



Nekaj pogledov na strokovni jezik v bibliotekarskih objavah: korpusna analiza strokovnih in znanstvenih člankov, objavljenih v reviji Knjižnica v letih od 1997 do 2016

*Some aspects of the technical language in library publications:
corpus analysis of technical and scientific papers published
in the journal Knjižnica from 1997 to 2016*

Ivan Kanič

Oddano: 17. 3. 2017 – Sprejeto: 11. 9. 2017

1.01 Izvirni znanstveni članek
1.01 Original scientific paper
UDK 001.891:[81'373.46:02]

Izvleček

Namen: Slovensko bibliotekarstvo je stroka, ki uspešno sledi najnaprednejšim razvojnim tokovom v svetu, to pa se dinamično odraža tudi v njeni terminologiji. Revija *Knjižnica* je osrednja slovenska znanstvena revija za področje bibliotekarstva, informacijske znanosti in knjigarstva, zato je najprimernejši vir za proučevanje sodobnega strokovnega izrazja. Namen raziskave je vzpostavitev korpusa strokovnih in znanstvenih člankov, ki so bili objavljeni v reviji *Knjižnica* v obdobju 1997–2016, kvantitativna in pomenska analiza dobljenega gradiva ter ugotavljanje rabe nekaterih izbranih terminov tudi primerjalno z nacionalnim referenčnim korpusom.

Metodologija/pristop: Po vnaprej določenih kriterijih je bilo izbranih 553 člankov, ki so prosto dostopni na portalu revije *Knjižnica*. Po tehnični pripravi in obdelavi besedil je bil s spletnim orodjem Sketch Engine vzpostavljen označeni korpus, ki obsega 2,4 milijona besed. Zaradi časovnih primerjav je razdeljen v dva podkorpusa, primerjalne analize pa so potekale tudi z drugimi korpusi, predvsem z nacionalnim referenčnim korpusom *FidaPLUS*, ki vsebuje 621 milijonov besed.

Analiza/rezultati: Raziskava je obsegala kvantitativne analize pogostnosti besed ter pomenske analize konkordanc, kolokacij in besednih skic v korpusu člankov in tudi

primerjalno z referenčnim korpusom. Potrjena je bila hipoteza, da vsebujejo članki v reviji *Knjižnica* bogat nabor tradicionalnih, pa tudi najnovejših terminov, katerih raba se v strokovnem jeziku po kolokacijah zelo razlikuje od rabe v splošnem jeziku. Kratka analiza je zajela tudi avtorske ključne besede z enostavno primerjavo po štirih petletnih kvartalnih.

Omejitve raziskave: Količina zajetih besedil omejuje pestrost nabora predvsem redkeje uporabljanih terminov, kar bi lahko brez velikih zadržkov rešila vzpostavitev mnogo večjega korpusa večine v reviji *Knjižnica* objavljenih člankov.

Uporabnost raziskave: Specializirani korpus, rezultati prvih analiz in v raziskavi pridobljena znanja so dobro izhodišče za podporo nadaljnjemu terminološkemu delu in dopolnjevanju slovarjev, veliko obeta tudi avtomatska ekstrakcija enobesednih in večbesednih terminov.

Ključne besede: *bibliotekarstvo, informacijska znanost, terminologija, korpusno jezikoslovje, besedilni korpusi, revija Knjižnica*

Abstract

Purpose: Slovenian library and information science is involved in most advanced international trends and development in the field, consequently rather dynamic changes are imposed on Slovenian LIS terminology. *Knjižnica* is the top-ranking Slovenian LIS journal so it proves to be the best source for terminology studies and tracking recent advances. Construction of a specialised text corpus which will include all professional and scientific articles, published in *Knjižnica* in the period between 1997 and 2016, will foster quantitative and semantic analyses and support studies related to the usage of individual terms and collocations, including comparative examination of the national reference corpus.

Methodology/approach: In compliance with the preset criteria 553 journal articles were selected and downloaded from the portal where they are published by *Knjižnica* in public access, they were subject to technical pre-processing and uploaded to the web-based Sketch Engine corpus tool. A tagged corpus comprising 2.4 million words was compiled and segmented into two subcorpora following a temporal scheme of publication. Comparative analyses involved other corpora as well, mostly the national reference corpus FidaPLUS comprising 621 million words.

Results: The study was focused on quantitative analyses regarding the frequency of words and semantic relations in concordances, collocations and word sketches in the frame of the special corpus and the reference corpus comparatively. The hypothesis has been confirmed proving that journal articles published in *Knjižnica* provide a rich selection of traditional as well as most recent LIS related terms, but occurring in rather different collocations than it is the case with the same words in the reference corpus of the general language. A short analysis dealt with author keywords and their frequency in five-year periods of publication to show the subject focus point shift.

Research limitation: The study was restricted by the moderate size of the corpus and the number of texts thus limiting the diversity and possibilities to extract less known and low frequency terms. The construction of a larger corpus containing all articles published by *Knjižnica* would be a feasible practical solution.

Practical implications: The recently constructed specialised corpus, practical results achieved and know-how resulting from the study are a solid start point and support for further terminology studies providing useful sources for the updating of the dictionaries, automatic term extraction routine being one of them.

Keywords: library science, LIS, terminology, corpus linguistics, text corpus, journal Library

1 Uvod

Ko je začel leta 1987, torej pred tridesetimi leti, nastajati slovenski *Bibliotekarski terminološki slovar* (Kanič, Leder, Ujčič, Vilar in Vodeb, 2009) in vzporedno z njim tudi prevajalni *Angleško-slovenski slovar bibliotekarske terminologije*, je bilo to pionirsko delo na povsem neraziskanem in neobdelanem področju, kjer ni bilo na razpolago nobenih osnov za delo in nobenih slovarjev, na katere bi se lahko slovaropisna skupina oprla. Ravnala se je po sodobnih načelih leksikografije, zato je delo temeljilo na ugotavljanju rabe izrazja v strokovnem jeziku in evidentiranju terminov z ročnimi izpisi bibliotekarskih izrazov iz obsežnega seznama slovenskih strokovnih besedil in na geslovniku, izdelanem po njih. Ta je bil osnova za izbor kasneje v slovarju obdelanih strokovnih pojmov in primerjanje z izborom v tujejezičnih strokovnih slovarjih na področju bibliotekarstva. Ročno izpisovanje izbrane slovenske bibliotekarske strokovne literature je zajelo 291 v celoti izpisanih besedil 140 slovenskih avtorjev, ki so bila objavljena v letih od 1940 do 1999.

Danes sodobni slovaropisci in terminografi ročnega izpisovanja ne poznajo več, na razpolago so jim različna orodja iz instrumentarija računalniških jezikovnih tehnologij, med katera sodijo tudi besedilni korpusi. Tudi za slovenski jezik obstaja več različnih korpusov, nekateri med njimi so zavirljivih razsežnosti in vsebujejo tudi več sto milijonov besed. Bibliotekarstvo je ena od redkih panog, ki ima tudi svoj specializirani korpus, to je *Korpus bibliotekarstva*,¹ ki je nastal z domačim znanjem in tehnologijo leta 2011 in že leto kasneje vseboval več kot 3,6 milijona besed iz 625 pretežno strokovnih in znanstvenih izvorno slovenskih besedil s področja bibliotekarstva in informacijske znanosti. Med njimi je tudi večje število člankov iz revije *Knjižnica*, ki je takrat že kako desetletje izhajala tudi v prosto dostopni spletni izdaji. *Korpus bibliotekarstva* se je izkazal kot izjemno zanimivo in koristno orodje za analizo bibliotekarskih objav in v njih

¹ Pridobljeno 5. 3. 2017 s spletne strani: <http://terminologija.blogspot.si/p/korpus.html>.

uporabljanega strokovnega jezika tako za bibliotekarsko slovaropisno skupino kot za nekatere posameznike, zato je bilo samo vprašanje časa, kdaj bo sledila njegova nadgradnja z najnovejšimi besedili, kjer je poleg potrditve že ustaljene terminologije seveda pričakovati tudi najsodobnejše izrazje, ki v strokovni jezik prihaja ali se kot neustaljeno šele oblikuje. Povsem po naključju leta 2017 sovpadata tudi dva za revijo *Knjižnica* pomembna mejnika, šestdesetletnica njenega nepretrganega izhajanja in dvajset let spletnega objavljanja v prostem dostopu, kar je bil pravšnji izziv za odločitve, da pripravimo ločeno korpusno analizo člankov, objavljenih v elektronski izdaji revije *Knjižnica* v letih od 1997 do 2016, ki lahko verodostojno pokaže nekatere jezikovne značilnosti rabe bibliotekarske terminologije in njenega razvoja.

2 Zasnova raziskave

Leta 2011 vzpostavljeni *Korpus bibliotekarstva*, ki vsebuje 3,6 milijona besed, se je izkazal kot odlično orodje za podporo pripravi bibliotekarskih slovarjev, omejene analize besedil in posameznih terminov ali frazeoloških gnezd ter pridobivanje znanj s področja praktičnega korpusnega jezikoslovja, vendar sodi med statične korpuse, torej se ne dopolnjuje, pa tudi njegov osnovni obseg ne zadošča več današnjim potrebam. Nujno je sledenje predvsem najnovejšim strokovnim besedilom ter dinamično nastajajoči terminologiji, ki se rojeva v izvirnih slovenskih besedilih in prihaja tudi s prevodi, zajemati pa je treba prvenstveno iz najkakovostnejših novejših objav.

Revija *Knjižnica* je osrednja slovenska znanstvena revija za področje bibliotekarstva, informacijske znanosti in knjigarstva, zato je najprimernejši vir za črpanje in proučevanje sodobnega strokovnega izrazja. Namen raziskave je z dovolj enostavnim in učinkovitim orodjem vzpostaviti korpus strokovnih in znanstvenih člankov, ki so bili objavljeni v reviji *Knjižnica* v zadnjem dvajsetletnem obdobju, izvesti kvantitativne in pomenske analize dobljenega gradiva ter pripraviti izhodišča za črpanje najnovejših terminov, kandidatov za vključitev v terminološki slovar. Analize naj zajamejo tudi ugotavljanje rabe nekaterih izbranih, že uveljavljenih terminov, v okviru korpusa člankov pa tudi primerjalno z že uveljavljenim *Korpusom bibliotekarstva* in nacionalnim referenčnim korpusom slovenskega jezika *FidaPLUS* (2000). Tako dobljeni specializirani korpus novejših besedil in v raziskavi pridobljena znanja naj nudijo podporo nadaljnjemu terminološkemu delu in dopolnjevanju slovarjev.

2.1 Metodologija

Za analizo besedil, rabe posameznih besed in terminov ter kolokacij in frazeoloških gnezd bomo uporabili sinhroni specializirani korpus, v katerem bodo zajeti vsi na spletu prosto dostopni strokovni in znanstveni članki revije *Knjižnica*, ki so izšli v obdobju 1997–2016 (letniki 41–60). Izhajali smo iz predpostavke, da za doseganje zelenih rezultatov potrebujemo najmanj vsa tista orodja in funkcije, ki jih zagotavlja že delujoči *Korpus bibliotekarstva*, vendar so nekateri dokaj zapleteni in časovno zamudni, zato bomo glede na razpoložljivo literaturo, možnosti testiranja in primerjave z referenčnim korpusom slovenskega jezika proučili možnosti uporabe drugih spletnih orodij. Orodje za vzpostavitev, vzdrževanje in uporabo korpusnih podatkov naj omogoči:

- večplasten in primerljiv prikaz sodobne rabe slovenskega jezika na področju bibliotekarstva,
- primerjavo rabe v določenih časovnih obdobjih,
- analize pogostnosti rabe nekaterih bibliotekarskih terminov,
- primerjavo značilnosti rabe terminov v strokovnem in v splošnem slovenskem jeziku,
- analize kolokacij in njihovih primerjav za posamezne termine in terminske sklope,
- primerjavo pogostnosti rabe in kolokacij z nekaterimi drugimi korpusi, predvsem:
 - z referenčnimi korpusi slovenskega jezika,
 - s specializiranim *Korpusom bibliotekarstva*,
- uporabo učinkovitega avtomatskega luščilnika za pripravo seznamov potencialno zanimivih terminov,
- pridobivanje specifičnega gradiva za dopolnjevanje *Bibliotekarskega terminološkega slovarja*,
- analitične informacije za objave na blogu *Bibliotekarska terminologija*,
- podporo terminološkemu svetovanju posameznikom in delovnim skupinam,
- koristne informacije za avtorje in urednike strokovne oziroma znanstvene revije *Knjižnica*,
- pridobivanje znanj za delo s korpusi, ki bodo omogočila:
 - kasnejše dodatne poglobljene analize,
 - dopolnjevanje korpusov in morebitno združitev obeh bibliotekarskih korpusov,
 - kasnejšo vzpostavitev podkorpusov za specifične namene.

Za pripravo korpusa je ključna dosledna selekcija člankov glede na zastavljene kriterije in ustrezna predpriprava besedil, ker je le tako mogoče zagotoviti pričakovano kakovost vsebine korpusa. Korpus bo razdeljen v dva podkorpusa, v katerih bomo primerjalno sledili časovnim spremembam rabe novejših in starejših

terminov, tako temeljnih kot redkeje uporabljanih. Nabor in obseg večine analiz je odvisen od ustreznih predpriprav korpusa, zato bo korpusno gradivo lematizirano in oblikoslovno-skladenjsko označeno. Primerjalne analize bodo potekale tudi z drugimi korpusi, predvsem s specializiranim *Korpusom bibliotekarstva* in nacionalnim referenčnim korpusom *FidaPLUS*, ki vsebuje 621 milijonov besed. Krajši analitični pregled avtorskih ključnih besed glede na pogostost in primerjavo po štirih petletnih kvartalih bo dopolnil prikaz značilnosti besedil in vsebinskih trendov objavljanja.

Raziskava bo temeljila na kvantitativnih analizah pogostnosti pojavljanja besed ter pomenskih analizah konkordanc in kolokacij v celotnem korpusu člankov in tudi primerjalno med podkorpusoma ter z referenčnim korpusom. Preveriti želimo hipoteze:

- članki v reviji *Knjižnica* vsebujejo bogat nabor tradicionalnih, pa tudi najnovejših terminov, katerih raba se v strokovnem jeziku spreminja tudi v krajših časovnih obdobjih,
- raba bibliotekarskih terminov se po kolokacijah zelo razlikuje od rabe teh besed v splošnem jeziku,
- raba ženskih in moških oblik poimenovanj poklicev in nazivov v bibliotekarstvu ni uravnotežena.

Tako vzpostavljen specializirani korpus in v raziskavi pridobljena znanja bodo tudi izhodišče za podporo nadaljnjemu terminološkemu delu in dopolnjevanju slovarjev, pričakujemo pa tudi orodje za učinkovito in kakovostno avtomatsko ekstrakcijo enobesednih in večbesednih terminov.

2.2 Vir besedil – strokovna in znanstvena revija *Knjižnica*

»Publicirana literatura je temelj vsakega strokovnega oziroma znanstvenega področja. Poleg neformalnih kanalov je publicirana literatura najpomembnejše sredstvo komunikacije. Tiskano gradivo je tako stalen in lahko dostopen vir znanja in razvoja. Izsledke objavljenih rezultatov novi strokovnjaki vgrajujejo v rezultate svojega dela in na ta način akumulirajo znanje nekega področja. Da bi lahko določili, kakšna je stopnja razvoja določenega strokovnega oziroma znanstvenega področja, lahko analiziramo publicirano literaturo tega področja.« (Likar, 1997) Odras razvitosti stroke ali znanstvene discipline, ki se zrcali v tej literaturi, je tudi strokovni jezik in v njem ustrezno jasna, sodobna ter urejena terminologija. Najrelevantnejši vir za proučevanje v strokovnih in znanstvenih besedilih uporabljane bibliotekarske terminologije je zagotovo revija *Knjižnica*.

Strokovna in kasneje tudi znanstvena revija *Knjižnica* je bila skoraj tri desetletja edina slovenska revija za področje bibliotekarstva. Občasno so ob njej izhajala najprej le Obvestila republiške matične službe NUK, ki so objavljala krajše strokovne prispevke, obvestila in predstavitve knjižnic, leta 1991 pa so se preimenovala v *Knjižničarske novice* ter tudi spremenila uredniško politiko. Prav tako leta 1991 je Zavod Republike Slovenije za šolstvo pričel izdajati revijo *Šolska knjižnica*, ki s krajšim premorom izhaja še danes, leta 1996 pa so začela izhajati *COBISS obvestila*, izdaja jih Institut informacijskih znanosti IZUM, ki so se leta 2002 preimenovala v *Organizacijo znanja*. Od leta 2006 izhajajo tudi *Potujoče novice*, ki so informativnega značaja, sem pa bi lahko šteli tudi *Štubidu*, ki ga od leta 1999 izdaja Študentska sekcija ZBDS.

Knjižnica je četrtletnik, praviloma na leto izideta dve samostojni in ena dvojna številka, občasno tudi peta, dodatna številka. Pokriva celotno bibliotekarstvo in vse panoge vključno z informacijsko znanostjo in knjigarstvom, saj v Sloveniji v nasprotju s tujino nimamo specializiranih revij za posamezna področja bibliotekarstva. V šestih desetletjih zorenja je postala *Knjižnica* sodobna znanstvena revija s prepoznavno zunanjo podobo in ustrezno znanstveno opremo, v primerjavi s prvim letnikom leta 1957 se je letni obseg revije petkrat povečal, revija je pridobila širok krog sodelavcev (Ambrožič, 2007). Uredniki dosledno upoštevajo standarde glede znanstvene opreme in urejenosti revije ter zahtevnosti posameznih prispevkov. Če hoče ohraniti status znanstvene revije, mora vsaj 76 % prispevkov v letniku dosegati zahtevne kriterije znanstvenega članka, zato se dosledno izvaja recenzijski postopek za vse strokovne in znanstvene prispevke ter skrbi za usklajenost recenzentov glede upoštevanja kriterijev za ocenjevanje in razvrščanje prispevkov. Razvojno naravnano uredniško politiko so uspešno načrtovali in izvajali vsi dosedanja glavni uredniki revije *Knjižnica*: Maks Veselko (1957, 1963–1964, 1970–1971), Branko Berčič (1958–1962), Jože Munda (1965–1969), Vlasta Pacheiner (1972–1973), Bruno Hartman (1974–1979), Jože Šifrer (1980–1985), Jelka Gazvoda (1986–1999), Melita Ambrožič (2000–2005), Eva Kodrič-Dačić (2005–2009), Branka Badovinac in odgovorni urednik Primož Južnič, (2010–2011), Ines Vodopivec in odgovorna urednica Alenka Kavčič-Čolić (2012–2015) ter od leta 2015 Damjana Vovk in odgovorni urednik Gorazd Vodeb. Na velik in pomemben vsebinski preskok kaže tudi sprememba podnaslova revije, ki je bila do leta 1999 »Glasilo Zveze bibliotekarskih društev Slovenije« (oziroma prej še Društva bibliotekarjev Slovenije), od leta 2000 pa je »Revija za področje bibliotekarstva in informacijske znanosti«.

Pomemben cilj uredništva v zadnjem desetletju je bil oblikovati revijo z mednarodno odmevnostjo, zato je bilo veliko pozornosti posvečene kakovosti prispevkov, ki so glavno merilo znanstvene odličnosti periodične publikacije. Revija *Knjižnica* je danes indeksirana in vključena v številne za stroko pomembne

mednarodne baze: LISA (CSA), LISTA (EBSCO), Proquest Science/Technology Databases (HT&AJ, Technology Journals, Illustrata Technology), Proquest Research Library, ProQuest Library Science, IBZ in DOAJ. Prispevki v celotnem besedilu so prosto dostopni na spletnem mestu revije, računalniškem katalogu COBISS/OPAC in v podatkovnih zbirkah dLib, Google Scholar ter LISA.²

Članki, objavljeni v reviji *Knjižnica*, so sistematično, redno in tekoče popisani v vzajemni bibliografski podatkovni zbirki Cobib, prispevki prvih tridesetih letnikov so zbrani v bibliografskem kazalu revije, ki ga je leta 1986 izdelal Martin Grum (Grum, 1987), za naslednjih dvajset let pa Mateja Ločniškar-Fidler in Damjana Vovk leta 2007 (Ločniškar-Fidler in Vovk, 2007). Prispevek objav v *Knjižnici* k slovenskemu bibliotekarstvu in informacijski znanosti je bil že večkrat kvantificiran, saj je bila *Knjižnica* predmet bibliometričnih analiz tudi v okviru študijskih programov Oddelka za bibliotekarstvo, informacijsko znanost in knjigarstvo Filozofske fakultete Univerze v Ljubljani. Podatke o avtorjih, citiranosti, tipologiji, jeziku objav in nekaterih drugih značilnostih ugotavljajo tudi dosedanje bibliometrične analize objav v reviji *Knjižnica* za posamezna obdobja (Popovič, Ambrožič in Južnič, 1984; Aparac-Jelušič, 1999; Likar, 1997; Vratarič, 2006; Česnik Ušaj, 2011), zato se v tej jezikovni analizi objavljenih besedil s temi podatki ne ukvarjamo.

Revija *Knjižnica* je sooblikovala in pospeševala razvoj slovenskega bibliotekarstva in informacijske znanosti in neprecenljiv je tudi njen prispevek k razvoju slovenske bibliotekarske strokovne terminologije. To je tudi eden od razlogov, da je Bibliotekarska terminološka komisija pri pripravi razlagalnega *Bibliotekarskega terminološkega slovarja* in tudi prevajalnega *Angleško-slovenskega slovarja bibliotekarske terminologije* pogosto črpala iz besedišča objav v *Knjižnici*.

Velik napredek pri zagotavljanju dostopnosti revije *Knjižnica* slovenskim bibliotekarjem je bila postavitev njene spletne različice leta 2000, najprej na lastnem strežniku, kasneje pa na spletnem mestu s podporo sistema Open Journal System (OJS), kjer so sedaj objavljene tekoče številke, ob njih pa še arhiv revije, ki sega do leta 2000.³ Del arhiva člankov za obdobje 1997–2011 je še vedno dostopen na starem spletnem mestu,⁴ ob tem pa tudi celotno bibliografsko kazalo po letnikih in številkah za obdobje 1957–2011. Digitalizirani so bili tudi članki iz obdobja 1957–1996, vendar so omejeno dostopni samo na posebnih računalnikih v čitalnici NUK, saj avtorskoppravna vprašanja z avtorji (še) niso ustrezno urejena. Prav

² Pridobljeno 15. 1. 2017 s spletne strani: <http://www.zbds-zveza.si/Knjiznica>.

³ Pridobljeno 15. 1. 2017 s spletne strani: <http://knjiznica.zbds-zveza.si/index.php/knjiznica/index>.

⁴ Pridobljeno 15. 1. 2017 s spletne strani: <http://revija-knjiznica.zbds-zveza.si>.

javna dostopnost elektronskih besedil je omogočila izgradnjo specializiranega korpusa kakovostnih strokovnih in znanstvenih besedil s področja bibliotekarstva in nekatere prve jezikovne in terminološke analize.

2.3 Avtorske pravice

Upoštevana besedila javnosti v korpusu niso dostopna v izvorni obliki ali v celoti, uporabljena so samo za izdelavo analitičnih in kumulativnih statističnih kazalcev jezika, npr. za sezname besed ali besednih zvez s pogostnostjo, besednih skic in kolokacij za posamezne izbrane izraze in v konkordančnih seznamih, vendar tudi tam le v obliki ožjega citata ne več kot treh povedi. Korpus torej ne posega v avtorske pravice piscev besedil ali založnikov revije *Knjižnica*, to je Zveze bibliotekarskih društev Slovenije in Narodne in univerzitetne knjižnice. Navedeni instituciji sta tudi izdajateljici *Bibliotekarskega terminološkega slovarja* in *Angleško-slovenskega slovarja bibliotekarske terminologije* ter tako prvi zainteresirani za uporabo izsledkov, temelječih na uporabi korpusa. Vsa dela so ustrezno citirana v seznamu zajetih besedil, izvorni članki pa so prosto dostopni v celotnem besedilu na portalu dLib.si in na spletišču revije *Knjižnica*. Zato posebnih dogovorov z avtorji in založnikoma nismo sklepali.

Glede avtorskopравnih vprašanj v zvezi z uporabo besedil za vzpostavitev in uporabo korpusa smo se zgledovali tudi po najnovejših prizadevanjih Evropskega parlamenta, ki je septembra 2016 sprejel predlog direktive za spremembo zakonodaje v članicah EU, s katero bi formalno potrdili pravice za besedilno in podatkovno rudarjenje v raziskovalne namene. Predlog navaja: »Uvesti bi bilo treba obvezne izjeme ali omejitve glede uporabe tehnologij za besedilno in podatkovno rudarjenje na področju znanstvenih raziskav, glede ilustracij pri poučevanju v digitalnem okolju in za namene ohranjanja kulturne dediščine ... Države članice določijo izjemo od pravic iz člena 2 Direktive 2001/29/ES, členov 5(a) in 7(1) Direktive 96/9/ES in člena 11(1) te direktive za reprodukcije in pridobivanje izvlečkov vsebine, ki jih opravijo raziskovalne organizacije, da bi izvajale besedilno in podatkovno rudarjenje na delih ali drugih predmetih urejanja, do katerih imajo zakonit dostop za namene znanstvenega raziskovanja.« (Direktiva, 2016)

Izhajali smo tudi iz v Sloveniji že sprejetih priporočil Akcijskega načrta za jezikovno opremljenost (2015): »Pri izdelavi korpusov znanstvenih besedil naj se iz obstoječih portalov, ob upoštevanju avtorskih pravic nad izvirnimi besedili, zajame znanstvena besedila in se jih pretvori v primerno jezikoslovno označeni korpus oziroma korpuse ter jih naredi prosto dostopne preko spletnih konkordančnikov za namene preučevanja slovenskega znanstvenega in strokovnega izrazja. Iz korpusa se naknadno lahko avtomatsko izlušči terminologijo, kjer je

to možno skupaj s tujejezičnimi ustreznici in termine ponudi v prosti in odprti dostop. Izluščene terminološke baze so nato uporabne za posamezne znanstvene skupnosti za namene upravljanja s terminologijami ...«

3 Korpusno jezikoslovje

Wolfgang Teubert, eminentni profesor jezikoslovja na Univerzi v Birminghamu, je strnil svoj pogled na pomen in vlogo korpusov in korpusnega jezikoslovja v kratko trditev: »Dandanes je korpus osnovni vir skoraj vsakomur, ki se ukvarja z jezikoslovjem. Nobena introspekcija si ne more lastiti verodostojnosti brez preverjanja v avtentičnem jezikoslovnem gradivu. Korpusna analiza je postala ključni element skoraj vseh jezikoslovnih študij.«⁵ (Teubert, 2005, str. 1)

Korpusno jezikoslovje je eno najmlajših področij jezikoslovja, ki se je začelo razvijati šele potem, ko so bili od šestdesetih let 20. stoletja naprej strokovni javnosti na razpolago prvi zmogljivejši računalniki, začelo pa se je seveda v angleškem jezikovnem prostoru. V slovenskem jezikovnem prostoru lahko za začetke korpusnega jezikoslovja štejemo aktivnosti, povezane z vzpostavitvijo korpusa *FIDA*, ki segajo v leto 1997, nekaj let kasneje je sledil predhodnik korpusa *Nova beseda* (2008). Korpusno jezikoslovje je raziskovanje reprezentativnega vzorca naravnega jezika na podlagi obsežnega nabora besedil iz vsakdanje javne jezikovne rabe (referenčni korpus) ali rabe v skupini po določenem kriteriju izbranih govorcev (specializirani korpus), kakršna sta *Korpus bibliotekarstva* in korpus člankov revije *Knjižnica*, o katerih bo govora v nadaljevanju. Redke korpuse so pred digitalno revolucijo izdelovali ročno, danes pa so to večinoma računalniške zbirke podatkov, shranjene na digitalnih medijih in najpogosteje javno dostopne na spletu.

3.1 Besedilni korpusi

V jezikoslovju je korpus oziroma besedilni korpus velika in strukturirana zbirka avtentičnih besedil, navadno grajena, hranjena in obdelana računalniško, nastala po vnaprej določenih merilih in z določenim ciljem ter opremljena z orodji, ki omogočajo večplastno iskanje jezikovnih podatkov. Korpuse se uporablja za

⁵ »Today, the corpus is considered the default resource for almost anyone working in linguistics. No introspection can claim credence without verification through real language data. Corpus research has become a key element of almost all language study.«

statistične analize pisanega in/ali govorjenega jezika, za preverjanje pojavitve besed in besednih zvez ali pa za potrditev lingvističnih pravil v določenem jeziku. So tudi nepogrešljivo in nadvse koristno orodje ob pripravi splošnih in tudi terminoloških slovarjev. Računalniški korpusi so konec 20. stoletja nadomestili zbirke listkovnega gradiva z izpiski, na podlagi katerih so v predračunalniški dobi navadno nastajali slovarji – med njimi tudi *Slovar slovenskega knjižnega jezika* (SSKJ). Tudi *Bibliotekarski terminološki slovar* je začel nastajati na podlagi ročnih izpiskov in kasneje več generacij specializiranega vzorčnega besedilnega korpusa.

Glede na nekatere osnovne značilnosti pri nastajanju korpusa in izboru besedil ter kasnejših obdelavah lahko govorimo o različnih vrstah korpusov.

- **Pisni in govorni korpusi** se ločijo glede na nosilec besedila, to je pisani in govorjeni jezik; govornih je manj in so novejšega datuma.
- **Sinhroni korpusi** zajemajo sodobna besedila svojega časa, **diahroni korpusi** hranijo starejša besedila, primerjava besedišč diahronnega in sinhronnega korpusa je lahko dober vir za ugotavljanje razlik in razvoja besedišča skozi čas.
- **Referenčni korpusi** so najboljšežnejši in zagotavljajo reprezentativen izbor besedil določene jezikovne skupnosti ali naroda. **Specializirani korpusi** predstavljajo jezik v točno določeni rabi, na primer v neki stroki.
- **Celoviti korpus** zajema celotna besedila, npr. celotne članke, besedila knjig v celoti, **vzorčni korpus** pa sestavljajo le fragmenti besedil, kar je zelo podobno karticam z izpiski v času ročne priprave slovarjev. Z razvojem tehnologije, ki omogoča zajemanje večjih količin besedil, so postali vzorčni korpusi manj aktualni.
- **Vzporedni korpusi** sopostavljajo poravnano izhodiščno besedilo in prevod oziroma prevode istega besedila v drug jezik oziroma druge jezike in so zanimivi predvsem za prevajalstvo.
- **Statični korpus** se, ko je enkrat zgrajen, ne spreminja več; takih korpusov je največ. **Spremljevalni korpus** se sprotno dopolnjuje in omogoča opazovanje jezika v spreminjanju. Takih je najmanj.
- **Enojezični korpus** zajema besedila zgolj v enem jeziku, taki so najpogostejši, **večjezični korpus** pa besedila v dveh jezikih ali več.
- **Neoznačeni korpus** prikazuje besedila in njihove dele (npr. stavke, besede, besedne zveze) le v obliki, kot so bili zapisani. **Označeni korpus** v besedilih dodatno vsebuje tudi jezikoslovne oznake, npr. oblikoslovne oznake, leme, skladske povezave, imena itd.

Te vrste med seboj seveda niso strogo ločene in se lahko med seboj prepletajo, naš je na primer enojezični pisni sinhroni in označeni specializirani korpus (Gorjanc in Fišer, 2013).

Referenčni korpusi so temeljna vrsta korpusov, ki naj bi predstavili celovito podobo nekega jezika. So večjega obsega, zanje je glede na tradicijo tudi najnatančneje izdelana metodologija gradnje, predstavljajo pa izhodišče za temeljne jezikoslovne raziskave predvsem s področja slovnice in slovarjev. Tudi za slovenski jezik obstaja več razmeroma velikih referenčnih korpusov (Logar Berginc idr., 2012) in na nekatere od njih se bomo oprli tudi za primerjave s korpusom člankov iz revije *Knjižnica*. Pri nas sodijo v to kategorijo trije besedilni korpusi: *FidaPLUS*, *Gigafida* in *Nova beseda*.

FidaPLUS je referenčna zbirka vsakdanje javne rabe slovenščine v pisnih besedilih v obdobju 1990–2006. Je uravnoteženi in reprezentativni, sinhroni in označeni korpus slovenščine, ki obsega približno 621 milijonov besed iz različnih virov jezika vsakdanje rabe, predvsem časopisov, revij, strokovne ter leposlovne literature in interneta. Kot referenčni korpus je namenjen slovaropisnim raziskavam in za pripravo eno- in večjezikovnih slovarjev, terminoloških slovarjev in drugih jezikovnih priročnikov, kot podatkovna infrastruktura pa za razne jezikovne tehnologije (Arhar Holdt in Gorjanc, 2007). FidoPLUS je sicer nadgradila in nadomestila *Gigafida* (B. l.), vendar se bomo v nadaljevanju večkrat sklicevali na korpus *FidaPLUS*, ker je ena njegovih različic nameščena v okolju Sketch Engine, kar omogoča različne primerjave z našim korpusom člankov.

Korpus Gigafida je nadgradnja, ki vsebuje skoraj celoten predhodni referenčni korpus slovenščine *FidaPLUS*. Vsebuje 1,2 milijarde besed, v njem so zajeta objavljena javno dostopna pisna besedila različnih zvrsti od 1991 do 2011, v tiskani obliki so to knjige, časopisi in revije, pri spletnih objavah pa je omejen na predstavitvene strani podjetij, državnih, pedagoških, raziskovalnih, kulturnih in podobnih ustanov ter na besedila novičarskih portalov. Prav tako sodi med uravnotežene in reprezentativne, sinhrono in označene korpusne slovenščine.

Nova beseda je besedilni korpus, namenjen predvsem slovaropisnim potrebam, čeprav se je začel s spletno postavitvijo elektronske zbirke leposlovnih besedil leta 1999, s skupaj tremi milijoni besed, zbranih in obdelanih v okviru doktorske disertacije. V naslednjih letih sta se obseg in zvrstnost korpusa postopoma povečevala do 318 milijonov besed leta 2010. Velika večina besedil je iz zadnjih desetih let. Prav tako sodi med sinhrono uravnotežene in reprezentativne korpusne, vendar ni označen.

Specializirani korpusi predstavljajo jezik v natančno določeni rabi, med njimi so najpomembnejši korpusi strokovnih jezikov, predvsem v okviru terminoloških raziskav in gradnje terminoloških slovarjev. Taka sta pri nas na primer *Korpus DSI* (Korpus zbornikov posvetovanja Dnevi slovenske informatike od 2003 do 2015 in revije *Uporabna informatika*), ki vsebuje tri milijone besed, in *KORP – korpus*

besedil odnosov z javnostmi, ki vsebuje skoraj dva milijona besed. Zanimiva sta tudi jezikoslovno označeni korpus Jos in *Evrokorpus* z dvo- in večjezičnimi vzporednimi korpusi besedil z različnih področij, v katerih slovenščina vedno nastopa kot eden od jezikov; skupaj več kot 240 milijonov besed. Nekaj prav posebnega pa je *Vayna* s četrtem milijona besed iz 360 časopisnih člankov, ki so v času od aprila do avgusta 1998 obravnavali t. i. »*verbalne napade na JLA*« (oziroma ozadje in potek procesa JBTZ). Nekaj slovenskih specializiranih korpusov navaja portal Jezikovna Slovenija.⁶

Specializirani besedilni korpusi se v terminologiji uporabljajo za inventarizacijo in preverjanje pojavitev besed in besednih zvez v strokovnih besedilih, torej za ugotavljanje, kateri izrazi in na kakšen način se pojavljajo v jeziku neke stroke. Iz njih je mogoče pridobiti zelo različne sezname besed in besednih zvez, z lematizacijo in oblikoslovno-skladenjskim označevanjem besed pa posegati tudi v analizo uporabe posameznih besednih vrst. Zato so nepogrešljivo in nadvse koristno orodje pri pripravi sodobnih terminoloških slovarjev. V ta namen smo si že dalj časa prizadevali vzpostaviti tak korpus tudi na področju slovenskega bibliotekarstva (Kanič, 2011c).

Korpus bibliotekarstva je tipičen predstavnik specializiranih korpusov, nastal je predvsem kot pomoč in orodje za delo Bibliotekarske terminološke komisije pri urejanju in dopolnjevanju *Bibliotekarskega terminološkega slovarja*, dobrodošel pa je tudi kot orodje in vir podatkov za avtorje prispevkov na blogu *Bibliotekarska terminologija*. Nastajal je v letih 2011 in 2012, zajel pa je 625 strokovnih in znanstvenih besedil s področja bibliotekarstva, praviloma slovenskih avtorjev in tudi nekaterih prevodov. Vsa besedila so bila objavljena v elektronski obliki in so praviloma prosto dostopna na spletu. *Korpus bibliotekarstva* sodi med specializirane enojezične uravnotežene in reprezentativne, sinhrono vendar neoznačene korpuse slovenščine, vsebuje pa okrog 3,6 milijona besed (Kanič, 2011b).

Če gre pri referenčnih korpusih za predstavitev »celotnega« jezika v vseh njegovih segmentih, predstavljajo specializirani korpusi jezik samo v določeni funkciji oziroma jezik v okviru določenega predmetno-pojmovnega področja, torej predstavljajo omejeni, glede na vrsto specializiranega korpusa točno določeni del jezika. S tem je pogojen izbor besedil in zagotavljanje njihove homogenosti, saj so le tako mogoči nadaljnja terminološka sklepanja in raziskave. Metoda gradnje takega korpusa je sicer podobna referenčnemu (Gorjanc in Logar, 2007).

⁶ Jezikoslovna Slovenija: Korpusi. Pridobljeno 15. 1. 2017 s spletne strani: <http://www.jezikovna-politika.si/opremljenost/jezikovni-opis/korpusi/>.

4 Korpus člankov iz revije *Knjižnica* v letih od 1997 do 2016

Z gradnjo korpusov v slovenskem jeziku smo dobili v zadnjih dvajsetih letih tudi Slovenci osnovno izhodišče za obsežnejše in raznovrstnejše jezikoslovne študije, zasnovane na podlagi podatkov o besedilnih realizacijah. Najpogosteje je bil v razpravi poudarek predvsem na pripravi in uporabi referenčnih korpusov *FIDA*, naslednika *Gigafide*, *Nove besede* in nekaterih vzporednih korpusov, sorazmerno zanemarjeni pa so ostali specializirani korpusi (Gorjanc in Logar, 2007). To vrzel skušajo vsaj deloma zapolniti specializirani korpusi, med njimi tudi korpusa z bibliotekarskega strokovnega področja.

Namen in cilji vzpostavitve korpusa so bili načrtani leta 2016 ob sovpadanju šestdesete obletnice rednega in uspešnega izhajanja tiskane izdaje znanstvene revije *Knjižnica*, zaključenem dvajsetem letniku te revije, prosto dostopnem na spletu, in tridesetih letih zelo uspešnega delovanja Bibliotekarske terminološke komisije. Zato naj bodo zajeti vsi na spletu prosto dostopni strokovni in znanstveni članki revije *Knjižnica*, logično pa se je postavila omejitev na obdobje 1997 do 2016 (letniki 41–60).

Pri načrtovanju korpusa člankov iz revije *Knjižnica* smo izhajali iz nekaterih osnovnih zahtev, ki dajejo korpusu in podatkom iz njega verodostojnost in ustrezno kakovost ter po možnosti tudi primerljivost z drugimi korpusi. Vnaprej smo mu določili naslednje karakteristike, iz katerih so sledili izbor besedil in vsa njihova nadaljnja obdelava. Korpus bo predstavljal in analiziral strokovne in znanstvene članke specializirane revije, zato je seveda pisen in specializiran, posvečen slovenski terminologiji, zajemal pa bo izključno članke v slovenskem jeziku. Glede na potrebe po evidentiranju in razvoju sodobne bibliotekarske terminologije in dopolnjevanju terminološkega slovarja, za kar bomo črpali gradivo tudi iz korpusa, bo korpus sinhroni, saj bo zajemal besedila zadnjih dvajset let, zadnja od njih so bila objavljena v času začetka gradnje korpusa. Zasnovan je kot statični korpus, ki se ne bo dopolnjeval, vendar ostaja odprta pot naprej, predvsem v smislu možnosti tekočega letnega dopolnjevanja (praviloma izideta dve enojni in ena dvojna številka revije *Knjižnica* na leto) in morebitne retrospektivne dopolnitve z letniki pred 1997, ki so bili digitalizirani, vendar so omejeno dostopni samo v NUK-u. Zaradi doseganja zastavljenih ciljev je bilo nujno zadostiti tudi pogoju, da mora biti korpus označen, kar je povzročilo tudi ne prav lahko odločanje o najustrežnejši programski podpori.

Značilnosti dobrih korpusov so najpogosteje opredeljene z naslednjimi zahtevami: avtentičnost, količina, kakovost, dokumentiranost in enostavnost (Erjavec, 2010). Avtentičnost smo zagotovili z natančno določitvijo kriterijev, ki jim mora korpus zadoščati, ter njihovim spoštovanjem ob vzpostavitvi korpusa. Na

količino ni bilo mogoče vplivati, ker je bil obseg nabora besedil omejen z njihovo pojavnostjo in dosegljivostjo (prosta spletna dostopnost), vendar je znano, da so specializirani korpusi praviloma mnogo manjši od referenčnih, večja je tudi homogenost besedil, frekvenčna razporeditev besed in besednih zvez, ki so v primeri specializiranih korpusov pogosto tudi termini, pa je drugačna. Na vsak način pa je dosežen obseg 553 člankov in z njimi 2.414.000 besed za zastavljene cilje solidna in jezikovno verodostojna velikost korpusa. Dokumentiranost gradiva je povsem zadovoljiva, saj sta na spletu dostopni bibliografski kazali (letno in po posameznih številkah) za celotno v korpusu zajeto obdobje, omogočen pa je tudi neposreden dostop do vsakega besedila v celoti, kar dovoljuje vpogled v širše so-besedilo in s tem boljše razumevanje termina ali celo dostop do nekaterih razlag. Tudi enostavnost računalniškega zapisa elementov in uporaba programskega orodja sta na povsem zadovoljivi ravni, saj je bilo izbrano orodje, na katerem deluje več sto korpusov, med njimi tudi nekaj slovenskih. Kakovost označevanja in obdelav (npr. konkordance, kolokacije, luščilnik itd.) je na visoki ravni in ni problematična, saj so bila uporabljena orodja, ki jih pozna tudi *FidaPLUS* in še nekateri slovenski korpusi. Zato pa s tehničnega vidika ni povsem zadovoljiva kakovost zajetih besedil, saj kljub ročnemu čiščenju vsebujejo še precej napak in šumov, ki se jim brez velikega delovnega vložka ni mogoče izogniti. O tem podrobneje tudi v razpravi o pripravi besedil.

4.1 Programska podpora

Pri pripravljanju besedil in za postavitev ter javno spletno uporabo prvega *Korpusa bibliotekarstva* je bila leta 2011 uporabljena domača programska oprema – urejevalnik besedil *EVA* in njegova internetna različica *NEVA*⁷ s specifičnimi funkcijami, ki že od leta 2000 omogočajo delovanje splošnega referenčnega korpusa slovenskega jezika *Nova beseda*,⁸ spletne različice *Slovarja slovenskega knjižnega jezika*⁹ in nekaterih drugih slovarskih in jezikoslovnih orodij. Vse priprave in obdelave besedil za *Korpus bibliotekarstva* so potekale na osebнем računalniku, prav tako izdelava številnih indeksov za konkordančno in besedno iskanje ter iskanje po besednih n-gramih, torej parih, trojčkih, četverčkih in peterčkih. Orodje za vzpostavitev, vzdrževanje in javno spletno uporabo korpusa se je izkazalo kot zelo priročno in je omogočilo serijo analiz, ki so vodile v dopolnjevanje slovarskega gradiva in terminološke objave (Kanič idr., 2011).

⁷ NEVA – interNet version of EVA. Pridobljeno 17. 11. 2016 s spletne strani: <http://www.laze.org/neva/>.

⁸ Nova beseda. Pridobljeno 17. 11. 2016 s spletne strani: http://bos.zrc-sazu.si/s_beseda3.html.

⁹ Slovar slovenskega knjižnega jezika. Pridobljeno 17. 11. 2016 s spletne strani: <http://bos.zrc-sazu.si/sskj.html>.

Ob zasnovi korpusa člankov iz revije *Knjižnica* leta 2016 je bilo treba ponovno razmisliti o programski podpori. Jasno je, da je za uspešno izrabo možnosti, ki jih nudi tako bogata zbirka strokovnih besedil, nujna tehnična podpora, ki mora od samega začetka slediti zahtevam tako glede strojne kot programske opreme ter biti sposobna oblikovati orodja za procesiranje zbranega gradiva. Prav pri obdelavi podatkov se je treba odločati tako, da jezikovnim podatkom s širokim spektrom različnih izhodnih oblik zagotovimo čim večjo uporabnost in primerljivost, izmenjavo ter trajnost z možnostjo prenosa podatkov v druge sisteme. Odločitev ni bila lahka, vendar so možnosti, ki jih nudijo že vgrajena lematizacija in oblikoslovno-skladenjsko označevanje, kolokacije, besedne skice, vizualizacija časovnih pojavitev posamezne besede, luščilnik in primerjava besedišča z drugimi korpusi, predvsem slovenskih referenčnim korpusom *FidaPLUS*, prevladale in izbrano je bilo orodje Sketch Engine.

Sketch Engine je javno dostopno spletno orodje za gradnjo, upravljanje in uporabo korpusov. Razvili so ga leta 2003 z namenom, da bi omogočili širšo podporo vsem, ki se ukvarjajo z obsežnimi zbirkami besedil, to je predvsem jezikoslovcem, leksikografom, prevajalcem in drugim raziskovalcem na področju jezikoslovja in jezikovnih tehnologij, ki potrebujejo zahtevne in jezikovno naravnane metode iskanja v besedilih. Orodje je dobilo ime po eni od svojih naprednih funkcij, imenovani Word Sketches (besedne skice), ki so enozaslonska predstavitev slovničnega in kolokacijskega obnašanja določene besede. Sketch Engine uporabljajo tudi nekatere največje založniške hiše, ki izdajajo slovarje, kot so Macmillan English Dictionary, Dictionnaires Le Robert in Oxford University Press. Različne institucije in posamezniki gradijo in vzdržujejo s tem orodjem nekaj sto enojezičnih korpusov v nekaj deset jezikih. Med njimi je tudi slovenščina. Kot pravi raziskovalec Simon Krek (Krek in Kilgarriff, 2006): »Zdaj je na voljo programski modul Sketch Engine, korpusno orodje, ki na vходу sprejme korpus kateregakoli jezika ter njegove slovnične vzorce, iz njih pa ustvari besedne skice za besede tega jezika. Hkrati avtomatsko generira tezaver in 'razlikovalne skice', ki izpostavljajo podobnosti in razlike med bližnjimi sopomenkami. V programski modul Sketch Engine smo naložili korpus *FidaPLUS*, oblikoslovno-skladenjsko označeni korpus slovenščine.« Program torej že »pozna« slovenski jezik in ustrezno uporabi za vse obdelave in analize vgrajena pravila za ta jezik. Tako pripravi sezname besed s pogostnostjo njihovega pojavljanja, besede lematizira in besednovrstno, oblikovno ter skladdenjsko označi, prepozna in analizira kolokacije, uvrsti besedo v tezaver in v strukturi zapiše istopomenke, protipomenke in podobne besede, iz besedil izlušči enobesedne in večbesedne termine, primerja besedišča dveh različnih korpusov ali pojavljanje iste besede v teh korpusih. Kljub množici funkcij in prilagoditev je delo s sistemom dokaj enostavno. Testno je Sketch Engine za določeno obdobje na razpolago zastoj, za več pa je treba nekaj doplačati, kar je odvisno od velikosti korpusa (v milijonih besed), ki ga želi uporabnik graditi,

in njegovega statusa. Trenutno so od večjih slovenskih korpusov tam dosegljivi *DGT*, *EUR-Lex Slovenian 2/2016*, *EUROPARL7*, *KAS-Dipl* (diplome), *KAS-Dr* (doktorati), *KAS-Mag* (magisteriji), *Kres*, *Lektor*, *OPUS2 Slovenian*, slovenski referenčni korpus *FidaPLUS v2* in *Slovenian Web 2015* (slTenTen15).¹⁰ Na javno dostopni odprtokodni verziji, imenovani noSketch Engine, je vzpostavljeno tudi večje število referenčnih, specializiranih in vzporednih korpusov slovenskega jezika (Erjavec, 2013). Mnogi med njimi so tudi javno dostopni,¹¹ vendar je funkcionalnost odprtokodne verzije dokaj omejena.

4.2 Vzpostavitev korpusa

Tehnični postopki vzpostavitve korpusa obsegajo izbor besedil v skladu z definicijo vsebine in obsega korpusa, pripravo datotek bodisi s pridobivanjem že obstoječih elektronskih besedil ali digitalizacijo in optičnim prepoznavanjem tiskanih besedil ter tvorbo datotek v ustreznem datotečnem formatu za nadaljnje obdelave, čiščenje besedil z odstranjevanjem motečih sestavin ter zaključno polnjenje korpusa s pripravljenimi besedili, ki mu sledijo lematizacija, oblikoslovno-skladenjsko označevanje in indeksiranja glede na zahtevano strukturo in funkcije korpusa.

4.2.1 Izbor besedil

Korpus člankov iz revije *Knjižnica* je specializirani enojezikovni sinhroni korpus, zato so vanj vključena samo besedila v slovenskem jeziku, tujejezični elementi pa se v korpusu lahko pojavijo le kot del slovenskega besedila. Sočasno je časovno omejen na najsodobnejša besedila zadnjih dvajsetih let, vendar samo tista, ki so javno dostopna na spletu. To so bile tudi tri glavne zahteve in filter pri izboru besedil iz bibliografskega seznama vseh objav revije *Knjižnica*.

Uredništvo opredeljuje svojo politiko objavljanja takole: »Revija *Knjižnica* objavlja znanstvene in strokovne prispevke v slovenskem in angleškem jeziku, izjemoma objavlja tudi znanstvene prispevke v nekaterih drugih jezikih (nemškem, francoskem, italijanskem, hrvaškem jeziku ...). Druge prispevke, kot so bibliografije, krajši članki, ocene, poročila z znanstvenih in strokovnih posvetovanj, objave dokumentov, obvestil, razpisov ZBDS, njenih strokovnih teles ali področnih

¹⁰ Sketch Engine: Slovenian text corpora. Pridobljeno 17. 12. 2016 s spletne strani: <https://www.sketchengine.co.uk/user-guide/user-manual/corpora/by-language/slovenian-text-corpora/>.

¹¹ Clarin.si. Pridobljeno 17. 12. 2016 s spletne strani: <http://nl.ijs.si/noske/index-en.html>.

društev, pa objavlja le v slovenskem jeziku.« (Splošna, B. l.) Zato so bili izbrani praviloma le strokovni in znanstveni članki, samo izjemoma tudi nekatere druge objave, pri katerih je bilo pričakovati za stroko zanimivo terminologijo (npr. pravilniki, obrazložitve stanovskih nagrad itd.). Zajeti so bili tudi prevodi v slovenski jezik, izvorni tujejezični članki pa ne. Prav tako so bili izločeni članki, ki iz kakršnega koli razloga niso bili prosto dostopni na spletu, saj z razpoložljivimi viri (en človek v prostem času) ni bilo mogoče zagotoviti skeniranja in dodatnih obdelav OCR. Pri starejših letnikih nekaterih del, čeprav so bila objavljena v elektronski obliki, ni bilo mogoče vključiti zaradi zaščite ali formata (pdf) objavljenih datotek, saj ta ne dovoljuje branja posameznih besed ali znakov v besedilu in jih kljub uporabi ustrezne programske opreme ni bilo mogoče ustrezno kakovostno prebrati.

Preglednica 1: Pregled objav v reviji *Knjižnica* v obdobju dvajsetih let (1997–2016)¹²

Leto (letnik)	Št. znanstvenih člankov	Št. strokovnih člankov	Št. drugih objav	Št. objav skupaj	Št. v korpusu analiziranih člankov	Št. besed
1997 (41)	*	*	4	49	41	149.800
1998 (42)	*	*	2	31	27	121.900
1999 (43)	12	25	5	42	31	117.700
2000 (44)	19	11	–	30	23	129.300
2001 (45)	16	12	–	28	22	110.000
2002 (46)	20	18	–	38	32	155.100
2003 (47)	19	4	9	32	26	119.000
2004 (48)	20	9	13	42	38	159.500
2005 (49)	18	19	8	45	31	131.800
2006 (50)	16	11	8	35	22	116.500
2007 (51)	20	13	11	33	33	135.700
2008 (52)	14	13	8	35	26	96.200
2009 (53)	11	8	2	21	12	47.200
2010 (54)	20	5	15	38	28	114.000
2011 (55)	27	12	8	45	39	180.800
2012 (56)	29	9	4	41	40	136.700
2013 (57)	13	6	3	22	17	94.100
2014 (58)	15	5	4	24	20	90.400
2015 (59)	19	5	5	29	26	122.500
2016 (60)	11	5	4	20**	19	68.100
Skupaj				680	553	2.423.700

* Do leta 1998 članki niso bili opremljeni s tipologijo objave.

** V času analize in priprave članka številka 4 za leto 2016 še ni izšla in zato ni upoštevana.

¹² Podatki za leta 2000–2015 so povzeti iz članka Alenke Kavčič-Čolič in Ines Vodopivec (Kavčič-Čolič in Vodpivec, 2015).

Na tak način je bilo od 680 objavljenih besedil za vključitev v korpus izbranih 553 besedil (glej Preglednica 1 in Slika 6). Zajete so vse številke letnikov od 1997 do 2016 z izjemo zadnje, saj v času priprave zaključne verzije analize¹³ zadnja številka letnika 2016 še ni izšla. Vsi članki od leta 1957 do 2015 so arhivirani na portalu dLib.si, ustrezno iskanje prikaže 2.479 zadetkov.¹⁴

4.2.2 Pridobivanje elektronskih besedil in priprava izhodiščnih datotek

Od seznama želenih in izbranih besedil do delujočega korpusa so nujne tri stopnje priprav: pridobivanje elektronskih besedil in priprava izhodiščnih datotek, čiščenje besedil in priprava vhodnih datotek za korpusno orodje. Besedila, ki smo jih izbrali, so bila glede na osnovne zahteve projekta že vsa v elektronski obliki in prosto dostopna na spletu. Črpali smo jih iz treh virov: starejši arhiv¹⁵ revije *Knjižnica* na spletišču ZBDS, novejši arhiv¹⁶ revije *Knjižnica* v sistemu Open Journal Systems in na arhivskem portalu dLib.si.¹⁷ Besedila so objavljena v datotekah različnih verzij formata pdf, ki praviloma niso zaščitene, redke med njimi pa so tudi zaščitene. Šele novejša postavitev¹⁸ portala dLib.si ponuja izbiro med datotekami formata pdf ali besedilnega formata txt, zato so bila vsa besedila prevzeta v datotečnem formatu pdf že pred tem. Prvi korak pretvorbe predstavlja ekstrakcija besedila s posebnim modulom Adobe Export PDF in preverjanje kakovosti pridobljenega besedila, predvsem njegove integralnosti, splošne čitljivosti in zapisa znakov č, š in ž. Besedilo se nato shrani v datotečnem formatu Word in poimenuje po vnaprej določenem sistemu tako, da ga je kasneje mogoče nezamenljivo identificirati; vzporedno se shranjuje delovne in arhivske kopije.

4.2.3 Čiščenje besedil

Čiščenje besedil je najbolj zamudna faza priprave korpusa, ki jo je treba za vsako od besedil opraviti ročno in ki odločilno vpliva na kakovost korpusa in rezultate korpusnih analiz. Najprej se iz besedila članka odstrani elemente, ki so v besedilih standardni in ki v korpus ne sodijo. To so angleški prevod naslova, angleške ključne besede in izvleček v angleščini. Sledijo sezname virov in literature, ki bi

¹³ Marec 2017.

¹⁴ Pridobljeno 1. 3. 2017 s spletne strani: <http://www.dlib.si/results/?query='srel%3DKnji%C5%BEnica'&pageSize=25>.

¹⁵ Pridobljeno 17. 11. 2016 s spletne strani: <http://revija-knjiznica.zbds-zveza.si/izvodi.html>.

¹⁶ Pridobljeno 17. 11. 2016 s spletne strani: <http://knjiznica.zbds-zveza.si/index.php/knjiznica/issue/archive>.

¹⁷ Pridobljeno 17. 11. 2016 s spletne strani: <http://www.dlib.si/>.

¹⁸ November 2016.

lahko s svojimi podatki bistveno vplivali na pogostnost nekaterih besed. Po podatkih nekaterih avtorjev bibliografskih in bibliometričnih analiz člankov, objavljenih v reviji *Knjižnica* (Vratarič, 2006; Česnik Ušaj, 2011; Kavčič-Čolić in Vodopivec, 2015), je bil v posameznih obdobjih delež citirane tuje literature od 36 % do 51 %, kar bi vneslo v korpus zelo veliko število neslovenskih besed. Standardna je tudi pagina viva s podatki o reviji, avtorju in naslovu članka, ki se v slovenskem in angleškem jeziku ponavlja na vsaki strani, in prav tako »smeti« gradivo za korpus. Isto velja za podatke o avtorju in njegovi instituciji. Ob samem pregledovanju besedila pritegnejo pozornost še nekateri drugi, nestandardni nezaželeni tujki, npr. daljše opombe ali citati v tujem jeziku in celo drugi pisavi (naleteli smo na dele besedila v cirilici in tudi grški pisavi), večje število URL-jev, e-poštnih naslovov ali deli besedila, ki jih Adobe konverter ni znal pravilno interpretirati in kjer so znaki neprepoznavni ali napačni. Tak je bil primer določenega števila letnikov, kjer je oblikovalec uporabljal nestandardno družino pisav in so bili poudarjeni (mastni) deli besedila, predvsem naslovi člankov in poglavij, precej deformirani in nečitljivi. Kljub tovrstnemu čiščenju besedil prihaja pri podatkih iz korpusa še vedno do določene stopnje korpusnega šuma. To so predvsem enkratni ali ponavljajoči se koščki besedil, ki jih je pred tem težko prepoznati in ki so iz različnih razlogov v korpusu vse pogostejši (o tem več v poglavju o pogostnosti besed, ker je v frekvenčnih seznamih te napake najlažje identificirati). Po zaključenem čiščenju besedila sledi še shranjevanje v datotečnem formatu txt, ki smo ga izbrali za vhodni format korpusa.

4.2.4 Generiranje korpusa

Pred vzpostavljanjem novega korpusa si je treba na portalu Sketch Engine zagotoviti ustrezen pomnilniški prostor glede na predvideno število besed. Želeli smo izvesti nekatere časovno pogojene analize besedil in primerjave med članki, objavljenimi v dveh zaporednih dekadah, zato smo zahtevali dvojno količino predvidenega prostora in namestili dve verziji korpusa. Enkrat celoten korpus za obdobje 1997–2016, drugič pa dva segmenta po deset let, prvega s članki 1997–2006 in drugega 2007–2016. Nastaviti je treba tudi nekatere spremenljivke glede jezika besedil, pravil lematizacije, oblikovnega in skladsenjskega označevanja itd., vendar je za potrebe korpusa zadoščalo posnemanje večine standardnih nastavitev drugih slovenskih korpusov (*FidaPLUS*). Sketch Engine lahko sprejme datoteke v različnih formatih; odločili smo se za besedilni format txt. Sistem omogoča nalaganje posameznih datotek, tudi dodajanje posamezne datoteke že delujočemu korpusu, vendar smo se zaradi velikega števila datotek (553) odločili za njihovo združevanje in nalaganje v enem kosu.

4.3 Uporaba in funkcije korpusa

Orodje za besedilni korpus je v celoti spletna aplikacija z lastnim oblakom za shranjevanje korpusa in delovnih podatkov oziroma rezultatov in ne potrebuje nalaganja nobenih komponent na uporabnikov računalnik, prav tako ni nobenih omejitev glede na vrsto in/ali verzijo spletnega brskalnika. Orodje Sketch Engine je dovolj prilagojeno, da lahko tudi v slovenskih besedilih identificira in analizira kolokacije, do neke mere tudi sopomenke in protipomenke, primere rabe v sobesedilu, ključne besede in termine, generira sezname pogostnosti besed ali lem, enobesednih ali večbesednih terminov, osnovna funkcija pa je seveda izdelava konkordanc (Slika 1), to je prikaza iskane besede ali besedne zveze v sobesedilu. Iskana beseda je grafično izpostavljena in ilustrira rabo v različnih situacijah in v povezavi z drugimi besedami v stavku. V bibliotekarstvu je bil *KWIC-indeks* poimenovanje za tovrstno kazalo že dolgo pred korpusi.

Query obogaten, resničnost 73 (24.49 per million) ⓘ

First | Previous Page 2 of 4 Go Next | Last

1	#2867604	pa je tudi naslednja, novejša definicija, ki obogateno resničnost celo defini- ra kot medij: obogatena resničnost
2	#2867613	obogateno resničnost celo defini- ra kot medij: obogatena resničnost je medij, v katerem digitalna informacija prekriva
3	#2867648	v realnem času (Craig, 2013). 2.1 Komponente obogatene resničnosti Osnovne komponente sistemov za obogateno resničnost
4	#2867654	obogatene resničnosti Osnovne komponente sistemov za obogateno resničnost se lahko delijo na dve sku- pini, in sicer na
5	#2867769	spletni uporabniški vnesnik posamezne storitve obogatene resničnosti), • server z digitalnimi vsebinami. 2.2 vrste
6	#2867781	, • server z digitalnimi vsebinami. 2.2 vrste obogatene resničnosti Obogatena resničnost se lahko deli po več različnih

Slika 1: Osnovna funkcija – konkordance ali KWIC-indeks

Zelo zanimiva in koristna funkcija je prikaz besedne skice (angl. word sketch), to je primerjalnih tabel kolokacij (sopojavnosti besed) in značilnosti delovanja dveh semantično povezanih besed v besedilu: v čem so podobnosti in v čem razlike. Sopomenke imajo na primer praviloma vedno nekaj skupnih kolokatorjev (sopojavnih besed), vendar ne vseh. Primerjava besednih skic zelo ilustrativno in z barvno shemo prikaže, s katerimi besedami se povezujeta sopomenki in s katerimi samo ena ali samo druga. V slovenščini je to na primer dobro pomagalo za analizo parov sopomenk, saj je ena od sopomenk pogosto domačega, druga pa tujega izvora, in besedna skica nazorno prikaže, kateri kolokatorji so jima skupni in kateri ne; iz tega lahko sklepamo, koliko sta v enem pomenu res sopomenki, v drugem pa morebiti ne.

Ilustracija prikazuje primerjavo besednih skic za dva para terminov – samostalnika *podatek* in *informacija* ter pridevnika *digitalen* in *elektronski* (Slika 2). Zaradi preglednosti primera smo izdvojili samo manjši del sicer mnogo obsežnejše skice. Barvna shema nazorno kaže, s katerimi besedami se najpogosteje povezuje prvi ali drugi termin. *Podatek* (zeleno) najpogosteje dopolnjujejo pridevniki, ki so v temno zelenem delu, manj pogosto tisti svetlo zeleni; *informacijo* (rdeče)

sami primerjamo in ovrednotimo z besednimi skicami (za publikacijo npr. ponudi podobne besede: revija, dokument, gradivo, članek, vir, knjiga, delo, literatura ... s pogostnostjo pojavljanja v korpusu), ob tem pa sestavi tudi ustrezen oblak pomensko povezanih besed (Slika 3).

V očeh slovaropisca in terminografa je zagotovo najbolj zanimiv luščilnik, to je modul, ki s pomočjo vgrajenih jezikoslovnih navodil iz besedil izlušči sezname enobesednih ali večbesednih terminov. Ti sezname so seveda zgolj groba ocena, kaj bi utegnili biti zanimivi termini, zato zahtevajo temeljito presojo človeka, ki izloči napačne in nepotrebne ter ustrezno popravi izbrane. Tak seznam (Slika 4) ponudi predlog termina s podatkom o njegovi frekvenci v korpusu (F), sama oblika pa potrebuje popravke samo v primeru, da v referenčnem korpusu slovenskega jezika ni našel podobne besedne zveze. Če jo je našel tudi tam (pogostnost v referenčnem korpusu je označena RefF), bo predlagani termin skoraj zagotovo oblikovan pravilno (predvsem ujemanje v spolu, sklonu in številu). To je zelo dobrodošla osnova za prepoznavanje in evidentiranje terminov v novejših besedilih, ki pogosto prinašajo najnovejšo, še ne ustaljeno terminologijo, brez velike mere natančne človekove vsebinske in pomenske presoje in dodatnega dela pa seveda ne gre. V pomoč je gumb z oznako »W«, ki za večino nabranih terminov ponuja neposredne hiperpovezave do strani Wikipedije s pomensko sorodnimi vsebinami (izbor pri terminu *področje bibliotekarstva* je povsem ustrezen).

Multi-word	Score	F	RefF
<input type="checkbox"/> osrednja knjižnica	<input type="checkbox"/> W 63.34	372	975
<input type="checkbox"/> digitalni arhiv	<input type="checkbox"/> W 62.56	221	199
<input type="checkbox"/> zakon o knjižničarstvu	<input type="checkbox"/> W 61.44	230	269
<input type="checkbox"/> informacijsko opismenjevanje	<input type="checkbox"/> W 57.63	225	324
<input type="checkbox"/> splošna knjižnica	<input type="checkbox"/> W 57.22	679	2,964
<input type="checkbox"/> delovanje knjižnic	<input type="checkbox"/> W 55.38	191	173
<input type="checkbox"/> področje knjižničarstva	<input type="checkbox"/> W 55.25	194	194
<input type="checkbox"/> uspešnost delovanja	<input type="checkbox"/> W 53.96	238	493
<input type="checkbox"/> področje bibliotekarstva			
<input type="checkbox"/> oddelek za bibliotekarstvo			
<input type="checkbox"/> iskanje informacij			
<input type="checkbox"/> knjižnično gradivo			
<input type="checkbox"/> knjižnična dejavnost			
<input type="checkbox"/> potreba uporabnikov	<input type="checkbox"/> W 34.09	242	1,304
<input type="checkbox"/> baza podatkov	<input type="checkbox"/> W 14.57	333	6,659
<input type="checkbox"/> mladinska knjiga	<input type="checkbox"/> W 13.34	204	4,158
<input type="checkbox"/> visoko šolstvo	<input type="checkbox"/> W 5.42	223	12,825
<input type="checkbox"/> kulturna dediščina	<input type="checkbox"/> W 3.50	281	25,894

Related Wikipedia articles

[Oddelek za bibliotekarstvo, informacijsko znanost in knjigarstvo, Filozofska fakulteta v Ljubljani](#)

[Mirko Popovič](#)

[Informacijsko-komunikacijska tehnologija](#)

[Narodna in univerzitetna knjižnica](#)

[Zveza bibliotekarskih društev Slovenije](#)

Slika 4: Avtomatska ekstrakcija večbesednih terminov


S kombinacijo uporabe konkordanc, kolokacij, besednih seznamov in skic ter tezavra ni težko pripraviti pregleda frazeološkega gnezda v povezavi z izbranim nosilnim terminom. Tak je na primer seznam sestavljenih terminov, ki so nastali okrog jedra **obdelava**, ta je po pogostosti na 241. mestu:

- avtomatska obdelava
- bibliografska obdelava
- enotna obdelava
- enotna obdelava knjižničnega gradiva
- formalna obdelava
- kakovost vsebinske obdelave
- kataloška obdelava
- nabava in obdelava
- način obdelave
- obdelava
- obdelava bibliografskih podatkov
- obdelava dokumentov
- obdelava gradiva
- obdelava in hranjenje
- obdelava in strukturiranje
- obdelava knjižničnega gradiva
- obdelava leposlovja
- obdelava podatkov
- obdelava serijskih publikacij
- obdelava v sistemu
- področje vsebinske obdelave
- postopki obdelave
- računalniška obdelava
- statistična obdelava
- strokovna obdelava
- strokovna obdelava knjižničnega gradiva
- v obdelavi
- vsebinska obdelava
- vsebinska obdelava gradiva
- vsebinska obdelava leposlovja

5 Rezultati raziskave in razprava

Po nalaganju besedil v sistem, opravljenih indeksiranjih in oblikovno-skladenjskem označevanju je korpus pripravljen za uporabo. Glede na njegovo relativno majhnost tehnični postopki končne vzpostavitve niso bili dolgotrajni. V sistemu so glavne značilnosti korpusa člankov iz revije *Knjižnica* predstavljene v poenostavljeni tabeli (Slika 5).

Knjižnica

Counts		General info		Lexicon sizes		Tags legend	
Tokens	2,981,005	Language	Slovenian	word	146,245	noun	S.* 
Words	2,414,011	Encoding	UTF-8	tag	1,212	verb	G.*
Sentences	97,279	Compiled	03/02/2017 08:20:17	lempos	74,967	adjective	P.*
Documents	4	Tagset	Description	gender_lemma	27,295	adverb	R.*
		Word sketch grammar	Definition	lemma	72,148	pronoun	Z.*

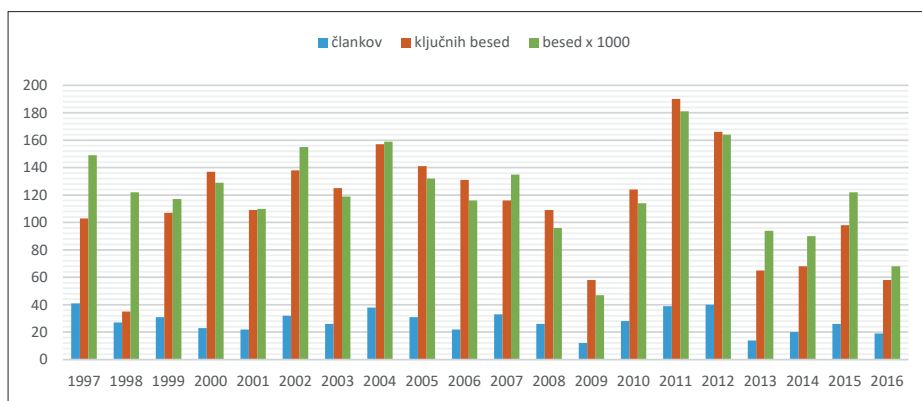
Slika 5: Osnovni podatki o strukturi in obsegu korpusa *Knjižnica* (Sketch Engine, izrez dela tabele)

Razpolagamo torej z gradivom, ki izhaja iz 533 slovenskih strokovnih in znanstvenih člankov, vsebuje skupaj okrog 2,4 milijona besed, od tega 72.000 različnih

besed, mnoge od njih bomo kasneje prepoznali kot termine ali elemente večbesednih terminov. Besede so besednovrstno označene in kasneje bomo obravnavali tudi delež besednih vrst (glej poglavje 5.3) in nekatere kolokacije med njimi.

5.1 Ključne besede

Vzporedno s pripravo korpusa smo zbrali tudi ključne besede, ki jih člankom praviloma določijo avtorji sami. Ključne besede so najboljši pokazatelj, kaj in kako pogosto so v posameznih obdobjih pisali slovenski strokovnjaki na področju bibliotekarstva. Od 553 zajetih člankov je bilo 466 člankov opremljenih s ključnimi besedami, saj predvsem v začetnem obdobju večina člankov ključnih besed ni imela, od leta 2000 naprej pa so praviloma obvezen del opreme članka. Ob tem je bilo zajetih tudi nekaj člankov strokovno-informativne vsebine, ki prav tako niso opremljeni s ključnimi besedami.



Slika 6: Število člankov, avtorskih ključnih besed in besed v besedilih, objavljenih v reviji *Knjižnica* med letoma 1997 in 2016 (letniki od 41 do 60)

Skupaj je bilo evidentirano 2246 ključnih besed ali povprečno po 4,8 ključne besede na članek. Zanimiv je vpogled v abecedni seznam vseh ključnih besed, ki kaže na njihovo relativno neenotnost, saj zanje ni normativnega seznama in so jih avtorji oblikovali po svoji presoji. Značilna je neenotnost glede

- uporabe ednine in množine, pojavljata se obe obliki, npr. anketa – ankete, bibliografija – bibliografije, knjižnična zbirka – knjižnične zbirke, knjižničar – knjižničarji, splošna knjižnica – splošne knjižnice itd.;
- uporabe sopomenk, npr. geslenje – gesljenje, kakovost – kvaliteta, marketing – trženje, management – menedžment, multikulturalizem – večkulturalnost, vrednotenje – evalvacija – ovrednotenje itd.

Ob tem kaže razmisliti, ali ne bi uredništvo revije *Knjižnica* avtorjem predlagalo uporabe svojega ali katerega od uveljavljenih geslovnikov ali tezavrov.

V naboru 2.246 uporabljenih ključnih besed je 1.238 različnih, med temi so seveda tudi zgoraj omenjene dvojnice, vendar taka razporeditev kljub temu kaže na bodisi izredno pestrost objav ali pa šibko rabo relevantnih ključnih besed. To je še toliko bolj očitno ob dejstvu, da se le 319 ključnih besed pojavlja več kot enkrat oziroma samo 148 več kot dvakrat, zgolj 18 ključnih besed pa presega mejo desetih pojavitev. Med enkratnicami je tudi nekaj imen in časovnih oznak, vendar je takih skupaj le kake tri ducate. Tolikšna disperzija relevantnih ključnih besed je nenavadna in z vidika sistema za poizvedovanje tudi problematična. Razporeditev najpogostejših 50 ključnih besed v posameznih petletnih obdobjih in kumulativno kažeta Preglednica 3 in Slika 7 v oblaku besed.

Preglednica 2: *Trend naraščanja, upadanja ali nihanje pogostnosti uporabe nekaterih izbranih ključnih besed ali njihovih delov – članki revije Knjižnica v štirih zaporednih obdobjih*

Ključna beseda	1997–2001	2002–2006	2007–2011	2012–2016
digitalen *	2	5	22	24
elektronski *	4	7	11	2
izobraževanje *	2	7	3	4
* uporabnik	14	17	5	5
iskanje informacij	8	1	1	0
* storitve	3	3	4	3
informacijska pismenost	3	5	2	1
založništvo, založniški	3	12	5	3
katalog	9	25	11	8

Med ključnimi besedami z najnižjim številom pojavitev so presenetljivo tudi e-knjige, FZBZ/FRBR, klasifikacija, matičnost, medknjižnična izposoja, meta-podatki, (narodne) manjšine, normativna kontrola, OA/odprti dostop, oddaljeni dostop, potujoče knjižnice, RDA, RFID itd., čeprav se nekateri od teh terminov z relativno nizko frekvenco v besedilih vendarle pojavljajo.

Preglednica 3: *Najpogostejših 50 ključnih besed v člankih po petletnih obdobjih in kumulativno 1997–2016*

Rang	1997–2001	2002–2006	2007–2011	2012–2016	1997–2016	Fre
1	knjižnice	Slovenija	Slovenija	splošne knjižnice	Slovenija	66
2	Slovenija	splošne knjižnice	splošne knjižnice	digitalizacija	splošne knjižnice	41
3	visokošolske knjižnice	bibliotekarji	knjižnice	Slovenija	knjižnice	35

Rang	1997–2001	2002–2006	2007–2011	2012–2016	1997–2016	Fre
4	bibliotekarstvo	biografije	obvezni izvod	kulturna zgodovina	bibliotekarji	25
5	iskanje informacij	katalogizacija	biografije	kulturna dediščina	visokošolske knjižnice	25
6	bibliotekarji	visokošolske knjižnice	Narodna in univerzitetna knjižnica	zasebne knjižnice	bibliotekarstvo	22
7	splošnoizobraževalne knjižnice	bibliotekarstvo	bibliotekarji	knjižne zbirke	digitalizacija	20
8	šolske knjižnice	knjižnice	digitalizacija	COBISS	katalogizacija	20
9	uporabniki	Narodna in univerzitetna knjižnica	digitalni arhivi	Dobreknjige.si	biografije	18
10	knjižnični fond	uporabniki	katalogizacija	knjižnice	Narodna in univerzitetna knjižnica	17
11	standardi	specialne knjižnice	uvodnik	razvoj	uporabniki	15
12	univerzitetne knjižnice	informatijska pismenost	bibliotekarstvo in informatijska znanost	arhivsko gradivo	knjižničarji	13
13	dijaki	založništvo	informatijski sistemi	bralna kultura	knjižnični fond	12
14	evalvacija	izobraževanje uporabnikov	Knjižnica	inventura	statistični podatki	12
15	kakovost	knjižničarji	poizvedovanje	leposlovje	informatijska pismenost	11
16	knjižničarji	knjižnični fond	revije	Narodna in univerzitetna knjižnica	standardi	11
17	marketing	serijske publikacije	statistični podatki	statistični podatki	šolske knjižnice	11
18	osnovne šole	spletne strani	trajno ohranjanje	analiza citiranja	založništvo	11
19	bibliometrija	bibliografije	visokošolske knjižnice	bibliografije	iskanje informacij	10
20	citiranje	bibliografski opis	zgodovina	bibliografski elementi	zasebne knjižnice	10
21	Evropa	Centralna tehniška knjižnica	bibliotekarstvo	bibliotekarstvo	knjige	9
22	informatijska pismenost	domoznanske zbirke	bralna kultura	branje	knjižnične storitve	9
23	Internet	elektronski viri	branje	brata Grimm	knjižnični katalogi	9
24	izobraževanje	Evropa	digitalne knjižnice	Digitalna knjižnica Slovenije	kulturna zgodovina	9

Rang	1997–2001	2002–2006	2007–2011	2012–2016	1997–2016	Fre
25	katalogizacija	Evropska unija	digitalni viri	digitalne zbirke	specialne knjižnice	9
26	knjižnična informacijska znanja	integrirni viri	elektronske knjige	informacijska znanost	bralna kultura	8
27	mladi raziskovalci	katalogizacijska pravila	knjige	izbrani članki	evalvacija	8
28	statistični podatki	kontinuirani viri	knjižničarji	knjige	izobraževanje	8
29	šolska knjižnica	Ljubljana	knjižnične storitve	knjižničarji	kakovost	8
30	zaposlovanje	medicinske knjižnice	knjižnične zbirke	knjižnične evidence	knjižničarstvo	8
31	antropologija	nabavna politika	knjižnični katalogi	knjižnične storitve	kulturna dediščina	8
32	avtorsko pravo	organizacija znanja	načrtovanje	knjižnični katalogi	obvezni izvod	8
33	bibliografski opis	posebne zbirke	sistemi za poizvedovanje	kulturna politika	univerzitetne knjižnice	8
34	digitalne knjižnice	strokovne revije	standardi	Maistrova knjižnica	vsebinska obdelava	8
35	domače strani	šolske knjižnice	vsebinska obdelava	Mestna knjižnica Ljubljana	COBISS	7
36	elektronske publikacije	univerzitetne knjižnice	barok	odprti dostop	digitalne knjižnice	7
37	elektronske revije	vrednotenje	dokumentalistika	osebne knjižnice	digitalni arhivi	7
38	etnologija	učbeniki	dolgoročna hramba	razstave	razvoj	7
39	geslenje	zgodovinski pregledi	elektronski viri	Scopus	serijske publikacije	7
40	gimnazija	arhiviranje	evalvacija	SICRIS	splošnoizobraževalne knjižnice	7
41	gradivo	avtomatizacija knjižnic	Filip Terpin	spletni portali	trajno ohranjanje	7
42	internet	avtorske pravice	Francija	založništvo	zgodovina	7
43	izposoja	bibliografska družina	Gornji Grad	zgodovinski pregledi	zgodovinski pregledi	7
44	kariera	bibliografska entiteta	informacijska pismenost	znanstveno publiciranje	bibliografije	6
45	katalogizacijska pravila	bibliografski odnosi	izobraževanje	anotacije	bibliometrija	6
46	kazalci uspešnosti	bibliografski zapisi	izposoja	Apple iBooks	elektronske knjige	6

Rang	1997–2001	2002–2006	2007–2011	2012–2016	1997–2016	Fre
47	knjige	bibliometrija	Janez Ludvik Schönleben	arheologija	Evropa	6
48	knjižničarstvo	bralna kultura	klasifikacija	arhitekturni posnetek stanja	izobraževanje uporabnikov	6
49	knjižnično gradivo	COBISS	knjižničarstvo	Aurasma	leposlovje	6
50	komunikacije	delo	knjižnični fond	avtorsko parvo	Ljubljana	6

Če izvzamemo besedo Slovenija, ki se pojavlja v 14 % člankov in v tem kontekstu nima velike informacijske vrednosti, so gotovo najpogostejše obravnavane splošne knjižnice (41), kamor moramo dodati še splošnoizobraževalne knjižnice (7), slovenske splošne knjižnice (2) in splošna knjižnica (1) (z novim Zakonom o knjižničarstvu je prišlo do formalnega preimenovanja). Zelo v ospredju so v vseh štirih petletnih obdobjih in število objav o njih od sedmega mesta narašča povsem v ospredje. Takoj na drugem mestu so knjižnice brez opredelitve, kar z vidika vsebinskega označevanja v zbirki o knjižničarstvu ni dobra oznaka. Do 25. mesta se zvrsti vseh pet tipov knjižnic, ob tem gre večji poudarek tudi strokovnim delavcem (bibliotekarji – 25, knjižničarji – 13) in nekaterim dejavnostim (bibliotekarstvo – 22, digitalizacija – 20, katalogizacija – 20, iskanje informacij – 10, evalvacija – 8, izobraževanje – 8) ter nekaterim drugim temam (biografije – 18, uporabniki – 15, knjižnični fond – 12, standardi – 11, knjižnične storitve – 9).



Slika 7: Oblak ključnih besed, uporabljenih v obdobju 1997–2016

5.2 Kako bogat je bibliotekarski besedni zaklad?

Po besedah dr. Primoža Jakopina, takrat vodje Laboratorija za korpus slovenskega jezika pri ZRC SAZU, obsega besedni zaklad slovenskega jezika okrog pol milijona besed (Todorovski, 2008). Gre za besede splošne slovenščine, ki jih najdemo v *Slovarju slovenskega knjižnega jezika* (93.500 gesel¹⁹), v Besedišču slovenskega jezika (178.000 gesel, ki jih niso uvrstili v slovar), ter druge besede, ki so se v minulih dveh desetletjih po nastanku slovarja pojavile v periodičnem tisku in knjigah, na televiziji, v diplomskih in magistrskih nalogah ter doktorskih disertacijah, pa še v drugih publikacijah, na primer letakih in reklamah, ter na internetu. V praksi pa predstavlja najpogostejših 1000 besed več kot 50 % v vsakodnevni slovenščini uporabljenih besed. Velika večina besed se pojavlja precej redko, večkrat tudi v omejeni skupini ljudi, taka so na primer strokovna besedišča.

5.3 Pogostnost besed in besedne vrste

Zato nas zanima, kakšen je jezik strokovnega komuniciranja v bibliotekarstvu, kolikšen in kako pester besedni zaklad uporabljamo, katere so najpogostejše besede, predvsem pa, kako je z bibliotekarskimi termini, ki jih avtorji uporabljajo v obravnavanih besedilih.

- V tem kontekstu je *osnovna enota besedila* vsak niz znakov v besedilu, ki je na obeh straneh razmejen s presledkom ali ločilom. Sem torej sodijo vse številke, oznake poglavij in tudi deli besed, če so te (npr. na koncu vrstice) deljene z vezajem itd. Take jezikoslovci imenujejo pojavnice (angl. token). Teh je v korpusu 2.981.000.
- Po ustreznih obdelavah in »čiščenju« besednjaka ostane 2.414.000 pravih besed (v našem primeru od »A« do »žvižgajoč«).
- Mnoge med njimi se ponavljajo, zato je po združevanju enakih oblik in izračunu njihove pogostnosti ostalo v slovarju korpusa še 146.245 različnih oblik besed ali različnic.
- Svoje k raznovrstnosti seveda prispeva tudi značilnost slovenščine, da je pregibni jezik. Bogastvo in kompleksnost jezika glede na pregibnost se izraža v tem, da je število izpeljanih oblik, ki jih imajo leme v povprečju, relativno veliko. Lema je kanonična, to je osnovna oblika besede (npr. prvi sklon ednine pri samostalniku, nedoločnik pri glagolu itd. – lema za besede »knjižničnega, knjižničnih, knjižničnemu, knjižničnim ...« je »knjižničen«, za »grem, greva, pojdiva, šli, šla, šle ...« pa »iti« itd.) iz katere z uporabo končnic nastanejo izpeljane oblike. Glede na bogastvo sklonov in števil ima na primer velika večina

¹⁹ K 93.148 slovarskim sestavkom prve izdaje je bilo v SSKJ² dodanih še 4521 novih (Snoj, 2014).

samostalnikov mnogo oblik. Šele z lematizacijo, to je računalniškim postopkom označevanja različnic in njihovim spreminjanjem v lemo, to je osnovno slovarsko obliko besede, je nastal pravi nabor vseh slovarsko različnih besed, ki so bile uporabljene v obravnavanih besedilih. Korpus vsebuje 72.148 lem.

- Pri tem je treba upoštevati, da je ostalo kljub »ročnemu čiščenju« še mnogo pravih besed iz tujejezičnih besedil (npr. navedki besedil, citirani naslovi itd.), imen in tudi besed, ki jih lematizator ni prepoznal in zato ni znal lematizirati, pa tudi neodstranjenih napak, npr. zatipkane ali iz formata pdf napačno prebrane besede, z vezajem deljene besede na koncu vrstice so razpadle v dve »besedi«, tujke (npr. daguerrotipija, kiberknjižnica, čekiranje, displej itd.), z vezajem sklanjane kratice (CD-ja) in še marsikaj nepredvidljivega. Število pravih besed slovenskega jezika v korpusu je zato mnogo manjše.
- Vse leme so tudi oblikoslovno-skladenjsko označene, zaradi česar je mogoče izvajati analize tudi v tem smislu. Uporabljen je bil označevalnik, ki ga orodje Sketch Engine že standardno ponuja za besedila v slovenskem jeziku, to je statistični označevalnik Obeliks.²⁰ Pripadnost posameznim besednim vrstam je korpusno orodje preštelo takole:

46.735	samostalnik	74	veznik
6.073	glagol	73	členek
17.714	pridevnik	48	medmet
2.377	prilastek	243	števnik
170	zaimek	118	okrajšava
81	predlog	259	drugo

Pogostnost uporabe besed v analiziranih 553 besedilih je seveda zelo različna, v skladu s pričakovanji vodi pomožni glagol **biti** s 141.189 pojavitvami, v vodilni skupini stotih mu sledijo predvsem funkcijske besede, ki bi v sistemih za poizvedovanje sodile med blokirane besede in informacijsko za sporočilnost besedila niso pomembne, prav tako seveda ne za stroko. Med praviimi bibliotekarskimi termini vodijo **knjižnica** (7. najpogostejša beseda s 33.712 pojavitvami), **gradivo** (20. mesto z 11.387 pojavitvami), **knjiga** (23. mesto z 9.698 pojavitvami) in **uporabnik** (26. najpogostejša beseda z 8.549 pojavitvami). Med petdesetimi najpogostejšimi besedami je tako 13 terminov, preostalo so funkcijske besede. Ta ugotovitev v celoti velja tudi za *Korpus bibliotekarstva* s 3,6 milijona besedami (glej Preglednica 8).

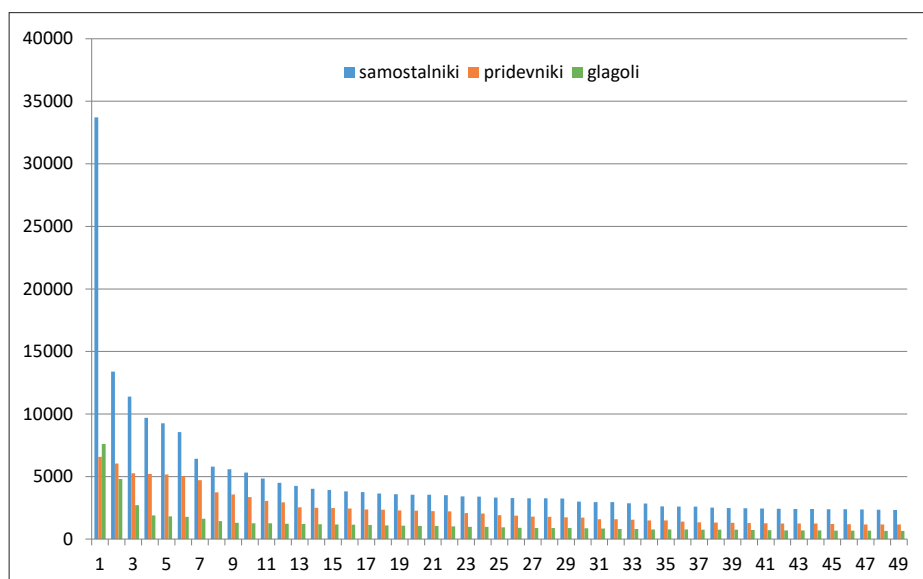
²⁰ Označevalnik. Pridobljeno 24. 8. 2016 s spletne strani: <http://www.slovenscina.eu/tehnologije/oznacevalnik>.

Preglednica 4: *Petdeset najpogostejših samostalnikov, pridevnikov in glagolov*

Zap. št.	Samostalniki	Fre.	Pridevniki	Fre.	Glagoli	Fre.
1	knjižnica	33712	knjižničen	6569	biti	141189
2	leto	13394	slovenski	6046	imeti	7593
3	gradivo	11387	velik	5263	morati	4802
4	knjiga	9698	informacijski	5204	uporabljati	2700
5	delo	9252	nov	5160	predstavljati	1894
6	uporabnik	8549	strokoven	5035	omogočati	1818
7	podatek	6420	različen	4716	iti	1772
8	zbirka	5787	splošen	3746	pomeniti	1625
9	sistem	5583	pomemben	3563	kazati	1437
10	področje	5312	posamezen	3358	začeti	1301
11	informacija	4837	elektronski	3057	najti	1258
12	vir	4491	določen	2934	dobiti	1254
13	uporaba	4249	univerziteten	2529	želeti	1226
14	del	4030	nacionalen	2492	ugotoviti	1208
15	oblika	3929	javen	2489	vsebovati	1181
16	katalog	3809	spleten	2443	moči	1166
17	publikacija	3755	potreben	2361	delovati	1157
18	primer	3646	bibliografski	2346	postati	1127
19	vsebina	3594	digitalen	2286	slediti	1087
20	naslov	3550	znanstven	2278	veljati	1071
21	število	3546	knjižen	2230	upoštevati	1049
22	knjižničar	3501	poseben	2206	uporabiti	1046
23	dejavnost	3412	kulturen	2089	izvajati	1007
24	stran	3403	sam	2040	vplivati	986
25	razvoj	3313	dober	1919	pokazati	970
26	Slovenija	3276	zaposlen	1866	navajati	934
27	avtor	3263	raziskovalen	1796	zahtevati	902
28	storitev	3262	deloven	1784	vkjučevati	876
29	čas	3246	visok	1733	potekati	876
30	raziskava	3003	visokošolski	1726	sodelovati	874
31	potreba	2955	šolski	1579	pridobiti	860
32	način	2949	ustrezen	1578	obstajati	847
33	zapis	2863	osnoven	1555	nastati	814
34	skupina	2850	majhen	1499	priti	813
35	članek	2606	star	1493	iskati	768
36	rezultat	2599	dostopen	1403	potrebovati	767
37	program	2599	vsebinski	1333	poznati	767
38	iskanje	2524	naslednji	1323	opravljati	754
39	enota	2472	zadnji	1296	voditi	742

Zap. št.	Samostalniki	Fre.	Pridevniki	Fre.	Glagoli	Fre.
40	slika	2458	mogoč	1274	govoriti	742
41	revija	2438	mednaroden	1267	pripraviti	728
42	organizacija	2424	bibliotekarski	1249	izhajati	714
43	vrsta	2408	tiskan	1238	oblikovati	690
44	mesto	2403	celoten	1235	obravnavati	689
45	dokument	2391	glaven	1197	pomagati	689
46	država	2386	skupen	1184	določati	674
47	naloga	2370	podoben	1174	hraniti	669
48	znanje	2341	računalniški	1172	ugotavljati	668
49	Ljubljana	2325	državen	1158	delati	659
50	dostop	2252	lokalen	1131	vedeti	651

Pogostnost strmo pada in beseda na 90. mestu že ima pogostnost manjšo od 3000, manj kot deset tisoč besed se pojavlja po desetkrat. Pogostnost 2 ali več ima **33.247 besed** in po pregledu preostalih enkratnic bi lahko trdili, da smo iz analiziranih besedil izluščili okrog 34.000 različnih slovenskih besed, mnoge so seveda samostojni termini ali deli terminov bibliotekarske terminologije. Luščenje teh enobesednih ali večbesednih terminov bo terminografsko in slovaropisno delo, zaradi katerega smo predstavljeni korpus člankov tudi vzpostavili.



Slika 8: Primerjava razporeditve najpogostejših samostalnikov, pridevnikov in glagolov (pomožni glagol biti ni upoštevan)

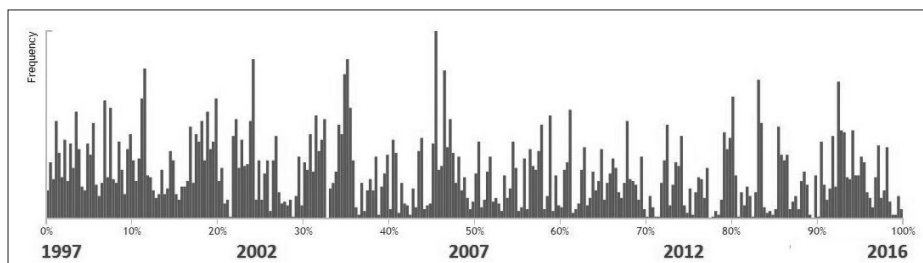
5.4 Pogostnost besed in njihova distribucija

Besede v korpusu lahko glede na njihovo pogostnost pojavljanja razvrstimo v tri skupine:

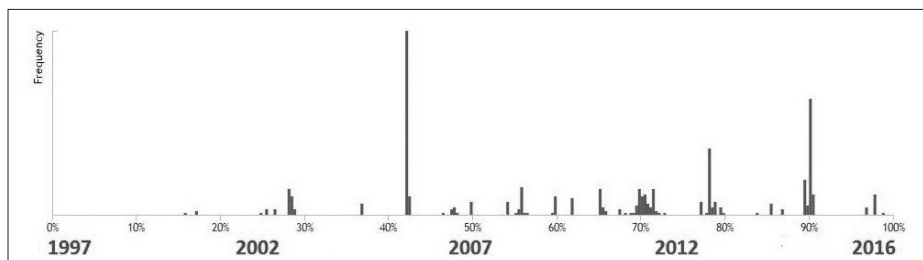
- *Zelo pogoste besede*, ki z redkimi izjemami ne prispevajo k strokovni oziroma informacijski sporočilni vrednosti besedila, med njimi so najpogosteje funkcijske besede, ki so v vseh besedilih na vrhu pogostnosti. V tej skupini ni dosti različnih besed, navadno sto ali nekaj več.
- *Zelo redke besede*, sem bi uvrstili tiste s pogostnostjo, manjšo od 5. Tudi te razen izjem ne prispevajo k strokovni oziroma informacijski sporočilni vrednosti besedila, med njimi je mnogo imen, tujih besed ali celo napak, nadaljujejo pa se v dolgi rep besed s pogostnostjo 1. V to skupino med drugim prištevamo pomensko zanimive besede, lahko tudi termine, ki se imenujejo *hapax legomena* in jim je posvečeno posebno poglavje (glej poglavje 5.4.3). Za njihov izbor je potrebno obsežno ročno delo in pregled, v našem korpusu je v skupini besed s frekvenco 1 in 2 okrog 47.000 besed ali njihovih fragmentov.
- Vmes je relativno ozko območje besed, ki so najpomembnejše nosilke vsebine in v našem primeru tudi najresnejše kandidatke za uvrstitev med strokovne termine. Teh je v korpusu člankov okrog 24.000.

5.4.1 Vzorci pojavljanja in časovna porazdelitev

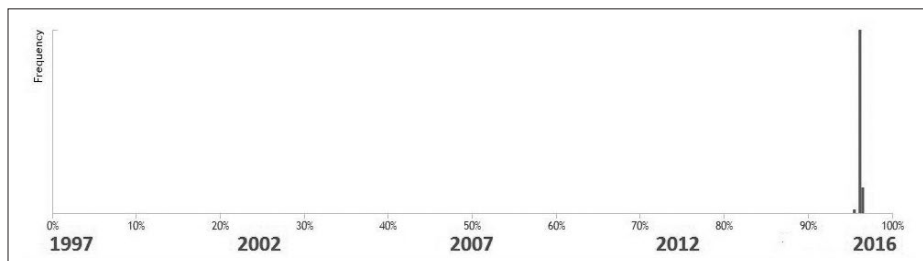
Poleg zelo različne pogostnosti pojavljanja posameznih besed v celotnem korpusu, ki je ni težko izračunati in primerjati v okviru lastnega korpusa ali tudi z drugimi korpusi, lahko pri posameznih besedah in terminih opazimo tudi nekatere značilne vzorce pojavljanja in časovne porazdelitve v obdobju, ki ga korpus obsega. Ti vzorci so v specializiranih korpusih zanimivi predvsem pri besedah in besednih zvezah, ki so prepoznane kot termini strokovnega področja, ker lahko njihovo proučevanje razkrije nekatere podrobnosti o njihovem pomenskem oziroma terminološkem razvoju in tudi razvoju same stroke. Raba termina lahko sčasoma zamre, ker ga nadomesti nov sodobnejši termin, ali pa se stroka neha ukvarjati s strokovnim področjem tega termina in ta iz besedil izgine. S pojavom novih znanj pridejo novi termini, ki jih prej ni bilo, nekateri temeljni termini stroke pa se seveda v besedilih pojavljajo vztrajno, neprestano in časovno dokaj enakomerno porazdeljeno. V nadaljevanju sledi nekaj grafičnih ponazoritev vzorcev pojavljanja izbranih bibliotekarskih terminov (Slike 9–13).



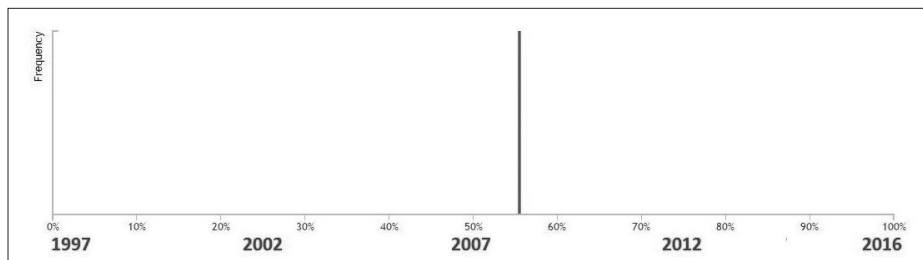
Slika 9: Temeljni termin se pojavlja pogosto, redno in v člankih časovno dokaj enakomerno porazdeljeno – primer: termin **knjižnica** (v proučevanem korpusu člankov frekvenca 35.082).



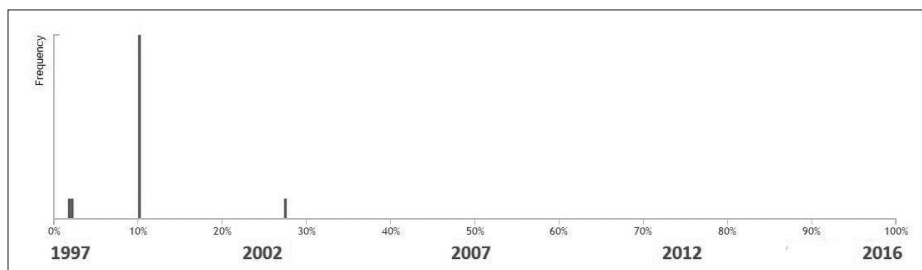
Slika 10: Termin se pojavlja redko, vendar skozi celotno obdobje in skoraj v pravilnih časovnih intervalih – primer: termin **metapodatki** (v proučevanem korpusu člankov frekvenca 24).



Slika 11: Termin se pojavlja redko, samo v manjšem številu člankov in le v zadnjem obdobju – primer: termin – **obogatena resničnost** (v proučevanem korpusu člankov frekvenca 73).



Slika 12: Termin se pojavlja samo v krajšem obdobju in v manjšem številu člankov (tokrat celo v enem samem članku) – primer: termin **obvezni izvod serijskih publikacij** (v proučevanem korpusu člankov frekvenca 6); pojav časovno sledi sprejetju novega zakona leta 2006.



Slika 13: Termin se pojavlja redko, le v manjšem številu člankov in samo v začetku proučevanega obdobja – primer: termin *bibliotekonomija* (v proučevanem korpusu člankov frekvenca 12).

Primerjava dveh časovno razmejenih podkorpusev omogoča vpogled v nabor besed, ki so uporabljene v besedilih enega segmenta, v drugem segmentu pa ne. Korpus smo v ta namen razdelili na obdobji 1997–2006 in 2007–2016. Primerjalna analiza je potrdila hipotezo, da se v novejšem obdobju zaradi razvoja in novih interesnih področij v bibliotekarstvu pojavljajo termini, ki so bili pred tem zelo redki ali jih sploh ni bilo mogoče zaslediti. Nekatere tovrstne razlike so lahko tudi naključne in rezultat interesa posameznega avtorja. Termini s frekvenco pojavljanja so navedeni v Preglednici 5.

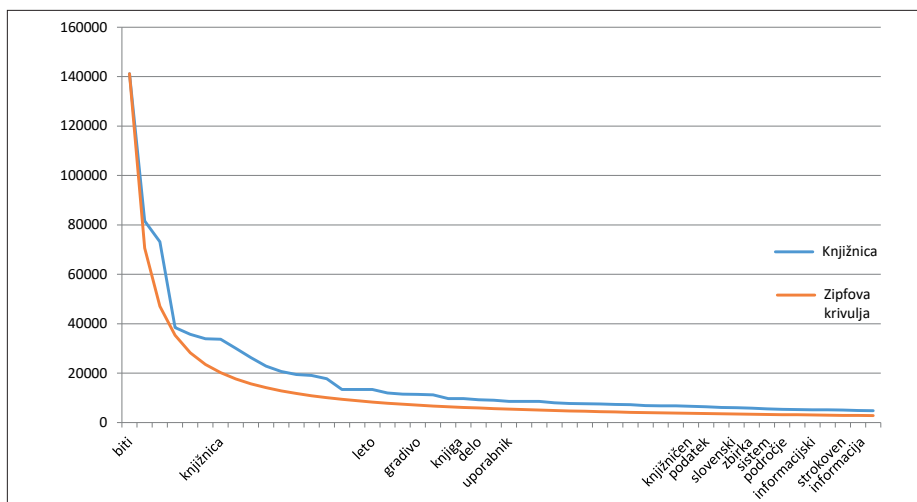
Preglednica 5: Nekaj terminov, ki so se v člankih pojavili šele v zadnjem desetletju (2007–2016), pred tem pa jih ne zasledimo

Termin	2007–2016	1997–2006
inventura	407	9
repozitorij	241	3
samocitat	168	6
vizualizacija	130	5
odprt dostop	113	0
DEDI	104	1
digitalizat	97	0
medkulturen	74	7
obogatena resničnost	73	0
RFID	69	1
knjižnično nadomestilo	60	0
nebralec	56	4
SICRIS	53	1
Kamra	49	6
dokumentarno gradivo	49	5
migrant	46	0
kazalnik	44	0
interoperabilnost	43	2

Termin	2007–2016	1997–2006
samocitiranost	42	0
zaupanja vreden	40	0
ETD	37	0
pravljničar	36	0
spletišče	36	3
e-hramba	34	0
EOD	32	0
e-vir	28	0
blog	27	0
e-vsebina	26	0
dLib	24	0
citaten	22	3
odprtokoden	21	0
skenogram	21	4
brajica	20	1
SVAROG	20	0
RDA	19	0
e-arhiviranje	15	0

5.4.2 Dejanska pogostnost besed v primerjavi s predpostavkami Zipfovega zakona o distribuciji besed v naravnem jeziku

Skokovito upadanje pogostnosti besed v naravnem jeziku je predmet Zipfovega zakona, zato preverimo, koliko ta velja za ugotovitve o pogostnosti pojavljanja besed v korpusu člankov iz revije *Knjižnica*. Zipfov zakon temelji na trditvi, da je majhno število besed uporabljeno zelo pogosto, mnogo drugih ali skoraj vse preostale pa zelo poredko. V svoji prvotni obliki označuje empirično ugotovitev harvardskega jezikoslovca Georja Kingsleya Zipfa, da je v vsakem naravnem jeziku pogostnost n -te najpogostejše uporabljane besede približno recipročno odvisna od n . Klasičen zgled Zipfove funkcije je funkcija $1/f$. Če množico po Zipfovem zakonu porazdeljenih pogostnosti uredimo od najpogostejše do najmanj pogoste, bo pogostnost druge najpogostejše ravno polovica pogostnosti prve, pogostnost tretje najpogostejše pa tretjina pogostnosti prve itd., tako da je pogostnost n -te najpogostejše $1/n$ pogostnosti prve. Zipfov zakon velja predvsem za splošni jezik, strokovna literatura pa seveda od tega odstopa, še toliko bolj, ker gre za ozek nabor strokovnih besedil, kjer nekateri poudarki strokovne terminologije, pa tudi razmišljanja avtorjev, obidejo značilnosti in zakonitosti živega jezika za vsakodnevno komuniciranje. To ponazarja odstopanje teoretične in empirične krivulje na prikazanem grafu (Slika 14), vendar je očitno, da se trenda krivulj dokaj ujemata. V srednjem delu krivulje podatki korpusa odstopajo navzgor, ker je koncentracija nekaterih besed, predvidevamo, da predvsem strokovnih terminov in njihovih stalnih kolokacij, pogostejša od normalne distribucije. Krivulja se v obeh primerih podaljšuje v dolgi rep besed s pogostnostjo 1.



Slika 14: Odstopanje krivulje pogostnosti petdesetih najpogostejših besed v besedilih člankov od teoretične Zipfove krivulje. Krivulja se nadaljuje v dolgi rep manj pogostih besed.

5.4.3 Dolgi rep najmanj pogostih besed

Zelo redke besede, sem bi uvrstili tiste s pogostnostjo, manjšo od 5, se v frekvenčnem kazalu nadaljujejo v dolgi rep besed s pogostnostjo 1. V našem korpusu je okrog 48.000 takih besed ali njihovih fragmentov. **Enkratnica** ali *hapax legomenon* (gr. (kar je bilo) rečeno samo enkrat; mn. *hapax legomena*) je beseda, ki se v določenem besedilu ali pri določenem avtorju pojavi samo enkrat. Termin enkratnica je menda uvedel dr. Jože Toporišič, čeprav njegova *Enciklopedija slovenskega jezika* (Toporišič, 1992) navaja še termin »enkratna beseda«. Pomen enkratnic je različen glede na okolje, v katerem se pojavljajo. Pri prevajanju, predvsem starejših besedil, lahko povzročajo težave, ker jih je težko ali skoraj nemogoče prevesti, saj njihovega pomena ni mogoče ugotavljati s primerjavo besedil, ki bi to besedo prav tako vsebovala.²¹ Nekateri literarni zgodovinarji se na enkratnice opirajo pri ugotavljanju avtentičnosti starejših besedil in potrjevanju ali zavračanju avtorstva.²² Pri podatkovnih zbirkah so enkratnice največkrat balast in nezaželene in nekateri sistemi jih pri indeksiranju izločijo, saj praviloma nimajo teže pri opisovanju vsebine niti velike vrednosti za poizvedovanje in malo je verjetno, da se bo kdo spomnil prav te besede za poizvedovanje, po drugi strani pa tudi obremenjujejo indekse in s tem povečujejo odzivni čas (Kanič, 2011a). Nasprotno pa so v besedilnih korpusih pogosto predmet raziskovanja.

Če so v bibliotekarstvu in informatiki t. i. *blokirane besede* praviloma najpogostejše besede v dokumentu in se jim izogibamo, so na drugem koncu porazdelitvene krivulje enkratnice, to je besede s pogostnostjo 1. Za evidentiranje izrazja, ki sodi v terminološki slovar in tam še ni zajeto, so zanimive tudi enkratnice, ki jih je treba prej temeljito ročno presejati in izbrati le zanimive in pomembne, ki pa se merijo prej v promilih kot v odstotkih. Te izbranke lahko prihajajo z ožjega strokovnega področja, kjer je malo strokovnjakov in/ali literature, pogosto so to manj uporabljane sopomenke ali pa novosti, ki se v literaturi še niso uveljavile.

V besednih indeksih korpusa člankov je kot enkratnica evidentiranih 37.922 pojavnic. Kazala smo se lotili s slovaropisno ambicijo najti nekaj manj znanih ali novih besed, ki pa vendarle sodijo v bibliotekarsko terminologijo. Ta nabor bo lahko osnova za razpravo redaktorjev slovarja, ki bodo presodili slovarsko vrednost in pomembnost posameznih terminov, poiskali morebitne sopomenke ali člane istega frazeološkega gnezda in termine uvrstili med slovarska gesla, ali pa jih kot nepotrebne zavrgli. Mnogi sodijo v skupino zastarelih ali zgodovinsko

²¹ Prav primerjava vzporednih besedil v treh jezikih in pisavah na kamnu iz Rosette je omogočila dešifriranje hieroglifov.

²² Znan primer so besedila W. Shakespearja.

pogojenih izrazov in jih mlajši bibliotekarji več ne poznajo, smo jih pa zasledili v *Bibliotekarskem terminološkem slovarju* s tako klasifikatorsko oznako. Med njimi so tudi ženska poimenovanja, ki so zapostavljena in se v bibliotekarski literaturi skoraj ne uporabljajo (več o tem glej poglavje 5.5.1). Nekaj primerov enkratnic iz prvega dela abecede: akviziterstvo, avtocenzura, avtorica, bibliстика, bibliofobija, broširanje, antifonar, avtograf, bibliofilija, bibliografinja, cardex, cenzuriranje, časnikarstvo, četverka, črkovalnik, desetinka, devede itd.

5.5 Spolno občutljiva raba jezika v bibliotekarskih besedilih

Vse od Asurbanipalovih knjižničarjev preko antičnih aleksandrijskih, srednjeveških samostanskih akademikov do akademikov novega veka pa vse v 19. stoletje so bili varuhi knjig, to je knjižničarji, praviloma moški. Izjeme se bile zelo redke, ena takih, morebiti prva knjižničarka, naj bi bila Hypatia, znanstvenica in filozofinja v Aleksandriji, ki je v nevarnosti prej reševala knjige kot svoje življenje. Prva pomembna slovenska knjižničarka je bila v prvi polovici 20. stoletja izjemna Melita Pivec-Stele, dvojna doktorica in dobitnica Čopove diplome. Kljub tako izključno moškemu poklicu v preteklosti pa vendar danes pogosto govorimo o knjižničarstvu kot *feminiziranem poklicu*. Feminizirani so tisti poklici, v katerih delež spolov ni uravnotežen, kar pomeni, da ni prisotnih vsaj 40 odstotkov pripadnikov drugega, to je moškega spola. Tudi strokovna revija *Knjižnica* je imela v svojih šestdesetih letih obstoja več kot polovico časa urednico in nekaj let manj urednika, trenutno pa je v uredniškem svetu ter domačem in mednarodnem uredniškem svetu skupaj šest strokovnjakov in 14 strokovnjakinj.

Enakopravno obravnavanje delavk in delavcev, strokovnjakov in strokovnjakinj se v spolno zaznamovanih jezikih izraža tudi v ustreznem poimenovanju njihovih poklicev, nazivov v ženskem spolu. Spolno občutljiv jezik oziroma spolno zaznamovan jezik je jezik, ki ima posebna obrazila, s katerimi lahko izraža, ali gre za osebo moškega ali ženskega spola. Slovenščina je tipičen spolno zaznamovan jezik, saj se spol vedno odraža v obliki samostalnika, pridevnika, glagola in zamka sočasno. Feminizacija v jeziku oziroma raba ženskih oblik v poimenovanju poklicev je v Sloveniji prvič omenjena leta 1995, tega leta je Evropski svet tudi začrtil smernice odpravljanja *seksistične rabe jezika*, kar je pomenilo tudi nepristransko naslavljanje oseb ženskega spola v delovnih razpisih in drugih uradnih dokumentih. Politični problem je tako v slovenščini hitro postal tudi jezikoslovni, saj veliko poklicev, posebej v primeru tujih besed, ni imelo ženske oblike (npr. informatik). Ženske oblike je bilo treba izpeljati in uzakoniti. Spolno zaznamovani nazivi v poklicih so v relativno kratkem času postali del uradnih dokumentov, zakonov in pravopisa. Standardna klasifikacija poklicev (SKP, 2000) z obema

oblikama poimenovanj je bila v Sloveniji sprejeta že leta 1997. Danes jezikovnih razprav o spolno zaznamovanem jeziku v Sloveniji skoraj ni več²³ (Kanič, 2016).

Z vidika besedotvorja za nova ženska poimenovanja v slovenskem jeziku ni ovir. Besedotvorni postopki omogočajo tvorbo ženskega para h kateremu koli moškemu poimenovanju, to velja tudi za poimenovanja poklicev, in zato Standardna klasifikacija poklicev za vse obstoječe poklice podaja tako moško kot žensko obliko prav zaradi uresničevanja načela o enakovredni uporabi oblik za moški in ženski spol v slovenskem jeziku. Nekaj predlogov za spolno nediskriminatorno («neseksistično») rabo samostalnikov, ki označujejo osebe, podajajo Interne smernice za spolno občutljivo rabo jezika *Komisije za ženske v znanosti* (Šribar, 2010), na spletu pa je tudi več zanimivih in koristnih objav z jezikovnimi nasveti.

5.5.1 Ženske oblike samostalnikov v besedilih strokovne revije *Knjižnica*

Ob pripravi *Bibliotekarskega terminološkega slovarja* je skušala slovaropisna skupina zajeti čim več poimenovanj poklicev, profilov in strokovnih nazivov, ki jih srečamo v knjižničarstvu kot stroki in dejavnosti. Pri tem si je prizadevala enakovredno izbirati in obravnavati za posamezna poimenovanja obe obliki, moško in žensko. Po teh kriterijih se je v izbor za slovar uvrstilo 65 poimenovanj za moške in 59 ustreznih poimenovanj za ženske, le šest poimenovanj je torej ostalo brez ženske oblike. To so *bibliotekarski tehnik*, *čitalniški nadzornik*, *informacijski posrednik*, *kustos*, *samostanski knjižničar* in *sistemski knjižničar*. V praksi pa je raba poimenovanj poklicev, profilov in strokovnih nazivov v bibliotekarstvu mnogo bolj naklonjena moškimi kot ženskimi oblikami. To se zelo značilno pokaže že pri osnovnih poimenovanjih *knjižničar* in *bibliotekar* v referenčnih korpusih (Preglednica 6), v strokovnih bibliotekarskih objavah pa je raba ženskih oblik skoraj zanemarljiva.

Preglednica 6: Pogostnost rabe ženskih oblik v primerjavi z moškimi oblikami terminov bibliotekar in knjižničar v treh korpusih

Termin	FidaPLUS		GigaFida		Knjižnica 1997–2016	
	Število	Procent	Število	Procent	Število	Procent
bibliotekar	459	70 %	720	70 %	1.367	96 %
bibliotekarka	193	30 %	310	30 %	52	4 %
knjižničar	1.867	60 %	2.849	59 %	3.617	95 %
knjižničarka	1.248	40 %	1.965	41 %	172	5 %

²³ Eden novejših dokumentov, ki resno opredeljujejo aktivnosti v zvezi s spolno občutljivo rabo jezika, je npr. Periodični načrt za izvajanje Resolucije o nacionalnem programu za enake možnosti žensk in moških za leti 2016 in 2016. Državni zbor RS, 2015. Pridobljeno 20. 11. 2016 s spletne strani: http://www.mdds.gov.si/fileadmin/mdds.gov.si/pageuploads/dokumenti__pdf/enake_moznosti/ResolucijaPeriodicni2016.pdf.

Enako je v korpusu člankov revije *Knjižnica* razmerje v rabi drugih nazivov: od 52 poimenovanj jih 28 niti enkrat ni uporabljenih v ženski obliki, od skupaj 7.473 evidentiranih primerov rabe je le 327 primerov rabe ženskega spola (4,3 %) (Preglednica 7).

Preglednica 7: Pogostnost rabe ženskih oblik v primerjavi z moškimi oblikami v 553 člankih revije *Knjižnica*

Termin	Moške oblike	Ženske oblike	Termin	Moške oblike	Ženske oblike
amaterski knjižničar	1	0	knjižničarski referent	8	2
analitik	23	0	knjižnični delavec	69	0
arhivar	26	1	konz(s)ervator	13	0
arhivist	58	0	kreator zapisa	4	0
bibliograf	83	2	ljudski knjižničar	5	0
bibliotekar informator	3	1	manager, menedžer	60	1
bibliotekar specialist	22	5	manipulant	26	0
bibliotekar	1.367	52	mladinski knjižničar	62	0
bibliotekarski specialist	1	2	otroški knjižničar	2	0
bibliotekarski svetovalec	16	3	potujoči knjižničar	28	0
diplomirani bibliotekar	33	0	referenčni bibliotekar	8	0
dokumentalist	42	7	referenčni knjižničar	10	1
geslilec	23	0	referent	81	13
informacijski menedžer (manager)	8	0	restavrador	12	0
informacijski specialist	39	0	sodelavec	407	30
informacijski znanstvenik	3	0	specialist	122	8
informatik	75	0	specialni bibliotekar	1	0
informator	127	4	specialni knjižničar	34	0
inventarizator	4	0	splôšni knjižničar	2	0
izposojevalec	70	0	šolski knjižničar	162	5
katalogizator	118	4	visokošolski knjižničar	105	0
klasifikator	20	2	višji bibliotekar specialist	9	1
knjižničar informator	7	0	višji bibliotekar	20	5
knjižničar	3.617	172	višji knjižničar	25	4
knjižničarski delavec	64	0	višji knjižničarski referent	4	2
knjižničarski manipulant	6	0	volonterski knjižničar	1	0

5.6 Primerjava korpusov

Preglednica 8: Primerjava stotih najpogostejših besed v splošnem referenčnem korpusu FidaPLUS in specializiranem korpusu besedil v reviji Knjižnica. Označene so besede, ki bi jih lahko opredelili kot polnomenške besede splošnega jezika oziroma termine v bibliotekarstvu.

Zap. št.	FidaPLUS	Knjižnica	Zap. št.	FidaPLUS	Knjižnica	Zap. št.	FidaPLUS	Knjižnica
1	biti	biti	35	nov	še	69	a	pomemben
2	v	in	36	ob	svoj	70	država	naslov
3	in	v	37	če	tako	71	iti	število
4	na	za	38	velik	knjižničen	72	zelo	en
5	se	na	39	kateri	podatek	73	več	knjižničar
6	z	z	40	drug	kateri	74	bolj	zato
7	za	knjižnica	41	veliko	slovenski	75	tisti	stran
8	da	ki	42	le	zbirka	76	prav	dejavnost
9	on	se	43	naj	sistem	77	Ljubljana	posamezen
10	ki	ta	44	pred	področje	78	vsak	saj
11	pa	da	45	morati	velik	79	ura	naj
12	ta	on	46	dan	informatijski	80	vendar	razvoj
13	tudi	tudi	47	čas	nov	81	sicer	glede
14	ne	pa	48	prvi	strokoven	82	stran	Slovenija
15	po	pri	49	dober	informacija	83	podjetje	ko
16	še	kot	50	Slovenija	morati	84	začeti	avtor
17	kot	leto	51	slovenski	le	85	dati	storitev
18	ves	o	52	ti	različen	86	kjer	čas
19	leto	ali	53	saj	kar	87	malo	ob
20	iz	gradivo	54	nekaj	str.	88	zadnji	elektronski
21	o	po	55	en	več	89	otrok	npr.
22	pri	ne	56	zaradi	že	90	priti	raziskava
23	imeti	knjiga	57	njegov	vir	91	vedno	potreba
24	od	delo	58	zato	njihov	92	program	način
25	jaz	lahko	59	mesto	uporaba	93	dobiti	določen
26	do	uporabnik	60	sam	del	94	k	zapis
27	ali	iz	61	kaj	oblika	95	njihov	skupina
28	že	drug	62	človek	katalog	96	zdaj	njegov
29	svoj	od	63	ker	prvi	97	konec	bolj
30	lahko	ves	64	dva	publikacija	98	trije	uporabljati
31	tako	imeti	65	ter	splošen	99	povedati	zaradi
32	med	med	66	del	če	100	tolar	oziroma
33	ko	ter	67	naš	primer	101	vedeti	predvsem
34	kar	do	68	svet	vsebina	102	brez	članek

Pogostnost besed, konkordance, značilnosti in pogostnost kolokacij, odnosi v tezavru in še nekatere druge tipične lastnosti besedil, ki smo jih analizirali, ostajajo v okviru enega samega korpusa brez relativizacije odnosov in primerjav z drugimi besedili in jezikovnimi okolji. Nekatero značilnost je mogoče dognati in ovrednotiti samo v primerjavi z drugimi besedilnimi sistemi, najboljše korpusi. Sketch Engine vsebuje nekaj orodij, ki omogočajo različne primerjave s korpusi, ki vsebujejo besedila v istem jeziku in so zgrajeni upoštevaje podobno strukturo in jezikovna pravila (Kilgariff idr., 2004). Na ta način smo opravili nekaj primerjav tudi za korpus člankov iz revije *Knjižnica*. Primerjalni seznam najpogostejših sto besed (lem) korpusa člankov in *FidePLUS* kaže prve velike razlike med splošnim slovenskim jezikom in strokovnim jezikom bibliotekarstva. Vrsto vodilnih funkcijskih besed prekinajo polnopomenske besede oziroma že takoj celo strokovni termini v bibliotekarstvu mnogo prej kot leksikalne besede v splošnem jeziku (Preglednica 8).

Bolj kot vzporedni sezname najpogostejših besed dveh korpusov je povedna neposredna primerjava deleža zastopanosti istih besed v dveh korpusih. Pri tem niso primerjane absolutne vrednosti, čeprav so informativno podane, pač pa normalizirane vrednosti pogostosti na milijon besed. Tako je omogočena objektivna primerjava zastopanosti besede, za ponazoritev zopet v *FidePLUS* in korpusu člankov revije *Knjižnica* (Preglednica 9). Beseda *knjižnica* se tako v reprezentativnem korpusu splošnega jezika pojavlja dvakrat toliko kot v bibliotekarskih člankih (64.306 proti 33.712), vendar je slika upošteva velikost korpusov povsem drugačna – v bibliotekarskih člankih se v resnici pojavlja 130-krat pogosteje kot v splošnem korpusu (11.308 proti 87 pojavljanj na milijon besed).

Pomembna razlika je razvidna tudi iz dojemanja pomenske širine in vsebinskih povzav pri uporabi nekaterih besed, ki so postale v bibliotekarstvu strokovni izraz. Za ilustracijo bomo uporabili dva v bibliotekarstvu temeljna termina, *knjižnica* in *knjiga*. Slikovna predstavitev z oblakom besed najnazorneje prikaže, s katerimi besedami se termina najpogosteje srečujeta v sobesedilu splošnega in strokovnega jezika (Slika 15). V splošnem jeziku je knjižnica pogosto omenjena v družbi z drugimi kulturnimi in izobraževalnimi ustanovami, kot so muzej, galerija, gledališče, šola, gimnazija, vrtec, zavod, društvo, klub, ustanova in/ali njihovo pripadnostjo (glasben, zdravstven, športen, turističen, mesten, cerkev) ali verjetno problemi s prostori (stavba, prostor, pisarna) in le redko s tistim, kar zanima bibliotekarja, to je dejavnost, knjiga, prireditve, osrednji. V bibliotekarskih besedilih so osrednje spremljevalke besede knjižnica predvsem tiste, ki poudarjajo predmet njihovega delovanja, to je knjiga, gradivo, katalog, publikacija, podatek, informacija, dokument, vir, avtor, ali pa namen in način delovanja, kot so uporabnik, storitev, sistem, izobraževanje, dejavnost in delo.

za predvidene analize in nadaljnje terminološko delo, nabor besedil pa je dovolj bogat vir verodostojnega, najnovejšega in najsodobnejšega strokovnega besedišča. S programskim orodjem, ki razpozna slovenski jezik in za vse obdelave in analize ustrezno uporabi vgrajena pravila za ta jezik, smo pripravili sezname besed s pogostnostjo njihovega pojavljanja, besede lematizirali in oblikoslovno ter skladdenjsko označili, izbrali in analizirali kolokacije in besedne skice nekaterih terminov vključno z uporabo funkcij tezavra in vizualizacije, nad pričakovanji uspešno izluščili sezname enobesednih in večbesednih terminov ter primerjali besedišča dveh različnih korpusov oziroma pojavljanje iste besede v teh korpusih.

Vzporedno s pripravo korpusa smo zbrali in analizirali tudi ključne besede, ki jih člankom praviloma določijo avtorji sami. Ključne besede so najboljši pokazatelj, kaj in kako pogosto so v posameznih obdobjih pisali slovenski strokovnjaki na področju bibliotekarstva, pa tudi, kako je jezikovno in sistemsko zasnovan sistem vsebinskih oznak člankov. Abecedni in frekvenčni seznam avtorskih ključnih besed kažeta na njihovo relativno neenotnost glede uporabe ednine in množine ter uporabe sopomenk, saj zanje ni normativnega seznama in so jih avtorji oblikovali po svoji presoji. V naboru 2.246 uporabljenih ključnih besed se le 319 ključnih besed pojavlja več kot enkrat oziroma 148 več kot dvakrat in zgolj 18 ključnih besed presega mejo desetih pojavitev. Tolikšna disperzija relevantnih ključnih besed je nenavadna in z vidika sistema za poizvedovanje tudi problematična. Ob teh ugotovitvah kaže razmisliti, ali ne bi uredništvo revije *Knjižnice* avtorjem predlagalo uporabe svojega ali katerega od uveljavljenih geslovníkov ali tezavrov.

Pogostnost pojavljanja besed v analiziranih besedilih je seveda zelo različna, v skladu s pričakovanji in teorijo o živih jezikih vodijo funkcijske besede, vendar se že v skupini stotih najpogostejših pojavljajo tudi nekateri temeljni bibliotekarski termini, npr. knjižnica, gradivo, knjiga, uporabnik, med petdesetimi najpogostejšimi besedami je tako kar 13 terminov. Na drugem koncu frekvenčnega seznama je množica besed, ki se pojavljajo samo enkrat ali dvakrat, med njimi je tudi nabor besed, ki so lahko osnova za razpravo redaktorjev slovarja in presojo slovarske in terminološke vrednosti ter pomembnosti.

Med termini z majhno pogostnostjo so tudi ženska poimenovanja. *Bibliotekarski terminološki slovar* je sicer skušal zajeti čim več poimenovanj poklicev, profilov in strokovnih nazivov, ki jih srečamo v knjižničarstvu, in enakovredno izbirati in obravnavati za posamezna poimenovanja obe obliki, moško in žensko. V praksi pa se kaže, da je raba ženskih oblik poimenovanj v strokovni literaturi zanemarljiva in mnogo bolj naklonjena moškim kot ženskim oblikam.

Skokovito upadanje pogostnosti besed in izredno poudarjen dolgi rep smo primerjali tudi s krivuljo Zipfovega zakona, ki sicer velja predvsem za splošni jezik,

strokovna literatura pa seveda od tega odstopa. Trenda krivulj se dokaj ujemata, vendar v srednjem delu krivulje podatki korpusa odstopajo navzgor, kar je posledica koncentracije nekaterih besed, predvsem strokovnih terminov in njihovih stalnih kolokacij, pogostejša od normalne distribucije. Krivulja se v obeh primerih podaljšuje v dolgi rep besed s pogostnostjo $f=1$.

Pri posameznih terminih so opazni tudi nekateri značilni vzorci pojavljanja in časovne porazdelitve v obdobju, ki ga korpus obsega. Ti vzorci so v specializiranih korpusih zanimivi predvsem zato, ker lahko njihovo proučevanje razkrije nekatere podrobnosti o njihovem pomenskem oziroma terminološkem razvoju in tudi razvoju same stroke. Primerjava dveh časovno razmejenih podkorpusov je omogočila potrditev hipoteze, da se v novejšem obdobju zaradi razvoja in novih interesnih področij v bibliotekarstvu pojavljajo termini, ki so bili pred tem zelo redki ali jih sploh ni bilo mogoče zaslediti.

Nekatere tipične lastnosti besedil, ki smo jih analizirali, ostajajo v okviru enega samega korpusa brez relativizacije odnosov, dognati in ovrednotiti jih je mogoče mnogo bolje v primerjavi z drugimi besedilnimi sistemi. Primerjalni seznam najpogostejših sto besed kaže prve velike razlike med splošnim slovenskim jezikom in strokovnim jezikom bibliotekarstva. Med vodilnimi funkcijskimi besedami se strokovni termini v bibliotekarstvu pojavijo mnogo prej kot leksikalne besede v splošnem jeziku. Bolj kot absolutna pogostnost besed dveh korpusov je povedna neposredna primerjava deleža zastopanosti istih besed v dveh korpusih. Primerjali smo normalizirane vrednosti pogostosti na milijon besed in dobili jasnejšo sliko objektivne primerjave zastopanosti besed. Beseda *knjižnica* se tako na primer v reprezentativnem korpusu splošnega jezika pojavlja 130-krat pogosteje kot v splošnem korpusu. Dodatno dimenzijo dojemanja pomenske širine in vsebinskih povezav pri uporabi nekaterih besed splošnega jezika, ki so postale v bibliotekarstvu strokovni izraz, smo analizirali s kolokacijami in oblakom besed iz sobesedila. Tako je na primer *knjižnica* v splošnem jeziku najpogosteje omenjena v sobesedilu z drugimi kulturnimi in izobraževalnimi ustanovami in okoljem, ki mu pripadajo, v bibliotekarskih besedilih pa predvsem z besedami, ki poudarjajo predmet njihovega delovanja. Beseda *knjiga* je v splošnem jeziku obdana predvsem z množico polnil brez posebne informacijske vrednosti, verjetno vrinki novinarjev in/ali politikov.

Tako lahko ugotovimo, da so analize besedil in posameznih terminov ali skupin terminov potrdile vse tri hipoteze:

- članki v reviji *Knjižnica* vsebujejo bogat nabor tradicionalnih, pa tudi najnovejših terminov, katerih raba se v strokovnem jeziku spreminja tudi v krajših časovnih obdobjih,

- raba bibliotekarskih terminov se po kolokacijah zelo razlikuje od rabe teh besed v splošnem jeziku,
- raba ženskih in moških oblik poimenovanj poklicev in nazivov v bibliotekarstvu ni uravnotežena.

Za raziskavo vzpostavljen specializirani korpus, rezultati analiz in v raziskavi pridobljena znanja bodo lahko izhodišče za podporo nadaljnjemu terminološkemu delu in dopolnjevanju slovarjev. Z večjim obsegom zajetih besedil bi lahko povečali pestrost nabora predvsem redkeje uporabljanih in starejših terminov, z rednim dodajanjem novih objav pa najnovejših, morebiti še neustaljenih terminov, kar bi lahko brez velikih zadržkov rešila vzpostavitev mnogo večjega korpusa večine v reviji *Knjižnica* objavljenih člankov.

Zahvala

Iskreno se zahvaljujem vsem avtoricam in avtorjem člankov, saj so njihova besedila omogočila pripravo te terminološke študije, predvsem pa sem se lahko od njih tudi na drugih področjih neizmerno dosti naučil.

Navedeni viri

Akcijski načrt za jezikovno opremljenost. (2015). Ljubljana: Vlada Republike Slovenije. Pridobljeno 20. 11. 2016 s spletne strani: http://www.mk.gov.si/fileadmin/mk.gov.si/pageuploads/Ministrstvo/raziskave-analize/slovenski_jezik/Akcijska_nacrta/ANJO.pdf

Ambrožič, M. (2007). *Knjižnica: od skromne publikacije do ugledne in pomembne revije. Knjižnica, 51(1), 11–17.*

Aparac-Jelušić, T. (1999). Znanstveni i stručni članci u knjižnici i Vjesniku bibliotekara Hrvatske: pokušaj usporedne analize stručnog i znanstvenog doprinosa u knjižničarstvu. *Knjižnica, 43(2–3), 179–196.*

Arhar Holdt, Š. in Gorjanc, V. (2007). Korpus FidaPLUS. *Jezik in slovstvo, 52(2), 95–110.*

Česnik Ušaj, J. (2011). *Revija Knjižnica – analiza vsebine in bibliometrijska analiza od leta 2005 do 2010.* Diplomsko delo. Ljubljana: Filozofska fakulteta.

Direktiva Evropskega parlamenta in Sveta o avtorskih pravicah na enotnem digitalnem trgu – predlog. (2016). Bruselj: Evropska komisija. Pridobljeno 22. 9. 2016 s spletne strani: <http://eur-lex.europa.eu/legal-content/SL/TXT/PDF/?uri=CELEX:52016PC0593&from=EN>

Erjavec, T. (2010). *Korpusno jezikoslovje: uvod* [prosojnice]. Ljubljana: Jožef Stefan InSTITUTE. Pridobljeno 24. 1. 2017 s spletne strani: <http://nl.ijs.si/et/teach/ung08-kj/ung08-kj-02.pdf>

Erjavec, T. (2013). Korpusi in konkordančniki na strežniku nl.ijs.si. *Slovenščina 2.0*, 1(1), 24–49. Pridobljeno 24. 5. 2017 s spletne strani: http://www.trojina.org/slovenscina2.0/arhiv/2013/1/Slo2.0_2013_1_03.pdf

FidaPLUS: korpus slovenskega jezika. (2000). Pridobljeno 20. 2. 2017 s spletne strani: http://web.archive.org/web/20150720092556/http://www.fidaplus.net:80/Index_eng.htm

Gigafida. (B. l.). Kamnik: Amebis. Pridobljeno 12. 9. 2017 s spletne strani: <http://www.gigafida.net/>

Gorjanc, V. in Fišer, D. (2013). *Korpusna analiza*. Ljubljana: ZRC SAZU.

Gorjanc, V. in Logar, N. (2007). Od splošnih do specializiranih korpusov – načela gradnje glede na njihov namen. V I. Orel (ur.), *Razvoj slovenskega strokovnega jezika* (str. 637–650). Ljubljana: Filozofska fakulteta, Oddelek za slovenistiko, Center za slovenščino kot drugi/tuji jezik. Pridobljeno 17. 8. 2016 s spletne strani: <http://centerslo.si/wp-content/uploads/2015/10/24-GorjancLogar.pdf>

Grum, M. (1987). Bibliografsko kazalo Knjižnice: 1957–1986. *Knjižnica*, 31(4), VII–X, 1–166.

Kanič, I. (2011a, 30. september). *Enkratnice* [blog zapis]. Ljubljana: Bibliotekarska terminologija. Pridobljeno 10. 10. 2017 s spletne strani: <http://terminologija.blogspot.si/search?q=enkratnice>

Kanič, I. (2011b). *Korpus bibliotekarstva* [blog zapis]. Ljubljana: Bibliotekarska terminologija. Pridobljeno 5. 3. 2017 s spletne strani: <http://terminologija.blogspot.si/p/korpus.html>

Kanič, I. (2011c). Slovenski besedilni korpus bibliotekarstva – najsodobnejša slovaropisna podpora bibliotekarski terminologiji. V M. Ambrožič in D. Vovk (ur.), *Knjižnica: odprti prostor za dialog in znanje: zbornik referatov* (str. 297–312). Ljubljana: Zveza bibliotekarskih društev Slovenije.

Kanič, I. (2016, 25. april). *Feminizacija in poimenovanja poklicev v knjižnici* [blog zapis]. Ljubljana: Bibliotekarska terminologija. Pridobljeno 17. 1. 2017 s spletne strani: <http://terminologija.blogspot.si/2016/04/feminizacija-in-poimenovanja-poklicev-v.html>

Kanič, I., Leder, Z., Ujčič, M., Vilar, P. in Vodeb, G. (2009). Bibliotekarski terminološki slovar [monografija]. *Knjižnica*, 53(3–4). Pridobljeno 10. 3. 2017 s spletne strani: <http://revija-knjiznica.zbds-zveza.si/Izvodi/K0934/BTS-2009.pdf>

Kanič, I., Leder, Z., Ujčič, M., Vilar, P. in Vodeb, G. (2011). *Bibliotekarski terminološki slovar*. Ljubljana: Bibliotekarska terminološka komisija. Pridobljeno 10. 3. 2017 s spletne strani: <http://www.termania.net/slovarji/85/bibliotekarski-terminoloski-slovar>

Kavčič-Čolič, A. in Vodopivec, I. (2015). Odsev razvoja bibliotekarske in informacijske znanosti na objave revije *Knjižnica* v obdobju 2012–2015. *Knjižnica*, 59(4), 117–138.

Kilgarriff, A., Rychly, P., Smrz, P. in Tugwell, D. (2004). *The Sketch Engine*. Berlin: ResearchGate. Pridobljeno 5. 3. 2017 s spletne strani: https://www.researchgate.net/profile/Adam_Kilgarriff/publication/260387608_ITRI-04-08_the_sketch_engine/links/54e0d1210cf24d184b0de48f.pdf

Krek, S. in Kilgarriff, A. (2006). Slovene word sketches. V T. Erjavec in J. Žganec Gros (ur.), *Jezikovne tehnologije: zbornik 9. mednarodne multikonference Informacijska družba IS 2006, 9. do 10. oktober 2006* (str. 62–67). Ljubljana: Institut »Jožef Stefan«. Pridobljeno 24. 1. 2017 s spletne strani: <https://www.kilgarriff.co.uk/Publications/2006-KrekKilg-Ljub-SloveneWS.pdf>

Likar, T. (1997). Analiza člankov in bibliografskih navedb v reviji knjižnica za obdobje 1990–1996. *Knjižnica*, 41(1), 107–119.

Ločniškar-Fidler, M. in Vovk, D. (2007). Bibliografsko kazalo revije *Knjižnica XXXI–L* (1987–2006). *Knjižnica*, 51(2), 7–110.

Logar Berginc, N., Grčar, M., Brakus, M., Erjavec, T., Arhar Holdt, Š. in Krek, S. (2012). *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko. Pridobljeno 12. 8. 2016 s spletne strani: <http://www.fdv.uni-lj.si/docs/default-source/zalozba/pages-from-logar-et-al--korpusi.pdf?sfvrsn=2>

Nova beseda. (2008). Ljubljana: ZRC SAZU. Pridobljeno 20. 2. 2017 s spletne strani: http://bos.zrc-sazu.si/s_beseda3.html

Popovič, M., Ambrožič, M. in Južnič, P. (1984). Nekaj značilnosti razvoja slovenskega knjižničarstva v novejšem obdobju. (Ob kvantitativni analizi člankov in referenc v reviji *Knjižnica za obdobje 1974–1983*). *Knjižnica*, 28(3–4), 167–198.

SKP – Standardna klasifikacija poklicev. (2000). *Uradni list RS*, št. 16/2000.

Snoj, M. (2014). Predgovor k drugi izdaji. V *Slovar slovenskega knjižnega jezika. Druga, dopolnjena in deloma prenovljena izdaja. Elektronska objava*. Ljubljana: ZRC SAZU. Pridobljeno 7. 12. 2016 s spletne strani: http://www.sskj2.si/Content/Site/Pdf/SSKJ2_Predgovori.pdf

Splošna načela ob predložitvi prispevka. (B. 1.). Ljubljana: Zveza bibliotekarskih društev Slovenije. Pridobljeno 22. 9. 2016 s spletne strani: <http://www.zbds-zveza.si/?q=knjiznica/avtorji>

Šribar, R. (2010). *Interne smernice za spolno občutljivo rabo jezika*. Ljubljana: Komisija za ženske v znanosti. Pridobljeno 20. 4. 2016 s spletne strani: http://www.arhiv.mvzt.gov.si/fileadmin/mvzt.gov.si/pageuploads/pdf/znanost/Zenske_v_znanosti/interne_jezikovne_smernice_za_komisijo_final.pdf

Teubert, W. (2005). My version of corpus linguistics. *International journal of corpus linguistics*, 10(1), 1–13.

Todorovski, I. (2008, 30. december). Slovenskih 500 tisoč. *Finance.si*. Pridobljeno 13. 5. 2012 s spletne strani: https://www.finance.si/233836/Slovenskih_500_tiso?cctest&

Toporišič, J. (1992). *Enciklopedija slovenskega jezika*. Ljubljana: Cankarjeva založba.

Vratarič, S. (2006). *Bibliometrijska analiza revije Knjižnica od leta 2000 do 2004*. Diplomsko delo. Ljubljana: Filozofska fakulteta.

Ivan Kanič

Centralna ekonomska knjižnica, Ekonomska fakulteta, Univerza v Ljubljani, Kardeljeva ploščad 17, 1000 Ljubljana
e-pošta: ivan.kanic@ef.uni-lj.si