Lorena Ninčević, University of Zagreb, Croatia

Analiza likova hrvatske književnosti pomoću računalnih alata

Abstract: Computer-aided Analysis of Characters in Croatian Literature

The paper deals with the computer-aided analysis of characters in the following texts of Croatian literature: Šuma Striborova (Ivana Brlić-Mažuranić), Oprava (Josip Kozarac), Krvavi most u Zagrebu (Mirko Bogović), Oko Lobora (Antun Gustav Matoš), Posljednji Stipančići (Vjenceslav Novak). In the theoretical part, digital humanities and its subfield digital philology are defined. Basic terms from the fields of natural language processing and narratology are explained, since they are the theoretical and practical basis of the research. The rest of the paper is research conducted on the texts above in the internet program called Sketch Engine. The research question is stated as follows: whether it is possible to obtain data that is useful for the interpretation of literary characters through computer-assisted analysis, based on narratology and the existing literary-theoretical elements of characterization. Based on the results, the affirmative answer is given to the formulated research question, and the potential and challenges of digital humanities are highlighted.

Keywords: digital humanities, natural language processing, Croatian literature, narratology, characterization, computational tools

1. Uvod

Razvojem digitalne tehnologije i sve većom dostupnosti tekstualnih (i ostalih vrsta) podataka ni područje humanističkih znanosti nije ostalo netaknuto. Dolazi do pojave interdisciplinarnog područja digitalne humanistike, koja proučava sve one objekte koji su u interesu i tradicionalne

humanistike, ali pritom koristi računalne alate, digitalne izvore te metode preuzete iz tehničkih i društvenih znanosti. Time se ističe njezina interdisciplinarna snaga i potencijal koji u tom aspektu posjeduje. S obzirom na to da se ovaj rad bavi književnošću, odnosno interpretacijom odabranih književnih tekstova, valja istaknuti digitalnu filologiju kao potpodručje digitalne humanistike koje se bavi proučavanjem tekstualnih objekata (uključujući i književne tekstove). Pritom treba spomenuti područje obrade prirodnog[1] jezika (engl. Natural Language Processing, NLP), koje se bavi obradom niza znakova koji nose određeno značenje jer upravo je značenje bitno za analizu. Obrada prirodnog jezika interdisciplinarno je područje koje pripada polju društvenih i tehničkih znanosti te se posebno oslanja na informacijske znanosti i metode preuzete iz statistike, a njezini pristupi redovito se koriste u istraživanjima digitalne filologije.

S obzirom na to da je rad interdisciplinarne naravi i želi se omogućiti jednako razumijevanje onima koji su tehnički jače i slabije potkovani, slijedi sažet prikaz najbitnijih postupaka, alata i pojmova važnih za shvaćanje ostatka rada.

Pojavnica: najmanja jedinica od koje se sastoji korpus (Sketch Engine). U programu Sketch Engine pojavnica može biti riječ, interpunkcijski znak, brojka, skraćenica, odnosno sve što je omeđeno razmakom.

Vertikalizacija: postupak razdvajanja pojavnica u zasebne redove tako da je krajnji rezultat „lista" pojavnica.

Označavanje vrsta riječi (engl. part-of-speech tagging, POS tagging): još se naziva i gramatičko tagiranje (Mrežnik); postupak pripisivanja oznake vrste riječi za svaku pojavnicu, s time da Sketch Engine (2023.) dodaje i mogućnost određivanja ostalih gramatičkih kategorija (rod, broj, glagolsko vrijeme itd.), ovisno o kojem se jeziku radi i koliko su razvijeni resursi za taj jezik.

Morfosintaktično označavanje (engl. morphosyntactic tagging, MSD tagging): gramatičko tagiranje koje daje dodatne informacije o različitim gramatičkim kategorijama, a ne samo vrstu riječi (Sketch Engine); program Sketch Engine ne navodi posebno ovaj tip označavanja, već ga objašnjava istodobno s POS tagiranjem složenije razine.

NER označavanje (engl. Named Entity Recognition): postupak pronalaženja i grupiranja dijelova teksta u određene unaprijed definirane kategorije, npr. lokacije, imena osoba, nazive proizvoda i sl.

Sketch Engine: internetski dostupan program koji se koristi za istraživanje jezika (njih više od 90) (Sketch Engine). Nudi različite mogućnosti, uključujući i izradu vlastitog korpusa, a u kontekstu ovog rada bitno je spomenuti GDEX (engl. Good Dictionary EXamples), sortiranje konkordancija[2] na temelju parametara koji određuju koliko je pojedinačna konkordancija prikladan rječnički primjer.

Izravna definicija: postupak u karakterizaciji likova u kojem pripovjedač ili drugi lik eksplicitno opisuju lika. Proza prije 20. stoljeća većinom je izravne definicije koristila tako da ih je izricao sveznajući pripovjedač (što je i slučaj u tekstovima koji se analiziraju u ovome radu) (Grdešić 69).

Neizravna prezentacija: umjesto da se izravno navodi kakav je lik, o njemu se može saznati i preko njegovih radnji, načina govorenja, vanjskog izgleda, okoline i sl., a to je onda neizravna prezentacija (Grdešić 71).

2. Povezana istraživanja

Postoje različita istraživanja koja se bave proučavanjem pojedinih elemenata književnih tekstova pritom se služeći računalnim alatima, statistikom, metodama iz obrade prirodnog jezika. Različiti su predmeti istraživanja takvih radova, a budući da se u ovome radu proučavaju likovi, u nastavku se donosi kratki pregled nekih istraživanja na sličnu temu.

Arthur M. Jacobs bavi se u svom radu, između ostalog, analizom sentimenta likova iz serijala knjiga o Harryju Potteru (5). Analiza sentimenta računalno je područje istraživanja mnijenja, emocija, stavova ljudi prema određenim entitetima (proizvodima, uslugama, temama, događajima, osobama) i njihovima svojstvima (Ligthart i sur. 4998). Stoga Jacobsa zanima emocionalnost i osobnost likova preko tri parametra: intenziteta, valencije i emocionalnog potencijala (6). Ovaj rad temelji se i na 5-faktorskoj dimenziji ličnosti, stoga pripada i području neuropoetike.

Zanimljivo je istraživanje koje su proveli Hardik Vala i sur., u kojem su pokušali detektirati likove u 58 romana (769). Pritom su istaknuli jedan problem koji se pojavljuje i u ovome radu, a to je pitanje kako detektirati one likove koji nemaju osobna imena, već se oslovljavaju isključivo općim imenicama. Oni su taj problem riješili upotrebom univerzalnih ovisnosti[3] (engl. universal dependencies, UD) tako što su gledali prototipsko ponašanje likova u samom tekstu, odnosno gledali su ovisnost između predikata i imenica s obzirom na to sadrže li predikati glagole za koje su

smatrali da su povezani s ponašanjem likova. Osim što su se bavili detekcijom likova, ispitali su i hipotezu da u urbanim okruženjima ima više likova nego u ruralnim, međutim u svome korpusu nisu naišli na neku znatnu razliku u broju likova. Drugo pitanje kojim su se bavili jest pitanje količine likova u različitim vremenskim periodima (orijentirali su se na romane nastale između 1800. i 1990. godine) te nisu pronašli znatne razlike u broju likova u romanima, neovisno o društveno-ekonomskim promjenama.

Nalisnick i Baird proučavali su Shakespeareove drame (479). Njih je također zanimala analiza sentimenta, odnosno dinamika odnosa između likova. Zbog vrlo jasne strukture dramskog teksta mogli su dobro pretpostavljati tko sudjeluje u kojem razgovoru. S obzirom na to kome se pojedini lik obraća, donosili su zaključke o sentimentu govora lika. To im je pomoglo u uočavanju bitnih emocionalnih trenutaka u razvoju lika te koji su mu likovi prijetnja, a koji mu pomažu.

3. O istraživanju

Glavni dio rada istraživanje je napravljeno na odabranim tekstovima hrvatske književnosti. Ideja koja stoji iza istraživanja jest pokušati računalnim alatima i metodama iz obrade prirodnog jezika izvući informacije o karakterizaciji likova koja se temelji na načelima naratologije.

Korpus[4] se sastojao od pet hrvatskih proznih tekstova: Šuma Striborova (bajka) Ivane Brlić-Mažuranić, Oprava (novela) Josipa Kozarca, Oko Lobora (putopis) Antuna Gustava Matoša, Krvavi most u Zagrebu (pripovijetka) Mirka Bogovića i Posljednji Stipančići (roman) Vjenceslava Novaka. Tablica 1. prikazuje veličinu korpusa. Odabir navedenih književnih tekstova dvojake je prirode. Jedan razlog praktične je naravi – izabrao se izvor (internetska stranica eLektire) koji sadrži već digitalizirane tekstove koji imaju slobodni pristup kako se ne bi narušila autorska prava. S druge strane, birani su tekstovi koji su dio kanona hrvatske književnosti kako bi se osigurala relevantnost provedenog istraživanja. Nadalje, svi su tekstovi prozni jer su korišteni alati trenirani na proznim tekstovima (pretežito novinski članci i sadržaji internetskih stranica), a književna proza strukturalno najviše nalikuje tim tekstovima, za razliku od lirike i drame. Naposljetku, tekstovi pripadaju različitim književnim vrstama kako bi se prikazala primjenjivost računalnih alata u različitim primjerima tekstova.Tablica 1. Veličina korpusa

	Šuma Striborova	Oprava	Oko Lobora	Krvavi most u Zagrebu	Posljednji Stipančići	ukupan korpus
postotak zastupljenosti u korpusu	2,83 %	8,86 %	3,86 %	10,80 %	73,65 %	100 %

Navođenjem tekstova koji čine korpus može se primijetiti da se radi o različitim književnim vrstama. Namjerno je došlo do takvog odabira jer se, osim klasične karakterizacije likova, htjelo vidjeti može li se iščitati kakva poveznica između načina prikazivanja likova i književne vrste kojoj tekst pripada, tj. nameće li vrsta neka pravila u oblikovanju likova. Za potrebe kasnije analize slijedi kratko definiranje pojedine književne vrste.

Bajke se u osnovi dijele na narodne i umjetničke bajke (Solar, Teorija književnosti 169). Milivoj Solar objašnjava da su narodne bajke više formulaične, kod njih je bitna struktura, ponavljajući motivi, likovi se ne mijenjaju kroz tekst te je prisutna jasna polarizacija dobra i zla (Teorija književnosti 169). Iz tih razloga narodna bajka pripada jednostavnim proznim oblicima. Umjetnička pak je bajka složeni prozni oblik koji dopušta puno više slobode u izražavanju, likovi se mogu mijenjati kako radnja napreduje (što je primjerice slučaj sa sinom iz Šume Striborove) i nije podložna strogim pravilima oblikovanja. Iz navedenog proizlazi da je Ivana Brlić-Mažuranić pisala umjetničke bajke s elementima hrvatskog folklora koje su bile kršćanskog nadahnuća (Pintarić 166). Bajke često kao mjesta radnje imaju fikcionalna mjesta koja ni u kojem smislu ne referiraju sa stvarnim izvanjskim svijetom, što se u bajci Šuma Striborova jasno vidi u izostanku naziva lokacija, pa čak i upotrebljavanju općih imenica kao imena likova. S druge strane, zazivanje Boga može se povezati s kršćanskim svjetonazorom, koji je nerijetko prisutan u djelima Ivane Brlić-Mažuranić. Time se bajka stavlja u hrvatski kontekst.

Mali broj likova i mjesta radnji tipični su za novelu kao književnu vrstu jer je naglasak većinom samo na jednom liku. Spada u kratku književnu vrstu koja donosi nešto novo i zanimljivo za čitatelja, kako joj i sam naziv govori (Solar, Ideja i priča 209). Upravo zbog svoje ograničenosti duljinom novela treba zaokupiti pažnju neobičnom temom i fokusirati se na jednog lika kojeg će razraditi u tančine (Solar, Ideja i priča 209). Novela Oprava, koja je predmet istraživanja ovoga rada, posjeduje navedene karakteristike jer se bavi unutarnjim stanjem glavnog lika Đure, njegovim emocijama i psihičkim promjenama, stoga se preciznije može reći da je riječ o psihološkoj noveli.

Za razliku od novele, pripovijetka je prozna vrsta koja sadrži više likova te nerijetko i više narativnih linija. Za nju bi se moglo reći da se po svojoj opsežnosti nalazi između novele i romana. Međutim, teško je povući jasnu granicu između, primjerice, dulje pripovijetke i kraćeg romana.

Putopis kao književna vrsta pripada području kulture putovanja. U njemu autor najčešće iznosi svoje iskustvo putovanja, isprepleteno različitim doživljajima i promišljanjima o viđenim krajevima. U sebi može sadržavati i elemente drugih književnih vrsta, kao što su dnevnici, memoari, pisma i sl.

Roman, premda dobro poznat profesionalnim i običnim čitateljima, jedna je od književnih vrsta koju je teško jasno definirati. Milivoj Solar dolazi do zaključka da se romani često nalaze na granici između romana i neke druge književne vrste (Ideja i priča 184). Ističe i da je nemoguće odrediti jedan kriterij po kojem će se romani međusobno razlikovati jer tijekom čitave povijesti književnosti postoji previše različitih primjera. Ni kvantiteta napisanoga teksta ne može se shvatiti kao vjerodostojan parametar razlikovanja romana od ostalih njemu sličnih književnih vrsta (npr. pripovijetke i novele) jer ponovno, postoje vrlo dugački, ali i vrlo kratki romani. Ipak, s obzirom na takozvane kraće pripovjedne vrste, roman većinom ima razgranatiju fabulu i više likova.

Karakterizacija likova radila se tako da su se upotrebljavali elementi karakterizacije koje donosi naratologija. Međutim, kao dio metodologije istraživanja nisu bili uključeni svi elementi karakterizacije (koje navodi Maša Grdešić u svojoj knjizi Uvod u naratologiju), nego samo oni za koje se procijenilo da se u ovoj početnoj fazi mogu računalno detektirati, a to su izravne definicije i neizravne prezentacije (od kojih rutinske i nerutinske radnje te govor). Donijela se odluka da se detaljnija karakterizacija likova radi samo za pojedine odabrane likove za koje se procijenilo da bi bili pogodni za analizu. Točnije, izabrani su oni likovi koji nose neku aktancijalnu ulogu iz aktancijalnih modela razrađenih za svaki tekst korpusa.[5] Riječ je o sljedećim likovima: stara majka, Domaći s Malikom Tintilinićem kao glavnim predstavnikom (Šuma Striborova), Lucija Stipančić i kanonik Vukasović (Posljednji Stipančići), župan Gardun i Pavao Slavinić (Krvavi most u Zagrebu), Đuro Vučetić (Oprava) i glavni lik iz putopisa Oko Lobora. Osim toga, gledalo se da odabrani likovi budu bitni za radnju tekstova jer je to značilo da se pojavljuju dovoljno učestalo da bi ih se moglo računalno analizirati. Teorijsko polazište istraživanja temelji se na strukturalizmu i naratologiji, budući da se analiziraju književni likovi. Strukturalizam književni tekst poima kao jednu

veliku strukturu unutar koje odnosi između njezinih sastavnica oblikuju vrijednost svake sastavnice. Ova književna teorija usko je povezana s lingvistikom, što je čini bliskom obradi prirodnoga jezika, ako se književni tekst promatra kao sustav, baš kao što se i jezik promatra. Jan Mukařovský na umjetničko djelo gleda kao na znak koje posjeduje mnogoznačnost smislova (162). Slijedom navedenoga, likovi su također jedna od sastavnica strukture koja se naziva književni tekst i oni svoje značenje dobivaju isključivo unutar teksta, u odnosu prema drugim likovima, radnji i umjetničkim postupcima (Mukařovský 164). Naratologija, kao opća teorija pripovjednog teksta, bavi se i likovima kao jednim od elemenata narativnog teksta i zbog svojih korijena u strukturalističkoj analizi pogodna je za primjenu u računalno potpomognutoj analizi likova. Razlog tomu jest to što način na koji računalo shvaća jezik, odnosno tekst (sustav znakova i pravila njihova kombiniranja), sličan je načinu kako naratologija promatra književni tekst (sustav elemenata i načini njihova povezivanja kojima se tvori zaokružena narativna cjelina). Međutim, kao i većina ostalih pristupa u književnosti, naratološka se analiza provodi pomnim čitanjem, koje se ipak razlikuje od udaljenog čitanja kakvo je karakteristično za analizu pomoću računala. Franco Moretti definira pomno čitanje kao ono koje je zaokupljeno samo kanonom, odnosno malom količinom tekstova, jer njegova narav ne dopušta ekstenzivno čitanje velike količine tekstova (48). Udaljeno čitanje Moretti objašnjava kao ono čitanje koje omogućava fokusiranje na puno manje ili veće jedinice od teksta, kao što su teme, tropi ili pak žanrovi i sustavi (49). Primjerice, bi li se pomnim i udaljenim čitanjem uočili isti pridjevi koji opisuju likove u romanima ruske književnosti (navode se pridjevi jer su oni vrsta riječi koja opisuje imenice, a likovi su u tekstu uglavnom prikazani imenicama)? Osobnost lika ono je što Gajo Peleš naziva psihemskom narativnom figurom i upravo za nju kaže da su u njezinu značenjskom zbiru česti pridjevi, koji su i u različitim primjenama obrade prirodnog jezika glavni izvor podataka (npr. u analizi mnijenja) (238). Ovime se želi reći da se ono što profesionalni čitatelj uhvati „golim okom” prilikom čitanja teksta može nadograditi udaljenim čitanjem pomoću računalnih alata, a da se pritom koriste jednaki književnoteorijski pristupi (u ovome slučaju naratologija). Time se omogućava dosljednije tumačenje jer se rezultati stavljaju u kontekst poznatih književnih kategorija (npr. izravna definicija lika). Budući da je ovo demonstracijsko istraživanje i da za hrvatske tekstove ne postoji slično, osnovno je istraživačko pitanje rada mogu li se računalno potpomognutom analizom, utemeljenoj na naratologiji, dobiti podaci korisni za interpretaciju književnih likova, vodeći se postojećim književnoteorijskim elementima karakterizacije likova.

Analizirani književni tekstovi preuzeti su u obliku .txt datoteka s portala e-Lektire,[6] a izrada korpusa i istraživanje provodili su se u već spomenutom internetskom alatu Sketch Engine.[7]

4. Računalne metode

Nakon što su se preuzele tekstualne datoteke, napravila se normalizacija – postupak izbacivanja suvišnih dijelova teksta, tj. sve ono što nije dio osnovnog književnog teksta, na primjer: naslovi, bilješke, kazala, rječnici i sl. Zatim je uslijedila vertikalizacija i morfosintaktičko označavanje u Sketch Engineu, ali i preko CLARIN.SI[8] stranice, kako bi se mogle usporediti točnosti klasifikatora.[9] Oba su postupka računalni alati napravili automatski kao sastavni dio procesa izrade korpusa, uz napomenu da se točnost vertikalizacije nije provjeravala, a točnost morfosintaktičkog označavanja jest.

Kako bi korpus bio što točnije označen za daljnju analizu likova, ručno su se ispravile greške u označavanju koje su se ticale naziva likova. Na početku valja istaknuti jedan problem u označavanju. Radi se o nazivima likova u bajci Šuma Striborova. Likovi u tom tekstu pretežito nemaju osobna imena, nego su oslovljeni općim imenicama (stara majka, guja, ubogo djevojče, momak). Zašto je to bilo problematično? Da bi se prilikom pretraživanja korpusa došlo do rečenica u kojima se javljaju likovi, oni moraju nekako biti „dohvatljivi", tj. moraju biti označeni kao vlastite imenice – kao što su označeni likovi u ostalim tekstovima. Stoga se taj dio morao napraviti ručno. Ovo je samo jedan primjer koji dobro pokazuje koliko se književni tekstovi razlikuju od ostalih vrsta tekstova te koliko je znanje književnog stručnjaka vrijedno i potrebno. Međutim, u ovom slučaju tehnologija pomaže stručnjaku da barem ima ispravno označene vrste riječi (imenica kao vrsta riječi), ako već ne i vrstu imenice. Nakon toga izračunao se udio osobnih zamjenica, a koje se odnose na analizirane likove, u nasumičnom uzorku osobnih zamjenica, te se isti postupak ponovio za udio vlastitih imenica. Htio se vidjeti udio osobnih zamjenica i vlastitih imenica jer se time može dobiti uvid koliko su likovi zastupljeni u pojedinom književnom tekstu, odnosno koliko im se daje prostora naspram ostalih elemenata kao što su mjesto radnje ili pripovjedač, koji se također realiziraju preko osobnih zamjenica (prvenstveno pripovjedač) i vlastitih imenica. Posljednji korak u prikupljanju podataka za karakterizaciju likova bilo je izvlačenje GDEX primjera, koji su

poslužili za dobivanje izravnih definicija likova, ali i analizu glagola koji se pojavljuju uz likove kako bi se proučila neizravna prezentacija. GDEX primjeri nastaju kad se u Sketch Engineu pretraži neki pojam, a čije se konkordancije zatim vrednuju s obzirom na to koliko su prikladni rječnički primjeri, ako se prilikom pregleda rezultata upita izabere GDEX opcija. Ovo je posebno korisno leksikografima jer se pokretanjem GDEX opcije sortiraju primjeri s obzirom na svoju kvalitetu, odnosno koliko su prikladni rječnički primjeri, ako se u obzir uzmu određeni unaprijed definirani parametri, što može skratiti vrijeme filtriranja „dobrih" i „loših" primjera. Parametri na temelju kojih se to vrši jesu: duljina rečenice, složenost vokabulara, koliko su prisutne takozvane kontroverzne teme (politika, religija i sl.), kontekst, referiranje na sadržaje drugih okolnih rečenica (npr. putem zamjenica), nazivi robnih marki itd. (Sketch Engine).

5. Rezultati

Za dobivanje udjela osobnih zamjenica koje se odnose na likove nasumično se iz korpusa u Sketch Engineu izvuklo ukupno 500 osobnih zamjenica (početna ideja bila je po 100 zamjenica iz svakog teksta) i gledalo se koliko se njih odnosi na likove koji se pojavljuju u tekstu. Budući da se pokazalo da u Oko Lobora u cijelom tekstu ima samo 96 zamjenica, odlučilo se da se iz preostala četiri teksta izvuče po 101 zamjenica kako bi ukupan broj uzorka bio 500. Do potrebne količine zamjenica koje su činile uzorak došlo se postupkom randomizacije, koji Sketch Engine nudi i koji slučajnim odabirom izvlači riječi iz teksta. Postotak osobnih zamjenica koje se odnose na likove mogu se vidjeti u tablici 2. Medijan pojavljivanja zamjenica za prva četiri teksta u tablici iznosi 87,12 %, a tekst Oko Lobora izdvojen je zbog ekstremnog rezultata koji odudara od ostalih.Tablica 2. Udio i broj osobnih zamjenica koje se odnose na likove

	Šuma Striborova	Oprava	Krvavi most u Zagrebu	Posljednji Stipančići	Oko Lobora
udio osobnih zamjenica	86,14 %	86,14 %	87,13 %	89,11 %	16,67 %

Sličan postupak napravio se i za vlastite imenice – izvuklo se po 100 vlastitih imenica iz svakog teksta (kako bi također ukupan uzorak iznosio 500) i izračunalo se koliko se njih odnosi na osobna imena likova. Rezultati su vidljivi u tablici 3. Također se izračunao medijan pojavljivanja za prva četiri teksta u tablici i on iznosi 72,50 %. Oko Lobora ponovno odudara od ostalih tekstova i zato nije uključen u izračun medijana.Tablica 3. Udio i broj vlastitih imenica koje se odnose na likove

	Šuma Striborova	Oprava	Krvavi most u Zagrebu	Posljednji Stipančići	Oko Lobora

udio vlastitih imenica	88 %	77 %	41 %	84 %	0 %

U tablici 4. mogu se vidjeti podaci o zastupljenosti ostalih kategorija unutar vlastitih imenica, a da se ne tiču osobnih imena likova.Tablica 4. Broj vlastitih imenica po različitim kategorijama

	Šuma Striborova	Oprava	Krvavi most u Zagrebu	Posljednji Stipančići	Oko Lobora
Np24318 (ukupno u tekstu)	188	117	543	3001	154
Np (uzorak)	100	100	100	100	100
Np (likovi)	88	77	41	84	0
Np (lokacije)	0	8	25	8	42
Np (povijesne osobe)	0	2	20	5	41
Np (narodi)	0	0	1	1	6
Np (institucije)	0	0	0	0	1
Np (ostalo)	7	1	1	0	6
pogrešno označene riječi25458	5	12	12	2	4

Dosad izneseni rezultati odnosili su se na općenitu situaciju s likovima u odabranim tekstovima. Rezultati koji slijede daju podatke za karakterizaciju točno određenih likova u pojedinim tekstovima (navedeni u poglavlju „O istraživanju”: stara majka, Domaći i Malik Tintilinić, Lucija Stipančić, kanonik Vukasović, Đuro Vučetić, Pavao Slavinić, župan Gardun).

Opcija GDEX u Sketch Engineu izbacila je primjere za pojedini lik, koji su bili vrednovani u rasponu od 0 do 1. Što je konkordancija bolji rječnički primjer, s obzirom na parametre koje GDEX opcija u sebi sadrži, to je njezin rezultat bliži 1. Što je konkordancija lošiji rječnički primjer, to je rezultat bliži 0. Ako je broj primjera koje je GDEX izbacio bio manji od 100, analizirali su se svi primjeri. Ako je broj bio veći od 100, analiziralo se prvih 100 primjera. Podaci o broju primjera dostupni su u tablici 5.Tablica 5. Podaci o GDEX primjerima u tekstovima

	baka (majka)	Domaći	Malik Tintilinić	Lucija Stipančić	Vukasović	Đuro Vučetić	Pavao Slavinić	Gardun
broj GDEX primjera	66	18	11	463 (uzorak: 100)	196 (uzorak: 100)	34	56	78
broj izravnih definicija	3	1	0	2	12	3	5	6

Izgenerirani GDEX primjeri poslužili su i za izvlačenje glagola koji se vezuju uz odabrane likove kako bi se promotrila njihova neizravna prezentacija. Za potrebe ovog istraživanja napravila se podjela glagola u četiri kategorije, a do koje je došlo proučavanjem glagola (izvučenih iz GDEX primjera) koji se vezuju uz odabrane likove: oni koji se odnose na: 1) dinamične radnje (otići, donijeti, poskočiti, popeti se itd.), 2) statične radnje (sjesti, gledati, pripovijedati, slušati itd.), 3) glagoli vezani uz emocije i razmišljanje (zamisliti se, plakati, nasmijati se, ciknuti itd.) te 4) glagoli

zbivanja (umrijeti, mladovati itd.). Zastupljenost četiriju navedenih kategorija za odabrane likove vidljiva je u tablici 6.Tablica 6. Udio pojedinih kategorija glagola za odabrane likove

	baka (majka)	Domaći i Malik Tintilinić	Lucija Stipančić	Vukasović	Đuro Vučetić	Pavao Slavinić	Gardun
ukupan broj glagola	40	31	83	62	43	62	95
% dinamičnih radnji u ukupnom broju glagola	37,5 %	61,29 %	8,43 %	19,35 %	0 %	50 %	29,47 %
% statičnih radnji u ukupnom broju glagola	32,5 %	22,58 %	53,01 %	59,68 %	58,14 %	24,19 %	46,31 %
% glagola vezanih uz emocije i razmišljanja u ukupnom broju glagola	25 %	12,90 %	31,33 %	16,13 %	39,53 %	20,97 %	23,16 %
% glagola zbivanja u ukupnom broju glagola	5 %	3,22 %	7,23 %	4,84 %	2,32 %	4,83 %	1,05 %

6. Diskusija

Postotak osobnih zamjenica u uzorku koji se odnosi na likove u većini tekstova iznosi više od 80 %. Izuzetak je jedino putopis Oko Lobora, u kojem postotak iznosi tek 16,67 %. Tomu je tako jer u ostalim tekstovima ima više likova, a još više jer su to pripovjedni prozni tekstovi, dakle radnja i likovi razgranatiji su, imaju veću ulogu, dok je Matošev tekst putopis, gdje je jedini lik ujedno i pripovjedač. U noveli Oprava također nema mnogo likova, dominira jedan glavni lik, ali za razliku od Matoševa putopisa taj glavni lik dobro je emocionalno i psihološki razrađen, naglasak je na njegovu unutarnjem svijetu i iako radnja nije bogata, postoji mnogo osobnih zamjenica koje se odnose na glavnog lika. Iz navedenog se može zaključiti da razrađena radnja i/ili razrađeni likovi vode do povećane uporabe osobnih zamjenica koje se tiču likova. Naravno, u budućnosti bi bilo korisno usporediti upotrebu osobnih zamjenica u nekom općem korpusu i vidjeti koliko književni tekstovi od njega odstupaju.

Sljedeći rezultati ticali su se udjela imena likova u vlastitim imenicama, a na općenitoj razini zastupljenost pojedinih tipova vlastitih imenica u analiziranim tekstovima. Rezultati pokazuju da po udjelu imena likova u nasumičnom uzorku vlastitih imenica odstupaju tekstovi Oko Lobora i Krvavi most u Zagrebu. U Matoševu putopisu nema pojave imena likova, što je razumljivo s obzirom na to da je jedini lik pripovjedač. Pripovijetka Krvavi most u Zagrebu ima najmanji udio imena likova u

odnosu na preostale tekstove, što može ukazivati na manje pridavanje važnosti likovima, odnosno njihovoj karakterizaciji.

Osim imena likova zanimljivo je vidjeti koja se još vlastita imena pojavljuju u pojedinim tekstovima te nameću li žanr ili tematika neke posebnosti (brojčani podaci o zastupljenosti pojedinih kategorija vlastitih imenica nalaze se u tablici 4.). Krene li se od Šume Striborove, jasno se uočava da su nazivi likova dominantni, a ono malo ostalih vlastitih imenica koje se pojavljuju u uzorku odnose se na imenicu Bog. To se može povezati s književnom vrstom kojoj tekst pripada – umjetničkom bajkom. Dominacija naziva likova može se tumačiti činjenicom da su za bajku bitni odnosi između likova, koji su zapravo brojni ako se uzme u obzir kratkoća samoga teksta, a što znači da će imati više pojavnica u tekstu za razliku od ostalih tipova vlastitih imenica.

Novela Oprava također ne obiluje raznolikošću vlastitih imenica. Dominantna su imena likova, a brojčano s osam pojavnica slijede nazivi lokacija – točnije Gleichenberg (7 pojavnica) i Zagreb (1 pojavnica). Gleichenberg je bitno mjesto jer ondje glavni lik odlazi na liječenje te ga se prisjeća nakon što se vrati kući u Slavoniju. Ako se usporedi broj pojavnica likova u Šumi Striborovoj i Opravi, razlika nije prevelika, ali očituje se u tome što se većina pojavnica u Opravi odnosi samo na Đuru Vučetića, dok bajka Ivane Brlić-Mažuranić raspodjeljuje pojavnice na više likova. Ono što se želi reći jest da je potrebno poznavati radnju svakog teksta kako bi se naizgled slični rezultati mogli pravilno tumačiti.

Krvavi most u Zagrebu, po vrsti pripovijetka, razlikuje se od prethodna dva teksta. Naime, uz likove, u zamjetnom broju pojavljivanja nalaze se lokacije i povijesne osobe. Ova je pripovijetka narativno razgranatija i kompleksnija u odnosu na Opravu i Šumu Striborovu, a i po tematici je drugačija. Velik dio fabule uključuje povijesna zbivanja i borbu za vlast između zaraćenih strana – pristaša biskupa Mihalja i pristaša župana Garduna. Uz to, postoji i ljubavna narativna linija koja prati odnos Pavla Slavinića i njegove Ružice, međutim pojavljivanje lokacija i povijesnih osoba postoji prije svega zbog povijesne tematike. Iz perspektive računalno potpomognute analize ono što se ovdje nameće kao problem jest razlikovanje imena likova od primjerice imena povijesnih osoba, vladara, političara i sl. Morfosintaktičko i NER (engl. Named Entity Recognition) označavanje ne prave distinkciju između takvih imena, zato je važno da istraživač bude upućen je li nešto ime lika ili nije.

Slična je situacija u Matoševu putopisu Oko Lobora. Naime, već je ranije spomenuto, a i jednostavno se primijeti iz tablice 4., u putopisu postoji samo jedan (glavni) lik, koji je istovremeno i pripovjedač, stoga se on jezično realizira preko osobne zamjenice ja, a ne preko vlastitog imena. Ipak, ono što se od vlastitih imenica pojavljuje, odnosi se, kao i u Krvavom mostu u Zagrebu, na lokacije i povijesne osobe. Za razliku od Bogovićeve pripovijetke, gdje se radnja zbiva u konkretnom povijesnom kontekstu i povijesne su osobe suvremenici likova, u Oko Lobora dolazi do prisjećanja prošlih vremena i u tim se dijelovima teksta imenuju pojedine povijesne osobe. Iz navedenog se može zaključiti da bilo kakvo narativno ili asocijativno odlaženje u prošlost potencijalno donosi učestaliju pojavnost ovakvih imenica. To saznanje moglo bi biti korisno u detekciji tematike ili žanra.

U najduljem tekstu ovog istraživanja, romanu Posljednji Stipančići, dominiraju osobna imena likova. U jako malom broju pojavljuju se imena povijesnih osoba i lokacija. Iako se roman bavi društveno-političkom tematikom, nema mnogo povijesnih osoba jer su likovi ti koji su svojevrsne povijesne osobe. Što se lokacija tiče, također ih ima malo jer je radnja usmjerena na senjsko područje.

GDEX opcijom u internetskom programu Sketch Engine htjele su se izvući izravne definicije odabranih likova. Odlučilo se za ovu opciju jer se njome mogu izvući dobri rječnički primjeri, odnosno definicije određenih pojmova. Dakle, ideja je bila da su izravne definicije likova svojim ustrojstvom slične rječničkim definicijama pojmova. Ako se pogleda vrednovanje GDEX primjera, primjećuje se da su najlošije vrednovani primjeri (oni najbliži vrijednosti 0) zapravo najbolje izravne definicije i obrnuto. U pravilu su jednostavnije i kraće rečenice imale najbolji rezultat, odnosno smatrane su dobrim rječničkim primjerom (bliži 1), ali za karakterizaciju likova puno su informativnije dulje rečenice, u kojima se više opisuje lik, ali koje zbog svoje opsežnosti nisu toliko dobre za rječničke definicije. U nastavku se može vidjeti primjer toga.

Primjer kraće rečenice: „Taj je čovjek bio Stipančićev daljnji rođak, kanonik Vukasović.” (GDEX vrijednost: 0,875).

Primjer dulje rečenice: „Glava pokreta, kanonik Vukasović , bio je uman čovjek, od prirode bogato

nadaren, a naukom prosvijetljen, te ne samo porijeklom i imenom nego i duhom svojim pravi narodni aristokrat.” (GDEX vrijednost: 0,09).

Tablica 5. donosi broj GDEX primjera za pojedine likove i koliko su od tih GDEX primjera izravne definicije. Pitanje koje se ovdje može javiti jest zašto svi GDEX primjeri nisu izravne definicije. Obrazloženje se nalazi u tome da su GDEX primjeri zapravo konkordancije kojima je pridružena određena vrijednost od 0 do 1 i time vrednovana njihova prikladnost za rječnički primjer. Stoga se trebala napraviti selekcija samo onih primjera koji odgovaraju izravnoj definiciji lika, koji nose neke informacije o liku i njegovu karakteru. Suprotno standardnom vrednovanju GDEX primjera, kod izravnih definicija slučaj je da što je vrijednost bliža nuli, to je izravna definicija kvalitetnija. Ostali primjeri GDEX opcije koji nisu izravne definicije pretežito se odnose na upravni govor i radnje likova, opisi vanjskog izgleda su nezamjetni. U nastavku se mogu vidjeti takvi primjeri za lik stare majke iz Šume Striborove.

„Bog mu oprosti, a majka mu nije ni zamjerila bila.”

„To se tako ražalilo majci, da odmah pođe na liticu po snijeg, jer joj nije bilo žao života.”

„Majka pak radosno pođe do sina govoreći: - 'Otpremi je, sine, otkud si je doveo, sad si na svoje oči vidio, koga u kući hraniš.'”

Prilikom generiranja GDEX primjera kao ključne riječi koristili su se nazivi likova i vjerojatno je to razlog zašto nema više primjera izravnih definicija. Naime, nije rijetko u književnosti da se na lika referira preko osobnih zamjenica, pogotovo ako se radi o nekom kraćem odlomku gdje se opisuje lik. Prvo ga se imenuje, a zatim se redaju rečenice u kojima se navode njegove karakteristike. Naravno, nije ni ovo pravilo koje uvijek vrijedi, ali osobne zamjenice često se referiraju na likove i takve bi se rečenice u nekim budućim istraživanjima trebale uzeti u obzir, uz napomenu da pretraživanje preko zamjenica može biti dosta složeno jer se u hrvatskom jeziku one ne koriste samo za osobe, nego i za ostale nežive imenice. Iz navedenog se može zaključiti da GDEX primjeri nisu donijeli prevelik broj izravnih definicija kako se prvotno očekivalo, ali pokazali su se korisnima kao izvor informacija za karakterizaciju likova pomoću neizravne prezentacije. U nekim budućim nadogradnjama istraživanja za pronalazak većeg broja izravnih definicija moglo bi se poslužiti CQL izrazima;[12] međutim taj korak zahtijeva detaljan uvid u strukturu rečenica za svaki

tekst, s obzirom na to da upit koji se postavlja alatu mora imati jasno definiran redoslijed vrsta riječi kako dobiveni rezultati ne bi bili preopćeniti.

Kao što je prije rečeno, nisu svi GDEX primjeri izravne definicije, ali preostali primjeri poslužili su kao baza za analizu glagola koji se vežu uz pojedinog lika. Detaljna raspodjela glagola po kategorijama vidljiva je u tablici 6. Odlučeno je da se gledaju glagoli jer se pomoću njih može steći uvid u osobine lika – koliko je statičan, dinamičan, refleksivan itd.

Likovi iz Šume Striborove, stara majka i Domaći (s Malikom Tintilinićem), najviše su prikazani kroz dinamične radnje. Majka tako donosi, potpaljuje, odlazi, poskakuje. Ona brine o obitelji, obavlja različite zadatke (koje joj zadaje i snaha-guja), aktivan je lik. Domaći su vrckavi i zaigrani te pokušavaju pomoći staroj majci. Također je i nastoje razvedriti i vratiti joj onaj nestašan dječji duh koji sami posjeduju. Osim dinamičnih radnji, kod stare majke pojavljuju se statične radnje u svega 5 % manjoj zastupljenosti. Majka je zabrinuta za sudbinu svoga sina, ima potrebu pripovijedati o tome Domaćima, dijeli svoju tugu i jad s ostalima.

Dvoje odabranih likova iz Posljednjih Stipančića, Lucija Stipančić i kanonik Vukasović, najviše uza sebe imaju glagole koji se vezuju uz statičnost. Iz toga bi se možda moglo zaključiti da su ta dva lika slična po svojim postupcima. Međutim, ako se pogledaju konkretni primjeri glagola, zamjećuje se da se radi o različitim oblicima statičnosti. „Lucijini glagoli" puno su suptilniji i mirniji: ona promatra, gleda, sjedi itd. Lucija Stipančić više je okrenuta prema introspekciji i proučavanju okoline. S druge strane, kanonik Vukasović onaj je koji huška, saziva, opire se, pripovijeda. „Njegovi glagoli" vezani su uz proaktivnost, moć i bunt. Vukasović kao politička osoba često mora uvjeravati svoje protivnike i pristaše u određene stvari, voli voditi glavnu riječ. Ono što dodatno može pomoći u razlikovanju ovih dvaju likova jest to što trećina (31,33 %) glagola vezanih uz Luciju Stipančić jesu glagoli emocija i misli. Čak i njezine statične radnje nisu „čisto" statične, odnosno više naginju prema ovom emocionalno-misaonom polu. Ponovno se pokazuje da nisu dovoljni samo statistički podaci da bi se došlo do ispravnih zaključaka. Kontekst i uvid u primjere od velike su važnosti.

Lik koji po razini emocionalnosti, statičnosti i promišljanja parira Luciji Stipančić jest Đuro Vučetić, glavni lik novele Oprava. Kod njega je slična situacija kao kod Lucije – 58,14 % čine statične radnje, 39,53 % analiziranih glagola vezani su uz emocije i promišljanja. Zanimljivo je da čak

nijedan glagol iz uzorka nije dinamične naravi. Sve to ide u prilog žanru kojem tekst pripada – psihološka novela. Naglasak je na jednome liku, njegovim osjećajima, psihičkim stanjima. Sama radnja također „tjera" Đuru da sve propituje. Naime, on sumnja da ga supruga vara i da on nije biološki otac svome sinu. Uz to je oslabljenog zdravlja, i duševnog i tjelesnog.

Dva glavna sukobljena lika u pripovijetki Krvavi most u Zagrebu, Pavao Slavinić i župan Gardun, ističu se po zastupljenosti statičnih i dinamičnih radnji, ali u različitim omjerima. Kod Pavla Slavinića najviše ima dinamičnih radnji (50 %), zatim slijede statične (24,19 %), a brojčano su vrlo blizu i glagoli vezani uz emocije i promišljanja (20,97 %). Dinamičnost ovog lika ponajprije je povezana sa sukobom sa županom Gardunom. Pavao Slavinić nalazi se na suprotnoj strani kao pristaša biskupa Mihalja i redovito dolazi u okršaje. Statične radnje najviše su realizirane kroz glagole govorenja jer, poput Garduna, Slavinić mora bodriti svoje ljude i s njima komunicirati (kao i sa svojim protivnicima). Emocije i promišljanja povezana su s obje narativne linije – ljubavnom i političkom, stoga je jasno da kod ovog lika ne nedostaje takvih glagola.

Župan Gardun prednjači u statičnim radnjama (46,31 %) koje su vrlo slične radnjama kanonika Vukasovića – Gardun zapovijeda, saziva, govori, hrabri itd. Obojica imaju publiku koja ih sluša te ih govorom žele uvjeriti u svoje ideje, a time pokazuju svoju moć. U dinamičnim radnjama uočava se sličnost s Pavlom Slavinićem – za oba lika koriste se vrlo slični glagoli i po tome bi se moglo zaključiti da se oni previše ne razlikuju, odnosno ne može se uočiti njihova polarnost u aspektu borbe dobra i zla. Pavao Slavinić pozitivan je lik, a župan Gardun negativan. Glagol koji prikazuje tu razliku jest glagol haračiti, koji se pojavljuje uz Garduna, ali ne i uz Pavla Slavinića, i ima negativnu konotaciju. Lik Gardun također ima glagola vezanih uz emocije i promišljanja (23,16 %) i podrijetlo je kao i u Pavla Slavinića – ljubav i sukob.

Glavni lik (pripovjedač) putopisa Oko Lobora predstavlja specifičnu situaciju. Kako je prije rečeno, GDEX primjera nema jer nema naziva lika, ali za potrebe proučavanja glagola napravio se klasičan konkordancijski upit za osobnu zamjenicu ja. Pokazalo se da pripovjedač nije aktivan lik i da jedino vrijeme kad se može smatrati da on vrši neku radnju jest u trenutku zamišljanja kako bi bilo da je posljednji gospodar Lobora. U većini primjera pripovjedač je objekt – u njega se gleda, njega se vodi i sl., a subjekti su neživa bića (sjena, noć itd.).

Osim što su se analizirale izravne definicije i tipovi radnji u kontekstu pojedinog lika, glagoli mogu poslužiti i za utvrđivanje rutinskih i nerutinskih radnji, što je dio neizravne prezentacije lika. Kao što i sam naziv kaže, rutinske radnje one su radnje koje lik ponavlja više puta, koje su dio njegove rutine. Nerutinske su pak radnje suprotno – radnje koje lik čini svega nekoliko puta ili samo jednom, ali to izlaženje iz rutine može biti važno za karakterizaciju lika. Vjerojatno najpoznatiji primjer nerutinske radnje, a koja je presudna u karakterizaciji lika, jest u romanu Zločin i kazna kad mladi Raskoljnikov sjekirom ubije staru lihvaricu. Računalno potpomognutom analizom može se doći do podatka o rutinskim ili nerutinskim radnjama, ako uzmemo u obzir da se radnje većinski jezično realiziraju glagolima. Broj pojavnica glagola kazuje koliko se neki čin često ili rijetko događa. Ono što ovdje predstavlja izazov jest „uparivanje" lika i njegovih radnji, odnosno nije dovoljno izvući samo statističke podatke o glagolima u tekstu jer se time ne dobiva informacija uz koga se ili što vezuje taj glagol. Stoga su se u analizi rutinskih i nerutinskih radnji u ovom radu koristili glagoli iz GDEX primjera, jer su oni sigurno vezani uz konkretnog lika. Već se iz analize tipova glagola moglo vidjeti da primjerice kanonik Vukasović i župan Gardun često drže govore, kazuju, nagovaraju, potiču i to se može shvatiti kao njihove rutinske radnje. Lucija Stipančić pak plače, crveni se, okreće glavu, sluša, što je u skladu s njom kao emotivnom, plahom djevojkom. Stara majka iz Šume Striborove pripovijeda, brine se, razgovara. To su sve radnje koje su kod nje najviše zastupljene i mogle bi se shvatiti kao rutinske. S ovih nekoliko primjera želi se reći da su glagoli koji su najzastupljeniji po podjeli iz tablice 6. zapravo dobra podloga za tumačenje rutinskih radnji likova. Što se tiče nerutinskih radnji, situacija je malo složenija. Naime, likovi koji se analiziraju u ovome radu pored sebe nemaju neke specifične glagole male pojavnosti za koje bi se moglo reći da su nerutinske radnje, točnije da pridonose njihovoj karakterizaciji. Moglo bi se zaključiti da pojavnost glagola može biti od koristi u pronalaženju rutinskih radnji, dok se za nerutinske moraju još istražiti mogućnosti koje računalni alati i programi nude. Svakako treba imati na umu da bi veći korpusi s više različitih tekstova bili puno pogodniji za donošenje općenitih zaključaka, a koji bi onda vrijedili za većinu tekstova određenog tipa.

Osim rutinskih i nerutinskih radnji, kao dio neizravne prezentacije može se proučavati i govor likova. Tijekom ove diskusije rezultata u nekoliko se navrata spominjao govor u kontekstu toga koji likovi pripovijedaju (npr. stara majka), nagovaraju i uvjeravaju (npr. kanonik Vukasović),

zapovijedaju (npr. Gardun), a koji pak slušaju (npr. Domaći i Malik Tintilinić), šute, ne progovaraju (npr. Lucija Stipančić). Na ovaj način dolazi do dodatnog raslojavanja značenja glagola, odnosno gledaju se glagoli koji se odnose na govorenje (ili negovorenje) te se pokušava okarakterizirati lika. Vidljivo je da se proaktivniji likovi kao što su Gardun, Vukasović te donekle Pavao Slavinić češće obraćaju, drže govore, bodre, uvjeravaju i potiču. Slično njima, i stara majka također često pripovijeda, ali već po izboru glagola (i sveukupnoj karakterizaciji njezina lika) vidi se da je njoj razgovor svojevrsna terapija, odnosno ona želi s drugima podijeliti svoju zabrinutost i kroz razgovor doći do rješenja. S druge strane, Lucija Stipančić i Domaći oni su koji slušaju, ali Lucija je više po strani u društvu i zapravo rijetko kad netko pita za njezino mišljenje. Domaći pak slušaju jer staru majku zanima kako joj mogu pomoći, dakle radi se o slušanju s drugačijom funkcijom nego kod Lucije Stipančić. Đuro Vučetić pak gotovo uopće uza sebe nema glagola koji bi izricali (ne)govorenje. To je u skladu s njegovim likom, koji uglavnom razmišlja o stvarima, emocije drži za sebe i sam sa sobom o njima raspravlja. Dodatna tome doprinosi i to što Oprava ima malo likova, pa Vučetić nema ni toliko prilika da stupi s nekim u verbalni kontakt.

7. Zaključak

Digitalna humanistika kao relativno novo interdisciplinarno područje nudi mnoštvo novih perspektiva u pristupanju književnosti. Posebno se to može reći za književnost na hrvatskom jeziku jer ima vrlo malo radova koji se time bave. Tome valja pridodati i činjenicu da ne postoje specijalizirani računalni alati koji bi bili napravljeni baš za književne tekstove. Alata za obradu (hrvatskog) prirodnog jezika ima, ali ne mogu se uvijek prikladno primijeniti na književnost jer nisu razvijeni na književnim tekstovima, a dio razloga leži i u tome što je književnost ipak umjetnost, kompleksno jezično ostvarenje puno skrivenih značenja, intertekstualnosti, stilskih figura zbog kojih se razlikuje od uobičajenih publikacija. Stoga je ovaj rad doprinos ovom slabo razvijenom području unutar hrvatske akademske zajednice.

Na konkretnim tekstovima i analizi njihovih likova moglo se pokazati što se sve računalnim alatima te statističkim metodama i metodama iz obrade prirodnog jezika može postići. Početno istraživačko pitanje bilo je mogu li se računalno potpomognutom analizom, utemeljenom na naratologiji, dobiti podaci korisni za interpretaciju književnih likova, temeljeći se na postojećim književnoteorijskim

elementima karakterizacije likova. Tumačenjem rezultata dobio se potvrdan odgovor na postavljeno istraživačko pitanje. Upotrebom elemenata karakterizacije iz naratologije htio se dati književnoteorijski okvir kako bi analiza bila što usmjerenija i književno relevantna.

Prilikom provođenja istraživanja uočeno je da postoje određeni problemi i pitanja koja se tiču primjene alata, odnosno kako računalno detektira pojedini element karakterizacije s obzirom na to da ne postoje specijalizirani alati kojima bi se primjerice jednostavno mogle izvući izravne definicije ili neizravne prezentacije. Bilo je potrebno empirijski razraditi teorijske postavke naratologije kako bi se računalni alati mogli koristiti na ispravan način. Javljali su se i problemi oko samog jezika – što učiniti kada su likovi nazvani općim imenicama ili ako se javlja više različitih imena pa treba odrediti koja se odnose na likove, a koja ne. Sve to zahtijevalo je fleksibilnost i prilagodljivost situaciji te eksperimentiranje s dostupnim alatima.

Prostora za daljnji rad i razvoj ima. Prije svega, karakterizacija likova mogla bi se proširiti na ostale elemente karakterizacije za koje se trenutno procijenilo da su prekompleksni za računalnu obradu. Na primjer: vanjski izgled, okoliš u kojem se likovi nalaze – umjetni i prirodni. Nadalje, analiza bi se mogla proširiti i na sve spomenute i korištene elemente karakterizacije, ali u primjerima gdje lik nije eksplicitno imenovan – dakle u dijelovima teksta gdje se na lik referira na neki drugi način (najčešće osobnim zamjenicama). S obzirom na to da za hrvatski jezik ne postoje nikakvi specijalizirani računalni alati za književnost, prostor za napredak je velik.

Bibliografija

Grdešić, Maša. Uvod u naratologiju. Leykam international, d. o. o., 2015.

Hrvatski jezični portal, hjp.znanje.hr/index.php?show=search. Posjećeno 2. prosinca 2022.

Hrvatski mrežni rječnik – MREŽNIK, ihjj.hr/mreznik/. Posjećeno 30. studenoga 2022.

Jacobs, Arthur M. „Sentiment Analysis for Words and Fiction Characters From the Perspective of Computational (Neuro-)Poetics.” Frontiers in Robotics and AI, vol. 6, 2019, doi.org/10.3389/frobt.2019.00053. Posjećeno 10. siječnja 2023.

Kasunić, Lorena. Računalno potpomognuta analiza likova u odabranim tekstovima hrvatske

književnosti. 2021. Sveučilište u Zagrebu, diplomski rad. ODRAZ – open repository of the University of Zagreb Faculty of Humanities and Social Sciences.

Ligthart, Alexander, et al. „Systematic Reviews in Sentiment Analysis: A Tertiary Study." Artificial Intelligence Review, vol. 54, 2021, str. 4998-5053, doi.org/10.1007/s10462-021-09973-3. Posjećeno 10. siječnja 2023.

Moretti, Franco. Distant Reading. Verso, 2013.

Mukařovský, Jan. „Strukturalizam." Suvremene književne teorije, ur. Miroslav Beker, Matica hrvatska, 1999, str. 158-68.

Nalisnick, Eric T., and Henry S. Baird. „Character-to-Character Sentiment Analysis in Shakespeare's Plays." Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), ur. H. Shuetze i sur., Association for Computational Linguistics, 2013, str. 479-83, aclanthology.org/P13-2085/. Posjećeno 11. siječnja 2023.

Peleš, Gajo. Tumačenje romana. ArTresor naklada, 1999.

Pintarić, Ana. Umjetničke bajke – teorija, pregled i interpretacije. Filozofski fakultet Osijek, 2008.

Sketch Engine, www.sketchengine.eu/. Posjećeno 12. siječnja 2023.

Solar, Milivoj. Ideja i priča: Aspekti teorije proze. LIBER, 1974.

Solar, Milivoj. Teorija književnosti. Školska knjiga, 1984.

Universal Dependencies, universaldependencies.org/. Posjećeno 17. siječnja 2023.

Vala, Hardik, et al. „Mr. Bennet, His Coachman, and the Archbishop Walk into a Bar but Only One of Them Gets Recognized: On The Difficulty of Detecting Characters in Literary Texts." Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisabon, Portugal, 17.–21. rujna 2015., Association for Computational Linguistics, 2015, str. 769-74, www.aclweb.org/anthology/D15-1088.pdf. Posjećeno 11. siječnja 2023.

[1] Prirodni jezik jest svaki jezik koji se razvio spontano s ciljem komunikacije među ljudima (npr. hrvatski, engleski, njemački).

[2] Konkordancija je pojavljivanje riječi s njezinim okolnim kontekstom, tj. s riječima koje joj prethode i slijede u određenom jezičnom nizu (npr. rečenici).

[3] „Univerzalne ovisnosti (UD) okvir su za dosljednu oznaku gramatike (označavanje vrsta riječi, morfološka obilježja i sintaktičke ovisnosti) za različite jezike" (Universal Dependencies, vlastiti prijevod). Riječ je o projektu koji razvija banke stabala (engl. treebank) kako bi se mogli proučavati ovisnosni odnosi između riječi.

[4] Izdanja tekstova koji čine korpus: Šuma Striborova iz zbirke Priče iz davnine (2018.), Oprava je dio zbirke Odabrane pripovijetke (n.d.), Oko Lobora nalazi se u Odabrani eseji i putopisi (n.d.), Krvavi most u Zagrebu (n.d.), Posljednji Stipančići (2018.).

[5] Više detalja o navedenim aktancijalnim modelima nalazi se u Kasunić 18-20.

[6] https://lektire.skole.hr/

[7] https://www.sketchengine.eu/

[8] http://clarin.si/services/web/login

[9] Vidi: Kasunić 28.

[10] Kratica Np označava morfosintaktičnu oznaku za vlastitu (p – personal) imenicu (N – noun).

[11] Riječi koje nisu vlastite imenice, a označene su nekom od oznaka za vlastite imenice.

[12] CQL (engl. Corpus Query Language) izraz jest tip upita kojim se pretražuju složeni gramatički ili leksički izraz u korpusu.

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License