost osamosvojena – gre za konkretistična dela, ki odstopajo od konvencij jezikovnega teksta. Visoka je stopnja vizualnosti v 8 % besedil, kar se približa vrednosti za delež vizualne in konkretne poezije v zvrstnem repertoarju (7 %). Vizualni eksperiment torej tekstov v obeh periodičnih publikacijah ni prevladujoče zaznamoval, je pa v njih že opazen, tako da obe publikaciji tudi v literaturi odstopata od tradicionalne puščobnosti prejšnjega periodičnega tiska.

4.5 Odkloni od slovenskega knjižnega jezika

Proces dekanonizacije in sekularizacije jezika književnosti, ki se kaže prek vdorov tujih jezikov in zvrsti neknjižne slovenščine, smo ocenjevali z analitičnimi oznakami v tabelarni bibliografiji. Ročno označevanje elementov neknjižne slovenščine v samem korpusu še poteka, tako da bo zanimivo primerjati interpretativno ocenjene vrednosti iz bibliografije z računalniško prešteti označenimi pojavnicami.

Elementi tujih jezikov so pri bralnem preletu besedil redko opaženi. Kar v 93 % besedil jih ni bilo zaznati. Med njimi so razmeroma številčnejša, a po odstotku zanemarljiva besedila z zaznano prisotnostjo angleščine (27) in srbohrvaščine (8). Kadar so tujejezične prvine že prodrle v knjižni jezik, po oceni prevladuje nizka stopnja njihove izrazitosti (5 %).

Leta 1964 so se elementi tujih jezikov pojavljali le na nizki stopnji in še to v 3 % vseh besedil. V času študentskega gibanja tujejezični elementi nastopijo že tudi na srednji in visoki stopnji izrazitosti (oboje po 1 %). Povsem brez zaznanih tujejezičnih elementov je v obdobju študentskega gibanja 92 % besedil.

Po analitičnih metapodatkih so odkloni od standardnega/knjižnega jezika v korpusu prav tako redki. Kar 90 % tekstov je brez opaženih primesi neknjižnih zvrsti slovenščine. Še največ je splošnega pogovornega jezika (9 %), elementi

mladostniškega slenga so bili pri hitrem branju prepoznavni v zgolj 5 tekstih. Odpiranje nestandardni slovenščini je količinsko torej še skromnejše od odpiranja tujim jezikom (13,6 %). Medtem ko v letu 1964 neknjižne jezikovne zvrsti v literaturi obeh publikacij skoraj ne puščajo sledi (brez teh primesi je 94 % besedil), se njihova opaznost v času študentskega gibanja vsaj nekoliko poveča: jezik literature se v našem korpusu prične odpirati splošnemu pogovornemu jeziku, saj elemente tega zaznamo v 10 % analiziranih enot, torej še enkrat toliko kot v letu 1964. A knjižni jezik v literaturi revolucionarnega časa še ni detroniziran.

Tudi podatke o tujejezičnih elementih in prisotnosti neknjižne slovenščine bo mogoče zanesljiveje presojati in primerjati glede na podatke, ki jih bomo izluščili po označevanju prvin tujih jezikov in neknjižne slovenščine v korpusu. Kvantitativna analiza metapodatkov, predstavljena v tem delu prispevka, je privedla do rezultatov, s katerimi je bilo hipoteze raziskave mogoče preveriti na gradivu korpusa *Maj68*. Zbrani in prešteti metapodatki iz bibliografije imajo torej vlogo podpore literarnozgodovinski in literarnosociološki interpretaciji korpusa. Toda ti rezultati, čeprav računalniško preštevni, so že sami odvisni od predhodnih interpretacij posamičnih besedil. Vsakemu besedilu je bilo treba pri pripravi anotirane bibliografije na podlagi (hitrega) branja najprej izbrati primerne metaoznake za določitev njegove zvrsti, modernizma in decentriranja knjižne slovenščine. Tu smo torej še na prehodu med analogno in digitalno humanistiko, med branjem od blizu in oddaljenim branjem. Morali smo narediti korak naprej, da bi ugotovili, kaj k interpretaciji literature iz let študentskega gibanja lahko prispeva digitalna paradigma s korpusno analizo.

5 Korpusna analiza besedil

Korpusna analiza literarnih besedil metodološko izhaja iz korpusnega jezikoslovja. McEnery in Wilson (2005: 1) postavljata začetek korpusne lingvistike v petdeseta leta 20. stoletja, in čeprav je v šestdesetih in sedemdesetih letih ta pristop naletel na precejšnje neodobranje, je pozneje postal eno obsežnejših področij jezikoslovnih raziskav, tako da ima tudi na Slovenskem že daljšo tradicijo. Slovensko korpusno jezikoslovje je usmerjeno predvsem v leksikografijo in opis ožje jezikoslovnih značilnosti (prim. Balažic Bulc in Gorjanc 2015), pa tudi v semantiko, pragmatiko (prim. Verdonik 2020; Šorli 2020), prevodoslovje (prim. Vintar 2017: 7) in analizo diskurza (prim. Fišer in Kalin Golob 2019; Gorjanc in Fišer 2020). V tovrstnih raziskavah so se oblikovali koncepti izdelave korpusa in orodja za ustrezno označevanje besedil v korpusu, pa tudi podlage za korpusne analize literature.

Naj poudarimo, da rezultatov deskriptivnega pristopa, predstavljenega v nadaljevanju, ni mogoče posplošiti na celotno literarno produkcijo obravnavanega obdobja, saj je naš korpus razmeroma majhen. Analize specializiranega literarnega korpusa *Maj68*, v katerem so količinska razmerja med zvrstmi besedil neuravnotežena, tako po številu besedil kot njihovem obsegu, so bile izvedene s spletnim orodjem noSketchEngine

(<https://nlp.fi.muni.cz/trac/noske>) in LancsBox.¹⁴ S tema prosto dostopnima orodjema smo opravili preliminarne leksiko-semantične in korpusnostilistične analize (najpogostejše besede, ključne besede, kolokacije), ki bi lahko nakazovale prevladujoče tematike in vrednotenjske pomene izbranih leksikalnih enot.

5.1 Besedni sezname in pogostnostna razmerja v korpusu

Z izdelavo besednih seznamov, posebej najpogostejših in ključnih besed, dobimo vertikalni vpogled v neko zbirko besedil. Z leksiko-semantičnimi analizami na podlagi izbranih seznamov smo ugotavljali, v kakšnih kategorialnih in pomenskih razmerjih ter kontekstih se pojavljajo leme literarnega korpusa *Maj68*.

Najprej smo se posvetili razmerjem deležev treh polnopomenskih besednih vrst med tremi literarnimi zvrstmi: poezijo, prozo in dramatiko. V korpusu je za poezijo v primerjavi s prozo in dramo značilna izrazito nadpovprečna raba samostalnikov (140,5 %);¹⁵ v prozi in dramatiki je relativni delež samostalnikov zelo blizu povprečne rabe v celotnem korpusu (z odstopanji za 2 % navzdol in navzgor). Zelo podobno velja za pridevnik: nadpovprečna raba v poeziji (za 26 %) ter manjše odstopanje od povprečja navzdol pri prozi (za 2,6 %) in nekoliko večje pri drami (15,3 %). Nasprotno pa je delež glagola v poeziji nekoliko pod korpusnim povprečjem (za 11 %), medtem ko je v prozi in dramatiki le nekoliko nad povprečjem (za 2,5 oz. 2,7 %). Izrazitejša odstopanja poezije glede na rabo samostalnikov, pridevnikov in glagolov v primerjavi z drugima zvrstema v korpusu kažejo na večjo statičnost in manjšo dogajalnost ter bolj samostalniški slog. Večjo prisotnost pridevnikov v poeziji glede na preostali zvrsti bi lahko navezali na izrazitejše niansiranje besedilnega sveta z lastnostmi, ki mu jih prisoja subjekt besedila. Kvantitativna distribucija besednih vrst vodi k interpretaciji, da se opaženi prevladi modernizma navkljub ohranjajo nekatere konvencionalne razlike med tremi literarnimi zvrstmi (subjektivnost – objektivnost, opisnost – pripovednost).¹⁶

Zaznavne so besednovrstne razlike med teksti po spolu: podpovprečna je raba samostalnika pri avtoricah (80,9 %) v primerjavi s povprečno rabo pri moških avtorjih. V našem korpusu so količinska razmerja besedil sicer premajhna, da bi ob tem rezultatu lahko tvegali dokončne ugotovitve, a pomenljivo je, da se rezultat ujema z ugotovitvijo iz analize metapodatkov, po kateri avtorice v korpusu *Maj68* v primerjavi z avtorji izraziteje preferirajo prozo.

¹⁴ Obe orodji omogočata standardne funkcije iskanja: po lemah, besednih oblikah in besednih zvezah ter v prvem tudi z jezikom CQL.

¹⁵ Rezultati so predstavljeni glede na relativni delež konkretnih pojavnih v celotnem korpusu. 100 % predstavlja povprečni delež na ravni celotnega korpusa, odstopanja višje ali nižje pomenijo odstopanja od povprečja v eno ali drugo smer.

¹⁶ To domnevo bi bilo zanimivo preveriti v primerjavi z večjim referenčnim korpusom.

5.3 Ključne besede in sledi političnega konteksta

Ključna beseda je v korpusni lingvistiki beseda, ki se v besedilih pojavlja nesorazmerno pogosto glede na neki drug korpus. Izbrano analitično orodje primerja frekvenčni seznam referenčnega korpusa¹⁸ s seznamom korpusa, ki ga obravnavamo. Tako dobimo seznam ključnih besed, od katerega pričakujemo vpogled v glavne teme našega korpusa, čeprav so v njem zaradi literarnosti tematike manj razvidne kot denimo v korpusih terminoloških besedil.

Spodnji seznam, zamišljen kot izhodišče raziskave prevladujočih tematik korpusa *Maj68*, prikazuje njegove ključne besede v razmerju do leposlovnega podkorpusa referenčnega korpusa *Gigafida v2.0*. Tabela 1 prikazuje prvih 20 ključnih besed, tj. lem, med katerimi izstopajo zaimki (osebni, kazalni in svojilni). Zlasti povedna je ključnost osebnih in kazalnih zaimkov v prvi in drugi osebi (*jaz, ti, tvoj, ta, tu*).¹⁹ Ta kaže na visoko stopnjo dialoščnosti, morda zaradi premega govora, a bi bilo treba to še raziskati. Visoko na seznamu je *človek*, v dejanskih kontekstih pogosto uresničen tudi v množinski obliki *ljudje*. Lastna imena na seznamu so deloma vezana na didaskalije, opozarjajo pa tudi na posebnost tega korpusa, ki bi zahtevala dodatno preučitev.

Stat	Freq	Freq_ref	Score
ti	4257	142475	1.200
ne	7975	376798	1.100
ves	4463	207714	1.100
jaz	7547	403925	1.100
že	2725	131253	1.100
človek	1543	64014	1.100
Dimitrij	431	340	1.100
tu	767	20462	1.100
tvoj	805	22886	1.100
brez	1009	36658	1.000
ta	8437	469722	1.000
naš	794	25303	1.000
samo	1664	76983	1.000
vesti	353	1204	1.000

¹⁸ To je lahko katerikoli izbrani korpus, običajno pa gre za večji, reprezentativni korpus, v našem primeru *Gigafida v2.0* (1990–2018), www.gigafida.net.

¹⁹ Vključuje tudi vikalne oblike (*vi, vam* itd.).

Andrej	350	1173	1.000
osredek	296	1	1.000
Milena	290	186	1.000
kaj	2567	134246	1.000
preveliko	267	3	1.000
zdaj	1478	71534	1.000

Tabela 1: Ključne besede po noSketchEngine.

Ker gre za korpus literarnih besedil, ustvarjenih v politično burnem zgodovinskem obdobju, nas preseneča, da med ključnimi ni besed *revolucija*, *upor*, *oblast*, *svoboda* in *protest*. Da bi izluščili kaj o političnosti literature iz korpusa *Maj68*, smo analizirali njihovo relativno pogostnost in kolokacije. Vse navedene besede so glede na majhen obseg korpusa relativno pogoste, vendar značilno manj razpršene. *Oblast* (85) denimo potrjuje predpostavko o politični motiviranosti, a je nesorazmerno pogosta v drami *Dolina neštetihi radosti*. *Revolucija* (s kolokatorji kulturna, identiteta, forma, revolucija) ima 84 zadetkov, *upor* (upor, pristati) 23 v 11 besedilih, *protest* le 3 v dveh besedilih in *svoboda* (smrt, živeti), sicer najenakomerneje zastopana, 82 zadetkov:

ohromelosti, znova je tu neznan občutek svobode , nobene teže ne čutim. ♪ Liza je bila že v pos Kavčič Konjar Dokler ♪ OD UTESNJENOSTI K SVOBODI Pogledi na integracijske procese ali Stolp v krogle kot kamen. ♪ V Ameriki imajo vsaj kip Svobode , drugod pa še tega ne! ♪ Če bi bil Adam turške k telesu nazaj, vsak mora najti sam svojo svobodo . Zakon ti je ne bo dal. Misliš, ker se ne ženiš t . Nikoli mi ne more nihče vzeti teh spominov, te svobode ..." Vedel sem, da govori resnico. Morda sem si že predolgo svobodni, zato pozabljate, kako se svoboda otrese strahu, ko niste plen skrbi sivih ličnic pred prestolom nesmisla in zapojem pesem o svobodi ; to pesem ste skoraj pozabili, ker se motorji v tistem času, ko so podtalno delovali za pojem svobode in veličino premagovanja strahu ... ♪ O zdaj, v da me kar preseneti, ko predsednik odbora zine: Svoboda zlata, kako si bila bogata žrtev ... ♪ Hitro sem , svetovnih. Tam, kjer je za vas bil morda višek svobode , se za nas vse spreminja v pretesno ječo. Vam v soncu in v snegu, do smrti opominja: ko sem za svobodo pasel po zelnikih svoje kože, krogle in uši, si rodov rodovom kazal pot iz suženjstva v svobodo ! Staroslavna vas drhti od ponosa. Svet bo , da si ravnal prav? Si vsaj odkril, kaj je svoboda ? ♪ - Preteklo je komaj nekaj dni ... Zaupaj mi) : Jutri bomo svet spet krojili po naši meri! Svoboda - po naši meri! Enakost - po naši meri! Pravica za novi red z bojem Slovencev za življenjsko svobodo . Jaz pa sem poudaril misel, da je bramba še napisati nekaj parol, "Dol z diktaturo" in " Svoboda univerzi". Vincent že ima fotokopije - zrl je ji damo popolnoma nagonsko naslov smrt fašizmu svoboda mavgliju ♪ na namizi je cvetje prepil ob vazi njimi ni bilo nikogar, združevala jih je ravno svoboda , samosvojost. V lastni božanskosti so si bili
--

Med najpogostejšimi kolokatorji *svobode* se znajde tudi *svoboda*, tj. jedro samo; gre za ponavljanje, ki za neliterarna besedila ni značilno:

Udarnik. Kino Svoboda. Kino Svoboda . Živela svoboda ! ♪ Na vogalu Nebotičnika in potem za toliko osvojenje – zakaj brez socialne in politične svobode je nemogoča kulturna svoboda Dokler bo in politične svobode je nemogoča kulturna svoboda Dokler bo ljudstvo suženj družbe, suženj tega v času in prostoru, nekaj osamljenega v svobodi z vsemi, v svobodi strojev in v suženjski
--

Majhnost korpusa ovira interpretacijo, ki stremi k univerzalnim zaključkom, a omogoča obvladljivo vzvratno navezavo korpusnih ugotovitev na metapodatke. Tako denimo ugotovimo, da se *revolucija* pojavi največkrat v prozi, a je relativno najpogostejša v hibridu, mladostniškem slengu, v *Problemih*, in sicer pri Rastku Močniku. *Oblast* je glede na metaoznake relativno najpogostejša v dramatiki, splošnem pogovornem jeziku, v *Tribuni* in pri Marku Švabiću; *upor* je relativno najpogostejši v hibridu, nato v poeziji, in sicer v knjižni slovenščini in v *Problemih*, daleč najbolj pri Močniku; *svoboda* nastopa relativno najpogosteje v dramatiki, splošnem pogovornem jeziku, v *Tribuni* in pri Žarku Petanu. Vse iskane besede se pojavljajo relativno najpogosteje prav v letu 1968.

Po drugi strani seznam ključnih besed zajema besede *človek, dan, noč, moči, zemlja, telo, črn, gledati, glas, bel, čas, noga, pot, beseda, nebo, cesta, tovariš, luč, piti, zelen, okno, praviti, pentlja, iti, kamen, ostati, živeti, stvar, prostor, hip, obraz, ljudstvo, veter, reka, senca, vojščak, šef, ogenj, žena, smrt, prazen, roka, veteran, kri, lev*. Če vzamemo za vzorec prvih 100 ključnih besed, v njem razločimo dve skupini: a) besede, ki konotirajo (mehko) poetičnost, npr. *noč, veter, senca*, in b) besede, ki zelo posredno, z ostrino nakazujejo politični boj in željo po svobodi, npr. *tovariš, ljudstvo, vojščak, smrt, kri*. Prvo pomensko skupino lahko potrdimo s konkordancami oz. konteksti:

volčje mesečine, psov podarjenih za praznik v nočeh ciganskih konj in kraj v rokah, očeh in ustih	noči	z nočjo metafora laži v resnici zamujenega
in vrbovja v kopnini jutra, v meglah pogovorov	noči	metafora laži v resnici zamujenega koraka ob
v kopnini jutra, v meglah pogovorov	noči	in strah v bokih predal darovanju. ¶ Vonj po
predmestja. Kupil sem utrujeno nespečnost	noči	ne bo nič. - Saj veš, nič tistega no.- ¶ -Dragi moji
je bilo čutiti v glasu, -njuna poročna	noč	našel stanovanje, kjer se edino lahko pripravim
neznanim kraju. Kam naj se obrne, da bo še pred	nočjo	, a zdaj vse to ni prav nič pomembno. Ne zanima me
pri nas, kakšni so bili poslednji dnevi, zadnja	noč	, vse življenje je bilo zaobjeto v enem samem
utrujenost, pozabljene so bile muke pretekle	noči	se zbiramo v šesti kleti "Pri črni večji".
nekaj ukreniti. Pridružiti se nam. Vsako	noč	je bila topla. Stisnil sem se ob skladovnico drv
da nekoga čaka. ¶ Nisem ga hotel motiti. Vetrovna	noč	. ¶ *** ¶ Drugi dan sva s starcem priredila majhno
, ki je v skrbeh za naju stal pred bajto in klical v	noč	je že. Starec je od utrujenosti zaspal.
o domu, otrocih in delu, ki naju čaka.	Noč	je. Ničesar ne pričakujeva. Vse je tako
ob reki, mimo slapov in še naprej. Topla poletna	noč	iz svojega brloga v majhni, zapuščeni kotlubi,
kmetije ob zavoju reke, pa se nisem ganil vse do	noči	se čisti in raste iz skalnih gmot, iz reke in jezu
... ¶ *** ¶ Jutro je sveže in temno. V naplavinah	noči	razburljivih samogovorov, obkrožen od
, onečaščen, ponižan, razkrinkan, prepoten po	noči	: To je živ spomin na mojega očeta, ki se ga od
ko mračen trak brajde, razpotegnjen tja v rosno	noč	sem sanjal o njem (izrinil je modro svetlobo),
z uvihnjenimi robovi motrim. ¶ Moj oče; še to	noč	, zdaj ti s svojo podobo izrini modro svetlobo;
telesu neke ženske z dobrimi nogami zapravljaj	noč	, venomer trajajoče, takšne, ki kar ne more
pregrada med mrzlo resničnostjo dolge deževne	noči	

Toda skupina b) se izkaže za manj utemeljeno. Besede *vojščak, šef, veteran* se namreč pojavljajo zlasti kot lastna imena, običajno v le enem literarnem delu, zato jih v resnici ni mogoče zares obravnavati kot ključne.

5.4 Medbesedna povezovalnost in kolokacije: k interpretaciji razlik med zvrstmi in družbenima spoloma

Kolokacija navadno zajema pogoste in statistično pomembne sopojavitve besed, ne glede na obstoj specifičnih pomenskih razmerij med njimi. Na primeru besede *dan*, najpogostejšega samostalnika v poeziji, bomo vzeli pod drobnogled najprej njegove najbolj tipične²⁰ prve leve pridevniškobesedne kolokatorje po literarnih zvrsteh.²¹

Za leve kolokatorje besede *dan* se pokaže, da se v poeziji poleg najpogostejšega zaimka *vsak*, ki se najpogosteje pojavi v vseh treh zvrsteh (vrednost logDice za poezijo 10,497; za prozo 10,131 in dramo 9,975), pojavljajo zlasti vrstilni števniki, ki označujejo zaporedje: najpogosteje *sedmi*, ki je na drugem mestu²² po tipičnosti kolokacije za poezijo, nato so znotraj petnajstih najbolj tipičnih še (navedeni zaporedno po seznamu) *šesti*, *prvi*, *četrti*, *tretji*, *osmi*, *peti*, *deseti*. Manj raznolike so pridevniške besede za trajanje (*trije* – na 7. mestu, *nekaj* – na 18. mestu), dokaj pogosta je nedoločnost (*nek* in *nekaj* – na 3. in 4. mestu po tipičnosti kolokacije). Šele v spodnjem delu seznama pridevniških besed k samostalniku *dan* se pojavita lastnostna pridevnika *črn* (na 19. mestu) in *bel* (na 28. mestu). Na mestih med njima se zvrstijo pridevniki, ki s samostalnikom *dan* običajno tvorijo stalne besedne zveze ali frazeme, na primer *rojsten*, *gospodov*, *deževen*. Nasprotno v prozi časovna razmerja med dnevi opredeljujejo zlasti *prejšnji* (na 3. mestu po tipičnosti), *naslednji* (5. mesto), *prvi* (11. mesto), *današnji* (12. mesto), nižje še *nov* (22. mesto) in *zadnji* (26. mesto). Trajanje označujejo pridevniške besede med 4. in 13. mestom po tipičnosti kolokacij: *cel*, *ves*, *nekaj* ali *trije*. Med lastnostnimi pridevniki po frekventnosti (na 7. in 9. mestu) nekoliko izstopata *dober* in *lep*, ki sta po večini del stalne besedne zveze (*dober dan* ali *lepega dne*). Nižje, od 14. mesta dalje so na seznamu v prozi *rojsten*, *jesenski*, *bel*, *deževen*, *deloven*, *gospodov*, *soden*, *sleheren*, *vroč* itd. Iz teh razlik med literarnimi zvrstmi je v korpusu prepoznavna konkretnjša časovna umeščenost dogajanja v proznih besedilih v primerjavi s poezijo. V dramskih besedilih, kjer je ravno tako pričakovana določnejša časovna umeščenost in opredelitev trajanja, je večja podobnost s prozo. Izrazitejša raba vrstilnih števnikov v poeziji bi lahko bila posledica sloga oz. rabe pesniških figur stopnjevanja in ponavljanja (pozornost zbuja tudi pojavljanje pravljličnih števil sedem in tri). Prav raba pesniških figur kot elementa sloga bi lahko bila tudi vzrok za nadpovprečno pojavnost besede *dan* v poeziji glede na celotni korpus.

Kolokativnost v širokem pomenu medbesedne povezovalnosti, ki vključuje tudi semantično bližino in nekatere vidike pragmatičnega pomena, smo izbrali tudi za izhodišče preliminarne raziskave družbenih pomenov in vloge žensk, kakor se kažejo

²⁰ Pri izračunu trdnosti povezav ni bila upoštevana absolutna frekvenca, ampak statistično merilo log-Dice.

²¹ *Dan* je visoko tudi na seznamu najpogostejših besed v prozi (med polnopomenskimi besedami na 21. mestu) in dramatiki (na 18. mestu).

²² Med tipičnimi kolokatorji prevladujejo pridevniške besede, na seznamu pa se pojavijo tudi druge besedne vrste, pa tudi *nekaj* ločil (skupno je na primer med prvimi 30 najbolj tipičnimi kolokatorji pri poeziji in prozi 16 % »nepridevniških« levih kolokatorjev in pri dramatiki 23 %).

v korpusu. Izdelali smo seznam 10 najpogostejših kolokatorjev za lemi »moški« in »ženska«, v razponu 5 levo in 5 desno od jedra, oba po pogostnostnem statističnem merilu (01 – Frequency, LancsBox).

Pod temi iskalnimi pogoji je seznam desetih kolokatorjev presenetljivo prekriven. Obe lemi namreč kolocirata oz. se povezujeta z istimi besedami, in sicer *biti*, *on*, *jaz*, *imeti*, *ves*, *ženska*. Sprememba v statistični meri (v našem primeru logDice namesto Frequency) privede do večje diferenciranosti že med prvimi 10 kolokatorji, kot vidimo v tabelah 2 in 3 (razpon kolokacij 1 levo in 1 desno od jedra). Seznama kažeta najtesnejše in zato tudi najbolj nedvoumne pomenske povezave z lemama *ženska* in *moški*.

ID	Položaj	Kolokator	LogDice	Pogostost kol.	Pogostost korp.
1	L	nek	8.22126	10	669
2	L	čudovit	8.02558	4	74
3	L	mlad	7.84012	5	286
4	R	ki	7.79160	37	5043
5	L	grd	7.75207	3	27
6	L	lep	7.70998	6	510
7	L	zame	7.37429	4	361
8	L	edin	7.360	3	169
9	L	star	7.336	5	585
10	L	moški	7.334	3	180

Tabela 2: Kolokatorji za lemo *ženska* v korpusu *Maj68*.

ID	Položaj	Kolokator	LogDice	Pogostost kol.	Pogostost korp.
1	L	mlad	7,730	3	286
2	L	visok	7,642	3	315
3	L	star	7,426	4	585
4	M	dva	7,345	4	629
5	R	ženska	7,307	3	444
6	L	drug	6,336	5	1851
7	R	ki	6,110	11	5043
8	L	mlad	7,730	3	286
9	L	visok	7,642	3	315

Tabela 3: Kolokatorji za lemo *moški* v korpusu *Maj68*.

Skupno seznamom prvih 10 kolokatorjev, ne glede na vrsto statističnega merila z izjemo frekvenčnega, je pri lemi *ženska* izstopanje besed, povezanih z vrednotenjem videza in starosti ženske (*lepa, mlada, grda, stara, čudovita, telo, imeti* itd.):

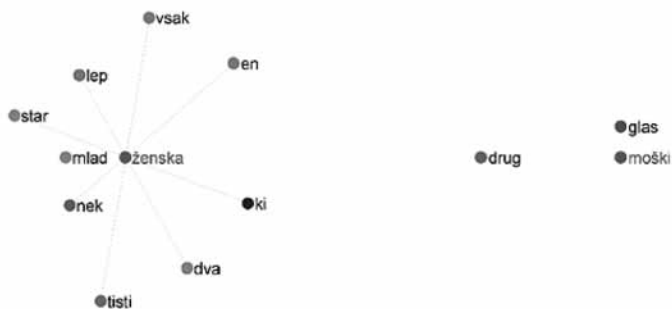
na cesti. Pod roko se ga je držala visoka, *lepa* ženska . ♀ Vsa njena toplota se je spremenila v en sar bisere jih grabijo, ubogi ideali na vratu *lepih* žensk - kitare odmevajo, sirena tuli, prispeli smo, junake zato, da jih lahko rešujejo. ♀ Nekatero ženske so premalo *lepe* , da bi lahko varale in preveč . Zavedaj se, da so noge najobčutljivejši del ženske , zlasti še tako *lepe* kot Naše Stvarnosti. Ob imeli najudobnejša stanovanja in *najlepše* ženske . Miku knjige se ne more upreti niti še tako dobro živi, nič slabega se mu ne godi: Tudi *lepih* žensk se naveličaš, pravi, ko stoji pred njim , pesniške idile, občuti gibanje v svoji duši. — Ženska mora biti *lepa* , močna, zdrava, roditi mora kot moškega, sir, še nisem videl *lepše* ženske . Toda kot vrtnar mislim, da je zelo prestrašen nakopal na glavo. ♀ Cola: Colombina je *najlepša* ženska v mestu in pika. ♀ Stoppino: Vejica, caro, vejica

Na videz analogno se tudi med kolokatorji za *moški* pojavljajo pridevniki, ki označujejo videz in telo (*mlad, visok, star, močen*). Pri raziskavi kontekstov rabe pa je mogoče ugotoviti, da gre večinoma za pripis lastnosti, ki nima izrazitega vrednotenjskega namena v smislu družbenih pričakovanj:

, ko je zagledal elegantnega *mladega* moškega , v sivi obleki, ki se je sprehajal ob vrtni ogovoriti, ko je izza vogala stopil *mlajši* moški in jo objel. Oba sta se obrnila v smer, od koder je Sedaj sedi tu na divanu objokana pred *mladim* moškim , ki jo je premamil z besedami in z na videz toplim

bogatega lepote, dovolj obložene z *močnimi* moškimi pogledi, zanesljivostjo in varnostjo krepih je lahko reči, Pandolfo! Inteligentem in *močan* moški si. Mene pa mučijo dvomi. Ne vem, če je prav, kar bodo še zvižgali škoric? ♀ Rad imam telesa, *močna* moška telesa, ki se tepejo za vsakdanji kruh,

Kljub nekaterim podobnostim je med nizi kolokatorjev za *ženska* in *moški* razvidno razhajanje. Lema *ženska* ima v različnih statističnih merah in v povprečju prepoznanih več (trdnih) kolokatorjev kakor *moški*, kot je razvidno iz kolokacijske mreže za obe lemi spodaj.



Slika 2: Kolokacijska mreža po LancsBox.

To statistično dejstvo se sklada z ugotovitvijo o podzastopnosti žensk kot avtoric besedil v korpusu, podpira pa tudi opažanje semantično-pragmatične analize, da se v korpusu *Maj68* o ženskah govori izrazito stereotipno (gl. Sliko 2).

6 Sklep

Običajno je korpusna analiza orodje, ki podpira interpretacijo literature, in tako jo v glavnem razumemo tudi mi. S preštevanjem bibliografsko-analitičnih metapodatkov o korpusu *Maj68* smo ugotovili nadpovprečno avtorsko zastopanost mlade generacije, podzastopanost žensk, prevlado poezije, dominantnost zmernega modernizma, pojavljanje novih eksperimentalnih zvrsti in vizualnosti ter začetek dekanonizacije knjižnega jezika. Naše interpretacije, podprte s preštevniimi metapodatki, pa bo pred posplošitvijo na celotno obdobje treba preveriti na širših primerjalnih korpusih.

Obenem korpusni pristop kot metodo oddaljenega branja, ki s hkratnim vpogledom v množico besedil pokaže tudi nepredvidljive vzorce, s pridom uporabljamo tudi pri preizpraševanju ustaljenih pojmovanj in snovanju novih interpretacij. Kieran O'Halloran (v Hoover idr. 2014: 146) denimo ugotavlja, da je mogoče strategijo korpusne analize – tj. prepoznavanje skladenjskih, leksikalnih, semantičnih in besedilnih vzorcev, med katere sodijo kolokacije in frazeološke enote – uporabiti za mobilizacijo (nove) interpretacije.

Naše prve ugotovitve o povezavah med takšnimi vzorci in družbenim pomenom v korpusu *Maj68* se opirajo na sezname najpogostejših in ključnih besed ter na raziskovanje njihovega ožjega besedilnega okolja. Če je po eni strani smisel takšne analize, da podpremo interpretacijo pomena, ki izhaja iz intuicije o smislu nenavadnosti ali pomembnosti besedilnih segmentov, je po drugi strani še pomembnejše odkrivanje kolokatorjev (pa tudi drugih vzorcev), ki bi jih bilo z branjem od blizu nemogoče predvideti, a lahko pripomorejo k »mobilizaciji« pomena. Interpretacija je nepogrešljiva še posebej pri analizi vrednotenjskega pomena. Naša analiza konkordanc in besedilne rabe lem *ženska* in *moški* pokaže, kako majhen in pretežno moški korpus govori o ženskah in obratno. Preliminarne raziskave obetajo, da bo korpus *Maj68* po nadgradnji omogočil zanesljivejše interpretacije družbenih pomenov v literarnih besedilih in izsleditev reprezentacij političnega konteksta, na katerega so se besedila odzvala in vanj posegla.

Na primeru primerjalne analize levih pridevniških kolokatorjev besede *dan* v korpusu *Maj68* se je izkazalo, da je beseda v prozi in dramatikii povezana s časovno umeščenostjo dogajanja, v poeziji pa s slogom. Sodeč po pogostosti samostalnika, pridevnika in glagola v korpusu, so literarne zvrsti tudi v dobi svojih radikalnih modernističnih transformacij ohranile nekatere tradicionalne razlikovalne poteze, npr. razliko med opisnostjo in dogajalnostjo.

Literatura

Allington, Daniel, Brouillete, Sarah, in Golumbia, David, 2016: Neoliberal Tools (and Archives): A Political History of Digital Humanities. *Los Angeles Review of Books*. 1. maj 2016. <<https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/>>. (Dostop 11. 11. 2021.)

Allison, Sarah, Gemma, Marissa, Heuser, Ryan, Moretti, Franco, Tevel, Amir, Yamboliev, Irena, 2013: *Style at the Scale of the Sentence*. Pamphlets of the Stanford Literary Lab 5. Stanford, Calif: Stanford University Press. <<https://eems.stanford.edu/workspace/druid:ty696mp4835/LiteraryLabPamphlet5.pdf>>. (Dostop 11. 11. 2021.)

Allison, Sarah, Heuser, Ryan, Jockers, Matthew Lee, Moretti, Franco, Witmore, Michael, 2011: *Quantitative Formalism: An Experiment*. Pamphlets of the Stanford Literary Lab 1. Stanford, Calif: Stanford University Press.

Balažič Bulc, Tatjana, Gorjanc, Vojko, 2015: The position of connectors in Slovene and Croatian student academic writing: a corpus-base approach. Starc, Sonja, Jones, Carys, Maiorani, Arianna (ur.): *Meaning making in text: multimodal and multilingual functional perspectives*. Basingstoke; New York: Palgrave Macmillan. 51–71.

Balžalorsky Antić, Varja, 2021: Maj '68 in vznik *écriture féminine*: Francoski center in slovenska periferija. Marko Juvan (ur.): *Med majem '68 in novembrom '89*. Ljubljana: ZRC SAZU, Založba ZRC. 317–336. <<https://doi.org/10.3986/9789610505587>>. (Dostop 11. 11. 2021.)

Clarke, Michael Tavel, in Wittenberg, David, 2017: Introduction. Michael Tavel Clarke in David Wittenberg (ur.): *Scale in Literature and Culture*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-64242-0_1>. (Dostop 11. 11. 2021.)

Da, Nan Z., 2019: The Computational Case against Computational Literary Studies. *Critical Inquiry* 45/3. 601–39. <<https://doi.org/10.1086/702594>>. (Dostop 11. 11. 2021.)

Dovič, Marijan, 2004: *Sistemske in empirične obravnave literature*. Ljubljana: Založba ZRC, ZRC SAZU.

During, Simon, 2021: 1968: Splošnemu sekularizmu naproti. Prev. Andrej Zavrl. Marko Juvan (ur.): *Med majem '68 in novembrom '89*. Ljubljana: ZRC SAZU, Založba ZRC. 33–46. <<https://doi.org/10.3986/9789610505587>>. (Dostop 11. 11. 2021.)

Fišer, Darja, in Kalin Golob, Monika, 2019: Corporate communication on Twitter in Slovenia: a corpus analysis. *Prispevki za novejšo zgodovino = Contributions to the contemporary history = Contributions à l'histoire contemporaine = Beiträge zur Zeitgeschichte* 59/1. 46–69.

Gorjanc, Vojko, Fišer, Darja, 2018: Twitter in razmerja moči: diskurzna analiza kampanj ob referendumu za izenačitev zakonskih zvez v Sloveniji. *Slavistična revija* 66/4. 473–495.

Hladnik, Miran, 1993: Preštevna določila slovenske povesti. *Slavistična revija* 41/1. 65–75.

Hladnik, Miran, 1995: Količinske in empirične raziskave literature. *Slavistična revija* 43/3. 319–340.

Hladnik, Miran, 2012. Digitalna humanistika na Slovenskem. *Air beletrina*. December 2012. <<http://www.airbeletrina.si/clanek/digitalna-humanistika-na-slovenskem>>. (Dostop 11. 11. 2021.)

Hoover, David L., Culpeper, Jonathan, in O'Halloran, Kieran, 2014: *Digital literary studies: corpus approaches to poetry, prose, and drama*. New York: Routledge.

- Jannidis, Fotis, Kohle, Hubertus, in Rehbein, Malte (ur.), 2017: *Digital Humanities: eine Einführung*. Stuttgart: J. B. Metzler Verlag.
- Jockers, Matthew Lee, 2013: *Macroanalysis: digital methods and literary history*. Urbana: University of Illinois Press.
- Juvan, Marko, Dokler, Joahim, 2015: Towards a GIS analysis of literary cultures: the making of the Slovenian ethnoscape through literature. *International Journal of Humanities and Arts Computing* 9/2. 196–218.
- Juvan, Marko, Habjan, Jernej, idr., 2018: *Maj 68 v literaturi in teoriji: Zadnja sezona modernizma v Franciji, Sloveniji in svetu*. Splet. <<https://maj68.zrc-sazu.si/>>. (Dostop 11. 11. 2021.)
- McEnery, Tom, in Wilson, Andrew, 2005: *Corpus Linguistics. An introduction*. Edinburgh: Edinburgh University Press.
- Moretti, Franco, 2011: *Grafi, zemljevidi, drevesa in drugi spisi o svetovni literaturi*. Izbor, prevod, spremna beseda Jernej Habjan. Ljubljana: Studia humanitatis.
- Moretti, Franco, 2018: *Canon/Archive: Studies in quantitative formalism from the Stanford literary lab*. New York: N+1 Foundation.
- Ogrin, Matija (ur.), 2005: *Znanstvene izdaje in elektronski medij: razprave*. Ljubljana: Založba ZRC, ZRC SAZU.
- Perenič, Urška, 2010: *Empirično-sistemsko raziskovanje literature: konceptualne podlage, teoretski modeli in uporabni primeri*. Ljubljana: Zveza društev Slavistično društvo Slovenije.
- Pretnar, Tone, 1997: *Iz zgodovine slovenskega verznege oblikovanja*. Ljubljana: Znanstveni inštitut Filozofske fakultete.
- Ramsay, Stephen, 2011: *Reading machines: toward an algorithmic criticism*. Urbana: University of Illinois Press.
- Rommel, Thomas, 2004: Literary Studies. Schreibman, Susan, Siemens, Raymond George, Unsworth, John (ur.): *A Companion to Digital Humanities*. Malden, MA: Blackwell Pub. 88–96.
- Rosenbloom, Paul, 2013: Toward a Conceptual Framework for the Digital Humanities. Terras, Melissa M., Julianne Nyhan, in Edward Vanhoutte (ur.): *Defining digital humanities: a reader*. Farnham, Surrey, England: Burlington, VT: Ashgate Publishing Limited; Ashgate Publishing Company. 219–233.
- Rybicki, Jan, Eder, Maciej, in Hoover, David L., 2016: Computational stylistics and text analysis. Crompton, Constance, Siemens, Ray, in Lane, Richard J. (ur.): *Doing digital humanities*. London & New York: Routledge. 123–144.
- Scherber, Peter, 1977. *Slovar Prešernovega pesniškega jezika*. Maribor: Obzorja.
- Schreibman, Susan, Siemens, Raymond George, Unsworth, John (ur.), 2004: *A Companion to Digital Humanities*. Malden, MA: Blackwell Pub.
- Siemens, Ray, 2016: Preface: communities of practice, the methodological commons, and digital self-determination in the humanities. Constance Crompton, Siemens, Ray, in Lane, Richard J. (ur.): *Doing Digital Humanities: Practice, Training, Research*. London & New York: Routledge. xxi–xxxiii.
- Siemens, Raymond George, in Susan Schreibman (ur.), 2007: *A companion to digital literary studies*. Malden, MA: Blackwell Pub.