Primerjalna književnost

TEMATSKI SKLOP / THEMATIC SECTION

Veliki podatki, male literature
Big Data, Small Literatures

RAZPRAVE / ARTICLES

TEMATSKI SKLOP / THEMATIC SECTION

Veliki podatki, male literature
Big Data, Small Literatures

Uredili / Edited by: Lucija Mandić, Ivana Zajc

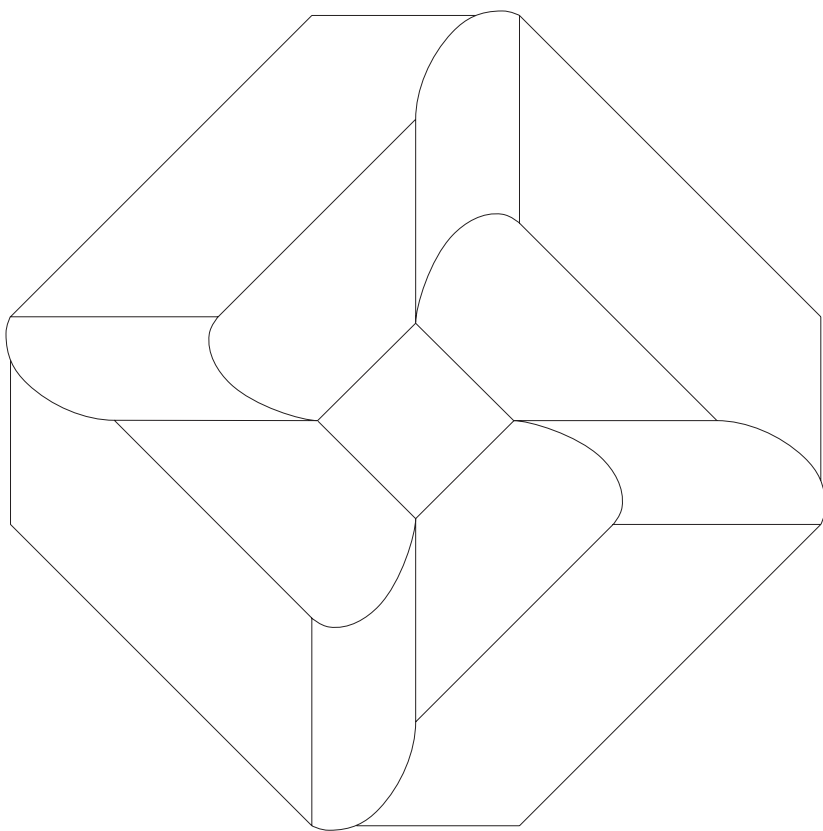# Tematski sklop / *Thematic section*

Veliki podatki, male literature
Big Data, Small Literatures
Uredili / *Edited by*: Lucija Mandić, Ivana Zajc

Veliki podatki, male literature (predgovor)

Lucija Mandić, Ivana Zajc

Digitalni obrat, ki posega v tako rekoč vsa področja našega vsakdana, je tudi v literarno vedo prinesel nove raziskovalne pristope. V času, ko so se napredne tehnologije, vključno z umetno inteligenco, znašle v središču globalnih razprav, postaja tudi za literarno vedo vse pomembneje, da smiselno izkoristi možnosti, ki jih ponuja tehnološki razvoj na področju literarnovedne metodologije. Uporaba računalniških orodij namreč odpira prostor za interdisciplinarne raziskave na preseku med humanistiko in informacijsko tehnologijo, kjer se vse intenzivneje razvija digitalna humanistika. Na področju literarne vede za računalniško podprte analize literature pogosto uporabljamo termin *oddaljeno branje*, uveljavili pa sta se tudi poimenovanji *makroanaliza* in *algoritemska kritika*. Čeprav paradigmatski obrat k oddaljenemu branju temelji na želji po decentralizaciji sistema svetovne literature, raziskovalci in raziskovalke s tovrstnimi pristopi zaradi dostopnosti digitalnih arhivov najpogosteje posegajo po literaturah v svetovnih jezikih. To je botrovalo tudi specifičnemu razvoju samih metod digitalne humanistike, ki so le redko prilagojene raziskovanju literatur v manjših jezikih, tudi v slovenščini. Pri oddaljenem branju gre za t. i. računalniško branje, ki literaturo analizira z uporabo jezikovnih modelov in strojnega učenja, pri čemer računalniki iz korpusov literarnih besedil pridobivajo podatke ter jih shranjujejo in obdelujejo s kvantitativnimi metodami. Tovrstne analize lahko zajemajo obsežne zbirke, ki vsebujejo po več tisoč literarnih del, lahko se ukvarjajo s posameznimi besedili ali njihovimi deli, včasih pa se osredotočajo tudi na metapodatke, povezane z literarnim sistemom. Z računalniškim branjem, ki temelji na statistiki in sorodnih vejah matematike, je mogoče izluščiti specifične formalne ali semantične informacije in v njih prepoznati vzorce, ki so s pomočjo natančnega branja nezaznavni. Natančno branje kot klasična metoda literarne vede in oddaljeno branje pa se ne izključujeta, temveč se dopolnjujeta, kot pokažejo tudi raziskave v pričujočem tematskem sklopu *Primerjalne književnosti*.

Članki tematskega sklopa »Veliki podatki, male literature« prinašajo nabor digitalnih pristopov k literarnovednim raziskavam v manjših literaturah, kakršne so na primer češka, romunska, baskovska in nenazadnje

1

tudi slovenska. S tem raziskovalke in raziskovalci zapolnjujejo vrzel v računalniških raziskavah književnosti, ki večinoma temeljijo na digitaliziranih besedilih, ta pa še vedno večinoma pripadajo literaturam v svetovnih jezikih. Pri raziskovanju literatur v manj razširjenih jezikih smo tako postavljeni pred izziv, kako zagotoviti digitalizacijo literarnih besedil in vso drugo potrebno digitalno infrastrukturo. Oddaljeno branje, katerega predmet raziskovanja naj bi bila svetovna literatura, se v praksi sooča z oviro vse bolj očitnega digitalnega jezikovnega razkoraka (ang. *digital language divide*), zaradi katerega več kot 3.700 jezikov v digitalnem okolju tako rekoč ne obstaja.

Tematski sklop obravnava problematiko digitalne literarne vede na področju malih literatur z dveh vidikov: članki v prvem delu se posvečajo izgradnji podatkovnih zbirk, njihovi anotaciji in analizam metapodatkov, v drugem delu pa so zbrane študije primerov, v katerih so metode strojnega učenja in obdelave naravnega jezika aplicirane na že sestavljene korpuse. V prvem članku **Vlad Pojoga** predstavi analizo romunskega literarnega časopisa *Convorbiri literare* kot preliminarno študijo, ki je nastala v procesu izgradnje obsežnega korpusa romunske poezije 19. stoletja. Ročno izluščeni metapodatki o lokalni literarni produkciji, prevodni literaturi in medliterarnih vplivih osvetljujejo proces kanonizacije romunskih avtorjev poznega 19. stoletja z vidika vpetosti v svetovno književnost. **Katja Mihurko**, **Ivana Zajc**, **Darko Ilin** in **Mila Marinković** prav tako proučujejo omrežja vpliva med pisatelji in pisateljicami, a težišče prenesejo na osebne korespondence. Na primeru metapodatkovnih in semantičnih analiz korpusa Pisma pokažejo, da so tudi neliterarni korpusi nepogrešljiv del digitalne literarne vede, saj ponujajo nov vpogled v produkcijo pa tudi recepcijo literarnih besedil. Članek **Ranke Stanković**, **Cvetane Krstev** in **Duška Vitasa** se vrača k izključno literarnemu korpusu, namreč ELTeC-srp, ki je v novi različici obogaten z metapodatki iz podatkovne zbirke Wikidata. Članek predstavi izzive, s katerimi so se soočali oblikovalci korpusa vse od digitalizacije besedil do avtomatske anotacije, ter ponuja inovativne rešitve, ki jih omogoča interdisciplinarno sodelovanje. Vezni člen med prvim in drugim delom sklopa je članek **Silvie Cinkove**, **Petra Plecháča** in **Martina Popela**, ki opisuje proces evalvacije avtomatskega označevalnika UDPipe na primeru češke poezije 19. stoletja kot nujen del procesa sestavljanja jezikovno označenega literarnega korpusa. Ker je za češko poezijo značilen slogovno zaznamovani besedni red, se je izkazalo, da je jezikovni model, ki je bil naučen na proznih besedilih, za poezijo pomanjkljiv.

V drugem delu tematski sklop prehaja s sestavljanja korpusov in analiz metapodatkov k tekstualnim analizam s pomočjo računalniških

orodij. **Dominika Werońska** aplicira stilometrično orodje Stylo, ki meri frekvence najpogostejših besed, na korpus baskovskih romanov. Z analizo pokaže, da sta spremembi sloga, ki se izkažeta za posledico prevajanja iz tujih jezikov ali vpliva narečij, prav tako močni kakor avtorski signal. **Ivana Zajc** prilagodi uporabo orodja Stylo analizi korpusa slovenske proze iz obdobja realizma in moderne, pri tem pa pokaže, da je avtorski signal močnejši kakor na primer žanrski. Podrobneje se posveti Ivanu Cankarju, pri katerem stilometrična analiza razkrije razvoj avtorskega sloga, pri čemer kot slogovno bolj homogeno izstopa predvsem Cankarjevo zgodnejše obdobje. Tudi **Andrejka Žejn**, **Marko Pranjić** in **Senja Pollak** pod drobnogled vzamejo slovensko prozo realizma in moderne, a za analizo avtorskega sloga uporabijo metodo kontekstualnih vektorskih vložitev besed. S pomočjo računalniške semantične analize zaznajo pomenske premike v delih Josipa Jurčiča in Ivana Cankarja ter ugotovijo, da so ti najizrazitejši pri splošnejših pomenskih poljih. V zadnjem prispevku tematskega sklopa **Lucija Mandić** razširi predmet analize s kanoniziranih avtorjev na celotno slovensko daljšo pripovedno prozo t. i. dolgega 19. stoletja. Semantična analiza vektorskih vložitev besed razkrije izrazito prekrivanje semantičnih polj kulture in politike tako v kanonizirani kakor v nekanonizirani literaturi, kar naveže na preddigitalne ugotovitve slovenske literarne vede o t. i. prešernovski strukturi.

Tematski sklop je nastal na Inštitutu za slovensko literaturo in literarne vede ZRC SAZU in na Raziskovalnem centru za humanistiko Univerze v Novi Gorici v okviru raziskovalnega programa »Literarnozgodovinske, literarnoteoretične in metodološke raziskave« (P6-0024) in raziskovalnega projekta »Transformacije intimnosti v literarnem diskurzu slovenske moderne« (J6-3134), ki ju financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije.

Big Data, Small Literatures (An Introduction)

Lucija Mandić, Ivana Zajc

The digital turn, which is affecting virtually every aspect of our daily lives, has introduced new research approaches into literary studies as well. In a time where advanced technologies, including artificial intelligence, dominate global discourse, it is increasingly important for literary studies to explore the methodological potential of the latest technological advances. The use of computational tools is able to create an interdisciplinary space at the intersection of the humanities and information technology, fueling the growth of digital humanities. In literary studies, computer-assisted analyses of literature are usually collected under the umbrella term of *distant reading*, with *macroanalysis* and *algorithmic criticism* serving as near synonyms. While the paradigmatic turn to distant reading aims to decentralize the literary world-system, researchers predominantly focus on major world languages due to the easy accessibility of digital archives in those languages. This trend has led to the development of methodologies of digital humanities, which are rarely adapted to the study of literatures in less dominant languages, such as Slovenian. Distant reading involves computational analysis of literature using language models and machine learning algorithms, whereby computers extract data from corpora of literary texts in order to store and process them using quantitative methods. Such analyses can encompass extensive collections comprising thousands of literary works, zoom in on individual texts or excerpts, or focus on metadata related to the literary system. Computer-assisted reading based on statistics and related mathematical branches extracts specific formal or semantic data and identifies patterns that would remain imperceptible to close reading. It is important to note, though, that close and distant reading, far from being mutually exclusive, are complementary approaches, as evidenced by the articles clustered in this issue of *Primerjalna književnost* under the title "Big Data, Small Literatures."

The cluster brings together various digital approaches to literature in smaller languages, such as Czech, Romanian, Basque, and Slovenian. The articles address a gap produced by computational literary studies as they predominantly work with digitized texts in globally spoken

languages, while research on less dominant languages faces the challenge of digitizing literary texts and establishing necessary digital infrastructure before it can analyze and interpret texts. Originally proposed as a method of studying world literature, distant reading is effectively faced with the obstacle of an increasingly widening digital language divide which has made more than 3,700 languages virtually non-existent in the digital environment.

As a cluster, "Big Data, Small Literatures" explores digital literary studies in small literatures from two perspectives: the first group of articles concentrates on database construction, annotation, and metadata analysis, while the second group presents individual case studies in which machine learning and natural language processing have been applied to already existing corpora. In the first article, **Vlad Pojoga** presents an analysis of the Romanian literary journal *Convorbiri literare* as part of a broader initiative to construct a corpus of nineteenth-century Romanian poetry. The examination of hand-extracted metadata sheds light on the canonization process of late nineteenth-century Romanian authors based on their integration into the networks of world literature. Similarly, **Katja Mihurko**, **Ivana Zajc**, **Darko Ilin**, and **Mila Marinković** delve into networks of influence, but shift the focus to personal correspondences to address issues of gender pertaining to individual writers. Their semantic analyses and exploration of metadata show that non-literary corpora are an indispensable part of digital literary scholarship, as they offer new insights into the production and reception of literary texts. **Ranka Stanković**, **Cvetana Krstev**, and **Duško Vitas** take us back to exclusively literary corpora as they examine the ELTeC-srp corpus and its recent upgrade with enriched metadata from the Wikidata database. Their article outlines the challenges encountered by the designers of the corpus, from text digitization to automatic annotation, and proposes innovative solutions facilitated by interdisciplinary collaboration. This is followed by the article by **Silvie Cinková**, **Petr Plecháč**, and **Martin Popel**, which provides a kind of transition between the first and second parts of the cluster. The article describes the process of evaluating the automatic annotator UDPipe using nineteenth-century Czech poetry as a case study—a crucial step in the development of a linguistically annotated literary corpus. Given the distinctive stylistic word order prevalent in Czech poetry, the linguistic model trained on prose texts proved partially insufficient for the analysis of poetry.

Transitioning from corpus construction to textual analysis, **Dominika Werońska** applies a stylometric tool Stylo, which measures

the frequencies of the most frequent words, to a corpus of Basque novels. Her findings reveal that stylistic fluctuations attributable to translation or dialectical influences rival the authorial stylistic signal. Stylo is used by **Ivana Zajc** as well, as she adapts the tool to analyze a corpus of Slovenian mid-nineteenth- to early-twentieth-century fiction, highlighting the dominance of the authorial style over genre indicators. Her analysis of works by Ivan Cankar reveals the development of the authorial style, with Cankar's earlier period standing out in particular with its stylistic homogeneity. In the next article, **Andrejka Žejn**, **Marko Pranjić**, and **Senja Pollak** examine the authorial style in Slovenian prose of the same period by utilizing contextual word embeddings. Their computational semantic analysis unveils semantic shifts in the works of Josip Jurčič and Ivan Cankar, demonstrating that they are particularly evident in broader semantic domains. In the final aticle, **Lucija Mandić** broadens the analytical scope to encompass the entire Slovenian long narrative prose of the so-called long nineteenth century. Semantic analysis using word embeddings reveals a distinct convergence of semantic fields related to culture and politics across both canonized and non-canonized texts, both confirming and expanding the findings of traditional Slovenian literary studies pertaining to the so-called Prešernian structure of Slovenian canonical literature.

The Last Bastion of European Romanticism: A Quantitative Analysis of Poetry in the Romanian Literary Magazine *Convorbiri literare* (1867–1916)

Vlad Pojoga

Lucian Blaga University of Sibiu, Faculty of Letters and Arts, Department of Romance Studies,
Bulevardul Victoriei nr. 5-7, 550024 Sibiu, Romania
https://orcid.org/0000-0002-5531-2488
vlad.pojoga@ulbsibiu.ro

The article presents a quantitative analysis of poetry published in the most popular literary periodical of modern Romania, Convorbiri literare *(*CL*), from its founding in 1867 to 1916, when Romania entered World War I. Often seen as the last major magazine of European Romanticism,* CL *appeared at a time when poetry was a privileged genre in Romanian literature, and it hosted in its pages both the national poet Mihai Eminescu and nineteenth-century's most prominent Romanian literary critic Titu Maiorescu. Arguing that the publication and theorization of poetry was an integral part of one of the last nation-building processes in Europe, the article indexes and quantitatively analyzes all forms of poetry and of what can be viewed as poetic networks in* CL. *By using ARCANUM's digital archive of more than 66,000 pages of* CL, *three strata of poetry-related metadata are extracted and examined: the local production of poetry (who published in* CL *and when, and how much did they publish); poetry imports (whose poetry was translated into Romanian in* CL *and when, and what were the authors' origins); and networks of influence (which foreigners were most often mentioned in relation to the two most talked-about Romanian Romantic poets of the century, Eminescu and Vasile Alecsandri). Thus, the article aims to chart the national regime of relevance that applies to Romanian modern poetry by uncovering the international network of authors who were at the center of literary debates in* CL.

Keywords: Romanian poetry / European Romanticism / poetry translation / Eminescu, Mihai / Alecsandri, Vasile / Maiorescu, Titu / quantitative analysis

Romanian literary studies go digital

If ten years ago anyone told Romanian scholars and even more so Romanian literary critics that they would be able to access, read (or go through), and analyze tens of thousands of pages, millions of words of curated corpora of literary or non-literary texts in a heartbeat, they would have shrugged it off at best, and laughed it off at worst.[1] What are the tenets that would guarantee such a reaction a decade or so ago? To begin with, the quasi-unanimous view of literary criticism in Romania was (and to some extent still remains) that every critical endeavor ought to be based in an aesthetic judgment formulated by a cultured individual's personal reading experience and history. Therefore, a clearly hierarchical view of the literary system was still taught, learned, and perpetuated. Such a view was in a direct relationship with the idea of the canon and the notion of literary relevance viewed as correlated with the position that these individual aesthetic judgements conferred (or not) on actual texts, and in direct opposition to the demand of the post-1990 local book market. It is no wonder then that a symptom of this exaggerated focus on high-brow literature, along with the fresh liberalization of the book industry, led to a steep increase in translated literature (see Terian, "Big Numbers" 67, for a statistic of the local versus translated novels published in Romania between 1990 and 2000). At the time, most Romanian writers sought canonization for themselves and therefore tended to write for the critics who they thought could help them achieve that, whereas casual readers turned more and more to imported commercial fiction. Consequently, as literature was approached as a way of imposing personal verdicts and as a vehicle of entering the canon, even scholars who were remotely interested in exploring "the great unread" (Cohen 23) of Romanian literature would not know where to start. Granted, this was in part because digital versions (either as scanned or as textual material) were both scarce and scattered, and corpora were either incomplete or reduced in size. And although "distant reading" (Moretti) was becoming a quite popular way of looking at literature, its actual implementation was much more difficult than in more technologically advanced countries where coherent digital resources where more readily available.

In 2015, few analytical attempts existed that used the resources of digital humanities, and even those were focused on English literature

(Nicolaescu and Mihai; Mihai), whereas the emerging debates were centered around questions of applying digital humanities in Romania (Ursa) and using quantitative tools in literary studies (Anon.). In fact, the major breakthroughs were not analytic at first, but rather concerned corpus and archive building, because there were no comprehensive digital primary sources to work on. The release of the Romanian ELTeC corpus by Roxana Patraș and her team (https://github.com/COST-ELTeC/ELTeC-rom), consisting of around 100 novels, and the parallel three-stage release of The Digital Museum of the Romanian Novel (https://revistatransilvania.ro/mdrr), consisting of around 1,300 novels, ushered in a new stage of digital humanities in Romania. For the first time, in 2019, there was a corpus comprehensive enough for large scale analysis. The same year, but prior to the project, Ovio Olaru (Olaru) and Patras and her team (Patraș et al.) published valuable overviews of the use of digital tools in Romanian humanities research. Just a few months later, the first results of the Digital Museum of the Romanian Novel project were published. In that first dossier, my colleagues and I (Pojoga et al.) provided a brief discussion of possible ways of analyzing the newly-build archive, and the two studies on the genres (Terian et al.) and the internal geography (Baghiu et al., "Geografia internă") of nineteenth-century Romanian novel set the stage for more detailed analyses by the same extended team (these were published in various Romanian and international journals, by various combinations of authors). Most of these contributions were focused on prose, since the main available corpora were novel corpora. We explored space (fictional and historical), time, subgenres, and character identities, among other topics. We extracted metadata, compiled maps, and built graphs. And at the end, when we reached the year 1947 and entered dangerous copyright territory, we started to wonder: what about other literary genres?

The digital museum of Romanian poetry

The choice was natural, given our previous preoccupation with the genre (see Vancu; Pojoga, "A Survey"): poetry. Thus, we tried to figure out how to build a Digital Museum of Romanian Poetry to mirror its twin, The Digital Museum of the Romanian Novel, but with the experience of archive building and cleaning (where we manually rewrote text that was not entirely recognized by OCR software) that enabled us to look at the process in a completely new way. At first, we surveyed the existing digital poetry archives, and, after moving past the

usual websites that can sometimes be great poetry archives but are often limited to English and American works, we found a few possibly noteworthy models for our venture. In terms of corpora, the ELTE Poetry Corpus (https://github.com/ELTE-DH/poetry-corpus), for instance, indexes poems authored by 50 Hungarian canonical poets, for a total of around 13,000 poems; A Gutenberg Poetry Corpus—extracted by Allison Parrish from the literary archive of Project Gutenberg (https://github.com/aparrish/gutenberg-poetry-corpus)—includes about 3 million lines of poetry; and the Finnish Folk Poetry Corpus brings together Finnish folk poetry from the sixteenth century to the 1930s (https://www.kielipankki.fi/corpora/skvr/). But all of these are restrictive one way or another, and we wanted more. We did not want to index and analyze canonical poets alone (like ELTE), or work with an eclectic corpus (like Gutenberg), or work on folk poetry, but rather we wanted to include all poetry written in Romanian until World War I.

A quite ambitious feat, as we are still finding out. Because all poetry written in Romanian until World War I meant volumes, brochures, pamphlets, composite volumes (prose and poems), but also poetry published in literary magazines and newspapers. Moreover, we were also interested in the post-archive building phase, namely the digital analysis of the extracted metadata and poems. In this regard, likely the most comprehensive project that goes beyond archive building is Poetry Standardization and Linked Open Data (POSTDATA) (https://cordis.europa.eu/project/id/679528), a European Research Council project coordinated by Elena González-Blanco. POSTDATA is "focused on poetry analysis, classification and publication, applying Digital Humanities methods of academic analysis—such as XML-TEI encoding […] in order to look for standardization, as well as innovation by using semantic web technologies […] to link and publish literary datasets in a structured way in the linked data cloud" (Gonzalez-Blanco et al.). Our modus operandi for the novel project involved, during the scanning phase, the manual extraction of certain metadata that were meant to provide either a basis for later digital analyses (country or city names providing a starting point for a series of studies that culminated with Baghiu et al., "Geografia romanului") or insights that at that point seemed difficult to extract automatically (main characters' occupation or social origin, time and space, etc.). Therefore, we needed to make important decisions about, say, what and how to extract, and we came up with a preliminary model. This model includes identification metadata (place of publication, author, translator, title), sizing metadata (number of stanzas, lines, words, characters without spaces), subgenre classification, main and secondary themes, space

and time metadata, and subject metadata. But since we wanted to test the process before applying it to the full archive we intended to build, we agreed that we would do a pilot run on a full archive of a magazine, before moving to full volumes.

In this article, I will thus work with the pilot identification and sizing metadata[2] for about 3,000 poems in The Digital Museum of Romanian Poetry, published between 1867 and 1916 in *Convorbiri literare* (Literary Conversations, henceforth *CL*), which is one of the most relevant Romanian literary magazines, and chart local production of poetry, poetry translations, and the two respective networks of influence for probably the most referenced Romanian poets of the nineteenth century, Mihai Eminescu and Vasile Alecsandri.

The last bastion of European Romanticism

The question that seems to arise now is why *CL* and not some other Romanian magazine? The short answer would stress the sheer impact *CL* had on the development of Romanian literary culture, but that is not necessarily a quantifiable and coherent marker. I will try to formulate a more elaborate answer in this section.

When attempting to delineate the precise epoch(s) of Romanticism across Europe, one invariably grapples with a nebula of years, volumes, and theoretical frameworks encompassing production, definitions of the movement, and its main features. For instance, the volume titled *A Companion to European Romanticism* is presented on the publisher's website as covering the national literatures of France, Germany, Italy, Poland, Russia and Spain. Even more confusing are assumptions such as "An era when millions of people flock to see *The Lord of the Rings* might forgivably be called Romantic still" (Ferber 8), which seems to merge cultural trends with artistic movements in a rather gratuitous manner. Duncan Wu, in another companion volume, situates the (English) Romantic movement broadly between 1790 and 1830, as the section "Contexts and Perspectives" shows (Wu 1–108). On the other hand, Virgil Nemoianu, an Eastern European critic, talks in his book *The Taming of Romanticism* about commenting on "some specific items of European literature between 1815 and 1848" (vii) and "these

---

[2] The project metadata (and subsequently the metadata for this article) were extracted by a project team based at the Lucian Blaga University of Sibiu, consisting of Geanina Giuhat, Ana-Maria Stoica, Hanna Han, Bogdan Contea, Andreea Popescu, Teodora Susarenco, Daniel Coman, and Crina Neacşu.

later stages" of Romanticism that go until 1848 (1). It has already been shown that the case of Central and Eastern European Romantic movements is quite different than their Western counterparts:

> The literatures of the region have often been concerned with their belatedness in relation to the Western literary cultures, and they occasionally measured their lag also in relation to neighboring cultures. Romanian literature, for example, experienced its "will to modernity" in the nineteenth century not only with an acute sense of belatedness but also with an "unhappy consciousness" about its need to choose "imitation" […] to recover the lost ground. (Neubauer 321)

Moreover, for most of Central and Eastern European literatures, Romanticism and its partial overlap with nineteenth-century nation-building processes were an opportunity to elevate one of their emerging poets to the status of the national poet; hence "the typical pattern in east-central Europe […]: picking a (Romantic) poet from the first half of the nineteenth century or mid-nineteenth century and canonizing him in the following decades" (Dović and Helgason, 63–64). In Romania, there were Romantic writers who published in the first half of the nineteenth century as well, and they founded several magazines, most notably *Dacia literară (The Literary Dacia)* by Mihail Kogălniceanu in 1840 and *România literară* (*The Literary Romania)* by Alecsandri in 1855, with both ending within a year, after only a few issues. The real Romantic boom does not happen until after 1864, when the Junimea literary society was founded in the city of Iași by Titu Maiorescu, P. P. Carp, Vasile Pogor, Iacob Negruzzi, and Theodor Rosetti. Sometime later, Maiorescu, the leading ideological and literary figure of the society, suggested "it would be wonderful to start a small literary magazine to print the writings read and approved by the Society," to which "everyone agreed without much debate" (Negruzzi 89–90; my translation). As a result, the inaugural issue of *CL* came out in 1867, and it was going to be published continuously (with the exception of the war years of 1916–1918) until its hiatus in 1944. Its importance to Romanian canonical literature and Romanian Romanticism is so great that, instead of seeing a continuity in his influential literary history, Nicolae Manolescu chooses a dual, parallel periodization, outlining Romanticism as a movement between 1840 and 1899, and Junimism as a period between 1867, when *CL* starts, and 1889, when Eminescu dies (see Manolescu 141–434). But rather than demonstrate further the national relevance of *CL* and the Junimea society, we should assume its local relevance and try to place it in an international landscape.

In lieu of a more refined metric, I propose a visual approach to situate *CL* within the broader timeline of European Romanticism. In Figure 1, I chart the birth and death years of 22 prominent Romantic poets hailing from 21 European countries spanning the geographical expanse from Western to Eastern Europe, encompassing spaces from England and France to Russia and Ukraine as well as from Denmark to Bulgaria. This visual representation serves as a comprehensive panorama of the Romantic literary timeline, encapsulating a diverse group of poets whose activity almost invariably starts in the first third of the nineteenth century.
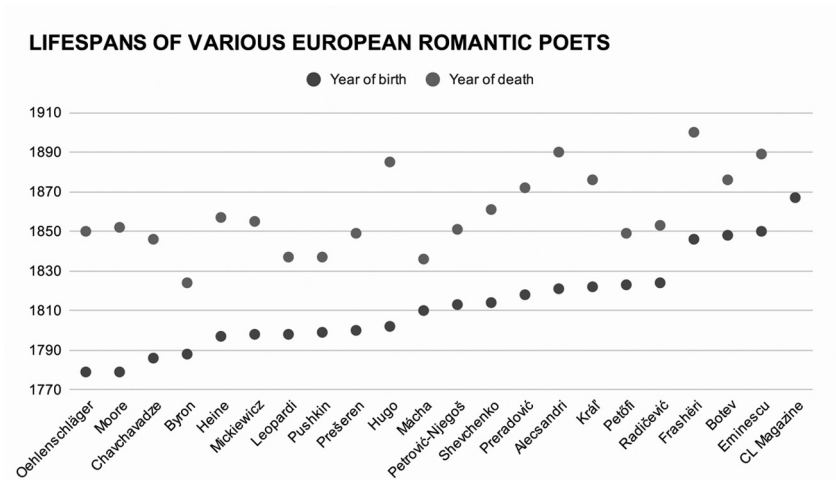
Figure 1: Lifespans of various European Romantic poets and the publication year of the first *Convorbiri literare* issue.

Within this chart, I added two Romanian representatives: Alecsandri, regarded as the original Romanian Romantic poet, and Eminescu, the so-called younger Romantic destined to usurp Alecsandri's potential position as Romania's national poet (Terian, "Mihai Eminescu"). As it can easily be seen, by the time of Eminescu's debut in 1866, most European Romantics were already dead. The same can be said about *CL*, the Romanian literary magazine that will be the main focus of this study, and whose first issue appears in 1867. But, as we have seen (and will see in more detail below), *CL*, as an appendix to and a popularization tool of the Junimea society, was the most important proponent of Romanian Romanticism, and the (chronologically) last bastion of European Romanticism.

An overview of poetry publishing in *Convorbiri literare*

Figures 2, 3, and 4 chart all poetry published in *CL* between 1867 and 1916; Figures 2 and 3 use the number of words as a metric, and Figure 4 uses lines. At first, I intended to simply count the number of words published as poetry every year in *CL*, but then I thought that trends would emerge more clearly if I looked at a five-year average (average number of words per five years, changing constantly) and a rolling average (average number of words per year from the beginning up to that year). This is shown in Figure 2. Besides the general trend, which is clearly downward, Figure 2 indicates that, after an initial rise (and peak), poetry publication drops in 1877—most likely because of the Romanian War of Independence in 1877–1878; the drop continues until the late 1880s when there seems to be a brief resurgence, and then returns until the twentieth century brings another rise. This calls for some explanation, which is what Figure 3 tries to achieve. Here, rather than focusing on the sheer number of words, I chose to split the number of words into three categories: local production (poetry written by Romanian poets, in Romanian), folk poetry (defined simply as authorless poetry), and translations (poetry written by non-Romanian poets, translated into Romanian). In the landscape of poetry published in *CL*, local production covers 53%, translations account for 36%, and folk poetry represents 11%. The data in Figure 2 led to possible explanations for all three peaks: the first and highest peak is mainly driven by local production (seconded by translations) and the need for the magazine to establish its direction and promote the authors associated with the Junimea society; the second peak, quite strangely given the generally low percentage of words attributed to the category (only 11%), seems to be driven by folk poetry (seconded by local production), due to a large number of folk poems collected by Alesiu V. (Alexiu Viciu) that are published between 1888 and 1890, mostly a type of traditional Romanian poetry called *doină*; and the third peak is clearly driven by translations, with the main culprit being the translation of Homer's *The Iliad*, translated by George Murnu over a six-year period (1900–1905). In spite of these peaks, the rolling average shows a clear image: *CL* gradually loses interest in poetry. But how does this change if, instead of looking just at numbers of words, we concentrate on the space occupied by poetry? By that I mean the percentage of lines (as the best metric to measure occupation, because words might be misleading in this case—for instance, one can have poems with very short verses, which means that less words occupy much more magazine pages than the much wordier prose) occupied by

poetry in each year of *CL*. The trend in Figure 4—which uses a formula that takes into account the number of lines per page, the number of columns per page, the number of pages per issue, and the number of issues per year—looks even harsher for poetry publication, going from an average of over 10% of total space prior to 1877, to a little over 5% prior to 1905, and less than 5% after 1905. But how and why does this happen? I will try to answer this question by analyzing local poetry production and poetry translations in the next two sections.
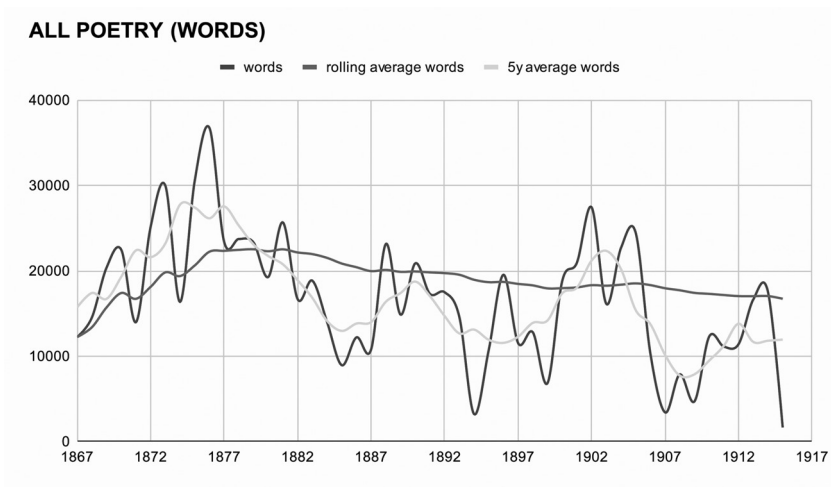
ALL POETRY (WORDS)

Figure 2: All poetry published in *Convorbiri literare* by yearly number of words (a).

ALL POETRY: WORDS/YEAR
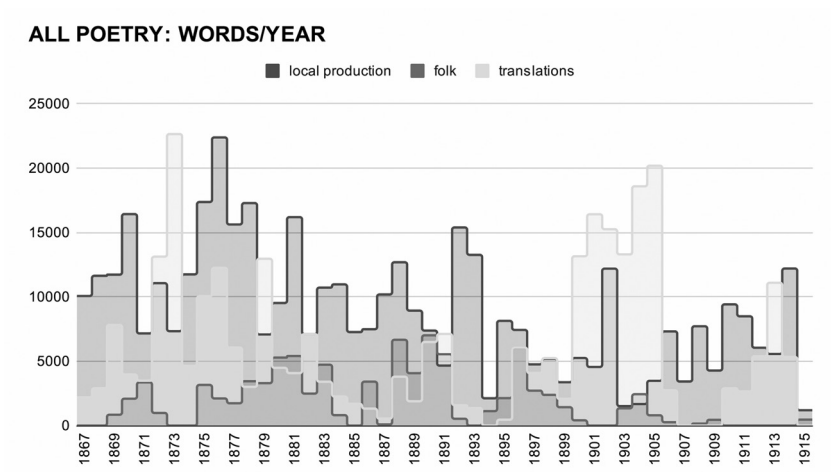
Figure 3: All poetry published in *Convorbiri literare* by yearly number of words (b)—local production, folk poetry, translations.

17

SPACE OCCUPIED BY POETRY IN CL (%LINES/YEAR)
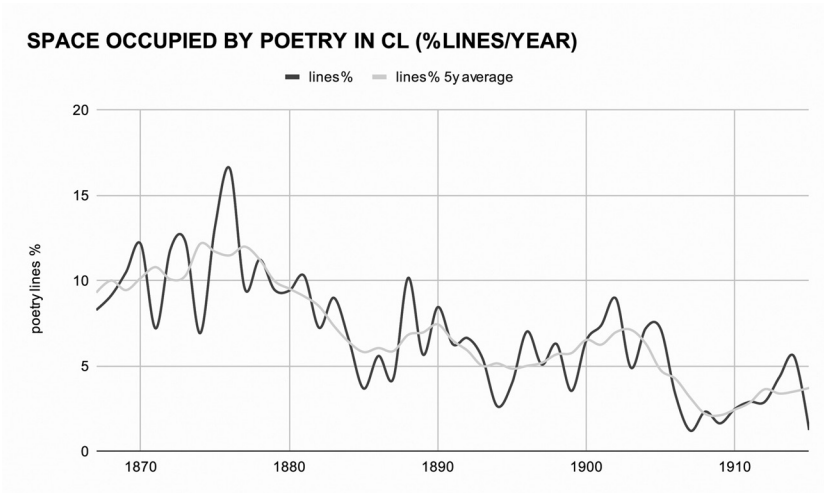
Figure 4: Space occupied by poetry in *Convorbiri literare* as percentage of all lines in the magazine per year.

Local production of poetry in *Convorbiri literare*

The slow start of the novel in Romania (see Terian, "Big Numbers") meant that there was an open creative space that was filled with other literary forms. Or at least that is what it meant in theory. For there was another gap between Romania and Western countries, a literacy gap. For instance, at the 1899 census, only 22% of the Romanian adult population (over the age of 15) was literate (Colescu 122). This low number suggests that the lack of original literary production in Romania may not have been a matter of literary genres, but rather a matter of the number of people who knew how to read and write. Indeed, based on the data that we have for both novels (Terian, "Big Numbers") and poetry, at least until World War I, novels and poetry seem to grow together on average, rather than any of them falling out. Nonetheless, as shown in the previous section, poetry publication in *CL* seems to get periodically lower, just as the poetry market grows.

In Figure 5, which only includes poems by Romanian authors, excluding translation and folk poetry, there are no other peaks above the rolling average after the starting surge, only two falls—one in 1877, and one after 1894, when Iacob Negruzzi leaves the magazine—slowly picking up again after 1907, when Simion Medehinți takes editorial control of the magazine. The totals in Figure 5 are corresponding to

80% of all poems and 89% of all words in poems published in *CL* with a Romanian origin. The rest is folk poetry, and from the extracted data, two stylistic characteristic specific to folk poetry (compared to poetry attributable to individuals) come to surface: folk poems have very few stanzas per poem (an average of two vs an average of seven for male-authored poems, and four for female-authored poems), and they have many more lines per stanza (an average of 16 vs an average of seven for male-authored poems, and four for female-authored poems). Of course, this is a rather expected result, given the nature of folk poetry and its predominant structure, but it nonetheless offers us a quantified confirmation. In the same vein, also expected is the gender imbalance, with 83% of all Romanian poets being male, and only 6% being female (almost half as many as the unidentified poets—signing with a pseudonym or initials—who make up 11% of the total number of Romanian poets). The disparity is even greater in terms of word totals, with 97% of all Romanian poetry words published by male poets, and only 2% by female poets (1% remaining for the unknown poets—because most of them only published a maximum of one or two poems). Hence, women publish less, but even when they do, their poems are much shorter than those of their male counterparts (the paradigmatic form seems to be stanzas of four lines each, for an average of four stanzas per poem).

In fact, the quatrain seems to be the paradigmatic stanza form for the whole corpus, with more than 40% of all poems authored by Romanian writers having four lines per stanza. This is the case even though, as mentioned above, the averages for male poets are seven stanzas per seven lines each, which is mostly due to the exceptionally long poems that they authored (all of the ca. 120 poems over 100 lines are authored by male poets, and only 3% of all poems over 50 lines are authored by female poets).

LOCAL PRODUCTION OF POETRY (WORDS)
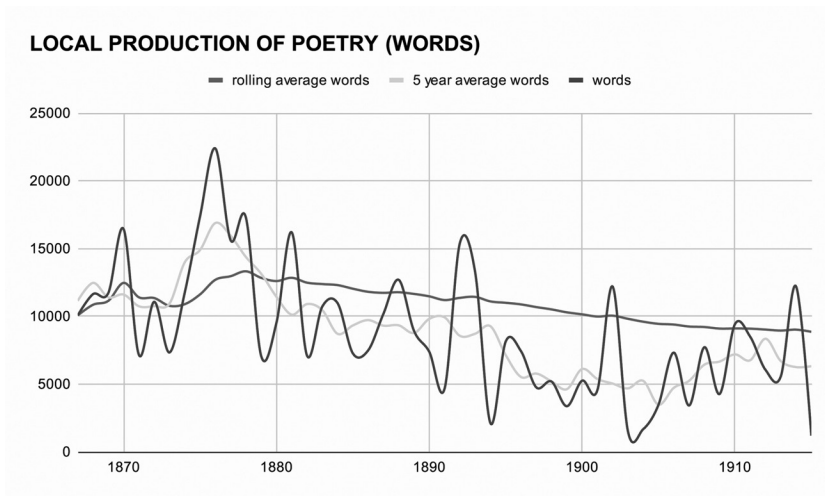
Figure 5: Local production of poetry (words/year) in *Convorbiri literare*.

In terms of best represented poets, our corpus includes 22 poets with over 20 poems, three of which are women. Yet these three female poets, namely Matilda Cugler, Veronica Micle, and Ana Conta-Kernbach, have authored 80% of all poems published by women in *CL* prior to World War I, with Matilda Cugler even being the fourth author overall, with 79 published poems. She is also the only woman to be included in the list of six put forward by Maiorescu in the aforementioned programmatic study, alongside Alecsandri, Eminescu, Samson Bodnărescu, Theodor Șerbănescu, and Dimitrie Petrino. Again, the data supports the notion that Maiorescu's direction is in fact *CL*'s direction (or the other way around), since, with the exception of Dimitrie Petrino, the other five authors also represent the top five in terms of the most poems published in the magazine. If we use a different metric, and look at the number of words instead of poems, only Alecsandri and Eminescu remain in the top five, alongside Iacob Negruzzi, Anton Naum, and Dimitrie C. Ascanio-Ollanescu. Needless to say, Alecsandri easily tops both categories.

One and a half centuries later, *CL* and Junimea seem to have achieved most of their goals, with the most important one probably being the elevation of Eminescu to the status of the national poet. In the process, however, they seem to have lost interest in other kinds of poetry and to have been unable to overcome Junimism, which in poetry meant a sort of combination of Classicism and Romanticism. Of their nineteenth-century poetry bets, only Eminescu has survived in

the high-school canon, although admittedly he also is the lone nineteenth-century poet on the list. After him, *CL* seems to begin to marginalize poetry, much as France (Sapiro 25) and Germany do in the late nineteenth century, but against the Romanian tide of poetry production.

Translations of poetry in *Convorbiri literare*
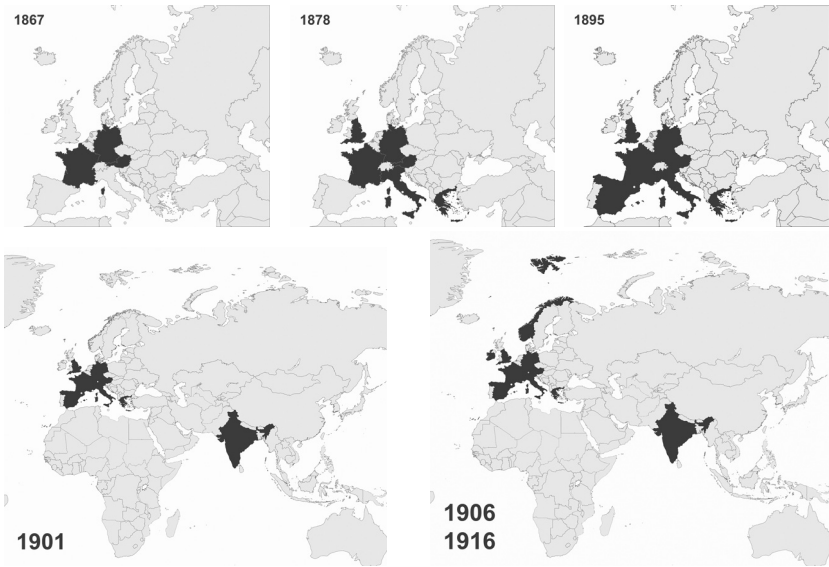
Figure 6: Composite landscape of five maps with the countries of origin for the poets translated in *Convorbiri literare* in 1867, 1878, 1895, 1901, 1906, and 1916 (the last two are the same because no changes occur).

36% of all poetry words in *CL* are translated. They belong to over 70 poets from 11 countries, plus Ancient Greece and the Roman Empire. The maps in Figure 6 chart the periods when *CL* puts a pin on a country, that is, when a poet from that country is translated into Romanian in its pages. I have not included the two state forms I mentioned—Ancient Greece and the Roman Empire—for historical reasons, and also because they would have overlapped with nineteenth-century Italy and Greece. Due to technical limitations, I had to use current borders of the countries, which I can defend by noting that the purpose of these maps is not to accurately represent historical borders, but rather to represent the translation reach of *CL*. In the first year, only the two continental

powerhouses, France and Germany, are represented, alongside a wandering Austria (through a single poem by Heinrich von Levitschnigg). Ten years later, England, Italy, and Greece appear on the map, and by 1895 Spain also joins the group. In 1901, the first (and only) non-European country, India (through several translations by George Coşbuc) appears, alongside Switzerland, and by 1906 they are joined by Norway and Ireland. The maps are relevant more because of what cannot be seen than of what can be seen. For while we could expect France, Germany, England, Italy, and Spain to emerge on the map, we should have expected to find more than just one non-European country. The absence of Romania's neighbors is unexpected as well, as is the status of Austria as the sole European country to the east of Germany.

At this point, some depth should be added to the maps; besides the geographical space targeted by translations, quantities are also relevant. Figure 7 shows the total number of poems translated from each country. I chose the number of poems for this graph rather than the number of words because the number of poems also plays a part in terms of how many times a certain poet appears in the magazine. Two clear frontrunners emerge here, and, without much surprise, they are France and Germany. Nonetheless, given that the general view of the era was that *CL* is decisively oriented toward German poetry and literature, it is quite remarkable to see both go head-to-head until the early 1900s, when France moves on top. This translation duo should be explored further; for instance, its fruit is one of the most peculiar events of literary translation in the so-called long nineteenth century. In 1871, *CL* published a poem signed by Gablitz, a supposedly German poet translated into Romanian by an unknown translator. In fact, "Gablitz is a poet who never existed, and by inventing this German name, I had in mind *CL*'s predilection for everything that comes from Berlin" (Hasdeu; my translation): this is the confession of Bogdan Petriceicu Hasdeu, one of the main opponents of Titu Maiorescu and Junimea. The impression that German literature was the go-to literature for *CL* was so strong, that a prank like this was thought of and, for that matter, pulled off. Even though the reality of poetry translation in *CL* contradicts Hasdeu, there is another reality that most likely molds his thinking and public discourse. That reality is the fact that Romanian culture was francophone during those years, and novel translations, for instance, reflect that, as shown by Baghiu (Baghiu 94), since less than 30% of all novels translated in Romanian prior to 1918 are non-French. Indeed, poetry translations are much more balanced, and directly contradict the francophone status quo of Romanian culture. And if we look even deeper, we discover that the general French

propensity has another consequence: granted, translations from French and German are similar in number and size, but the eras they cover are vastly different, with 21% of French translations rendering pre-Romantic poets, 58% Romantic poets, and 21% post-Romantic poets, whereas 94% of German translations render Romantic poets, distinctively showing the main movements of interest for *CL*. Moreover, the next two cultures of origin are the Roman Empire and Ancient Greece, insofar as Ireland, although fourth, is an anomaly, due to 51 translations from Thomas Moore published in one year, 1905, in a sort of retrieval effort, which is another possibly fruitful avenue of interpretation.

Since I have already suggested some possible categories to chronologically integrate all translations, Figure 8 shows (this time by number of words) the evolution of the publication of poetry translations in *CL*, with all poems separated in five groups: (Ancient) Greek, Latin, pre-Romantic, Romantic, and post-Romantic. Besides the fact that Romantic poetry seems to be all-present, and clearly dominates the first third of the graph, something strange happens afterward. Following the publishing of many Romantic poems up to the early 1880s, one would expect *CL* to turn to, or at least attempt to turn to, the post-Romantic era; instead, contributors begin to intensively translate Latin poetry (until the late 1890s), followed by the Ancient Greek poetry that dominates all translations published after 1900. Therefore, just as *CL* does not seem to be able to leave Junimism behind when it comes to local production, its conservative stance turns even more conservative, and the journal turns (radically) backward instead of going forward.
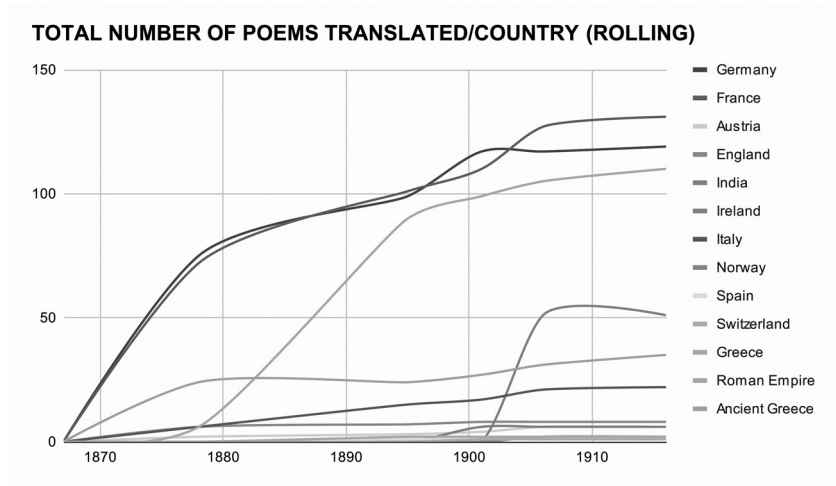
Figure 7: Total number of poems translated per country (rolling sum).
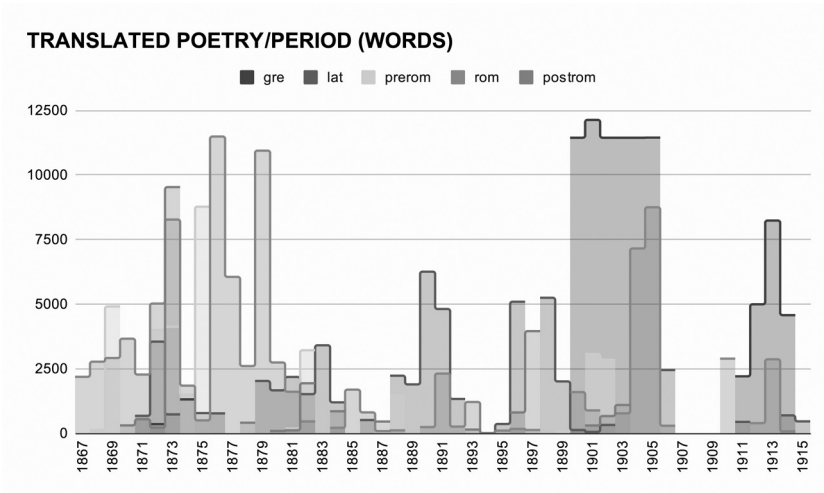
Figure 8: Translated poetry per designated period (words).

Vasile Alecsandri and Mihai Eminescu in world literature networks

As shown above, if we consider the number of poems and the number of words at the same time, two poets emerge as central to *CL*: Alecsandri and Eminescu. But even so, Alecsandri leads (by far) in both categories. Even though he was already an established poet by the time the first issue of *CL* appeared, he was appreciated by the Junimea society. At every step, however, his name seemed to be deployed as a tool to push forward Eminescu's name, starting with a nationally read article that Maiorescu published in *CL* in 1872; in that article, Maiorescu mentions Eminescu directly after naming Alecsandri and thus starts a twofold process: the canonization of Eminescu and the replacement of Alecsandri as the prominent Romanian poetic figure. This leads to Eminescu's "quick recognition as a top literary figure and even 'national poet'," which "must have appeared odd because the position seemed to have been filled already by an older and more distinguished writer, namely Alecsandri," who "surely fit the job description" and, in fact, "was overqualified" (Terian, "Mihai Eminescu" 38). The relationship between Maiorescu and Eminescu seemed to mirror another relation that led to the eventual canonization of a national poet, namely that between Slovenian critic Matija Čop and Slovenian poet France Prešeren, in which "Čop offered his friend Prešeren philosophical, historical and

comparative aesthetic knowledge as well as the fund of his private library and that of the lyceum" (Juvan 43). Junimea and Maiorescu funded Eminescu's studies abroad, arranged jobs for him, and published his works. But how did Maiorescu and *CL* canonize Eminescu? This last part of my study tries to quantitatively prove a hypothesis proposed by Andrei Terian in his contribution to the collective volume *Romanian Literature as World Literature*, namely that

> it is not by accident that Karel Hynek Mácha has been called "the Czech Byron," and that Hristo Botev has been dubbed "the Bulgarian Victor Hugo," whereas Adam Mickiewicz was exported to the United States during the First World War sometimes as "the Polish Goethe" and sometimes as "the Polish Shakespeare." All these authors gained legitimacy as national poets also through "at-distance" associations of various kinds with authors belonging to other cultures—if you think about it, the very granting of the title of national poet implies or, to my mind, should imply that the author in question does double duty as a transnational poet, that his or her work and his or her overall figure are a kind of business card one literature offers to the others. (Terian, "Mihai Eminescu" 36)

In brief, the hypothesis is that the Romanian national poet is canonized by means of linking him to foreign cultures that legitimize him on the local literary scene. Therefore, my colleague Ana-Maria Stoica and I extracted the names of non-Romanian poets and philosophers (to cover both literary production and thought production) who appear on the page before, on the same page, and on the page after the appearance of the names of Alecsandri and Eminescu. From a methodological point of view, I used ARCANUM's database and OCR (so clearly there will be errors, as not all instances of names were found) to search for the names—using "Eminescu" for Eminescu and "Alecsandri" plus "Alexandri" for Alecsandri. During the manual extraction, names were eliminated in various instances: they appeared in the contents section; they appeared after the end or before the beginning of the contribution containing the name; they appeared in contributions about other literary figures, and they did not specifically refer to Alecsandri or Eminescu. The results are in the table below and in Figures 9 and 10.

	mentions	links	unique links
Alecsandri	1679	336	85
Eminescu	1203	1121	207
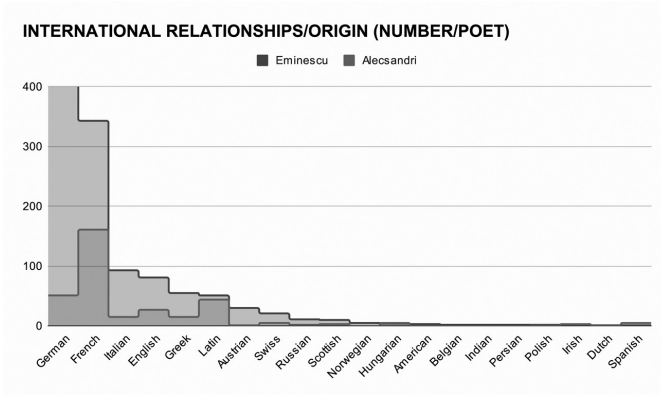
Figure 9: Number of international relationships per origin for Mihai Eminescu and Vasile Alecsandri.
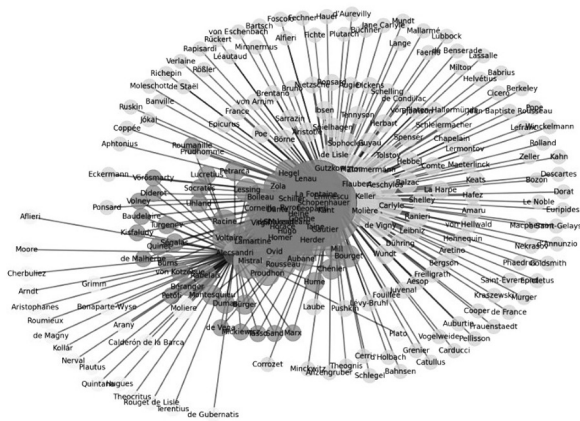
Figure 10: Network visualization of international relationships per origin for Mihai Eminescu and Vasile Alecsandri (generated with GPT-4, based on a links .csv).

The discrepancy in links and unique links is immediately striking. Eminescu has almost three times more international connections with over two times more unique links. And yet, the difference only intensifies if we calculate the average ratio between the number of mentions and the number of links, as Alecsandri is at 0.2, while Eminescu is at 0.93. That means that every five times Alecsandri's name appears, one

foreign name accompanies it, whereas almost every time Eminescu's name appears, it does so alongside a foreign name. Terian's intuition seems to be duly confirmed.

But if we look at the data more closely, and operate with categories rather than just numbers (Figure 9), we will see that Alecsandri's network of relationships is mainly in the French world (48%), whereas Eminescu's is mainly in the German world (36%), but with the French world being a close second (31%). This echoes the place of their formal education, with Alecsandri being educated partially in Paris, and Eminescu partly in Vienna and Berlin. But it also reveals a pattern of preference that associates Alecsandri only with the French canon (the next spaces of origin are the German one at 15% and the Latin one at 13%), while associating Eminescu with both the French and the German canons. And as we have already seen, the poetry scene is not nearly as francophone as the prose scene, which suggests that this system of canonization greatly favored Eminescu at Alecsandri's expense. Moreover, the network in Figure 10 (color-coded by singular and shared relationships: on the left are Alecsandri's singular relationships, in the middle the shared relationships, and on the right Eminescu's singular relationships) shows the sheer difference in breadth and scope between the two names. Around 20 names would disappear from the network if Alecsandri's name also disappeared, whereas more than 130 would disappear if Eminescu's name also disappeared.

Conclusions

Drawing everything from the analytical parts of this study together, a few insights were gained that support the idea that *CL* is the last bastion of European Romanticism. Firstly, the local poetry published in the magazine centered around the poetic output of Junimism, a late blend of Classicism and Romanticism, and was unable to overcome it, thus almost abandoning poetry publication by World War I despite the general growth in poetry production. Secondly, the translations published in *CL* were mainly from French and German, but after the exhaustion of the European Romantic poets the magazine did not turn to post-Romantic poetry, but rather circled back to Latin and Ancient Greek translations. And thirdly, Eminescu was canonized as the Romanian national poet through the critical internationalization of his poetry and his integration into world literature networks rather than through a better national grounding.

WORKS CITED

Anon. "How Do We Happily Marry Statistics and Literature?" *Metacritic*, vol. 1, no. 1, 2015, pp. 230–248.

Baghiu, Ştefan. "Translations of Novels in the Romanian Culture During the Long Nineteenth Century (1794–1914): A Quantitative Perspective." *Metacritic*, vol. 6, no. 2, 2020, pp. 87–106.

Baghiu, Ştefan, et al. "Geografia internă a romanului românesc în secolul al XIX-lea." *Transilvania*, no. 1, 2019, pp. 29–43.

Baghiu, Ştefan, et al. "Geografia romanului românesc (1933–1947)." *Transilvania*, no. 9, 2021, pp. 1–9.

Blakesley, Jacob. "Reading Distantly Poetry Translation: Modern European Poet-Translators." *Sociologies of Poetry Translation*, edited by Jacob Blakesley, Bloomsbury Academic, 2018, pp. 197–220.

Cohen, Margaret. *The Sentimental Education of the Novel*. Princeton University Press, 1999.

Colescu, Leonida. *Analiza rezultatelor recensământului general al populației române dela 1899*. Bucharest, Institutul Central de Statistică, 1944.

Dović, Marijan, and Jón Karl Helgason. *National Poets, Cultural Saints: Canonization and Commemorative Cults of Writers in Europe*. Brill, 2017.

Ferber, Michael, "Introduction." *A Companion to European Romanticism*, edited by Michael Ferber, Wiley, 2005, pp. 1–9.

Gonzalez-Blanco, Elena, et al. "Poetry and Digital Humanities Making Interoperability Possible in a Divided World of Digital Poetry." *EADH 2018*, https://zenodo.org/records/2203807. Accessed 26 Feb. 2024.

Hasdeu, Bogdan Petriceicu. "Unu remasiagu." *Familia*, no. 31, 1871, p. 368.

Hobbs, Andrew. "Five Million Poems, or the Local Press as Poetry Publisher, 1800–1900." *Victorian Periodicals Review*, vol. 45, no. 4, 2012, pp. 488–492.

Juvan, Marko. "World Literature in Carniola: Transfer of Romantic Cosmopolitanism and the Making of National Literature." *Interlitteraria*, no. 7, 2012, pp. 27–49.

Manolescu, Nicolae. *Istoria critică a literaturii române*. 2nd ed, Piteşti, Paralela 45, 2019.

Mihai, Adriana. "The Politics of Online Reproduction of Shakespeare." *Transilvania*, no. 8, 2015, pp. 15–21.

Moretti, Franco. "Conjectures on World Literature." *New Left Review*, no. 1, 2000, pp. 54–68.

Negruzzi, Iacob. *Amintiri din Junimea*. Bucharest: Humanitas, 2011.

Nemoianu, Virgil. *The Taming of Romanticism*. Harvard University Press, 1984.

Neubauer, John. "Introduction." *History of the Literary Cultures of East-Central Europe: Junctures and Disjunctures in the 19th and 20th Centuries*, vol. 1, edited by Marcel Cornis-Pope and John Neubauer, Johns Benjamins, 2004, pp. 321–324.

Nicolaescu, Mădălina, and Adriana Mihai. "Teaching Digital Humanities in Romania." *CLCWeb*, vol. 16, no. 5, 2014, http://docs.lib.purdue.edu/clcweb/vol16/iss5/15. Accessed 26 Feb. 2024.

Olaru, Ovio. "What Is Digital Humanities and What's It Doing in Romanian Departments?" *Transilvania*, no. 5–6, 2019, pp. 30–37.

Patraş, Roxana, et al. "The Splendors and Mist(eries) of Romanian Digital Literary Studies." *Hermeneia*, no. 23, 2019, pp. 207–222.

Pojoga, Vlad. "A Survey of Poetry Translations in Romanian Periodicals (1990–

2015).” *The Culture of Translation in Romania*, edited by Maria Sass et al., Peter Lang, 2018, pp. 99–121.

Pojoga, Vlad, et al. “Tehnici digitale pentru analiza romanului românesc.” *Transilvania*, no. 10, 2019, pp. 9–16.

Sapiro, Gisèle. “Publishing Poetry in Translation: An inquiry into the Margins of the World Book Market.” *Sociologies of Poetry Translation*, edited by Jacob Blakesley, Bloomsbury Academic, 2018, pp. 23–42.

Terian, Andrei. “Mihai Eminescu: From National Mythology to the World Pantheon.” *Romanian Literature as World Literature*, edited by Mircea Martin et al., Bloomsbury, 2018, pp. 35–54.

Terian, Andrei. “Big Numbers: A Quantitative Analysis of the Development of the Novel in Romania.” *Transylvanian Review*, Supplement no. 1, 2019, pp. 55–71.

Terian, Andrei, et al. “Genurile romanului românesc în secolul al XIX-lea.” *Transilvania*, no. 1, 2019, pp. 17–28.

Ursa, Mihaela. “Is Romanian Culture Ready for the Digital Turn?” *Metacritic*, vol. 1, no. 1, 2015, pp. 80–97.

Vancu, Radu. *Elegie pentru uman: O critică a modernității poetice de la Pound la Cărtărescu.* Bucharest, Humanitas, 2016.

Wu, Duncan, editor. *A Companion to Romanticism.* Blackwell, 1999.

Zadnji branik evropske romantike: kvantitativna analiza poezije v romunski literarni reviji *Convorbiri literare* (1867–1916)

Ključne besede: romunska poezija / evropska romantika / pesniški prevodi / Eminescu, Mihai / Alecsandri, Vasile / Maiorescu, Titu / kvantitativna analiza

Članek predstavi kvantitativno analizo poezije, objavljene v najbolj priljubljeni romunski literarni reviji druge polovice 19. stoletja *Convorbiri literare* (*CL*) od njene ustanovitve leta 1867 do leta 1916, ko je Romunija vstopila v prvo svetovno vojno. *CL*, ki velja za zadnjo večjo literarno revijo evropske romantike, je začela izhajati v času, ko je bila poezija privilegirana zvrst v romunski književnosti. Na svojih straneh je predstavljala tako narodnega pesnika Mihaia Eminescuja kot najvidnejšega romunskega literarnega kritika 19. stoletja Tituja Maiorescuja. Članek indeksira in kvantitativno analizira objavljeno poezijo in podatke, ki jih lahko razumemo kot pesniška omrežja. Rezultati analize podpirajo tezo, da je bilo objavljanje poezije in teoretiziranje o njej sestavni del enega zadnjih državotvornih procesov v Evropi. Iz ARCANUM-ovega digitalnega arhiva z več kot 66.000 stranmi *CL* so izluščeni in preučeni trije sloji metapodatkov, ki so povezani s poezijo: lokalna

produkcija poezije (kdo je objavljal v *CL* in kdaj ter koliko); tuja produkcija poezije (čigava poezija je bila v *CL* prevedena v romunščino in kdaj ter od kod so bili avtorji prevodne literature) in mreže vpliva (kateri tujci so bili najpogosteje omenjeni v zvezi z dvema najpopularnejšima romunskima romantičnima pesnikoma, Eminescujem in Vasilom Alecsandrijem). Z odkrivanjem mednarodne mreže avtorjev, ki so bili v središču literarnih razprav v *CL*, si članek prizadeva očrtati nacionalno shemo relevantnosti, ki velja za romunsko poezijo v drugi polovici 19. stoletja.

Analize omrežij v elektronski zbirki Pisma: vidik metapodatkov in semantičnih povezav besedišča

Katja Mihurko, Ivana Zajc, Darko Ilin, Mila Marinković

Univerza v Novi Gorici, Raziskovalni center za humanistiko, Vipavska 13, 5000 Nova Gorica
https://orcid.org/0000-0003-4225-0422
katja.mihurko@ung.si

Univerza v Novi Gorici, Raziskovalni center za humanistiko, Vipavska 13, 5000 Nova Gorica
https://orcid.org/0000-0002-5502-6772
ivana.zajc@ung.si

Univerza v Novi Gorici, Raziskovalni center za humanistiko, Vipavska 13, 5000 Nova Gorica
https://orcid.org/0000-0002-4200-6983
darko.ilin@ung.si

Videm 53, 1262 Dol pri Ljubljani
mm9136@student.uni-lj.si

Članek predstavlja elektronsko zbirko Pisma, procese njenega nastajanja in izzive, s katerimi se je raziskovalna skupina na Raziskovalnem centru za humanistiko Univerze v Novi Gorici soočala pri zbiranju in organiziranju več kot 1.600 pisem iz 19. in 20. stoletja, ki jih zbirka trenutno obsega. Članek izpostavlja vlogo metapodatkov tako za orientiranje po tej zbirki kakor za računalniško podprto oddaljeno branje v literarni vedi nasploh. Sistematična klasifikacija pisem skupaj z obsežnimi metapodatki ponuja številne možnosti raziskovanja. Spletna stran zbirke omogoča iskanje številnih tem pa tudi posameznikov in posameznic ter poljubnih izrazov iz njihovih pisem. Z metodo digitalne analize omrežij, ki je eden od tipov oddaljenega branja zgodovinskih dokumentov, članek prikaže računalniške analize različnih tipov omrežij z vidika povezav med tokovi različnih korespondenc v zbirki. Poleg tega predstavlja tudi z delotokom Clowdflows podprto analizo semantičnih povezav besedišča, povezanega z intimnim čustvom ljubezni v pismih avtoric slovenske moderne.

Ključne besede: slovenski pisatelji / pisemska korespondenca / ljubezen / metapodatki / analiza omrežij / digitalna humanistika

31

Uvod

Elektronska zbirka Pisma izhaja iz že uveljavljene tradicije elektronskih arhivov pisem v digitalni humanistiki in raziskovanja pisemskih korespondenc znanih literatov ter intelektualcev (prim. Scherber; Vidmar).[1] Zbirka, ki nastaja na Raziskovalnem centru za humanistiko Univerze v Novi Gorici, vsebuje več kot 1.600 pisem iz 19. in 20. stoletja. Obsega pisma avtoric in avtorjev slovenske moderne pa tudi druge korespondence, na primer pisma družine Kremenšek. Kot opaža Marija Dalbello, igrajo zgodovinske zbirke pisem in njihove digitalne prilagoditve ključno vlogo v različnih digitalnohumanističnih pobudah, pri čemer služijo kot dokumentarni dokazi, poudarjajo pa tudi pomen pisem kot medija in oblike pisnega izražanja; prispevajo k analizi besedil pa tudi k osvetlitvi njihovih družbenih kontekstov (Dalbello). Deljenje in s tem demokratizacija kulturnega kapitala s pomočjo digitalnih in spletnih medijev raziskovalni skupnosti omogoča lažji dostop do različnih predmetov raziskovanja. Kljub temu nekatere digitalne zbirke gradivo zgolj objavijo brez dodajanja metapodatkov, pri čemer lahko gradivo beremo in proučujemo, ne moremo pa ga uporabljati na noben drug način. Naš cilj ni bila zbirka, kjer bi bilo mogoče le brati kopijo pisma, ampak zbirka, ki bi omogočala dejavno interakcijo s pismi, iskanje različnih tem in raziskovanje kulturne dediščine, pa tudi preizkušanje in uporabo novih metodologij s pomočjo digitalnih orodij. Odgovoriti smo želeli na vprašanje, kako s sodobnimi pristopi s področja digitalne humanistike obdelati zgodovinske vire, da postanejo dostopni za širšo javnost. V članku predstavljamo proces nastajanja tovrstne zbirke in nekatere digitalne raziskave, ki so bile izvedene na podlagi elektronske zbirke Pisma.

Raziskave korespondenc za razumevanje (literarne) zgodovine

Raziskave korespondenc so pomembne za razumevanje literarnih pojavov, svetovnih nazorov in idej predstavnikov določenih skupin pa tudi literarne in kulturne zgodovine (prim. Gilroy in Verhoeven). Raziskovanje pisem predvideva sodelovanje skupine raziskovalcev, saj je ponovno združevanje razkropljenih pisem že v primeru enega samega posameznika iz zgodovine zamudno delo, ki včasih zahteva obsežno

---

delovanje celotne skupine znanstvenikov (Hotson in Wallnig 9). Pisma ponujajo intimen in avtentičen vpogled v razmišljanje posameznih avtorjev, saj običajno niso namenjena širšemu občinstvu. Vsebujejo informacije o ljudeh, publikacijah in dogodkih pa tudi o pisateljevem življenju in delovnem okolju (prim. Dumont). Vsebujejo literarne reference, odnose, bralne navade in intertekstualne navedbe (Scherber 279–280). Kot vir so ključna pri razumevanju subjektivnih izkušenj ljudi iz preteklosti (Eiranen 78, 82). Pisma niso le komunikacijski kanal, ampak vplivajo tudi na različne načine razumevanja literature in kulture v času svojega nastanka. Poleg tega lahko pisma iz preteklih obdobij razkrijejo genealogije naših kulturnih identifikacij, njihovo proučevanje pa nam lahko pomaga, da se z njimi soočamo.

Besedila pisem so vir za njihovo prepisovanje, razlaganje, prevajanje in opremljanje; toda za navigacijo, analiziranje in vizualizacijo ogromnih količin kompleksnih podatkov, ki so potrebni za pregled tega pojava, potrebujemo druge metode (Hotson in Wallnig 9). Za to je nujno interdisciplinarno raziskovanje, ki združuje pristope s področij literarne zgodovine, računalništva, oblikovanja, obdelave naravnega jezika in korpusnega jezikoslovja (Pressman in Swanstrom). Računalniške metode so v zadnjih dveh desetletjih pomembno obogatile študij literarne zgodovine, vendar nove pristope digitalne humanistike v veliki meri določa literarna veda (Pressman in Swanstrom). Digitalna humanistika poleg kritičnega ponovnega pretresa obstoječih metodologij v literarnih vedi vključuje metode medijskih in komunikacijskih študij, kadar prispevajo k interdisciplinarnim pristopom; trenutno primerjalna literarna veda in sorodne vede prispevajo k t. i. kritičnemu obratu v digitalni humanistiki, ki poleg kritičnega ponovnega proučevanja dolgoletnih metodologij v literarni vedi vključuje tudi relevantne vidike medijskih in komunikacijskih študij (Berry in Fagerjord).

V zadnjih letih je bilo izvedenih več raziskovalnih projektov na temo pisemskih korespondenc. Ena od njih je COST akcija na temo t. i. republike pisem iz let 1500–1800, ki je utemeljila digitalne raziskave tega vidika intelektualne zgodovine Evrope. Projekt ePistolarium ponuja virtualno raziskovalno okolje za analizo besedil približno 20.000 digitaliziranih pisem vodilnih intelektualcev 17. stoletja. Projekt Mapping the Republic of Letters se je osredotočil na vizualno in zlasti kartografsko predstavitev podatkov na tem področju. Projekt Visual Correspondence uporablja računalniške vizualizacije informacij in omogoča nova dognanja o zgodovinskih osebnostih s pomočjo njihovih korespondenc; poleg zagotavljanja orodij za vizualizacijo obsežnih zbirk korespondence, tj. 164.196 pisem iz 53 zbirk, projekt

omogoča nov način raziskovanja pisem, in kjer je mogoče, so dostopne tudi povezave do celotnega besedila pisem in dodatne informacije, ki pisma kontekstualizirajo. Projekt Letters 1916–1923 je po zaslugi digitaliziranih pisem omogočil intimen vpogled v življenje na Irskem z začetka 20. stoletja, pokazal pa je tudi, kako so na irski prostor gledali od zunaj. Projekt Early Modern Letters Online združuje številne rokopise, tiskane in elektronske vire na enem mestu, pri tem pa ne samo povečuje dostopnost do njih in zavedanje o njih, ampak omogoča tudi prečno iskanje, združevanje, analizo in vizualizacijo najrazličnejših povezanih korespondenc. Projekt, ki se osredotoča na posamezno avtorico, je digitalna izdaja vseh znanih pisem Belle van Zuylen oziroma Isabelle de Charrière (1740–1805). Study Platform on Interlocking Nationalisms, ki jo razvijajo na Univerzi v Amsterdamu, je ustvarila nabor metapodatkov več kot 38.000 pisem iz 19. stoletja, pri čemer so bili podatki vneseni ročno in le delno samodejno uvoženi.

Historiat elektronske zbirke Pisma

Prva različica baze podatkov, ki se je razvila v elektronsko zbirko Pisma, je bila rezultat študentskega projekta, ki ga je financiralo Ministrstvo za izobraževanje leta 2020. V tem projektu je desetčlanska interdisciplinarna skupina študentk in študentov pod vodstvom projektne vodje Katje Mihurko ustvarila bazo podatkov in vanjo vnesla 500 pisem, ki so bila opremljena z metapodatki. Projekt se je med drugim zgledoval po elektronski zbirki Letters 1916–1923, ki je prvi participativni digitalnohumanistični projekt na Irskem. Elektronska zbirka Pisma se je nato razvijala v okviru programske skupine Historične interpretacije 20. stoletja in v okviru raziskovalnega projekta Transformacije intimnosti v literarnem diskurzu slovenske moderne, ki si je kot enega od ciljev zadal raziskovanje pisem slovenskih literatk.

Prvi korak je zahteval pretipkavanje arhivskih izvirnikov pisem, nato je sledil sistematičen vnos pisem v elektronsko zbirko. Ker smo želeli vnesti večje število pisem, smo se odločili, da posamezno besedilo vnesemo neposredno v bazo podatkov. Strateško smo predvidevali, da bo ključna faza v procesu digitalizacije vključevala natančno klasifikacijo teh pisem. V okviru našega raziskovanja tematik intimnosti je ta proces vključeval analizo osrednjih tem pisem. Poleg tega smo kurirali bistvene metapodatke, povezane z vsakim pismom, vključno s podatki o pošiljatelju, prejemniku ter izvoru in namembnosti dopisovanja. Med razvojem elektronske zbirke smo razširili nabor tem, saj

so posamezne raziskovalke in raziskovalci prepoznali potrebo po novih temah v procesu vnašanja pisem. Po vsaki spremembi so uporabniki preizkusili zbirko in skupini nudili povratne informacije. Poleg tega smo ugotovili, da je smiselno med metapodatke vključiti določene informacije, ki jih sprva ni bilo, med drugim podatek o tem, ali je določeno pismo poslal pošiljatelj ali pošiljateljica. V zadnjem delu leta 2022 je elektronska zbirka dobila svežo podobo in uvedli smo nove funkcionalnosti iskanja. Tehnološka osnova elektronske podatkovne baze pisem je podatkovna baza MySQL, ki je bila nadgrajena s platformo odprtokodnega CMS-ja (ang. *Control Managment System*) Laravel, ki omogoča pregledno in enostavno urejanje vsebine. Metapodatke in pisma je mogoče tudi izvoziti. Trenutno elektronska zbirka Pisma vsebuje skoraj 1.600 pisem, ki segajo od leta 1878 do 1946. Uporabnikom omogoča brezplačen dostop do intimne korespondence najrazličnejših posameznikov iz obdobja, ko so bila pisma pomembno sredstvo komunikacije. Pri vključevanju pisem se ravnamo po kriteriju, da ni pomembno, ali je pismo napisala običajna oseba ali znano ime, ampak je ključno to, kako lahko pismo uporabimo za razumevanje osebnega doživljanja intimnih dogodkov v življenju posameznikov pa tudi osrednjih zgodovinskih trenutkov.

Sistematična klasifikacija gradiva skupaj z obsežnimi metapodatki je odprla vrsto možnosti za raziskovanje različnih razsežnosti zgodovinskih virov. Spletna zbirka poleg iskanja številnih tem omogoča iskanje posameznikov in poljubnih izrazov, tako da je iskanje lahko osredotočeno na določeno področje. Primer so teme, povezane z intimnostjo, ki odpirajo možnosti raziskovanja zgodovine čustev in izkušenj ljudi iz zgodovine, in sicer tako običajnih ljudi kakor znanih literarnih ustvarjalk in ustvarjalcev. Spletna zbirka omogoča različne vpoglede: prikaže na primer korespondenco, kjer sta zbrana tako pošiljatelj kakor naslovnik; prikaže pisma izbranega posameznika različnim naslovnikom, pri čemer lahko raziskujemo vso korespondenco, ki jo je vodila določena oseba; in prikaže posamezno pismo z vsemi pripadajočimi metapodatki.

Elektronske zbirke Pisma ne uporabljamo le za raziskave, temveč tudi za poučevanje. Na primer pri predmetu Spol v dobi moderne na Fakulteti za humanistiko Univerze v Novi Gorici študentje in študentke berejo pisma, povezana z različnimi temami, in razpravljajo o različnih pojmovanjih spola, kot so izražena v pismih, ki so jih napisali tako običajni ljudje kakor slovenski intelektualci. Poleg tega je projekt del slovenske in evropske mreže občanske znanosti, saj tudi občani prepisujejo in urejajo pisma ter jih raziskujejo, s tem pa se seznanjajo s kulturno dediščino. Marsikdo iz Slovenije in tujine je že stopil v stik z

35

uredništvom zbirke in nam ponudil pisma za prepisovanje; leta 2023 smo na ta način v zbirko dodali več kot 500 pisem. S skrbno načrtovano komunikacijo dogodkov in dosežkov projekta v javnih ustanovah, medijih in družabnih omrežjih širimo ozaveščenost o pomembnosti pisem kot pomembnega vira za razumevanje preteklosti in sedanjosti.

Raziskava omrežij pisem obdobja moderne

Kakšni so odnosi oziroma povezave med različnimi vidiki korespondenc, ki so vključene v digitalno zbirko Pisma, smo raziskali z metodo digitalne analize omrežij. Tovrstno oddaljeno branje zgodovinskih dokumentov obsega kvantitativno in računalniško podprto analizo različnih tipov omrežij na literarnovednem področju, ki omogočajo boljše prepoznavanje vzorcev, povezav in tematik ter so informativni temelj za kvalitativne analize. Omrežje je vizualizacija odnosov med izbranimi metapodatki in prikazuje, kako so določeni elementi medsebojno povezani v določenem času in prostoru. To vključuje različne tipe podatkov: razmišljanje o literaturi z vidika omrežij je interdisciplinarno, saj presega posamezna področja raziskav (Anhert idr.). Ti pristopi bogatijo literarno vedo z novim zornim kotom, v nasprotju s hierarhijami in strukturami, ki imajo jasno središče, pa prinašajo decentralizacijo (Jagoda). Analize omrežij v književnosti in literarni zgodovini predvidevajo sodelovanje različnih ved in nam omogočajo posredovanje velike količine podatkov hkrati (Anhert idr.). S tem v besedilih prepoznavamo vzorce, ki nam omogočajo nov vpogled v zgodovinsko dogajanje. To je pomembno tudi za analize s področja primerjalne literarne vede; med drugim lahko preverimo, ali držijo teorije o posameznem viru, ki ima neposreden vpliv na druge avtorje, saj lahko digitalne analize odkrijejo številne druge navezave (Shore). Z računalniki lahko raziskovanje literarnih vplivov poglobimo in sistematiziramo ter ga s tem bolje razumemo, saj lahko pokažemo, da je proces literarnih vplivov zelo pluralen (Shore).

Analize omrežij pisem je raziskovalna skupina na Raziskovalnem centru za humanistiko Univerze v Novi Gorici izvedla s sodelovanjem študentke Fakultete za računalništvo in informatiko Univerze v Ljubljani Mile Marinković na podlagi projekta INSPIRO. Raziskava je temeljila na metapodatkih v zbirki, ki je bila za te potrebe razdeljena na tri osrednje korpuse: pisma družine Kremenšek, omrežje slovenske moderne in celotna zbirka pisem. V članku predstavljamo rezultate, vezane na pisma avtoric in avtorjev slovenske moderne. Vizualne predstavitve podatkov,

kot so grafi, so bile skrbno izbrane, preliminarne raziskave pa so bile nadgrajene na podlagi povratnih informacij raziskovalne skupine, pri čemer smo se osredotočili tako na vsebino kakor na obliko vizualizacij. Prek vizualizacij smo uvideli vzorce, ki jih pred tem nismo zaznali, s tem pa smo pridobili pomembne informacije o spletni zbirki, ki so nam pomagale tudi pri prihodnjih načrtih za njene nadgradnje.

Pisemsko omrežje avtorjev in avtoric slovenske moderne trenutno obsega nekaj manj kot 2.000 pisem: od tega je 937 prejetih pisem (prejemnic je 490, prejemnikov pa 447), 957 pa je poslanih pisem (prejemnic je 561, prejemnikov pa 396). Največ pisem je bilo poslanih iz Ljubljane, z Dunaja, iz Zagreba, Trsta in Prage. Spodnja slika prikazuje komunikacijsko mrežo korespondenc.
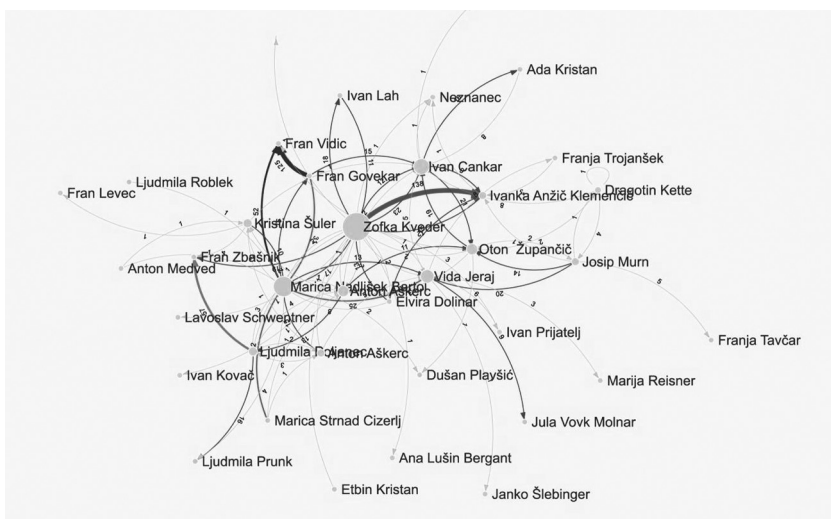
Slika 1: Komunikacijska mreža pisem literarne moderne.

Vsako od pisem so raziskovalci ob vnosu v bazo prepisali in ročno vnesli tematike posameznih besedil. S t. i. oddaljenim branjem teh metapodatkov smo ugotovili, da so bile najpogostejše teme pisanja uredniško delo, kultura in umetnost, literarna kritika, literarne revije in družabnost, sledijo pa potovanja, literarna kritika, ženske revije, zdravje, literarne revije, osebna razmerja, žensko prijateljstvo, finance, izobraževanje, delavnost idr. Spodnji sliki prikazujeta najpogostejše teme pri avtorjih pisem na eni strani in pri avtoricah na drugi. Iz tematik je razvidno, katere teme so bile aktualne v intelektualnih debatah tega časa.
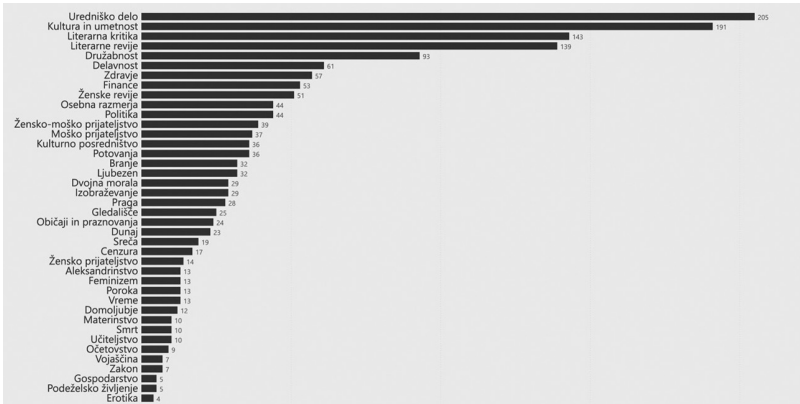
Slika 2: Najpogostejše teme avtorjev pisem.
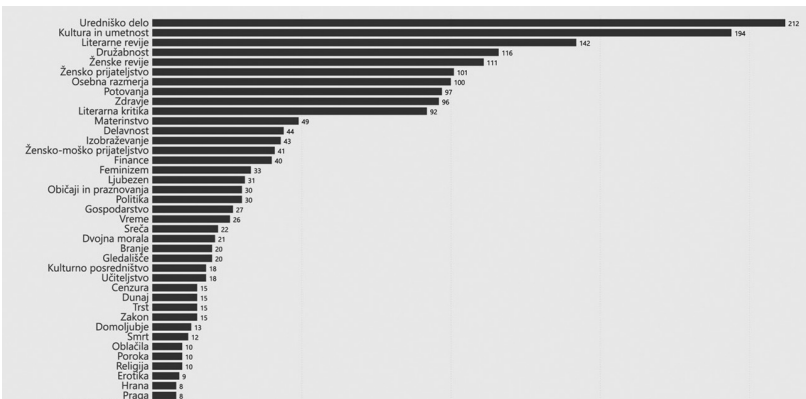
Slika 3: Najpogostejše teme avtoric pisem.

Tako pri piscih kakor pri piskah je najpogostejša tema uredniško delo, kar kaže, da je bilo literarno uredništvo v času moderne pomembno. Tudi tematika kulture in umetnosti je tako pri piscih kakor pri piskah na drugem mestu. Zatem pri avtorjih najdemo teme literarne kritike, literarnih revij in družabnosti, pri avtoricah pa se podobne teme ponovijo v nekoliko drugačnem zaporedju, pri čemer je poudarek na ženskih revijah. Ugotovitve kažejo na pomembno vlogo, ki jo je imela v času moderne literarna periodika, v njenem okviru pa tudi literarna kritika. Med najpogostejšimi so tudi teme, ki zadevajo družabno življenje in medčloveške odnose. To ne čudi, saj so bili znotraj literarnih krogov v obdobju moderne poleg poklicnih vezi pogosto v ospredju tudi prijateljski in drugi medosebni odnosi.

Raziskava tematike ljubezni

Po t. i. emocionalnem obratu v literarni vedi so čustva, ki so eden ključnih elementov bralnega izkustva v književnosti, v ospredju številnih literarnovednih raziskav, vključno s študijami s področja digitalne literarne vede. V tem okviru poteka proučevanje čustev v literarnih delih, raziskovanje njihovega izražanja s pomočjo različnih literarnih postopkov ter povezovanje z etičnimi in političnimi vprašanji, tesno povezanimi s čustvenimi temami (Hogan idr.). Raziskovalci lahko s pomočjo digitalnih orodij proučijo, kako se čustvena stanja odražajo v izbranih literarnih ali neliterarnih besedilih. V članku se ob upoštevanju predpostavke, da so čustva vedno vpeta v zgodovino (Champion in Lynch x), osredotočamo na analizo pisemske korespondence pisateljev in pisateljic iz obdobja moderne. Ob pismih kot primarnem viru literarne zgodovine nas zanimajo subjektivne izkušnje in interpretacije pisateljev in pisateljic, in sicer zlasti subjektivna narava literarnozgodovinskih virov.

Tematika ljubezni v pismih v primerjavi z drugimi temami

Analiza metapodatkov o pismih, prikazana v spodnji sliki, je pokazala, da se tematika ljubezni v korespondencah literarnih ustvarjalk in ustvarjalcev moderne povezuje s številnimi drugimi temami, med drugim s cenzuro, zakonom, spolnostjo, družabnostjo in prijateljstvom.
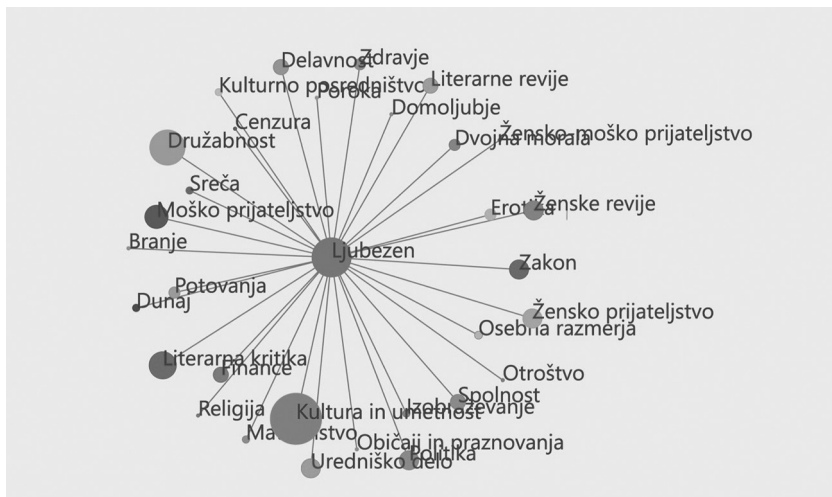
Slika 4: Povezave tematike ljubezni z drugimi temami v pismih obdobja moderne.

Pisma torej kažejo na kompleksno razumevanje ljubezni, zlasti v primerjavi s pismi drugega obravnavanega korpusa, korespondence družine Kremenšek, katere člani niso bili del tedanjih literarno-kulturnih krogov, temveč delavci in kmetje. V tej korespondenčni mreži se tematika ljubezni povezuje le s temami deklištva, zdravja, delavnosti in živali.

Tematika ljubezni v pismih v primerjavi z drugimi temami

S pomočjo študije primera smo raziskali tudi povezovanje ljubezni z drugimi besedami v pismih. Za ta del raziskave, ki jo je izvedla Ivana Zajc, smo najprej ustvarili besedilni korpus, ki vsebuje vsa pisma ženskih pisateljic iz obdobja slovenske moderne, ki so vključena v elektronsko zbirko Pisma. Besedila pisem smo združili v enoten dokument in odstranili metapodatke. Za njihovo obravnavo smo uporabili delotok ClowdFlows, ki so ga razvili na Oddelku za tehnologije znanja na Inštitutu »Jožef Stefan« v okviru mednarodnega projekta Embeddia. Cilj raziskave je bil najti pomene, ki so v pismih povezani z intimnostjo.

Za namen raziskovanja teme ljubezni s pomočjo delotoka ClowdFlows je bil korpus pretvorjen v reprezentacijo globokega učenja. ClowdFlows ima integrirano orodje za vektorske vložitve besed, ki omogoča pridobivanje semantičnih povezav izbranih izrazov z drugimi besedami v korpusu. Za analizo sta bili izbrani besedi 'ljubezen' in 'srce', delotok pa je identificiral besede v okolici teh besed. Tako so postali razvidni vzorci pomena, ki kažejo, kateri izrazi so semantično povezani med seboj. Računalniško branje besedil je na primer pokazalo, da je beseda 'jubezen' povezana z besedo 'skrivnost', beseda 'srce' pa z besedo 'skrivati'. Ta proces nam je omogočil, da smo pridobili dodatne informacije o pomenu izbranih besed in o semantičnih povezavah, ki so jih pisateljice iz obdobja slovenske moderne ustvarile v svojih pismih: izkazalo se je, da je proces samocenzure intimnih tem potekal na besedilni ravni posameznih pisem.
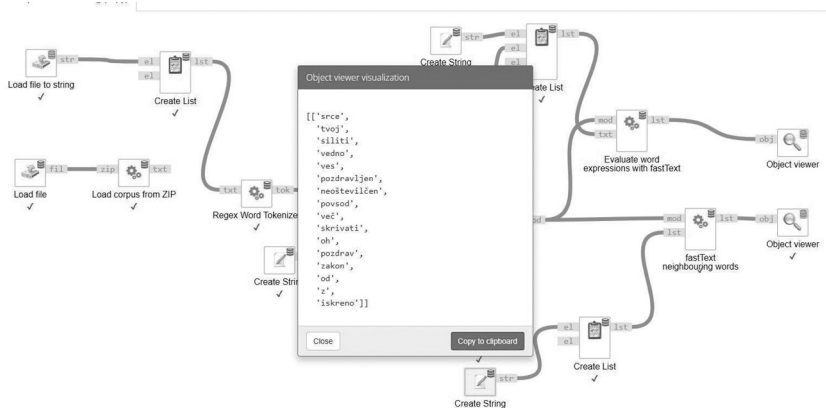
Slika 5: Analiza semantičnih povezav besede 'srce' v delotoku ClowdFlows.

Namen tega dela raziskave je bil t. i. oddaljeno branje pisem avtoric moderne, ki pa smo ga nadgradili s t. i. natančnim branjem. Izsledki oddaljenega branja so nas namreč napotili k vprašanju, kako avtorice pisem pišejo o ljubezni; gre torej za primer natančnega branja, ki ga je informiralo prvotno oddaljeno branje. Pisma avtoric moderne namreč vsebujejo tudi različne izpovedi čustev, čeprav ima sentimentalnost v njih pogosto negativno konotacijo. Ob tem so razvidne družbene konvencije, povezane s tem, kako naj bi bilo primerno izražati čustva. Za raziskavo smo izbrali čustvo ljubezni, ki je imelo v 19. stoletju v primerjavi z drugimi čustvi pomembno vlogo; po Ireni Selišnik je ljubezen celo »ključno čustvo stoletja«: »[r]omantična ljubezen, ki prisega na vznesena čustva, zahteva žrtvovanje in končno doseže stalno ljubezen (poroko)« (Selišnik 68). Kot ugotavlja Katja Mihurko, postaneta v 19. stoletju zakon iz ljubezni in zakonska ljubezen enovito načelo, ki naj bi izpopolnjevalo človeka (Mihurko Poniž 190–191).

Z natančnim branjem pisem smo ugotovili, da je v obravnavanih pismih ljubezen pogosto tematizirana z vidika omejitev na eni strani in svobodnega izražanja na drugi. Izražanje ljubezni v pismih pogosto prekinjajo samokritični komentarji, uperjeni v neprimernost te tematike, ki jim sledi sprememba tematike; na primer Marica Nadlišek Bartol v pismu Franu Vidicu piše o svojem novem partnerju in razlaga, da do njega čuti strast in da jo je ta odnos spremenil, nato pa svojo pripoved nenadno prekine in zapiše, da noče biti sentimentalna in da bo prešla na novo tematiko, ter pismo nadaljuje z vprašanjem, kako se je naslovnik imel na Dunaju (Marica Nadlišek Bartol Franu Vidicu, 11. marec 1899,

Ms 1834). V drugem pismu avtorica istemu naslovniku zaupa, da težko najde moškega, s katerim bi lahko bila v zvezi, zatem pa to komentira z besedami, da ji bo najbrž naslednji dan žal, da mu je v pismu to zaupala, in da upa, da njena pisma uničuje ali vsaj skriva (Marica Nadlišek Bartol Franu Vidicu, b. d., Ms 1834). Vida Jeraj Marici Nadlišek Bartol opiše svoja ljubezenska čustva do moškega, ki ga pogreša, ker je šel na Dunaj, ob tem pa se spominja njegovih lepih oči; misel v pismu avtorica prekine z mislijo, da so ženske otročje, nato pa zamenja temo (Vida Jeraj Marici Nadlišek Bartol, 1. januar 1929, Ms 703).

Ugotavljamo, da je bila ljubezen v omrežju korespondenc avtoric in avtorjev slovenske moderne pogosta tematika, ki se je povezovala z mnogimi drugimi temami predvsem z vidika njihovega intelektualno-umetniškega učinkovanja. Z računalniško semantično analizo smo ugotovili, da se tema ljubezni pogosto povezuje s prikrivanjem in skrivanjem. To potrjujejo tudi izsledki natančnega branja pisem, ki je odkrilo številne primere, ki pričajo o tem, da je bilo čustvo ljubezni v korespondencah iz obdobja moderne zaznamovano.

VIRI

Marica Nadlišek Bartol Franu Vidicu, b. d., IV Korespondenca, Marica Nadlišek – Bartol (53, 1896–1937 in b. d.), Zapuščina Frana Vidica (Ms 1834), NUK. https://pisma-rch.ung.si/pismo/686.
Marica Nadlišek Bartol Franu Vidicu, 11. marec 1899, IV Korespondenca, Marica Nadlišek – Bartol (53, 1896–1937 in b. d.), Zapuščina Frana Vidica (Ms 1834), NUK. https://pisma-rch.ung.si/pismo/682.
Mihurko, Katja idr., ur. *Elektronska zbirka Pisma*. Raziskovalni center za humanistiko, Univerza v Novi Gorici. https://pisma.org.
Vida Jeraj Marici Nadlišek Bartol, 1. januar 1929, II. Korespondenca, Vida Jeraj (29, 1896–1920 in b. d.), Zapuščina Marice Nadlišek Bartol (Ms 703), NUK. https://pisma-rch.ung.si/pismo/42.
Vida Jeraj Mariji Reisner, 3. julij 1905, Korespondenca, Vida Jeraj (2, 1905), Zapuščina Marije Reisner (Ms 1161), NUK. https://pisma-rch.ung.si/pismo/54.

LITERATURA

Anhert, Ruth, Sebastian Anhert, Catherine Coleman in Scott Weingart. *The Network Turn*. Cambridge University Press, 2020.
Berry, David M., in Anders Fagerjord. *Digital Humanities: Knowledge and Critique in a Digital Age*. Polity, 2017.
Champion, Michael, in Andrew Lynch. »Understanding Emotions: 'the things they left behind'«. *Understanding Emotions in Early Europe*, ur. Michael Champion in Andrew Lynch, Brepols, 2015, str. i–xii.

Dalbello, Marija. »Digitality, epistolarity and reconstituted letter archives«. *Proceedings of CoLIS 8, Copenhagen, Denmark, 19-22 August, 2013*, posebna številka revije *Information Research*, let. 18, št. 3, 2013, https://informationr.net/ir/18-3/colis/paperC26.html. Dostop 2. 2. 2024.

Dumont, Stefan. »Connecting Scholarly Editions of Letters«. *Journal of the Text Encoding Initiative*, št. 10, 2016, https://journals.openedition.org/jtei/1742. Dostop 9. 2. 2024.

Eiranen, Reetta. »The Narrative Self: Letters and Experience in Historical Research«. *Private and Public Voices: An Interdisciplinary Approach to Letters and Letter Writing*, ur. Karin Koehler in Kathryn McDonal-Miranda, Inter-Disciplinary Press, 2015, str. 77–99.

Gilroy, Amanda, in W. M., Verhoeven, ur. *Epistolary Histories: Letters, Fiction, Culture*, University Press of Virginia, 2000.

Hogan, Patrick Colm, idr., ur. *The Routledge Companion to Literature and Emotion*. Routledge, 2022.

Hotson, Howard, in Thomas Wallnig, ur. *Reassembling the Republic of Letters in the Digital Age*. Göttingen University Press, 2019.

Jagoda, Patrick. »Literature: Networks«. *Oxford Research Encyclopedias: Literature*. https://oxfordre.com/literature/display/10.1093/acrefore/9780190201098.001.0001/acrefore-9780190201098-e-135. Dostop 1. 2. 2024.

Mihurko Poniž, Katja. *Labirinti ljubezni v slovenski književnosti od romantike do II. svetovne vojne*. Založba Sophia, 2008.

Pressman, Jessica, in Lisa Swanstrom. »The Literary and/as the Digital Humanities«. *Digital Humanities Quarterly*, let. 7, št. 1, 2013, http://www.digitalhumanities.org/dhqdev/vol/7/1/000154/000154.html. Dostop 9. 2. 2024.

Scherber, Peter. »Cankarjeva pisma v kontekstu njegovega celotnega opusa: nove oblike in možnosti elektronskih edicij pisem«. *Starejši mediji slovenske književnosti*, ur. Urška Perenič in Aleksander Bjelčevič, Center za slovenščino kot drugi in tuji jezik, Oddelek za slovenistiko FF UL, 2018, str. 279–284.

Shore, Daniel. »Was It for This? Influence, Archive, and Network«. *Modern Philology,* let. 113, št. 3, 2016, str. 398–421.

Selišnik, Irena. »Ljubezen in prijateljstvo v drugi polovici 19. stoletja: ljubezen je burja življenja, a prijateljstvo njega pokoj«. *Nečakov zbornik: procesi, teme in dogodki iz 19. in 20. stoletja*, ur. Kornelija Ajlec idr., Univerza v Ljubljani, Filozofska fakulteta, 2018, str. 57–67.

Vidmar, Luka. *Zoisova literarna republika: vloga pisma v narodnih preporodih Slovencev in Slovanov*. Založba ZRC SAZU, 2010.

Zajc, Ivana. »Becoming a Woman Writer: The Self-representations in the Letters of Women Writers of the Slovene moderna«. *Literaturna misʼl*, let. 66, št. 2, 2023, str. 139–161.

Network Analysis in the Electronic Collection Letters: The Perspective of Metadata and Semantic Connections of Vocabulary

The article presents the electronic collection Letters, the processes of its creation, and the challenges faced by the research team at the Research Center for Humanities at the University of Nova Gorica in collecting and organizing more than 1,600 letters from the nineteenth and twentieth centuries, which the collection currently encompasses. The article highlights the role of metadata both in navigating this collection and in computer-assisted distant reading in literary studies in general. The systematic classification of letters, along with extensive metadata, offers a range of research possibilities. The website enables users to search for numerous topics, as well as for individuals and individual terms used in the letters. Using the method of digital network analysis, which is a type of distant reading of historical documents, the article demonstrates computational analyses of various types of networks in terms of connections between different correspondences within the collection. Additionally, it presents an analysis of the semantic connections of vocabulary related to the intimate emotion of love in the letters of Slovenian women writers from the modernist period, supported by the Clowdflows workflow.

SrpELTeC: A Serbian Literary Corpus for Distant Reading

Ranka Stanković, Cvetana Krstev, Duško Vitas

University of Belgrade, Faculty of Mining and Geology, Djusina 7, 11000 Belgrade, Serbia
https://orcid.org/0000-0001-5123-6273
ranka.stankovic@rgf.bg.ac.rs

Language Resources and Technologies Society, Studentski trg 3, 11000 Belgrade, Serbia
https://orcid.org/0000-0003-3328-9392
CvetanaJK@gmail.com

Language Resources and Technologies Society, Studentski trg 3, 11000 Belgrade, Serbia
https://orcid.org/0000-0003-4194-692X
dusko.vitas@gmail.com

The article presents SrpELTeC, a corpus developed within the COST action Distant Reading for European Literary History (CA16204). All novels in SrpELTeC were selected, prepared, and annotated using the common principles established for all language collections in the European Literary Text Collection (ELTeC). The challenges and solutions in preparing SrpELTeC from scratch are outlined. All novels were manually encoded in TEI with rich metadata and structural annotation. The automatic annotation included POS-tagging, lemmatization, and named entities, relying on Natural Language Processing resources developed and maintained by the JeRTeh Language Resources and Technologies Society. The integration of SrpELTeC with Wikidata was supported with a set of SPARQL queries for the retrieval of metadata with different visualization options. Recent activities within the COST Action NexusLinguarum–European Network for Web-centred Linguistic Data Science (CA18209) are related to the linked data version of SrpELTeC using the NLP Interchange Format. All versions of SrpELTeC are freely available under the CC-BY license.

Keywords: digital humanities / Serbian literature / text corpora / distant reading / linked data / named entity recognition / text analytics

Introduction

The paradigm of distant reading involves the use of computer methods for the analysis of large collections of literary texts. The goal of these analyses is to complement the methods used in literary theory and history. Franco Moretti proposed reading methods that involve works outside the

45

established literary canon, which, following Margaret Cohen, he refers to as the "great unread" (Moretti 55). A novelty that Moretti suggests for the study of literature is the use of patterns, statistics, paratexts, and other properties that literary studies tended to disregard.

Applying methods of distant reading necessitates a careful selection of works based on firm criteria and an equally careful preparation of the selected works. The focus of the COST action Distant Reading for European Literary History (CA16204), which ran from 2017 to 2022, was the preparation of a multilingual resource dubbed the European Literary Text Collection, or ELTeC (Odebrecht et al.; see also Burnard et al.). The core of ELTeC contains a hundred novels first published between 1840 and 1920, covering twelve languages commonly spoken in Europe, each forming linguistic sub-collections (Schöch et al.). In addition, for nine languages partial collections of under a hundred novels were developed, while for six languages extended collections were developed as well.

The mandatory criteria for selection demanded that each work belong to narrative prose (as a novel or a long story), include a minimum of 10,000 words, and appear in first edition between 1840 and 1920. In order to be included in a certain language sub-collection, the work had to be originally written in that language, since translations were not foreseen. Preference was given to works that were published as books, rather than in installments in serial publications.

Additional conditions were set for the composition of each sub-collection, with the idea to ensure, on the one hand, the diversity of the works represented and, on the other hand, a comparative analysis of sub-collections and the application of key methods for the statistical analysis of texts. These additional criteria for the desirable corpus balance involved each sub-collection size, genders of authors, the lengths of texts and the number of their editions, even coverage of the period 1840–1920, and the number of novels per author. A sub-collection had to contain a hundred works that qualify as novels according to the mandatory criteria; optimally, it would also strike a balance between male and female authors. Canonical works as well as unknown and forgotten works were to be represented, where the number of editions of a work was used as a measure of its canonicity. The selected time period of first editions divided into four periods lasting 20 years each had to be evenly represented in each sub-collection, and each of these periods had to be represented by 20–25 works. According to their length, works were divided into short (10,000–50,000 words), medium-length (50,001–100,000), and long texts (more than 100,000 words). A sub-collection

was to contain at least 20% works of all lengths, ideally 30–40%. Additionally, a sub-collection had to contain nine to 11 authors represented by exactly three works (which would, for example, allow testing automatic authorship checking systems), while all other works had to be written by different authors to ensure a sufficient level of diversity.

The sub-collections for Czech, German, English, French, Swiss German, Hungarian, Polish, Portuguese, Romanian, Slovenian, Spanish, and Serbian are completed with 100 works. The additional nine languages, namely Greek, Irish, Croatian, Italian, Lithuanian, Latvian, Norwegian, Ukrainian, and Swedish, have incomplete sub-collections.

The Serbian sub-collection of novels (SrpELTeC)

The Serbian sub-collection of novels, SrpELTeC, was created by a research team led by Cvetana Krstev. Given the aforementioned set selection criteria and the demand for balance, it comes as no surprise that creating a sub-collection of Serbian novels was not a trivial task, and that its development required much more effort than, say, in the case of English or French. First and foremost, prose writing in the Serbian language, and especially novel writing, appeared later than in most European countries, namely with the emergence of realism, which prevailed as a direction in the last three decades of the nineteenth century (Deretić, *Istorija* 362). This practically means that the set of works from which works could be chosen in order to better satisfy the balance criteria for Serbian was not as rich as for many other languages.

A number of Serbian literary works from this period, mostly canonical ones, were already digitized. Unfortunately, the way they were digitized did not allow us to include these digital editions in the SrpELTeC sub-collection. The library that contains the greatest number of Serbian literary works from the nineteenth and early twentieth centuries is called Antologija srpske književnosti (The Anthology of Serbian Literature) and was developed by the Faculty of Education at the University of Belgrade in cooperation with Microsoft. However, as these digital editions lack metadata, we cannot tell which editions of works included in this Anthology were used for digitization. It was therefore necessary to create the Serbian sub-collection from scratch. This took place in several steps, which will be briefly described below.

The first task was the compilation of a list of works that meet the selection criteria. The initial list with significant works and novelists was compiled from scholarly books by Jovan Deretić and Živan

Milisavac (Deretić, *Srpski*; Milisavac). The real challenge was to find so-called marginal works of Serbian literature, that is, books that were published only once or twice and whose authors are mostly forgotten today. The list was further enhanced with works that were found by searching the common catalogue of Serbian libraries COBISS+ by selecting appropriate values for the type of content or literary genre (novel, short story, short prose, etc.) and applying necessary restrictions on the language and the year of publication. Some works were retrieved from Belgrade dealers of antique books, while some suggestions were obtained from already acquired old books which contained appended lists of published works by the same publisher. Finally, a list of more than 150 candidates for the SrpELTeC sub-collection was obtained.

The next issue was to find and scan the books themselves, using the first editions whenever possible (Trtovac et al.). Books were mostly found in Belgrade national and university libraries and private libraries of project participants. Metadata were assigned to each novel of SrpELTeC, including data on their authors and publication, data on editions used for the collection, data concerning balance criteria, as well as data on institutions and individuals acknowledged for their help with the production of SrpELTeC (Krstev, "Serbian Part"). The metadata enables analyses of, for instance, titling practices of Serbian narrative literature in the years 1840–1920 (Patras et al.). In order to enable efficient metadata exploration, metadata are not only prepared as Text Encoding Iinitiative novel headers and Comma Separated Values files, but are also stored in Wikidata (Ikonić Nešić et al., "Serbian ELTeC"). This will be illustrated below.

The digitization pipeline

The pipeline for the digitization process is presented in Figure 1. Scanning and optical character recognition of a chosen work (step 2) was followed by the manual correction and annotation (steps 3 and 4) in which a number of volunteers helped. The basic annotation of chapters, paragraphs, footnotes, parts in foreign language, parts in italics, or otherwise highlighted parts was introduced following XML/TEI specification accepted for the whole ELTeC collection, the level-1 annotation (https://distantreading.github.io/Schema/eltec-1.html). At level-1, besides some basic TEI structural element, namely <front>, <body>, <back>, <div>, <head>, <p>, <milestone>, and <pb>, some textual elements were introduced as well: <hi>, <foreign>, and <title>.

All these tags were manually introduced during text reading and correction, except for <p>, which was introduced automatically.
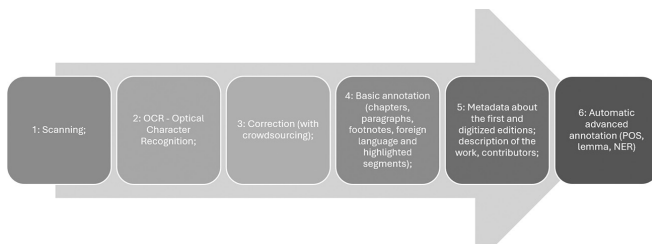
Figure 1: The digitization pipeline for the novels included in SrpELTeC.

Additionally, each novel was equipped with a TEI header with the following obligatory XML elements: <fileDesc> including <titleStmt>, <extent> (size in the number of pages and words), <publicationStmt> (availability and licensing), and <sourceDesc> (source[s] from which it was derived, with the obligatory information about the first edition whether it was used or not for SrpELTeC), <profileDesc> including <langUsage> (language[s] the text was written in), and <textDesc> (text characteristics that serve to check balance criteria of the whole sub-collection). The element <revisionDesc> was added to record all changes made to the file.

Having consistent sub-collection headers enabled the production of various statistics, such as the gender and age of each author (in the time of the publication) as well as statistics related to publishers, publication places, and so on (Krstev, "Serbian Part").

Automatic annotation of the SrpELTeC Level-2 collection

The ELTeC collection is multi-layered, as the level-2 is built upon level-1 and contains sentence segmentation elements <s>, token elements <w> for words, and <pc> for punctuation. The mandatory attributes for a word token are part-of-speech (@pos), lemma (@lemma), and information about space after a token (@join); since space is not considered a token, its presence or absence after a token is indicated with this attribute. Added to the general XML attribute for the unique identification (@xml:id) is the optional attribute for the more detailed morphosyntactic description (@msd).

The annotation pipeline (Stanković et al., "Annotation") built upon various language resources and tools for Serbian was designed and developed within the JeRTeh Language Resources and Technologies Society. Sentence boundaries were recognized, and sentences were accordingly delimited between <s> and </s> tags using the Serbian specific Unitex transducer (Krstev, *Processing*). The next step was named entity recognition, for which the rule- and lexicon-based SrpNER system was used to enable us to recognize different classes of NEs, such as dates, time, monetary and measurement expressions, geopolitical and personal names, events, and organizations (Krstev et al., "System"). The level-2 tagset contained the following seven classes: PERS, ROLE, LOC, ORG, DEMO, EVENT, and WORK. To map SrpNER using more detailed tags to the corresponding ones from this tagset (Šandrih et al.), various existing and newly developed NER-related tools were integrated into the NER&Beyond online platform (Šandrih Todorović et al.).

TXM tool (Heiden) was used for lemmatization and POS-tagging, adding new information to each token while keeping the existing XML structure intact. The parameter file for TreeTagger (Schmid) was used for the part-of-speech tagging and lemmatization within TXM. The TreeTagger model for the Serbian SrpKor4TaggingTreeTagger (https://live.european-language-grid.eu/catalogue/ld/9296) using the Universal Dependencies tagset (https://universaldependencies.org/u/pos/) was trained on a dataset created from several annotated Serbian texts, SrpKor4Tagging (https://live.european-language-grid.eu/catalogue/corpus/9295). TreeTagger also requires a lexicon and a list of open classes for the training procedure. For this purpose, Serbian morphological dictionaries SrpMD (Krstev, *Processing*) were used to produce the lexicon SrpMD4Tagging in the required format. More about the pipeline can be found in Stanković et al. ("Annotation").

Some statistics about SrpELTeC

The 100 novels of the core SrpELTeC collection were written by 66 authors: 62 male authors wrote 92 novels, while four female authors (Isidora Sekulić, Jelena Dimitrijević, Draga Gavrilović, and Milica Janković) wrote eight novels. One author (Jaša Ignjatovic) is represented in SrpELTeC with five novels, 12 authors are represented with three novels, 6 authors with two novels, and 47 authors with one novel.

Short novels prevail in SrpELTeC, as 55 texts in the collection have 10,000–50,000 words, followed by 39 medium-length novels

(50.000–100.000 words), while the collection contains only six long novels. The Serbian sub-collection contains 38 canonical novels with a high number of reprints, while the remaining 62 novels were published few times, many of them only once. The first period, 1840–1859, is represented by only two novels, and 18 novels were first published in the years 1860–1879, while the last two periods, 1880–1899 and 1900–1920, are both represented by 40 novels.

The core SrpELTEC contains 4,931,503 words distributed in 368,156 sentences and 149,522 paragraphs across 20,851 pages and 2,329 chapters. The collection also contains 518 quoted segments, 2,873 verses, 853 footnotes, 840 cited works, and 754 phrases in foreign languages.

The average number of words per paragraph is 40, while the average number of words per sentence is 14. The novel with the longest average sentence length, namely 26 words, is *Zločin jedne svekrve* (The Crime of a Mother in Law), while the shortest sentences were used in the novel *Hajduk Stanko* (Haiduk Stanko), the average length being seven words (Stanković et al., "Distant").

A total of 250,340 named entities were tagged in SrpELTeC. Personal names are the most frequent category (PERS 124,338), followed by the persons' professions, positions, or titles (ROLE 82,483), and geopolitical and other urban names (LOC 21,318). The frequencies of other categories are as follows: names of organizations (ORG 1,539), names of inhabitants and ethnic groups, including adjectives derived from geopolitical names (DEMO 19,409), events (EVENT 769), and titles of artistic or professional works (WORK 484). The most frequent first names for men are *Miloš*, *Boža*, *Milan*, *Radiša*, *Stojan*, *Micko*, and *Pera*, while the most frequent names for women are *Jelica*, *Mara*, *Ljubica*, *Darinka*, *Ana*, and *Danica*. The most frequent geopolitical names in SrpELTeC are *Srbija* 'Serbia,' *Beograd* 'Belgrade,' and *Kosovo*. Frequently mentioned countries include *Bosna* 'Bosnia,' *Rusija* 'Russia,' *Turska* 'Türkiye,' *Austrija* 'Austria'; the most frequent inhabited places are *Beograd*, *Carigrad* 'Istanbul,' *Golubac*, *Beč* 'Vienna,' *Niš*, *Skoplje* 'Skopje,' while the most frequent rivers are *Dunav* 'Danube,' *Morava*, *Sava*, and *Drina*.

Some examples of SrpELTeC usage

In this section we will give some illustrative examples of the use of SrpELTeC in the various domains of research. We will start by demonstrating how comprehensive electronic dictionaries of the Serbian

language can be used to help analyze the content of SrpELTeC. This will be followed by a presentation of a textometric analysis, and finally we will show some different ways of implementing the concept of linked data with the SrpELTeC corpus.

The SrpELTeC corpus was processed using the Unitex/GramLab Multilingual Corpus Processing Suite with the help of a system of electronic morphological dictionaries for the Serbian language (Krstev, "Processing"). One of the research topics was related to eating habits and the language of food. The examples from the corpus show not only what kinds of food were used in the narratives included in SrpELTeC, but also the attitude of the narrated local population towards food, as well as elements of their taste in food as part of their collective identity (Vitas).

One topic of research dealt with the use of alcoholic drinks. Contrary to expectations, *vino* 'wine' was more frequently mentioned than *rakija* 'rakia,' namely 1,181 versus 637 occurrences. If various varieties of wine and rakia are taken into consideration, the ratio remains similar: 1,263 versus 820. Also, whereas less than ten specific brands of *rakia* were mentioned, such as *šljivovica* 'slivowitz,' wine came in more than 30 varieties, including, for instance, *malvasija* 'malvasia.' This type of queries was enabled by semantic markers assigned to entries in the SrpMD e-dictionaries, such as +Drink for drinks. The varieties of wines were categorized in broad categories: *belo vino* 'white wine' (12), *crno vino* 'red wine' (21), and *ružica* or *crveno vino* 'rosé' (2). This shows that red wine was traditionally named *crno vino* 'lit. black wine,' contrary to the current tendency to rename it into *crveno vino* under the influence of English and French. It was also interesting to note that champagne was not unknown to (some) Serbians at that time: it was mentioned 28 times under different names.

Another topic of interest was the evidence of literary works read at the time of SrpELTeC. This information could be retrieved due to the information about the domain assigned to e-dictionaries' personal name entries (in this case, the marker <Dom=Lit> in e-dictionaries). The tag <title> annotating the cited works was also used. Figure 2 presents a few concordance lines: the first column gives the author's name, the novel's title, and publication year, the second column gives the retrieved concordance segment (in Serbian), while its automatic English translation is given in the third column.

Šišković, Dragomir: *Jedan od mnogih*, 1920	Čitala je bez izbora i razbora, sve ruske pisce koji su joj došli do ruku. **Tolstoj, Gogolj, Dostojevski, Gorki, Arcibašev [Sanjin]** ređali su se pred njenim očima. Nije prosto znala šta pre da čita.	She read all the Russian writers she could get her hands on without choice or judgment. **Tolstoy, Gogol, Dostoevsky, Gorky, Artzybashev [Sanjin]** lined up before her eyes. She simply did not know what to read first.	
Jevtić, Stevan J.: *Danica* 1891	Ti neprestani sanjao postao si sad najedanput drugi **Demokrit**. Od mene se opet načinio neki novi Epaminonda, „špikovan" **Hajneovom** melanholijom i **Bajronovom** mizantropijom	You never stopped dreaming, now you have suddenly become another **Democritus**. A new **Epaminondas** was made of me again, "spiked" with **Heine's** melancholy and **Byron's** misanthropy.	
Komarčić, Lazar: *Prosioci*,1905	Znao je „Branka" napamet , i iz „Gorskog Venca" mogao je čitave strane odeklamovati . On vam je mogao, s kraja na kraj , ispričati „Sirotu Bosiljku", „Alpisku Pastirku", „Kasiju Caricu"; „Ezopove Basne" i Dositejeva naravoučenija, tako isto. On je u narodu važio kao neka retka pojava od bistrine, upravo kao neki — seljak filosof.	He knew "**Branka**" by heart, and he could recite whole pages from "**The Mountain Wreath**". He could tell you, from end to end, "**Poor Bosiljka**", "**Alpine Shepherdess**", "**Cassia the Empress**"; "**Aesop's Fables**" and **Dositej's moral lessons**, as well. He was regarded among the people as a rare phenomenon of clarity, precisely as a peasant philosopher.	
Dimitrijević, Jelena: *Nove*, 1912	One tri nove nađoše u Fatminoj biblioteci mnoge svoje mile poznanike, pored **Šatobriana** i **Lamartina** — **Gij de Mopasana, Marsela Prevoa, Pola Buržea**, i druge. Nađoše i **Flobera**, i **Lotija**...	The three newcomers found many of their dear acquaintances in Fatima's library, in addition to **Chateaubriand** and **Lamartine**—**Guy de Maupassant, Marcel Prévost, Paul Bourget**, and others. They also found **Flaubert** and **Loti**...	

Figure 2: Examples of literary works and their authors mentioned in the SrpELTeC corpus.

Figure 3 lists the most frequently mentioned authors in SrpELTeC, ordered by the frequency: the author (or the title of the work in case of anonymous authors) is given in the first column, the novel that mentions the author is given in the second column, the third column gives an example sentence from that novel, while the fourth column gives the automatic English translation of the sentence. One should note that Goethe, as the author of *Werther*, was by far the most frequently mentioned author, due to the fact that SrpELTeC includes a kind of parody of *Werther*, namely the novel (or long story) *Verter (Werther)* by Laza Lazarević.

Goethe (and his *Werther*)	Lazarević, Laza: *Verter*, 1881	Pravi proizvod nemačke poezije. **Gete** video **Rusovljevu** „Novu Eloizu", pa i on napisao „Vertera".	A true product of German poetry. **Goethe** saw Rousseau's "**The New Heloise**", so he also wrote "**Werther**".
Dositej (Obradović)	Gavrilović, Draga: *Babadevojka*, 1887	Još mi je bilo slobodno čitati neke prevode od **Fenelona** i **drugih stranih pisaca**, kao i **bibliju** i zlatoustove besede, od **Dositeja Obradovića**, koje su bile još u rukopisu .	I was still free to read some translations by **Fénelon** and other foreign writers, as well as the **Bible** and Zlatoustove besedes, by **Dositej Obradović**, which were still in manuscript.
The Bible	Kosta Barunčić: *Pastir kralj ili Oslobođenje Srbije*, 1879	može pesnik naći u dičnoj Srbiji verne obrasce običaja iz **Biblije** i **Omira**.	can a poet find in beautiful Serbia true patterns of customs from the **Bible** and **Homer**.
Tolstoy	Janković, Milica: *Pre sreće*, 1918	Čika kaže da je **Mopasan** možda jači, ali je **Tolstoj** bolji.	Uncle says that **Maupassant** may be stronger, but **Tolstoy** is better.
Plutarch	Popović-Šapčanin, Milorad: *Sanjalo*, 1888	Učitelj Maksa ode u sobu da ostavi svoga **Plutarha** u školski orman.	Max's teacher went to the room to leave his **Plutarch** in the school closet.
Heine	Matijašević, Stevan: *Grofica Agneša Janković*, 1897	— Čitali ste **Hajnea**,... ja proklinjem čas kada sam u ruke uzeo njega i *Vertera*.	— You read **Heine**, ... I curse the hour when I took him and **Werther** in my hands.
Hugo	Ljubiša Branković: *Pred Zoru*, 1878	Razgovarahu se o *Higovim „Jadnicima"*.	They talked about Hugo's "**Les Misérables**".
	Stevan Sremac: *Pop Ćira i pop Spira*, 1894	— Ако желите **Работнике на мору** или	— If you want **Workers at Sea** or
Dumas	Jakov Ignjatović	— Čitala sam od *Dima „Draj Musketire"*	— I read "**Die drei Musketiere**" by Dumas
	Svetolik Ranković: *Seoska* učiteljica	— Јеси читала **Монте-Христо**?	— Have you read **Monte Cristo**?

Figure 3: The most frequently mentioned authors in SrpELTeC.

Textometry

Besides adding different morphosyntactic annotations, the TXM tool (Heiden) can be used to calculate various text statistics (Krstev et al., "Analysis"), including textometry as a powerful technique for the analysis of text corpora. TXM provides the following qualitative tools: KeyWord in Context concordances of word patterns based on Corpus Query Language; word pattern frequency lists based on tokens, lemmas, parts-of-speech, or structural annotations, including named entities; and word pattern progression graphics. The quantitative analysis tools are based on R packages: factorial correspondence analysis, cluster analysis, specific word patterns analysis, and collocations analysis.

One of the research questions we tackled by using the TXM tool were typical professions of female characters as evidenced in SrpELTeC and their use over time and by different authors. We were able to do that thanks to existing annotations of professions, positions, and titles (tag <role>), and we also had the possibility to include structural tags in queries when using TXM. We began by analyzing the 100 most frequently used roles in the collection and we noticed that 83 refer to men, while 17 refer to women. Among these most frequent roles of men there are numerous professions, including *pop* 'priest,' *kapetan* 'captain,' *doktor* 'doctor,' *učitelj* 'teacher,' *seljak* 'peasant,' *vojnik* 'soldier,' *kaluđer* 'monk,' *sluga* 'servant,' *majstor* 'artisan,' and *pisar* 'registrar,' while roles ascribed to women include only two professions, *sluškinja* 'servent maid' and *učiteljica* '(woman) teacher.' Next, we selected all professions ascribed to women and grouped them in TXM in 7 broad categories, as presented in Figure 4. The figure shows that professions that belong to the category of (manual) workers are the most frequent ones, followed by teachers, where (nursery) governesses are mentioned besides elementary school teachers.

women's professions per categories

category	count
workers	657
teachers	287
artists	159
peasants	126
medical workers	112
students	85
clerks	79

workers:
sluškinja
kuh(v)arica
služavka
aščika
izmećarka

teachers:
učiteljica
učiteljka
nastavnica
vaspitateljka
profesorka

artists:
glumica
pevačica
igračica
sviračica
baletkinja

peasants:
seljanka
beračica
čobanica
pastirka
žetelica

medical workers:
bolničarka
doktorka
Babica

students:
učenica
pitomica
studentkinja
maturantkinja
đačica

clerks:
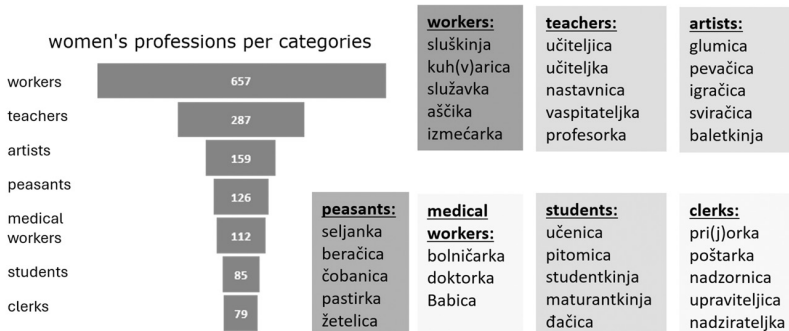pri(j)orka
poštarka
nadzornica
upraviteljica
nadzirateljka

Figure 4: Professions of female characters per categories in SrpELTeC.

Implemented in TXM, progression graphs enabled us to gain insight into the changes in professions of female characters during the period covered by SrpELTeC (see Figure 5). One can observe that workers' professions for female characters were used more or less uniformly throughout the collection. The same holds for peasants, but less frequently. On the other hand, women teachers appeared later (there is a steep rise after one million corpus words), which is due to three novels written by Draga Gavrilović in which teacher maids are the main characters. After that, teacher maids did not occur in novels for a long time (hence the straight line after approximately one million words). There is a slight rise of women artists, clerks, and medical workers by the end of the period; however, in the same period, female students did not seem to be interesting to novelists.
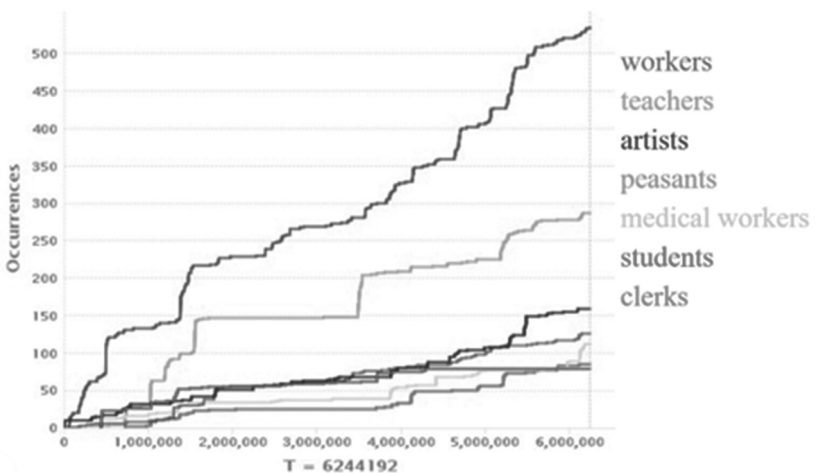
Figure 5: Changes in professions of female characters over time in SrpELTeC; the numbers of corpus words are given on the *x*-axis, but as the novels in SrpELTeC are sorted according to the year of the first publication, this axis also refers to time.

With the textometric approach to the corpus it is possible to recognize some specific entities or characteristics along with their high or low representation in certain parts of the corpus. The specificity score based on the hyper-geometric distribution shows the probability of a lexical unit occurring in a particular part of the corpus. The TXM also provides a graphic representation of the specificity distribution of the selected units. Specificity score values higher (positive) or lower (negative) than expected express a more or less represented lexical unit or pattern

(Heiden). This is illustrated in Figure 6. The corpus was separated in two partitions: works written by women and works written by men. The occurrence of professions of female characters in the two partitions was analyzed, showing that women teachers tended to be written about by female authors, while women peasants were more often thematized by male authors.
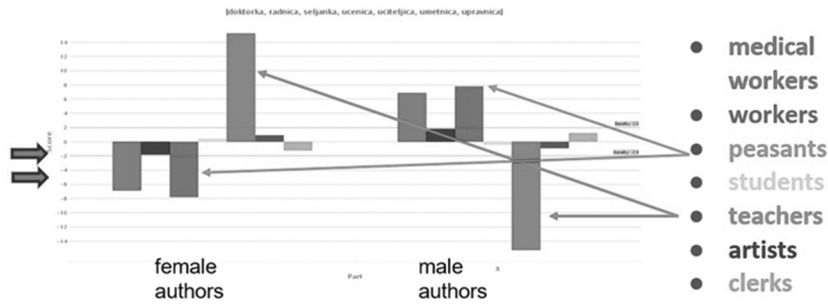
Figure 6: The specificity of the appearance of professions of female characters by the gender of authors in SrpELTeC.

Having in mind that SrpELTeC will be used for various kinds of lexical and linguistic research, the use of diverse tools and methodologies was provided to make it available through various channels, thus hopefully meeting the needs of different types of users. Three platforms on which these novels are published, namely, Udaljeno čitanje, Aurora, and Sketch Engine, are presented in Stanković et al. ("SrpELTeC"). The Udaljeno čitanje platform is intended for readers who would like to see the original print as a picture while reading the digitized version. The Aurora portal provides researchers of Serbian literature and other interested users with a detailed inspection of the novel's vocabulary, enabling text browsing, concordances, and frequency lists. The Sketch Engine (Kilgarriff et al., "Sketch Engine"; Kilgarriff et al., "Sketch Engine: Ten Years On") is a platform for corpora management and exploration, as well as for analyzing texts to identify what is typical in a language and what is a rare, unusual, or emerging usage. A NoSketch Engine node is installed and maintained by JeRTeh, offering access to several monolingual and bilingual corpora. The SrpELTeC corpus can be freely accessed and searched using Corpus Query Language, and registration is possible without any special conditions.

Linguistic Linked Open Data

With the development of Linguistic Linked Open Data, interest in formalizing the bridge between digital humanities and web-centered linguistic data science has been intensified, although mainly with a focus on lexical data. In the digital humanities community, TEI XML-based standards represent the prototypical publishing approach and have been criticized for not establishing a sufficient degree of interoperability and synchronization with formal semantics and web standards such as RDF and OWL.

Preparing the data on the SrpELTeC novels for Wikidata and linking Wikidata to various applications started as a manual process. The opportunity to speed up this process was seen in using information already encoded in the TEI header of each novel. Data on 700 novels in seven languages from the ELTeC collection were introduced in Wikidata as part of the WikiELTeC project (Ikonić Nešić et al., "From ELTeC"). Since the automation of the process of data preparation and import was envisaged, the different solutions were analyzed, and finally the synergy of OpenRefine and QuickStatements tools was chosen as the best choice (Ikonić Nešić et al., "Serbian ELTeC"). WikiELTeC was semi-automatically populated from <TeiHeader> using OpenRefine, QuickStatements, and custom-made procedures; after the extraction of metadata from headers, the mapping with Wikidata schema was defined in OpenRefine and predicates (properties) that connected subjects and objects in RDF triples were specified in the OpenRefine table header. Each statement for a subject has a property and a value that can become a Wikidata item, an external URL, or a literal (string). After consolidation in OpenRefine, RDF triplets were imported in Wikidata using QuickStatements.

Several groups of data were added or improved, including authors, publishers, metadata about novels, the novels' printed and electronic editions, including SrpELTeC, main characters and their relations, and novels' settings. As a result, 71 authors and 120 novels from the core and the extended SrpELTeC collections were represented in Wikidata, comprising, together with associated items for first editions and digital SrpELTeC editions, approximately 3,500 statements. Metadata about the novels were automatically imported, while main characters and their relations were added manually by volunteers, mostly students at the University of Belgrade and members of JeRTeH.

Each novel's metadata item is linked with an appropriate metadata instance for electronic edition (Q59466853), first edition (Q10898227), print edition (Q59466300), and digital edition (Q1224889), using the

property (P747: has edition or translation), and every item of an edition must be connected with a corresponding item for a novel with the inverse property (P629: edition or translation of). The list of all properties is documented in WikiProject_ELTeC.

A more detailed description of SrpELTeC is provided by introducing main characters and the places mentioned in the novels. This was done within the Serbian WikiProject WikiELTeC (https://sr.wikipedia.org/wiki/Википедија:Википројекат_WikiELTeC), in the scope of which numerous SPARQL queries were provided for the retrieval, analysis, and diverse visualization options of the SrpELTeC data. Main characters of the novels were linked not only with their family and social relations, but also with actors who interpreted them in films and TV series. The understanding of the corpus got a new dimension with provided wikidata supported by SPARQL queries. For example, browsing novels written by Borisav Stanković (Q370392), one can start with the following query and continue by browsing the graph as presented in Figure 7:

```
#defaultView:Graph
SELECT DISTINCT ?author ?authorLabel ?authorImage ?novel ?novelLabel ?image
WHERE {
VALUES ?author { wd:Q370392}
?novel wdt:P50 ?author;
wdt:P18 ?image;
wdt:P747 ?edition.
?edition wdt:P1433 ?collection.
OPTIONAL {?author wdt:P18 ?authorImage}
SERVICE wikibase:label
{ bd:serviceParam wikibase:language «sr,[AUTO_LANGUAGE],en». }}
```
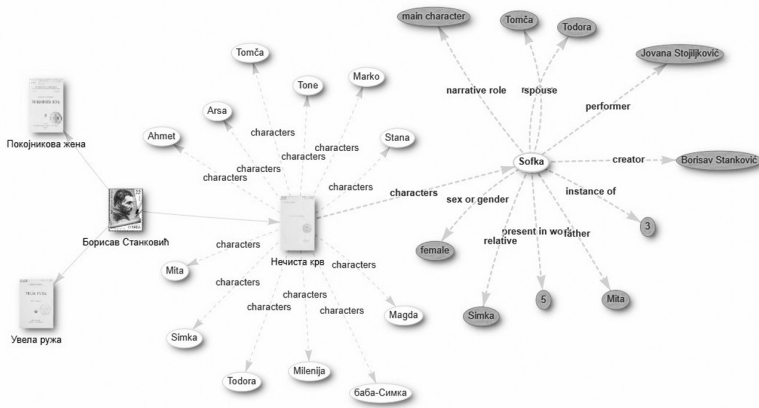
Figure 7: The knowledge graph for one author (Borisav Stanković) obtained by the presented query (at the top), his novels, and one novel's (*Nečista krv* [Impure Blood]) related data.

Introducing characters and places mentioned in novels in WikiData enables linking their occurrences in the texts of SrpELTeC. The INCEPTION tool was used for this, where one can start with the manual annotation while the system learns to make predictions as annotation advances. An example is presented in Figure 8: rivers *Nišava* (Q583062) and *Dunav* (Q1653) are annotated as locations (LOC) and linked to their corresponding Wikidata QIDs. The same figure demonstrates that professions and titles, annotated with the ROLE label, are also linked to WikiData, namely *trgovac* 'merchant' (Q215536) and *pisac* 'writer' (Q36180). Current activities are focused on training a model for automatic linking of recognized named entities.
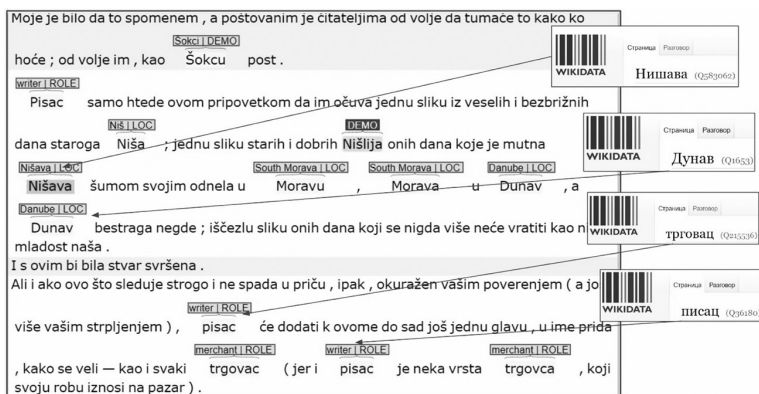
Figure 8: A text in the INCEPTION platform with examples of entity linking with Wikidata.

The linked data version of SrpELTeC using NLP Interchange Format (NIF) was produced within the COST Action NexusLinguarum—European Network for Web-centred Linguistic Data Science (CA18209). NIF is designed to facilitate the integration of NLP tools in knowledge extraction pipelines; it provides support for part-of-speech tagging, lemmatization, and entity annotation, enabling ELTeC level-2 layer transformation (Stanković et al., "Towards"). Several ontologies were consulted to use equivalents of named entity types: OLIA, DBpedia, Wikidata. For Wikidata, the following mapping was specified: wd:Q5 (PERS), wd:Q7884789 (LOC), wd:Q43229 (ORG), wd:Q1656682 (EVENT), wd:Q28640 (ROLE), wd:Q217438 (DEMO), and wd:Q386724 (WORK).

We should note that recognized named entities have not been linked with Wikidata or DBpedia items yet, as they are only marked and classified into one of seven predefined types. The Apache Jena Fuseki server was used for testing the Serbian ELTeC NIF corpus SPARQL at the JeRTeh site (http://fuseki.jerteh.rs/#/dataset/SrpELTeC/query). A SPARQL query presented in Figure 9 illustrates the retrieval of the most frequently used nouns in novels written by Jakov Ignjatović (wd:Q570913), that is, *kuća* 'house' (275), *otac* 'father' (208), *dan* 'day' (144), *mati* 'mother' (140), *godina* 'year' (127), and *ruka* 'hand' 123.

Figure 9: The Fuseki node with the SrpELTeC edition as linked data using NIF: an example of a query.

Conclusion

It is our belief that the developed platforms and different formats and editions of digital versions of novels will contribute to raising the visibility of SrpELTeC as a valuable Serbian language resource for linguists and literary scholars. Additionally, they will shed light on a relatively unknown period of Serbian literary history. The SrpELTeC corpus can help paint a portrait of life in Serbia in the second half of the nineteenth century and early twentieth century. Besides investigating eating habits, literature read, and professions practiced, more complex issues, such as the position of women in the Serbian society, education of children, cultural habits, medical treatments, means of travel, or attitudes towards so-called others, could be explored, opening possibilities for further analysis of the corpus. To this end, multi-word expressions and other additional layers will be added to the collection. Finally, valuable data should be added to the header information by literary scholars, including the dialects and pronunciations used in novels, the backgrounds of the authors (e.g., whether an author's birthplace was in the Turkish Empire or Austro-Hungarian Empire), and the generic features of the novels.

WORKS CITED

Burnard, Lou, et al. "In Search of Comity: TEI for Distant Reading." *Journal of the Text Encoding Initiative*, vol. 14, 2021, pp. 1–9, https://doi.org/10.4000/jtei.3500. Accessed 24 Jan. 2024.

Deretić, Jovan. *Istorija srpske književnosti*. Belgrade, Nolit, 1983.

Deretić, Jovan. *Srpski roman 1800–1950*. Belgrade, Nolit, 1981.

Heiden, Serge. "The TXM Platform: Building Open-source Textual Analysis Software Compatible with the TEI Encoding Scheme." *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation. Vol. 2. No. 3*, edited by Ryo Otoguro et al., Institute for Digital Enhancement of Cognitive Development, Waseda University, 2010, pp. 389–398, https://aclanthology.org/Y10-1044. Accessed 24 Jan. 2024.

Ikonić Nešić, Milica, et al. "From ELTeC Text Collection Metadata and Named Entities to Linked-data (and Back)." *Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference*, edited by Thierry Declerck et al., European Language Resources Association, Paris, 2022, pp. 7–16, https://aclanthology.org/2022.ldl-1.2/. Accessed 24 Jan. 2024.

Ikonić Nešić, Milica, et al. "Serbian ELTeC Sub-Collection in Wikidata." *Infotheca*, vol. 21, no. 2, 2021, pp. 60–87, https://doi.org/10.18485/infotheca.2021.21.2.4. Accessed 24 Jan. 2024.

Kilgarriff, Adam, et al. "The Sketch Engine." *Proceedings of the Eleventh EURALEX International Congress*, Université de Bretagne-Sud, Faculté des lettres et des sciences humaines, 2004, pp. 105–115, https://euralex.org/publications/the-sketch-engine/. Accessed 24 Jan. 2024.

Kilgarriff, Adam, et al. "The Sketch Engine: Ten Years On." *Lexicography*, vol. 1, no. 1, 2014, pp. 7–36, https://doi.org/10.1007/s40607-014-0009-9. Accessed 24 Jan. 2024.

Krstev, Cvetana. *Processing of Serbian: Automata, Texts and Electronic Dictionaries*. Belgrade, Faculty of Philology of the University, 2008.

Krstev, Cvetana. "The Serbian Part of the ELTeC Collection Through the Magnifying Glass of Metadata." *Infotheca*, vol. 21, no. 2, 2021, pp. 26–42, https://doi.org/10.18485/infotheca.2021.21.2.2. Accessed 24 Jan. 2024.

Krstev, Cvetana, et al. "A System for Named Entity Recognition Based on Local Grammars." *Journal of Logic and Computation*, vol. 24, no. 2, 2014, pp. 473–489.

Krstev, Cvetana, et al. "Analysis of the First Serbian Literature Corpus of the Late 19th and Early 20th Century with the TXM Platform." *DH_BUDAPEST_2019*, Eötvös Loránd University, Centre for Digital Humanities, 2019, pp. 36–37.

Milisavac, Živan, editor. *Pripovedači.* Novi Sad / Belgrade, Matica srpska / Srbska knjizhevna zadruga, 1972.

Moretti, Franco. "Conjectures on World Literature." *New Left Review*, vol. 1, 2000, pp. 54–68.

Odebrecht, Carolin, et al. *European Literary Text Collection (ELTeC): April 2021 Release with 14 Collections of at Least 50 Novels (v1.1.0).* Zenodo, 2021, https://doi.org/10.5281/zenodo.4662444. Accessed 24 Jan. 2024.

Patras, Roxana, et al. "Thresholds to the 'Great Unread': Titling Practices in Eleven ELTeC Collections." *Interférences Littéraires / Literaire Interferenties*, vol. 25, 2021, pp. 163–187, http://interferenceslitteraires.be/index.php/illi/article/view/1102. Accessed 24 January 2025.

Schmid, Helmut. "Improvements in Part-of-speech Tagging with an Application to German." *Natural Language Processing Using Very Large Corpora*, edited by Susan Armstrong et al., Springer Netherlands, 1999, pp. 13–25.

Schöch, Christof, et al. "Creating the European Literary Text Collection (ELTeC): Challenges and Perspectives." *Modern Languages Open*, vol. 1, 2021, https://doi.org/10.3828/mlo.v0i0.364. Accessed 24 Jan. 2024.

Stanković, Ranka, et al. "Annotation of the Serbian ELTeC Collection." *Infotheca*, vol. 21, no. 2, 2021, pp. 43–59, https://doi.org/10.18485/infotheca.2021.21.2.3. Accessed 24 Jan. 2024.

Stanković, Ranka, et al. "Distant Reading in Digital Humanities: Case Study on the Serbian Part of the ELTeC Collection." *LREC 2022 Conference Proceedings*, edited by Nicoletta Calzolari et al., European Language Resources Association, Paris, 2022, pp. 3337–3345, https://aclanthology.org/2022.lrec-1.356/. Accessed 24 Jan. 2024.

Stanković, Ranka, et al. "SrpELTeC on Platforms: Udaljeno čitanje, Aurora, noSketch." *Infotheca*, vol. 21, no. 2, 2021, pp. 136–153, https://doi.org/10.18485/infotheca.2021.21.2.7. Accessed 24 Jan. 2024.

Stanković, Ranka, et al. "Towards ELTeC-LLOD: European Literary Text Collection Linguistic Linked Open Data." *Language, Data and Knowledge 2023*, edited by Sara Carvalho et al., NOVA CLUNL, Lisbon, 2023, pp. 180–191.

Šandrih, Branislava, et al. "Development and Evaluation of Three Named Entity Recognition Systems for Serbian–The Case of Personal Names." *RANLP 2019: Natural Language Processing in a Deep Learning World: Proceedings*, edited by Galia Angelova et al., INCOMA Ltd., Shumen, 2019, pp. 1060–1068, https://aclanthology.org/R19-1122/. Accessed 24 Jan. 2024.

Šandrih Todorović, Branislava, et al. "Serbian NER&Beyond: The Archaic and the Modern Intertwinned." *RANLP 2021: Deep Learning for Natural Language Processing Methods and Applications*, edited by Galia Angelova et al., INCOMA Ltd., Shumen, 2021, pp. 1252–1260, https://aclanthology.org/2021.ranlp-1.141/. Accessed 24 Jan. 2024.

Trtovac, Aleksandra, et al. "The Serbian Part of the ELTeC–From the Empty List to the 100 Novels Collection." *Infotheca*, vol. 21, no. 2, 2021, pp. 7–25, https://doi.org/10.18485/infotheca.2021.21.2.1. Accessed 24 Jan. 2024.

Vitas, Duško. "From Onions to Champagne–Food and Drink in the SrpELTeC Corpus." *Infotheca*, vol. 21, no. 2, 2021, pp. 88–118, https://doi.org/10.18485/infotheca.2021.21.2.5. Accessed 24 Jan. 2024.

SrpELTeC: srbski literarni korpus za oddaljeno branje

Ključne besede: digitalna humanistika / srbska književnost / besedilni korpusi / oddaljeno branje / povezani podatki / prepoznavanje imenskih entitet / besedilna analitika

V članku je predstavljen korpus SrpELTeC, ki je nastal v okviru COST akcije Distant Reading for European Literary History (CA16204). Vsi romani v SrpELTeC-u so bili izbrani, pripravljeni in označeni po skupnih načelih, ki veljajo za vse jezikovne zbirke v European Literary Text Collection (ELTeC). Opisani so izzivi in rešitve pri pripravi SrpELTeC. Vsi romani so bili ročno kodirani skladno s parametri XML-TEI in opremljeni z bogatimi metapodatki ter strukturnimi opombami. Avtomatično označevanje je vključevalo oblikoskladenjske oznake, lematizacijo in imenske entitete, ki so temeljile na virih za obdelavo naravnega jezika, ki jih je razvilo in jih vzdržuje Društvo za jezikovne vire in tehnologije JeRTeh. Integracija SrpELTeC z Wikidata je bila podprta z nizom poizvedb SPARQL za pridobivanje metapodatkov z različnimi možnostmi vizualizacije. V okviru nedavnih dejavnosti v okviru COST akcije NexusLinguarum – European Network for Web-centred Linguistic Data Science (CA18209) je bila ustvarjena povezana podatkovna različica SrpELTeC z uporabo NLP izmenjevalnega formata. Vse različice SrpELTeC so prosto dostopne pod licenco CC-BY.

Rhymes and Syntax: A Morpho-Syntactic Analysis of Czech Poetry

Silvie Cinková, Petr Plecháč, Martin Popel

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Malostranské nám. 25, 118 00 Praha 1, Czechia; Czech Academy of Sciences, Institute of Czech Literature of the CAS, Na Florenci 3/1420, 110 00 Praha 1, Czechia
https://orcid.org/0000-0003-4526-3915
cinkova@ufal.mff.cuni.cz

Czech Academy of Sciences, Institute of Czech Literature of the CAS, Na Florenci 3/1420, 110 00 Praha 1, Czechia
https://orcid.org/0000-0002-1003-4541
plechac@ucl.cas.cz

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Malostranské nám. 25, 118 00 Praha 1, Czechia
https://orcid.org/0000-0002-3628-8419
popel@ufal.mff.cuni.cz

A linguistically informed distant reading presupposes an adequate performance of Natural Language Processing tools. This article describes our evaluation of the UDPipe parser on a manually annotated sample of nineteenth-century Czech poetry in the following steps: (1) creation of a documented data set for this domain (poetry, nineteenth century, Czech); (2) domain-specific annotation decisions; (3) error analysis. The sample consisted of 29 randomly selected poems which were first automatically tagged and parsed with the UDPipe parser and then manually checked word by word. The following features were checked: word segmentation (chunking), lemmatization, part of speech assignment, assignment of more fine-grained morphological details, the position in the syntactic dependency tree (selection of the syntactic parent), as well as the label of the syntactic relation between the word and its parent. The findings were analyzed. The most typical parser errors are associated with complex noun phrases that contain other noun(s) as modifier(s), especially when these occur in a poetry-specific word order, that is, preposed to the governing noun. On the other hand, neither archaic orthography nor neologisms posed substantial issues.

Keywords: Czech poetry / distant reading / text corpora / Universal Dependencies / natural language processing / treebanks

65

Introduction

Some text-mining use cases benefit from reaching beyond the bag-of-words approach to extraction of lexical or grammatical patterns.[1] This is made possible by automatic morphological tagging and syntactic parsing wherever such a tool is available for the given language and achieves adequate performance within the given domain. Most parsers are run with language models that have been trained on contemporary non-fiction, and their performance is likely to decrease by the same measure that input texts deviate from those models' domains.

UDPipe, the largest Czech language model used by the best-performing Czech parser (Straka et al.), was trained on the 1990s daily Czech press (Hajič). At first glance, the main differences between this domain and that of nineteenth- and twentieth-century Czech poetry have to do with vocabulary, orthography, and word order. However, the effect of these differences on the parser performance is not predictable. The parser performance can be measured and the most typical errors can only be detected by manual annotation of a random sample and its comparison to the automatic output. While this work is time-consuming, the domain-specific annotated data could be added to the original model to increase performance on this new domain in the future—considering that this goal may turn out to require several iterations of additional annotation. In our experiments, we use the largest model, czech-pdt-ud-2.12-230717, and a smaller model based on fiction, czech-fictree-ud-2.12-230717.

Data

The data set is comprised of 29 random Czech poems from PoeTree (Plecháč et al.; Plecháč and Kolár), with a total of 6,643 tokens and 2,687 types (unique words). Most of the poems were written at the turn of the nineteenth century. About half of the represented poets belong to the Czech high-school literary canon. Most poems are rhymed. Figure 1 shows the publication dates of each poem along with its author's lifespan.
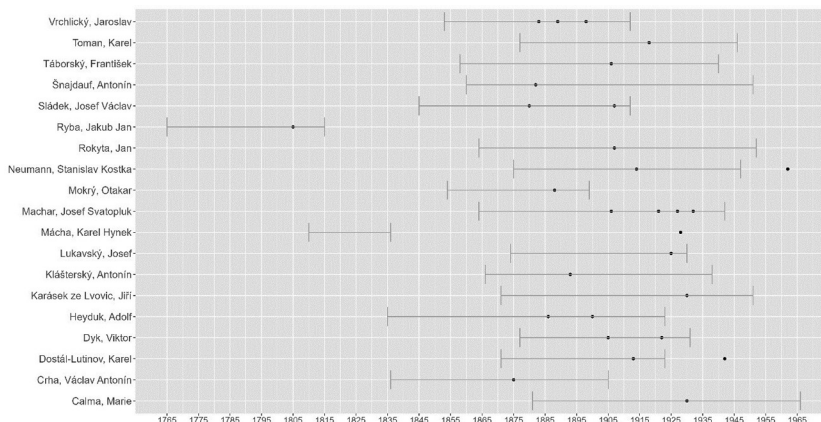
Figure 1: The PoeTree Czech sample: authors' life spans and poems' publication years.

Diachronic and stylistic language perspective

The oldest poem (1805) was written during (and in the language of) the Czech National Revival, and is therefore quite different from the later poems. Most of the nineteenth-century poems are written in somewhat modern Czech, that is, in the Czech language as it was re-established after more than a century of Germanized education and at an advanced stage of efforts to integrate the norms of a written Czech no longer in use with the spoken vernacular of the time, which was naturally perceived as low standard. The twentieth-century poems can be considered representative of (a very marked stylistic register of) truly modern Czech. The entire nineteenth century saw competing progressive as well as regressive normative trends, with the variation in poetry furthermore augmented by a rapid increase in poetic experimentation and manneristic personal style distinctions (Šlosar). Habitual modes of linguistic periodization, as a consequence, are not very helpful in the case of this poetry sample. Despite all this variation, we can still track several recurring differences between contemporary Czech prose and the language observed in this sample. This section lists a few of the resulting annotation decisions.

Spelling

Spelling variation can be found in both word stems and morphemes. In order to enable searching across different diachronic layers without altering words, we preserved the token forms while normalizing

67

lemmas, wherever possible, to contemporary spelling variants. For instance, rather than transcribing *nervosníma* as *nervózními*, we lemmatized using the current term *nervózní*. Whenever the modern equivalent was not instantly apparent or the word had undergone more substantial morphological changes, such as *s křeku* (*z keříku* 'from a bush') or *junoše* (*jinoch* 'lad') in the 1805 poem, we left the lemmas intact.

A prominent feature of Czech word formation—and one that presents a particularly difficult and longstanding obstacle for language processing—is compound function words. The compounding of prepositions with other parts of speech, especially nouns and adverbs, produces adverbs, particles, conjunctions, and prepositions that are written at times as discrete words and at others as prefixes, according to numerous rules with numerous exceptions (Osolsobě), thus posing challenges and spelling issues for Natural Language Processing—indeed, even for educated native speakers (Žižková). Many of these words can be found in the basic vocabulary, such as *na shledanou* 'good bye' and *zpočátku* 'initially.' Throughout the nineteenth century, little attention was paid to graphical word boundaries in general, although partial and mutually contradicting recommendations existed in grammar books. This had various consequences. For instance, the first generation of revivalists treated compound function words with complex rules depending mostly on the word formation type of the noun or adjective that followed (Dobrovský), while a later generation of grammarians tended to decompose them into discrete words (Kampelík).

The unmanageable spelling variations in the compound function words in our sample hampered lemma normalization. Whenever a compound function word consisted of two tokens, it was annotated as a syntactic relation between two words.

Punctuation and sentence splitting

The sample displays certain punctuation peculiarities: some poems combine the usual punctuation principle (syntactic segmentation) with the highlighting of rhetorical pauses (see Examples 1 and 2 below) in the manner of classic public speakers' speech notes (Pavel Kosek and Jana Pleskalová).

(1) Ale ty oči! oči smilnící!

(But those eyes! those fornicating eyes!)[2]

(2) Deset let už o tom píše, pan Vejr v Švandě Dudákovi.

(For a decade he has been writing about this, Mr. Vejr in Schwanda the Bagpiper.)

Even if the punctuation determines clause boundaries fairly well, sticking to the syntactic (and not rhetorical) principle, that is, to separate clauses (as well as conjuncts and appositions), sentence boundaries remain blurry. This applies both to poems without enjambement, where clause boundaries do not tend to cross verse boundaries (typically trochee verses, such as the extract from K. H. Mácha's *Prolog k pouti Krkonošské*, transl. *Prologue to the Riesengebirge Pilgrimage*, in Example 3 below), and to poems with long-winded clauses (often verses in prose, such as J. Karásek's ze Lvovic *Nad obrazem Marie Magdaleny v hradčanské Loretě*, transl. *Over the Painting of Mary Magdalene in the Hradschin Loretta*, in Example 4).

The short paratactic clauses evoke a swift narration pace and sentence boundaries do not play a role, while the syntactically long-winded clauses evoke an agitated stream of consciousness, which nevertheless unfolds within a solid syntactic scaffolding.

(3) Víc a více světnice se plní,
Hovor hlučí, kouř se z dýmek vlní;
Při stropu ho plamínku zář zlatí.

(Gradually the room is getting crowded,
The talk is loud, smoke is curling from pipes;
The glow of small flames gilds the ceiling.)

(4) V starobných ambitech, kde ztuhlá světic ctnost
V škrobených límcích španělských se vztyčuje,
Kde marně Šebestián sličný Kypící nahotu,
Drážděné šípem genitálie ukazuje,
Aby sváděl ctnostně odvrácené zraky,
Ty potměšilou vlnu ňader,
Tak měkce tajících,
Teď svůdně rozléváš,
Své tělo zase obnažuješ,

---

[2] All translations of poem samples are by Peter Gaffney.

> Tolika tknuté milenci,
> A dráždíš pletí tolika ústy zlíbanou
> A klínem, tolika vášněmi rozrývaným,
> A očima samice, očima smilnýma.
>
> (In the age-old ambits, where the stiff virtue of the saint rises
> In starched Spanish collars,
> Where in vain Sebestian, handsome,
> Vibrant nudity,
> His arrow-teased genitals puts on display,
> To tempt the virtuously averted eyes,
> You luscious wave of breasts,
> So softly melting, now seductively spill,
> Your body is exposed again, touched by so many lovers,
> And you tease with your skin kissed by so many mouths
> And a lap, torn by so many passions,
> And with the eyes of a female,
> With fornicating eyes.)

The strategy for annotating sentence segmentation was set to make the sentences as coherent as possible, that is, with the fewest possible stand-alone clauses with major ellipses.

Lexical perspective

Concerning vocabulary, some patterns of differences between PoeTree and the relevant Czech treebanks (UD_Czech-PDT and UD_Czech-FicTree, henceforth PDT and FicTree) were predictable, namely archaic word forms (*jest* 'is'; *kdys* 'long ago'), archaic words (*junoše* 'young lad'), Latin words (*Ave*; *absolvo*), and neologisms (čaroskvělý 'miraculously magnificent').

From the quantitative perspective, the overlap between case-insensitive types (unique word forms) in the PoeTree sample and the training data sets of PDT and FicTree is approximately 59 and 47 respectively, excluding proper nouns, punctuation, and symbols. That means the UDPipe parser has never seen about one half of the words that occur in PoeTree, using either model.

To allow for more qualified guesses about domain-adaptation requirements, we extracted a frequency list of all PoeTree tokens missing in PDT and a frequency list of all PoeTree tokens missing in FicTree. We compared the distributions of these types, as well as the parts of speech they belong to. In both groups, the top-ranking

PoeTree-specific tokens belong to the following parts of speech as defined by the Universal Dependencies (UD) tagging scheme: nouns, verbs, adjectives, adverbs, determiners, and pronouns. Even though the lower-ranking parts of speech ranked differently, there was no statistically significant difference between their distributions (Fisher's exact test for count data, p-value = 1).

In the next step, we extracted the symmetric difference of both lists (PoeTree-specific types that were missing either in PDT or FicTree but not in both), corresponding to 21% of their union (PoeTree-specific types missing either in PDT or FicTree or in both). From a total of 557 PoeTree-specific types, 136 were missing from PDT and 421 from FicTree. 370 of these types only occurred once in PoeTree (Figure 2). At this point we resorted to qualitative analysis.
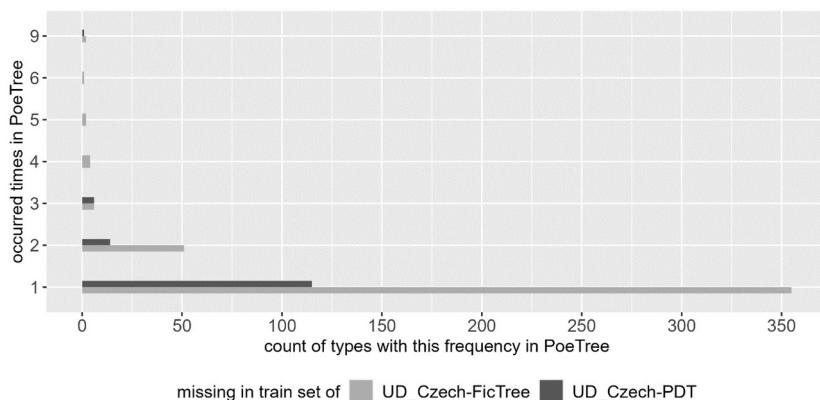
Figure 2: In which reference corpus are these PoeTree-specific words missing?

The list of missing types contained 20 types that occurred at least four times in PoeTree. Only one of them was also missing in PDT: *tobě* ('you' in the dative singular). The FicTree corpus was missing the archaic form of the third person singular of *jest* 'to be' and the vocalized form of the preposition *ku* 'towards.' Other words with minimum frequency 2 (down to Rank 87) were mostly missing in FicTree, probably because FicTree is smaller than PDT (166K tokens vs. 1M tokens in PDT). They did not seem to follow any interesting lexical or morphological pattern that would help distinguish FicTree from PoeTree. By contrast, a particularly striking pattern emerges in PoeTree compared to PDT. Here, PDT appears to have a bias against the second person singular: of 44 verb types in PoeTree that were specifically missing in PDT, six were in second-person singular form, as opposed to only two

from 143 verb types missing in FicTree. Even more strikingly, of the nine PoeTree-specific pronouns and determiners, five were second-person singular words and all were absent in PDT. They even turned out to be among the most frequent PoeTree types, which is not typical for pronouns in a pro-drop language.

Indeed, a search through the entire PDT suggested a noticeable difference in the frequency of the second person singular in PDT and in PoeTree: it detected only 45 occurrences of the singular *ty* 'you' (compared to 77 in PoeTree), nine occurrences of the singular *tvůj* 'your,' also nine in PoeTree, and 79 occurrences of the verb *být* 'to be' in the second person singular (compared to 30 in PoeTree). (It should be noted that the conjugated *to be* acts as auxiliary verb in the past and imperfective future tense.) It also detected 261 verbs in the present tense or imperative in the second person singular (101 in PoeTree), of which 95 were the fixed expression *viz* ('see,' as in 'cross reference' or 'cf.') in PDT.

Finally, we listed types missing in both PDT and FicTree. Among the most frequent types (four to six occurrences) were the archaic forms *kdys* (*kdysi* 'long ago'), *přec* (*přece* 'yet' or 'nevertheless'), *by* (*aby*, a polysemous subordinator), *chcem* (*chceme* 'we want'), and *jich* (*jejich* 'their'). The most frequent universal parts of speech (UPOS) among the hapaxes was noun, followed by verb and adjective (420, 297, and 261 occurrences respectively). Many of them were rare words or neologisms, and those belonging to common vocabulary were often either in archaic or otherwise marked forms (or second person), forming no other apparent pattern.
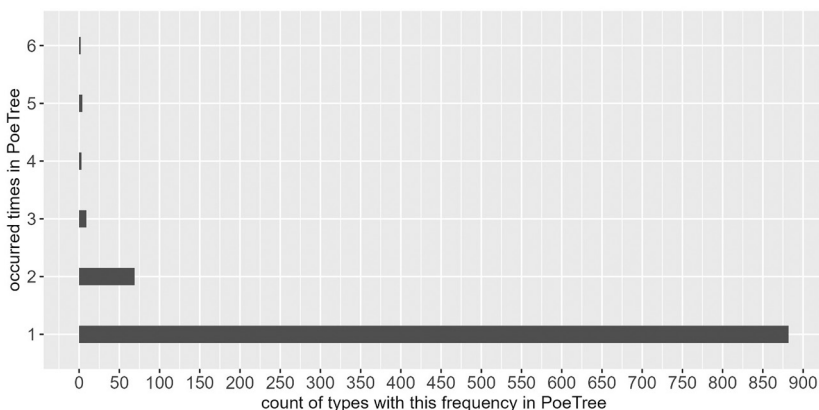
Figure 3: Distribution of PoeTree-specific words missing in both PDT and FicTree.

Syntactic perspective

While we did not make any a priori decisions concerning syntactic dependencies, we did make assumptions about how word order was likely to diverge from what is usual in modern prose or non-fiction treebanks. The observed differences are described in detail in Section 7.

The annotation process

We pre-processed the sample with the most recent version at the time of the largest Czech language model, czech-pdt-ud-2.12-220711 (Straka), employed in the UDPipe parser (Straka et al.; Straka and Straková). One annotator edited the automatic annotation node by node to come as close as possible to a manual annotation made from scratch. It was published under the title *UD_Czech-Poetry* in the Release 2.13 of the UD corpora in the LINDAT-CLARIAH repository (http://hdl.handle. net/11234/1-5287).

Evaluation results

We evaluated UDPipe's performance on the sample by comparing them to the UDPipe models based on PDT and FicTree, and then drilled into more detail using several analytical scripts in Udapi (Popel et al.). We also carried out manual error analysis.

Figure 4 presents the performance of UDPipe-PDT and UDPipe-FicTree operationalized by ten standard metrics (Kübler et al. 79–86; Zeman et al., "CoNLL"). Their values are measured as Precision (percentage of correct instances predicted by the parser), Recall (percentage of instances of gold annotation correctly predicted by the parser), and F1 (harmonic mean of Precision and Recall). They are plotted as the points of three different shapes. The first six metrics are self-explanatory, with AllTags showing the performance on morphological tagging (disregarding syntactic dependencies). The metrics UAS, LAS, MLAS, and BLEX consider each token in relation to its parent. UAS (Unlabeled Attachment Score) concentrates purely on the tree topology, which means that it only observes whether the given token is governed by the right parent. LAS (Labeled Attachment Score) considers the dependency label of the given token as well (that is, the relation to its parent). MLAS (Morphology-Aware Labeled Attachment

Score) adds UPOS and Features to the considerations. BLEX (Bilexical Dependency Score) combines content-word relations with lemmatization (but not with tags or features). The plot also shows the performance of the respective models (F1 Score) on their regular test data sets as colored bars. The performance values of both models on their regular test data sets are well above 95%. On PoeTree, the performance is generally worse, by the largest margin in Sentences and MLAS.

UDPipe-PDT and UDPipe-FicTree perform very similarly on PoeTree, with UDPipe-PDT scoring slightly better than UDPipe-FicTree in general and even substantially better in Sentences, UFeats, and MLAS. Therefore, UDPipe-PDT appears to be the parser of first choice for PoeTree and we limit manual error analysis in the next sections to the output of UDPipe-PDT.

Error analysis

Figure 4 reveals that the lowest scoring metric is Sentences, that is, the recognition of sentence boundaries. This is indeed neither surprising, given the a priori observations of punctuation and sentence splitting, nor extremely interesting, since sentence boundaries in poetry are often disputable even to a human annotator. For further error analysis, we have therefore aligned the manual and automatic word-to-word and re-segmented the automatic annotation to matching chunks of text. In this setup, we counted and classified the mismatches between manual and automatic annotation.

The most frequent error is the choice of parent (546, that is, ca. 8% of tokens), of which 395 are not combined with any labeling error. This also corresponds to Figure 4, where the second lowest scoring metric is MLAS, the combination of tree topology (choice of parent) and the syntactic and morphological labels in the given token. It also confirms that topological errors are not to blame on sentence splitting alone.

Of the 50 most frequent errors listed by the official UD evaluation script (Straka and Popel), 26 are dependency-labeling (deprel) errors, 13 are tokenization errors in thus far unseen contracted forms with unstable orthography, 11 are feature-labeling errors (Ufeats), four are part-of-speech errors, and four are lemma errors.

The lemma errors revolve around the so-called canonic number for the base form in pronouns (e.g., *náš* 'our' as *náš* 'our' vs. *můj* 'my') and reveal the general need for permanent data harmonization against the ultimate morphological lexicon (Hajič et al., "MorfFlex").

The most frequent UPOS error (24 occurrences, or 0.4%) concerned the blurry distinction between adverbs and particles (also suggesting inconsistencies in the manual annotation of different corpora), and the similarly blurry distinction between coordinating conjunctions used within a single sentence compared to those used across sentence boundaries (to be marked as sentential adverbs).

We also found that the most frequent features errors were not really errors but innovations encouraged by the Czech UD coordinators: to date, homonymous word forms have not been disambiguated in the PDT and FicTree data sets (e.g., Gender=Fem,Neut), unlike the PoeTree sample (Gender=Fem or Gender=Neut). These two approaches differ by the extent of contextualization. While the earlier approach deliberately relied on as little context as possible, the more recent developments in machine learning are likely to master context-based morphological disambiguation even across sentence boundaries. A prominent example of this change might be the disambiguation of active verb participles (used to form past tense): the neutral plural is homonymous with feminine singular, and Czech is a pro-drop language, which means that the coreferential antecedent of the dropped subject must often be tracked back across sentence boundaries.
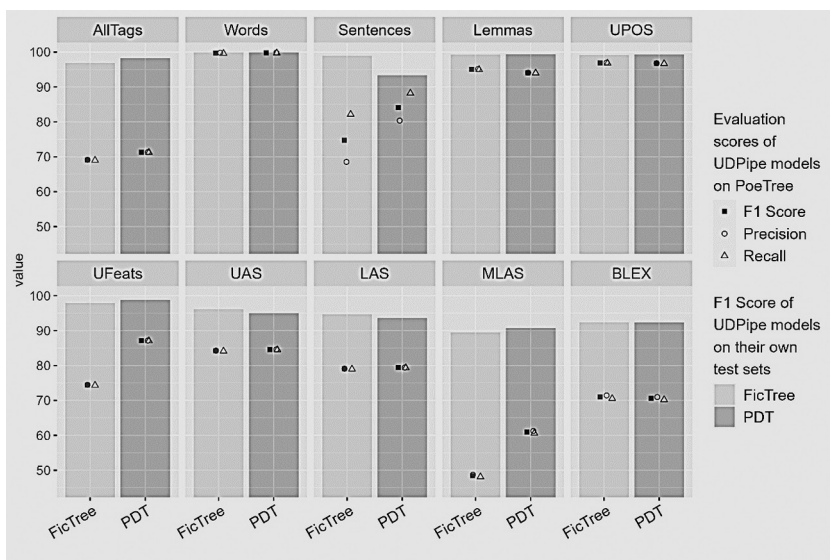
Figure 4: Model evaluation.

Since the aforementioned errors are not entirely errors, occur only rarely, or can be automatically corrected in the model training data, after which they are likely to present themselves correctly, what remains is tree topology and dependency labeling (syntactic parsing). Focusing on dependencies also makes sense, given that syntactic dependencies represent one of the advantages of extracting information from treebanks with comparison to carrying out linear searches. In the context of information-extraction use case, we find it appropriate to emphasize errors in phenomena that are likely to hamper the extraction of relevant patterns (such as convoluted attributive structures) over errors that may be frequent but do not necessarily affect rule-based extraction of noun modifiers or predicates and their arguments and adjuncts. Such largely irrelevant errors may involve punctuation, coordination vs. parataxis mismatches, or inconsistent labeling of prepositional noun modifiers, such as we find with nmod (noun modifier) vs. obl (oblique case).

Most prominent parsing errors

Labeling confusion as weighted centrality degree in a network of labels

The LAS results are best explained as a directed network graph (Figure 5) of dependency labels (deprels), with emphasis on their weighted degree centrality. Each edge connects a source node (human-assigned deprel) with a target node (deprel automatically assigned by UDPipe) on the same token, with the frequency of the given deprel combination in the same source-target direction increasing the edge weight. The number of outgoing edges along with their weights constitutes the weighted out-degree centrality of each deprel.

In this scheme, deprels can have out-degree centrality only when used in the gold annotation, whereas they only have in-degree centrality when they are used in the automatic parsing. Hence, the top-ranking deprels listed in Table 1 and highlighted in Figure 5 are gold-annotation deprels that UDPipe labeled with the wrong deprel, in addition to attaching them to the wrong parent. Each node in this graph represents one syntactic label. The nodes are connected with directed edges (arrows). Each arrow starts in the gold annotation and points to its mismatched label in the automatic annotation, respectively. Dotted edges connect the top 20% of gold annotation with the most frequent mismatches (totaled across all mismatched labels).

The thick highlighted source-target edges in Figure 5 convey which deprels are frequently confused in both directions, such as, for instance, nmod and obl (ranking 1 and 2 in Table 1). Both denote a noun or noun group, possibly even introduced by a preposition. This modifier is called nmod (noun modifier) when modifying a noun, such as *John* in *letter to John*, but obl (oblique case) when modifying a verb, such as in *write to John*. In a vague context such as *write a letter to John*, the modifier *John* can be attached to either, while in *give the letter from/by Mary to John*, we would rather attach *Mary* to the noun *letter* as nmod than to the verb *give* as obl.

It does not come as a surprise that conj (conjunction) and root are strongly interconnected in both directions: in complex sentences with several clauses, the parser easily fails to identify the main predicate. Quite symptomatically, root is also connected with advcl (adverbial clause, subordinate clause), parataxis (coordination of two main clauses without a conjunction), and orphan (clause with an elided predicate). The strong associations of conj with obl, nsubj (nominal subject), and, to a lesser extent, obj (direct object), indicate misrecognized coordinations of nouns.

The second strongest association with root is nsubj, which can be easily accounted for by the fact that the UD scheme prefers content words as parents of function words (e.g., nouns govern prepositions), while at the same time regarding copula verbs as auxiliary words. In copula predicates, therefore, the root is the predicate noun (Figure 6), which may be confused in turn with the subject (nsubj).

Ultimately, the strongest confusion emerges between the aforementioned nmod and obl. Since our statistic considers only labeling mismatches on incorrectly attached nodes, we can generally assume that nmod cases in the automatic sample are governed by nouns (since the parser has learned that nmod only modifies nouns) and obl cases are governed by verbs. This implies that a fraction of nouns in attributive positions or positions of verb arguments or adjuncts will be systematically lost when searching the poetry data with syntactic corpus queries. It is worthwhile to investigate whether this confusion occurs randomly or according to a pattern.

deprel	weighted out-degree centrality	parent UPOS NOUN	parent UPOS VERB
nmod	68	+	-
obl	47	-	+
root	43	-	-
nsubj	32	-	+
conj	31	+	+
advcl	29	-	+
amod	28	+	-

Table 1: Highest weighted out-degree centrality. The last two columns illustrate the possible distribution of the deprels among nouns and verbs as parent tokens.

Gold(out) vs. auto(in) labeled node attachment
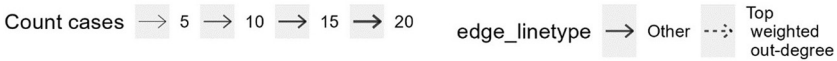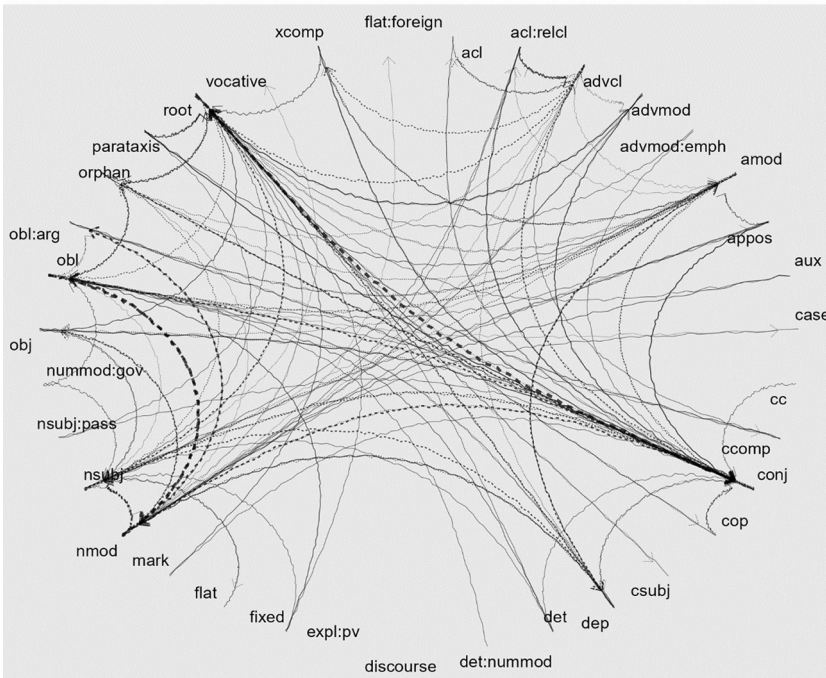
Figure 5: The most prominent labeling errors in a network graph of tokens with the wrong parent as well as the wrong dependency label.
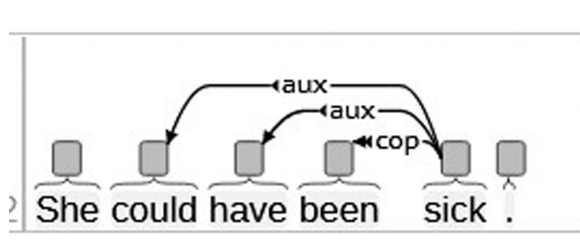
Figure 6: Copula predicate. The predicate noun and hence the sentence root is *sick*, while *she* is the subject (Marneffe et al.).

Labeling errors on nouns in attributive position

The high weighted out-degree centrality of nmod means that nmod UDPipe kept assigning other labels to nodes that should have been nmod. It hence makes sense to examine errors from the perspective of gold data, that is, to concentrate on nouns and their noun attributes.

When concentrating on nouns and their attributes, we get the following picture: the PoeTree sample contains 501 cases of attributive nouns (i.e., nmod). Of those, 169 (34%) were attached to an incorrect parent. Of 117 attributive nouns in a prepositional case, 49 (41%) were misrecognized. Of 384 attributive nouns in a direct case, 120 (31%) were misrecognized. However, when the case was genitive and the attribute noun was preposed, as many as 51 of 52 cases (98%) were misrecognized.

Comparing that with adjectival attributes, we observed only 117 of 877 incorrect parent attachments (13%). When the adjective preceded the noun, 63 of 575 attributes (11%) were incorrectly attached; when it followed the noun, the number was 54 of 302 (17%). Parser performance decreased in proportion to the distance between tokens. However, the data was too sparse to be statistically significant (for a token distance of 3 or more, the number was 19 of 25 errors if postposed, and 6 of 36 if preposed).

Preposed genitive attributes

The analysis has shown that the parser failed most dramatically with preposed genitive attributes, apparently because it had never spotted them in the training data.

In current Czech prose, it is not uncommon for noun attributes to take the form of another noun in the genitive case. The noun in genitive often denotes the agent or patient of an event (*hledání odpovědi* 'the search for an answer'), the owner or bearer (*planeta opic* 'planet of the apes'), or a quantified mass or set (*pytel brambor* 'sack of potatoes'). Nevertheless, in all these cases, the genitive follows the head noun. In the entire PDT, there are only two cases of a preposed genitive attribute. One is the lexicalized expression *svého druhu* 'of sorts'; the other concerns attributive nouns modified by a cardinal numeral, which in Czech requires the genitive of the governing noun (Figure 7). In this last case, one could argue that word order is slightly marked, emphasizing the amount, whereas in the unmarked order the genitive noun follows the head noun (see Section 8.1).

In poetry, on the other hand, the preposed genitive attribute is a legitimate structure, given its 10% proportion of all attributive nouns in our sample. This alone–its 98% error rate–calls for a domain adaptation of the language model to poetry.

A boring case in current prose
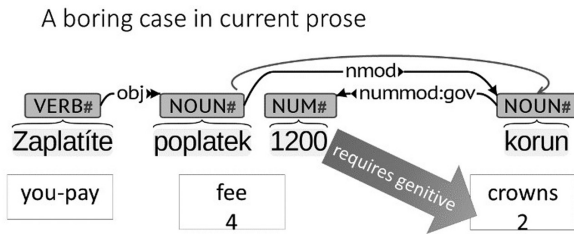
Figure 7: Preposed noun genitive in current Czech. The numbers 4 and 2 mean accusative and genitive.

Even with current Czech, the parser gets confused (Figure 7), regarding both *poplatek* 'fee' and *korun* 'crowns' as verb arguments, rather than two direct objects (obj), which the annotation scheme does not allow (a verb clause can only have one instance of a subject and object). Errors of this kind appear systematically in the poetry data (Figure 8).
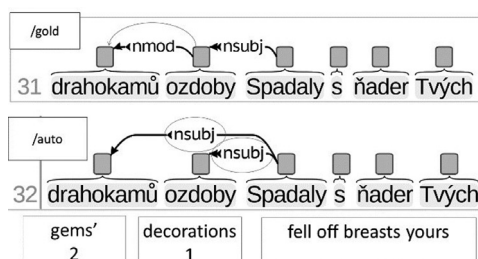
Figure 8: Preposed genitive attribute in poetry.

As Figures 9, 10, and 11 show, virtually any clause chunk can land between the preposed genitive attribute and head noun, resulting in additional parsing errors in the vicinity. This is why UDPipe again mistakes the genitive attribute, in Figure 10, for a second subject apart from its head noun (the true subject). In Figure 9, UDPipe (bottom line) has not recognized any syntactic dependency relation between the genitive (*Madonna's*) and head noun (*face*). The same applies to Figure 11, where it has missed the relation between *hair* (in the genitive) and *flood*.
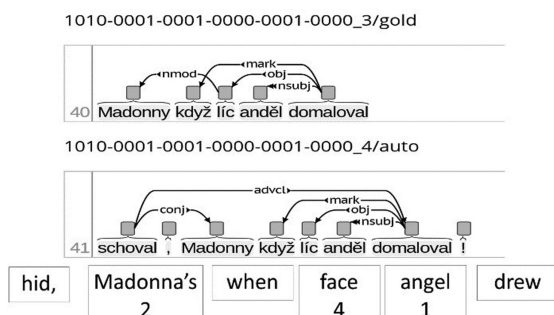
Figure 9: Discontinuous attributive sequence.
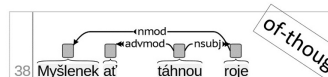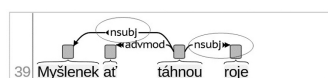
Figure 10: A whole clause between attributive genitive and its head noun.
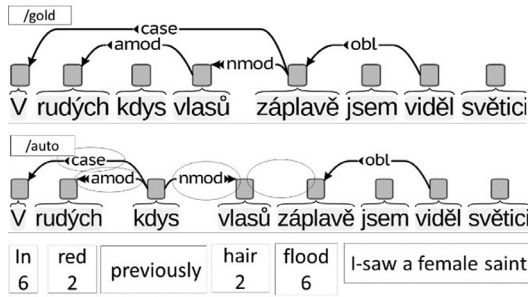
Figure 11: Very disrupted parsing.

Comparison of parsing errors in PoeTree and PDT

Qualitative findings from the previous section suggest that the specific constraints on prosody and meter may require poetic texts to allow longer distances between tokens and their modifiers (edge lengths, measured in tokens), as well as specific word order patterns. This section investigates the distance between several frequent syntactic dependencies, the order of their members, and the performance of the UDPipe-PDT model on two data sets: the PoeTree sample and PDT-test set (on which the performance of UDPipe-PDT was measured).

Performance on preposed genitive attributes

As Figure 12 shows, most preposed genitive attributes occur immediately before the head noun, within the maximum distance (edge length) -6 in PDT-test and -5 in PoeTree. The red bars in the blue-red pairs are lower than the blue ones in both PDT-test and PoeTree, but the difference is smaller in PDT than in PoeTree, which implies higher recall in PDT than in PoeTree. At the same time, orphaned red bars occur to the right of zero in both data sets. These are precision errors, and they are markedly fewer in the PDT-test sample.

By and large, the distributions of edge lengths are almost identical. At this point we should note that in PDT, unlike PoeTree, the preposed genitives are the product of grammatical congruence with a genitive-requiring cardinal numeral (denoting containers, substances, currencies, or metric units, see Section 7.3). Therefore it comes as no surprise that UDPipe processes them much better in PDT than in PoeTree. Also disregarding the lexical patterns and looking at raw

frequencies, the preposed genitives are clearly overrepresented in poetic texts compared to PDT, given that the PDT-test set is almost 27 times larger than the PoeTree sample and its occurrence of preposed genitives is only approximately double.
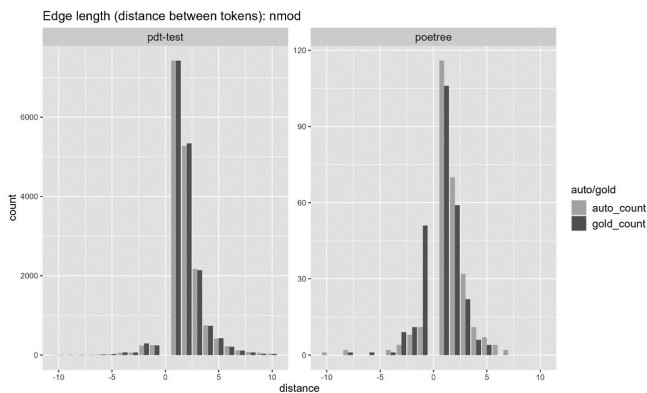
Figure 12: UDPipe-PDT's performance on the preposed genitive attribute in PoeTree and PDT-test.

Performance on any noun attribute

Generally speaking, noun attributes preceding nouns are overrepresented in poetic texts, and UDPipe has a precision issue with postposed noun attributes in PoeTree and a recall issue with preposed noun attributes (slightly above half of the approximately 50 items immediately before head nouns would be genitives).
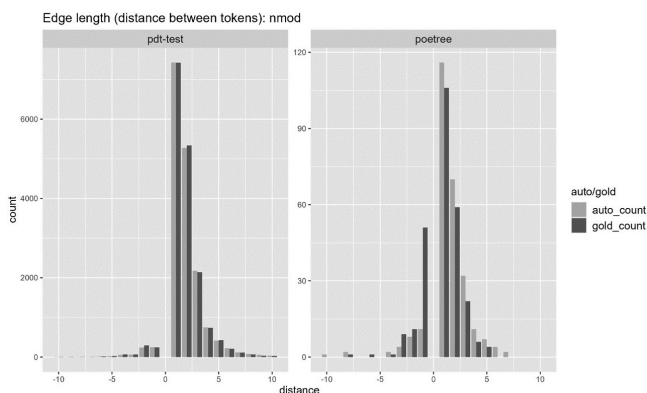
Figure 13: UDPipe-PDT's performance on noun attributes in PoeTree and PDT-test.

Performance on adjective attributes

The adjective attributes apparently are relatively more frequent in PoeTree than in PDT, but UDPipe-PDT processes them well, although the overall performance of UDPipe-PDT on PoeTree is slightly lower than on PDT-test, with errors both in precision and in recall, and both left and right from the governing noun.
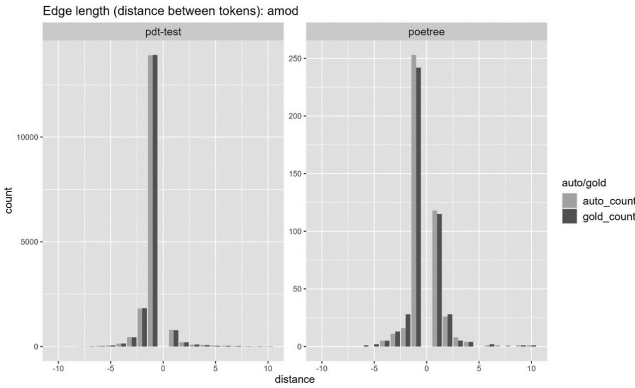
Figure 14: UDPipe-PDT's performance on adjective attributes in PoeTree and PDT-test.

Performance on clause subjects

The distribution of edge lengths for the clause subject is apparently identical in both data sets. On PDT-test, UDPipe-PDT tends to produce precision errors, while on PoeTree both error types occur. Interestingly, the overall performance appears slightly higher on PoeTree.
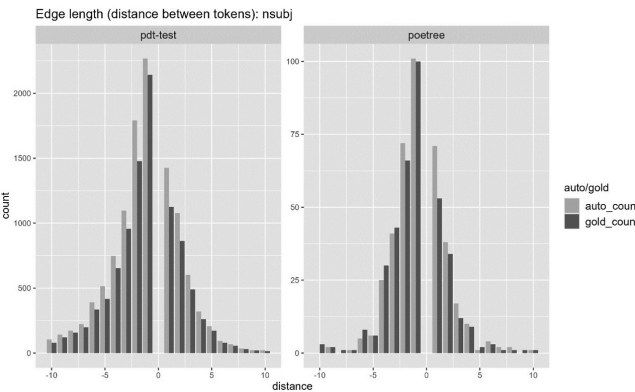
Figure 15: UDPipe-PDT's performance on subjects in PoeTree and PDT-test.

Performance on direct objects

Direct objects occur apparently more often immediately before their governing verb in PoeTree than they do in PDT-test. UDPipe-PDT performs slightly worse on PoeTree than on PDT-test, in all positions, but the difference is not dramatic.
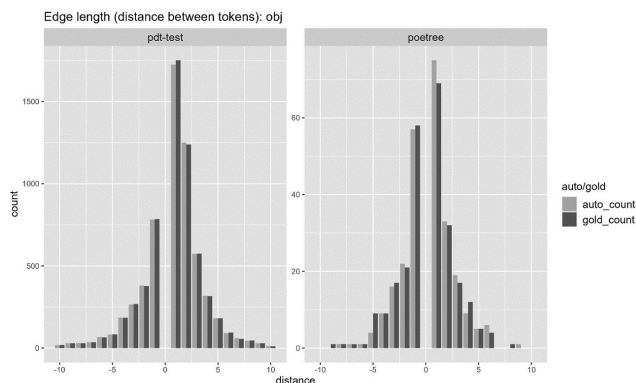
Figure 16: UDPipe-PDT's performance on direct objects.

Performance on prepositional objects and adverbials

Distributions are similar for prepositional objects and adverbials, with one interesting observation: objects immediately following the verb are rather rare, especially in PDT-test, and UDPipe-PDT has a severe precision problem (too many false positives) on both datasets, across positions.
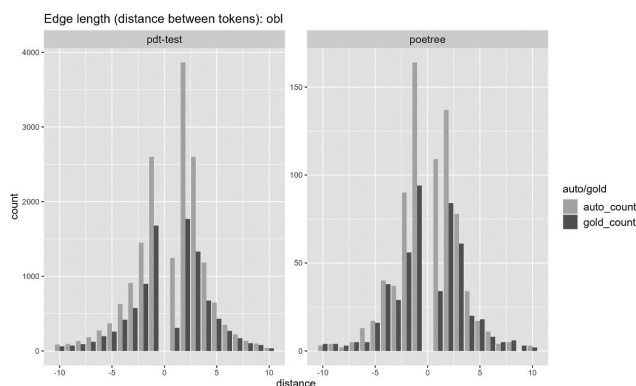
Figure 17: UDPipe-PDT's performance on nouns with prepositions modifying verbs.

Discussion and conclusion

We have evaluated the performance of the UDPipe parser with the largest Czech model based on the Prague Dependency Treebank (Hajič et al., "Prague") converted to Universal Dependencies (Zeman et al., "Universal"), and performed a semi-manual error analysis focused on parts of speech and dependency relations that are most likely to occur in corpus queries to extract information from texts in text-mining or distant reading research tasks.

Czech poetry makes ample use of the free word order that is a feature of the Czech language. Hence, PoeTree contains structures that do not normally occur, and UDPipe-PDT fails to parse them correctly because it has never spotted them in the training data. These structures are not random but recurrent, and therefore it is important to, first, identify and tackle them as parsing issues, and second, provide manually annotated data to the UDPipe model training pipeline to improve UDPipe's performance on poetry.

WORKS CITED

Dobrovský, Josef. *Ausführliches Lehrgebäude der Böhmischen Sprache, zur gründlichen Erlernung derselben für Deutsche, zur vollkommenern Kenntniß* für Böhmen. Prague, Johann Herrl, 1809.

Hajič, Jan. "Complex Corpus Annotation: The Prague Dependency Treebank." Jazykovedný ústav L. Štúra, *SAV*, 2004, https://ufal.mff.cuni.cz/pdt2.0/publications/Hajic2004.pdf. Accessed 24 Jan. 2024.

Hajič, Jan, et al. "MorfFlex CZ 2.0." *LINDAT/CLARIAH-CZ*, 2020, http://hdl.handle.net/11234/1-3186. Accessed 24 Jan. 2024.

Hajič, Jan, et al. "Prague Dependency Treebank 2.0." *Linguistic Data Consortium*, 2006, https://ufal.mff.cuni.cz/pdt2.0/. Accessed 24 Jan. 2024.

Kampelík, František Cyril. *Čechoslovan, čili národní jazyk v Čechách, Na Moravě, ve Slezku a Slovensku*. Prague, Jan Hostivít Pospíšil, 1842.

Kübler, Sandra, et al. *Dependency Parsing*. Springer, 2009.

Marneffe, Marie-Catherine de, et al. "Syntax: General Principles–The Status of Function Words." *Universal Dependencies Guidelines*, 2017, https://universaldependencies.org/u/overview/syntax.html#the-status-of-function-words. Accessed 24 Jan. 2024.

Osolsobě, Klára. *Česká morfologie a korpusy*. Prague, Karolinum, 2014.

Kosek, Pavel, and Jana Pleskalová. "Spřežkový Pravopis." *CzechEncy–Nový encyklopedický slovník češtiny*, edited by Petr Karlík et al., Brno, Masarykova univerzita, 2017, https://www.czechency.org/slovnik/SPŘEŽKOVÝ PRAVOPIS. Accessed 24 Jan. 2024.

Plecháč, Petr, and Robert Kolár. "The Corpus of Czech Verse." *Studia Metrica et Poetica*, vol. 2, no. 1, 2015, pp. 107–118, https://doi.org/10.12697/smp.2015.2.1.05. Accessed 24 Jan. 2024.

Plecháč, Petr, et al. *PoeTree. Poetry Treebanks in Czech, English, French, German, Hungarian, Italian, Portuguese, Russian and Spanish. 0.0.1*. Zenodo, 2023., https://zenodo.org/records/10008459. Accessed 24 Jan. 2024.

Popel, Martin, et al. "Udapi: Universal API for Universal Dependencies." *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies*, edited by Marie-Catherine de Marneffe et al., Northern European Association for Language Technology, 2017, pp. 96–101.

Straka, Milan. "Universal Dependencies 2.12 Models for UDPipe 2." *LINDAT/CLARIAH-CZ*, 2023, http://hdl.handle.net/11234/1-5200. Accessed 24 Jan. 2024.

Straka, Milan, and Martin Popel. "Eval.Py. 1.2." *GitHub*, 2023, https://github.com/UniversalDependencies/tools/blob/master/eval.py. Accessed 24 Jan. 2024.

Straka, Milan, and Jana Straková. "UDPipe 2." *LINDAT/CLARIAH-CZ*, 2022, http://hdl.handle.net/11234/1-4816. Accessed 24 Jan. 2024.

Straka, Milan, et al. "UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing." *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, edited by Nicoletta Calzolari et al., European Language Resources Association, Paris, 2016, pp. 4290–4297, https://aclanthology.org/L16-1680. Accessed 24 Jan. 2024.

Zeman, Daniel, et al. "CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies." *Proceedings of the CoNLL 2018 Shared Task*, edited by Daniel Zeman and Jan Hajič, Kerrville (TX), The Association for Computational Linguistics, 2018, pp. 1–21, http://www.aclweb.org/anthology/K18-2001. Accessed 24 Jan. 2024.

Zeman, Daniel, et al. "Universal Dependencies 2.12." *LINDAT/CLARIAH-CZ*, 2023, http://hdl.handle.net/11234/1-5150. Accessed 24 Jan. 2024.

Žižková, Hana. "Compound Adverbs as an Issue in Machine Analysis of Czech Language." *Journal of Linguistics / Jazykoedný* Časopis, vol. 68, no. 2, 2017, pp. 396–403, https://doi.org/10.1515/jazcas-2017-0049. Accessed 24 Jan. 2024.

Rime in skladnja: oblikoskladenjska analiza češke poezije

Ključne besede: češka poezija / oddaljeno branje / besedilni korpusi / Universal Dependencies / obdelava naravnega jezika / odvisnostne drevesnice

Oddaljeno branje, ki upošteva jezikoslovna spoznanja, predpostavlja ustrezno delovanje orodij za obdelavo naravnega jezika. Članek prikaže evalvacijo raz-členjevalnika UDPipe na primeru ročno označenega vzorca češke poezije 19. stoletja v naslednjih korakih: (1) ustvarjanje dokumentiranega nabora podat-kov za to področje (poezija, 19. stoletje, češčina); (2) odločitve o označevanju,

specifične za področje; (3) analiza napak. Vzorec je obsegal 29 naključno izbranih pesmi, ki so bile najprej samodejno označene in razčlenjene z razčlenjevalnikom UDPipe, nato pa so bile oznake ročno preverjene za vsako posamično besedo. Preverjene so bile naslednje značilnosti: segmentacija besed (razdelitev), lematizacija, dodelitev oblikoskladenjskih oznak, dodelitev natančnejših morfoloških oznak, dodelitev položaja v skladenjskem drevesu (izbor nadrejenega elementa) in oznaka skladenjskega razmerja med besedo in njenim nadrejenim elementom. Ugotovitve smo analizirali; najpogostejše napake razčlenjevalnika so povezane s kompleksnimi samostalniškimi besednimi zvezami, ki vsebujejo druge samostalnike kot modifikatorje, še posebej, če se ti pojavijo v besednem redu, specifičnem za poezijo, npr. kot določilo samostalniškega jedra. Po drugi strani niti arhaični pravopis niti neologizmi niso predstavljali bistvenih težav.

A Stylometric Glance at Novels in Euskara

Dominika Werońska

Jagiellonian University, Doctoral School in the Humanities, Pałac Spiski, Rynek Główny 34, 31-010 Kraków, Poland
https://orcid.org/0000-0002-1053-9056
dominika.weronska@doctoral.uj.edu.pl

While Basque has been posited as possibly the oldest existing language on the European continent, it appears in written form only in the sixteenth century. The first Basque novel emerges over 300 years later and to this day the genre lacks exhaustive research. The article sets as its aim a stylometric analysis of selected twentieth- and twenty-first-century Basque novels, sourced from the online platforms Armiarma and Booktegi. These are analyzed based on the frequency of the most frequent words measured using cluster analysis and set against a backdrop of foreign novels translated into Euskara. The results show that the originals in Euskara remain distinct from translated works, pointing to the unique linguistic character of the Basque novel. Some linguistic patterns potentially responsible for this distinction are presented. The results are visualized on a map revealing the chronological evolution and the contribution of the Basque novel to the broader literary landscape.

Keywords: Basque literature / Euskara / text corpora / cluster analysis / stylometric map

Introduction

The Basque novel, though over 300 years old, remains relatively under-researched in terms of stylometry, a gap which this article aims to bridge through a stylometric analysis of selected Basque novels. The goal is to identify distinctive stylistic patterns through the application of various stylometric methods such as cluster analysis and the construction of bootstrap consensus trees. Prior to presenting the analysis, the linguistic and literary context of the Basque novel are outlined.

Before there was literature there was language: A few words on Basque

The Basque language, also known as Euskara, boasts a rich history spanning over two millennia (Jansen 7). Where Basque originated and how

89

it came to exist on the European continent remains unknown. Despite numerous attempts to establish linguistic connections between Basque and other language families, including Celtic, African, Caucasian, and indigenous American languages, no definitive evidence has as yet emerged to support any hypothesis (8–9).

Attempts were made as early as the seventeenth century, when poet and historian Arnaut Oihenart suggested a connection between Basque and Iberian in his 1656 historiographical work *Notitia utriusque Vasconiae, tum Ibericae, tum Aquitanicae* (*A Survey of the Two Basconias, the Iberian and the Aquitaine*). While the theory was disproven, it was widely embraced in the nineteenth century by, for example, Spanish Jesuit and personal confessor of Queen Maria Anna of Neuburg, Manuel de Larramendi; according to his study *De la antiguedad, y universalidad del bascuenze en España* (*On the Antiquity and Universality of Basque in Spain*), Basque is a linguistic isolate and descendant of Iberian: "Now it is easily concluded that the Basque language was a universal language; […] that being the language of the first settlers, they would have to extend it along with the settlements they were founding, and […] would speak the same language, until foreigners came, who began to introduce their different languages." (Larramendi 43; my translation)

In his essay from 1818, Danish linguist Rasmus Rask noted a similarity between Basque and Greenlandic; he claimed that "verbs, in particular, have an extremely complex inflection, viz. eleven moods and six tenses in each of the first six moods, all of which seems rather to resemble Greenlandic than any language of the Gothic class" (Rask 90). Yet upon further deliberation, Robert L. Trask dismissed any notable similarity between the two languages, pointing instead to Basque's plausible North African origins in either Mauretania or Gaetulia (Trask, "Origins" 91).

In the twentieth century, research focused on finding a potential historical link between Basque and the Celtic languages (Trask, "Origins"; Trask, *History*) or Basque and Aquitanian (see Michelena), as well as on establishing a Proto-language which would have been a common ancestor for Proto-Basque and Proto-Indo-European (Tovar 1970). While Koldo Mitxelena (also known as Luis Michelena) could be credited with elucidating the most probable theory, the remaining studies were acknowledged to lack conclusive evidence.

At the same time, it must be admitted that Basque does have analogues in many languages across the globe. For example, like Hungarian, it is agglutinative and relies heavily on suffixation (Trask, *History* 201). Like Inuit, Mam, and Jacaltec, it is an ergative absolutive language with the

subject of a transitive verb appearing in the ergative case (marked with the suffix *-k*) and the subject of an intransitive verb showing up in the absolutive case (unmarked). Finally, like Japanese and many Indo-Iranian languages, Basque follows subject-object-verb sentence order (89, 109).

However, establishing a conclusive linguistic relationship necessitates non-linguistic evidence and a sharing of more than just one or two coincidental features. Progress in this regard remains limited and the prospect of establishing a definitive genealogical link remote (Jansen 9).

Dialects of Basque

Spoken today by no more than 806,000 inhabitants of the Basque Country, Basque speakers are further divided according to various dialects. Most scholars assert that variants of the Basque language were spoken as early as the second millennium BC, and point to the mountainous terrain and to the low prestige of the language as the two main reasons for dialectal fragmentation (Trask, *History* 5). Mitxelena offers a different perspective, arguing that this linguistic fragmentation commenced in the Middle Ages, post the decline of the Roman Empire (Michelena 300).

While earlier classifications in the nineteenth and early twentieth centuries partitioned the Basque-speaking area into six to nine major dialects, including the now-obsolete Roncalese (Jansen 8–9), more recent investigations indicate that over the last five decades, the dialects have converged, with no more than five identifiable varieties (Zuazo 22). These comprise the central dialects of Bizcayan (spoken in the area of Bilbao), Gipuzkoan (spoken in Donostia, Tolosa, Bergara, and Zarautz), and Upper Navarrese (spoken in the high areas of Navarre), as well as the two peripheral dialects of Navarro-Lapurdian and Zuberoan (spoken in the French region). Furthermore, in the 1960s, the Royal Academy of the Basque Language succeeded in cultivating a standardized literary Basque, known as Euskera Batua, to confront pressures from state-supported French and Spanish entities. This standardized form of the language provided writers with an orthographic and morphological framework amenable for employment irrespective of their native dialects and strengthened Basque's resilience against external pressures. However, this also meant that individual varieties of Basque received less space to flourish as means of cultural expression, also in the domain of literature. The corpus selected for the present analysis comprises of works written from 1897 to 2022. Of the 32 works published before 1960, the majority were written in Guipuzkoan and Bizcayan.

In contrast, the majority of the works published from the 1970s to the twenty-first century predominantly employ Euskara Batua.

Basque literature as a small literature

Basque literature can be described as a small literature. The adjective *small* refers to both quantitative and qualitative aspects. Quantitatively, it encompasses measurable factors of the literary system such as the number of authors, yearly publications, sales, potential readers, and publishing houses. Qualitatively, the term reflects not only the size of a literary corpus but also its influence on literary production and reception, its targeted audience, internal and external perceptions of its scale, its political autonomy or lack thereof, and so on.

With about 300 authors, a readership of around 806,000 Basque speakers, and an annual publication range of 1,500 works, Basque literature can well be classified as a small literature in quantitative terms (Kortazar 11–12). For a comparison with English and Slovenian literature, refer to Table 1.

Literature	Authors	Publications (per year)	Male/Female Writers	Speakers (potential readership)	Publishing houses
Basque	≈ 300	≈ 1,500	82% / 18%	806,000	226
Slovenian	269	≈ 6,000	42% / 58%	2,100,000	1400
U.S.	54,010	4,000,000	49.5% / 50.5%	400,000,000 (natives) 1,500,000,000 non-native)	2840

Table 1: Statistics behind the literary market (a comparison of Basque, Slovenian, and American literature).

Basque literature also meets the qualitative conditions for a small literature. In terms of typology, the Basque literary canon belongs to the group of literatures in peripheral languages without an external reference such as Estonian, Latvian, or Welsh. As is the case with most literatures produced at the boundaries of a major language in a small language, Basque literature finds itself in a dialectic between ideology and literary autonomy (Kortazar 12). On the one hand, it is expected to commit to the national idea and reinforce the identity of its people; on the other hand, it seeks to be cosmopolitan and express the personality

of its individual authors. The tension it encounters is integral to its survival, as language is closely tied with identity, and often results in a collective appropriation of the discourse with a preference for philologized language, as the use of language itself holds intrinsic literary value.

It is interesting to note that a similar observation was made by the Slovenian poet France Prešeren in a letter to his friend Matija Čop dated 5 July 1832, in which the poet remarked that "the tendency of our carmina and similar literary activities is no other than to cultivate our mother tongue" (qtd. in Juvan 185). The expectations put upon literature to be subservient to language are undoubtedly a characteristic trait of small literatures and a cause for tension, amplified by having to resist potential attacks of opposing forces on both political and cultural grounds. This is certainly the case with the Basque people who have, in the words of Mark Kurlansky, "stubbornly fought for their unique concept of a nation without ever having a country of their own" (Kurlansky 5).

The oppression faced by the Basques certainly took its toll on the development of their literature, and of their novel in particular. While the first known words in Basque were most probably engraved on a bronze relic already 2,100 years ago, and the first work of Basque literature, the collection of poems *Linguae Vasconum Primitiae* (*The First Fruits of the Basque Language*) by Bernard Etxepare, was printed in 1545, almost 300 years had to elapse before the first Basque novel appeared, and then it was only a proto-novel in dialogue form (Benzine). The work most scholars recognized as the first Basque novel was *Auñemendiko lorea* (*The Flower of Auñemendi*) by Domingo Agirre. It appeared only in 1897 and began a series of historical novels in the Romantic style soon overtaken by *costumbrista* prose. However, according to Jesús María Lasagabaster the novel never really had the chance to develop:

> It was always my opinion that the Basque novel, which was born so late, never had—and could never have had—the same kind of historic development Basque poetry had and was forced to jump instead, without logical progression or continuity, from these anachronistic novels of manners of blatant idealistic and romantic roots to experimentation and the avant-garde, without going first through realism, the current that—in European literatures at least—marked the coming of age and the establishing of the novel as a genre. (Lasagabaster 16)

Yet Lasagabaster claims that the lost ground has been recovered and that recovery is visible in the number of prizes awarded to Basque authors in recent years (16). While this may be the case, he is forced to admit that literature in Basque is "'contaminated' (in the etymological sense of the word)" (18) with linguistic and sociocultural influences current to France and Spain, the political places where it evolves.

A stylometric map of Basque literature

The idea of presenting a corpus of literature on a map is far from innovative. Critics of literature have often resorted to using elements of cartography to visualize texts in space, to outline the development of genres, or to illustrate the stylistic similarity between works (Brooks; Moretti; Bulson).

One early example is Cleanth Brooks's mapping of the setting and geography of William Faulkner's novels. In *William Faulkner: The Yoknapatawpha Country*, this American literary scholar explored the Mississippi writer's fictional county and the important role it played in so much of his work.

In *Novels, Maps, Modernity: The Spatial Imagination, 1850–2000*, Eric Bulson examined the depiction of place in nineteenth-century works of such writers as Charles Dickens and Herman Melville all the way through twentieth-century fiction. He likened the experience of reading a novel to examining a map, balancing between the immediate experience and a broader structural understanding.

Literary mapping has also been explored and popularized by Franco Moretti. In *Graphs, Maps, Trees: Abstract Models for a Literary History*, Moretti argues that these materialist concepts of form can often reveal more about literature than the texts themselves. An advocate of distant reading, that is, studying world literature devoid of direct textual analysis where one relies solely on studies done by other researchers, Moretti posits that seeing less is actually seeing more. He claims that maps possess "emerging" qualities which allow us to see aspects not visible at the lower level (Moretti 61).

In line with Moretti's distant reading, scholars within the discipline of stylometry have begun to produce graphical representations of literature in order to explore the relationships between individual works as well as entire literary canons (Rybicki, "Pierwszy rzut"; Rybicki, "Second Glance").

Stylometry, as the term suggests, measures authorial style using the statistical analysis of distinct features such as n-grams, grammatical categories, and most frequent words (MFWs). Methods relying on the frequency of the MFWs have proven particularly successful at resolving authorship issues, delineating the chronology of a set of works, or even distinguishing between different translators.

The following study applies stylometric methods such as cluster analysis and the construction of bootstrap consensus trees to a corpus of novels originally in Basque as well as to novels translated from various languages into the Basque language. The objective is to construct a

literary map for these works and to explore: (1) the efficacy of stylometry in authorship attribution for a so-called small literature within a highly inflected, agglutinative language; (2) the potential visibility of chronological evolution of the Basque novel and the extent to which the purported lack of continuity and progression in the genre's development is reflected on the stylometric map; (3) the distinction between novels originally written in Euskara and those translated from other languages; (4) the capability of stylometry to identify features of translationese; and (5) the impact of culling on the results of stylometric analysis and the nature of this impact.

Method

Perhaps the most influential stylometric method is Burrows's Delta. To quote Jan Rybicki, this "frankeinsteinish" approach consists in chopping up the pieces of analyzed literature into individual words, selecting from among them the least significant ones (MFWs) which constitute a kind of connective tissue of the works, and calculating their frequency for each text (Rybicki, "Second Glance" 7). The frequencies are then ordered in a frequency table and compared using some type of distance measure. Evert et al. have shown the Cosine Delta variant to be the most successful in terms of authorship attribution, hence its use in the present study. Thanks to the Cosine Delta measure, a distance matrix for each pair of texts is created. The smaller the distance between each text, the greater the similarity between the texts analyzed. For greater reliability, the calculations can be repeated for different numbers of the MFWs.

The results obtained are processed with consensus network analysis and visualized using the open-source software Gephi (Bastian et al.), which makes it possible to create a map of the given texts.

In the following study, seven types of stylometric tests relying on Cosine Delta were conducted on three different corpora. Below is a more detailed description of each test with the results obtained.

The first three tests were a series of cluster analyses performed on a corpus of works originally in Basque. Tests IV and V consisted in using the function Oppose to detect translationese in works by an individual author, Nikolas Ormaetxea (a.k.a. Orixe). Tests VI and VII were conducted using cluster analysis and bootstrap consensus trees, respectively, on a mixed corpus of originally Basque novels and novels translated into Basque from other languages. Below is a more detailed description of each test with the results obtained.

95

Test I

The first test was a cluster analysis performed in R using the stylo packet.

Corpus

Works selected for the cluster analysis were downloaded from Armiarma (37 works) and Booktegi (20), two online platforms which provide access to PDF versions of literary works in Basque, spanning from the eighteenth to the twenty-first century. Selection criteria included a minimum length of 25 pages (50 KB) and the inclusion of another work by the same author in the study. Consequently, 57 texts by 20 authors met these criteria.

The corpus ranged from Azkue's 1897 novel *Batxi Guzur* to Ortega's novels from 2022. Prominently featured authors, with at least four novels each, comprised Domingo Agirre, Jean Barbier, Txema Arinas, and Pedro Urruzuno. In contrast, authors such as Tomas Agirre, Itxaro Borde, and Manuel Etxeita contributed only two texts each.

The works, initially in PDF format, were converted to text files for analysis in R.

Method

The first cluster analysis was performed using Cosine Delta on 100 MFWs. Culling was not used, meaning that all words were considered for the analysis.

Results

Figure 1 presents a dendrogram illustrating the results of the cluster analysis. The root of the tree is on the right and branches extend to the left, culminating in the individual works. The *x*-axis represents the distance or dissimilarity between clusters, with larger distances implying less similarity. Encircled works signify instances of misattribution. As can be seen from the dendrogram, 48 of the 57 works are grouped correctly by author, yielding an authorship attribution accuracy of 84%.

Notably, this is below the 94% accuracy achieved by Eder in his analysis of 66 English novels (Eder 54). Misattributions include works by Azkue, Agirre, Borde, Orixe, Erkiaga, and Etxeita. Considering the

dendrogram's sensitivity to the number of variables (MFWs), a bootstrap consensus tree was applied in Test 2 to enhance reliability.
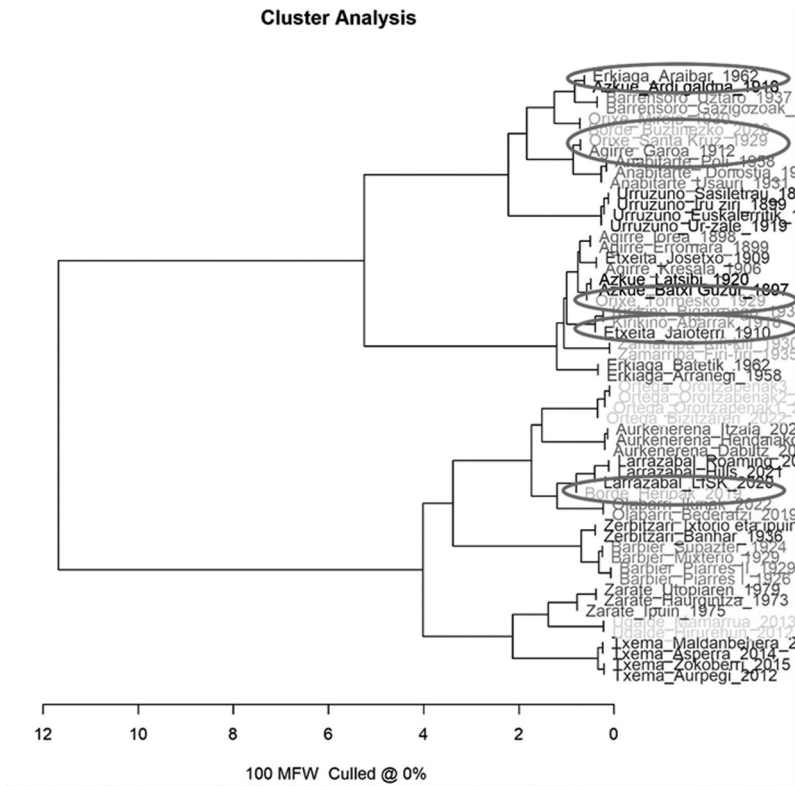
Figure 1: Results of Test I (cluster analysis performed on 57 novels originally in Basque).

Test II

The second test was a series of cluster analyses performed with the bootstrap method in R using the Stylo packet.

The bootstrap consensus tree is a revised, more reliable approach than cluster analysis (Eder 62). It is a statistical technique which involves repeatedly resampling the dataset to improve reliability by filtering out local disturbances. Links indicate consensus strength derived from multiple snapshot dendrograms instead of stylometric distances between individual texts. The corpus remained the same as for Test 1.

Method

Multiple cluster analyses were conducted on the corpus using a range of 100 to 1,000 MFWs. Culling was not employed. Cosine Delta measure was used for each cluster analysis. From the multiple bootstrap samples, a consensus tree was constructed (see Figure 2).

Results

The tree in Figure 2 represents the most stable and recurring relationships found across the various bootstrap samples. The tree indicates that authorship attribution was largely successful, with 50 out of 57 works attributed correctly (almost 88%). Seven works were misattributed, as identified by the circles in the tree.
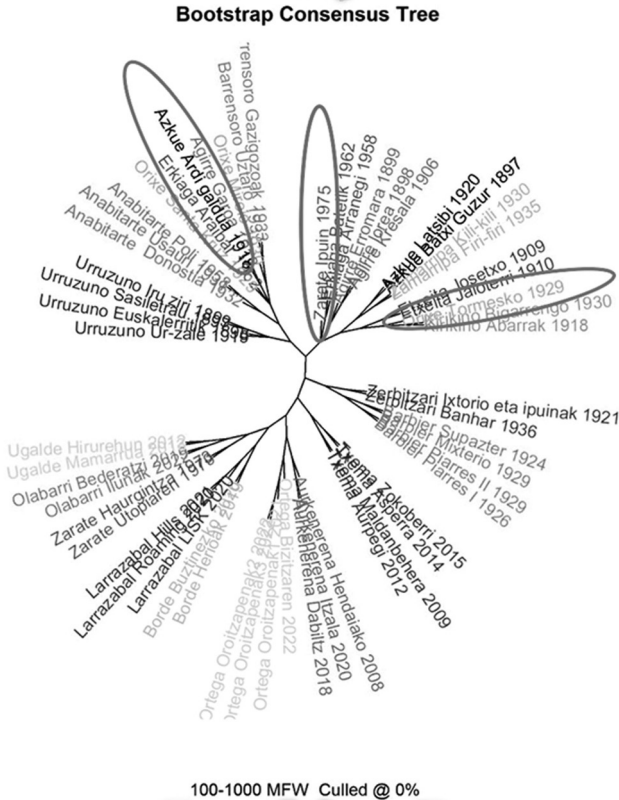
Figure 2: Results of Test II (a bootstrap consensus tree on 57 Basque novels, without culling).

As can be seen in Figure 2, the number of authors with misattributed works had decreased to Orixe, Agirre, Azkue, and Erkiaga. Additionally, Zarate's *Ipuin antzeko alegi mingotsak* (*Bitter Fables in the Guise of Stories*) clustered with Erkiaga's works. In order to decrease dimensionality with the hope of improving discrimination, culling was employed in Test III.

Test III

The third test was a series of cluster analyses performed using the bootstrap method and culling. Culling allows users to set a threshold for how frequently a feature must appear in the corpus to be considered in the analysis (Eder et al. 111). Features not meeting the specified occurrence threshold across the samples are excluded. For example, at 100% culling rate, only words appearing in at least 100% of the samples are included. Culling has been shown to improve discrimination leading to more accurate clustering. The corpus remained the same as for Tests I and II.

Method

The attribution test was run several times with different vectors of MFWs and with different settings for culling. Cosine Delta was performed using 100 to 2,000 MFWs, with increments of 100. The culling was performed from 0–100% with an increment of 20.

Results

The tree in Figure 3 represents the most stable and recurring relationships found across the various bootstrap samples. With 53 out of 57 works attributed correctly (almost 93%), authorship attribution was the most successful of all three tests.

Bootstrap Consensus Tree

100-26 MFW Culled @ 0-100%

Figure 3: Results of Test III (a bootstrap consensus tree on 57 Basque novels,
with culling).

The four works which were not clustered according to their authors comprised Resurreción Maria Azkue's *Ardi galdua* (*The Lost Sheep*), Eusebio Arriaga's *Araibar zalduna* (*Sir Araibar*), as well as Orixe's *Tormes'ko itsu-mutila* (*Lazarillo de Tormes*) and *Mireio*. These four works were consistently misattributed across the three tests. Hence, they warranted further literary investigation relying on traditional literary scholarship.

The first of these works, *Ardi galdua*, was written by Resurreción Maria Azkue, a priest, poet, and academic (Olaziregi 147). As the first head of Euskaltzaindia, Azkue worked on finding a literary standard for the Basque language. He proposed Gipuzkera Osotua, a codified and expanded form of the Gipuzkoan dialect, as the possible contender for euskara batua. *Ardi galdua* was his first work written in Gipuzkera

Osotua. Erkiaga's *Araibar zalduna,* which clusters with *Ardi galdua*, is also written in the Guipuzkoan dialect. While Erkiaga experimented with Navarrese and other dialects, *Araibar zalduna* was his attempt at Guipuzkoan. Incidentally, Agirre's *Garoa* (*Fern*), which in Test II clustered with the two works just mentioned, was also written in Gipuzkera Osotua. Agirre, a Carmelite priest who formed part of Euskaltzaindia, happened to be a friend of Azkue. While he wrote *Kresala* (*Saltpeter*) and *Auñemendiko lorea* in the Basque language of Bizkaia, both about life at sea, *Garoa*, focusing on the agricultural world, was written in Guipuzkoan (Olaziregi 145). It would seem, then, that stylometry was indicating dialectal variations. This could imply its potential utility in dialectometric studies.

The remaining texts for which authorship attribution was unsuccessful were all Orixe's works. Although all three were procured from the section of the Armiarma website containing works originally in Basque, only *Santa Kruz apaiza* (*The Priest Santa Cruz*) was actually originally written in Basque. The remaining two were translations, namely of *Mireio* from Provencal French and of *Lazarillo de Tormes* from Spanish. This aligns with Rybicki's observation that stylometry often groups texts by author rather than by translator (Rybicki, "Stylometric" 203–204). It is possible that Orixe's translatorial signal was too weak for stylometric analysis to recognize him. At the same time, cluster analysis indicated anomalies in authorship, revealing some works as translations. This raises the question of stylometry's potential in detecting translationese.

Tests IV & V

A potentially suitable method for addressing the question of characteristic textual features of translationese is the Oppose function. This function is typically used to identify which features are most characteristic of different authors (Eder et al. 117). In the following two tests, it is employed to differentiate between translated and original texts. Oppose operates by dividing a group of texts into two sets, the primary and the secondary one, and quantifying word frequencies within each. Once it identifies words with significant frequency disparities between sets, Oppose applies statistical measures, such as z-scores or Craig's Zeta, to allocate a distinctiveness score to each term. It then produces a list of words ranked by their scores, indicating which words are characteristic of the primary set (positive

scores: words preferred) and which are characteristic of the secondary set (negative scores: words avoided).

Test IV uses the Oppose function to contrast Orixe's translation of *Mireio* from French (primary set) with his work in Basque (secondary set). Test V compares Orixe's translation of *Lazarillo de Tormes* from Spanish (primary set) with his original work (secondary set). The results of these tests are presented in Figures 4 and 5.
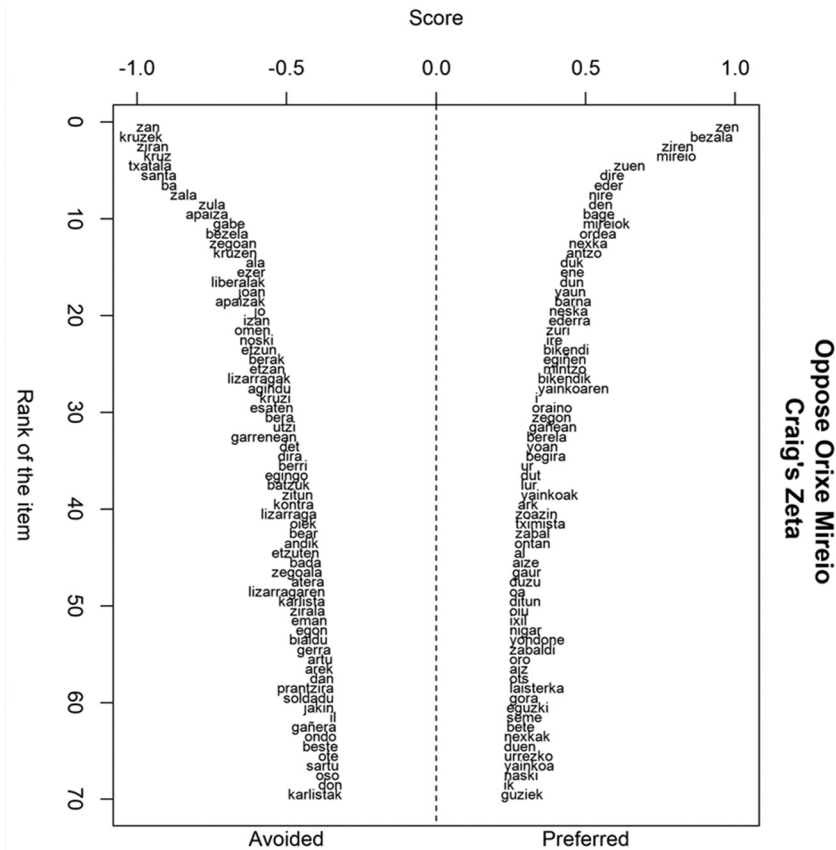
Figure 4: Results of Oppose on *Mireio* (words preferred) and *Santa Kruz apaiza* (avoided).
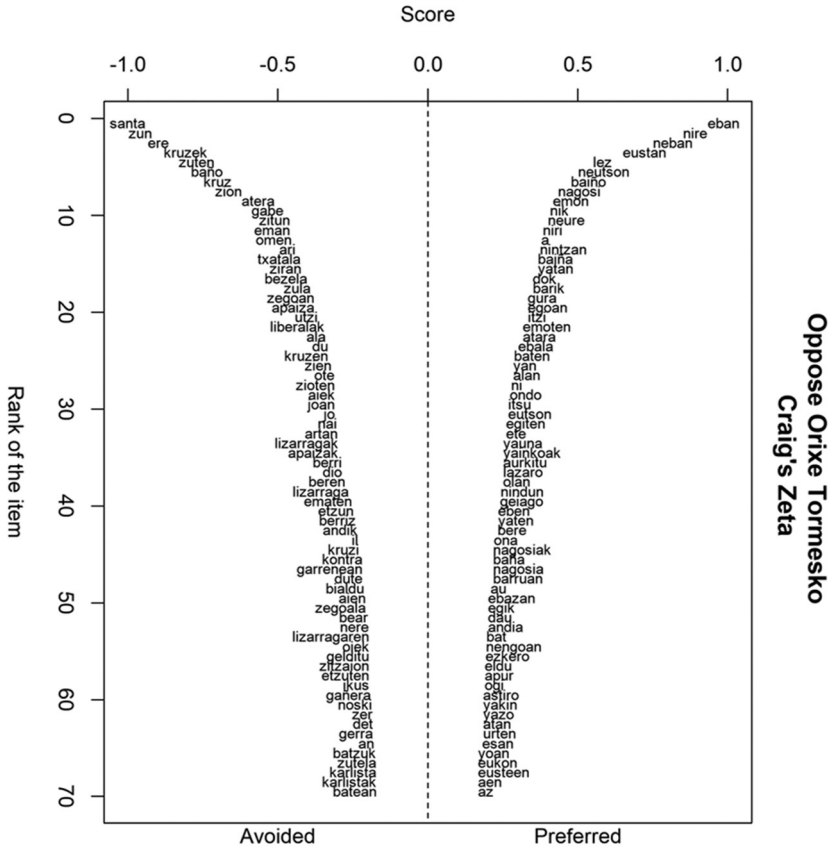
Figure 5: Results of Oppose on *Tormesko itsu-motila* (words preferred) and
Santa Kruz apaiza (avoided).

The words presented on the right side of each Figure are the ones pref-
erentially used in Orixe's translations. Meanwhile, the words on the
left are preferred in *Santa Kruz apaiza.* The farther away we move from
the center line, the stronger the association with the corresponding set
of texts.

The words presented on the graphs belong to various grammatical
categories. The bulk of these are nouns and verbs, but there are also
adjectives, adverbs, prepositions, conjunctions, and interjections. Nouns
mainly indicate the subject matter of the works analyzed. It is of little
surprise that the words "kruz" 'cross,' "apaiza" 'priest,' "santa" 'holy,'
"errege" 'king,' "bandera" 'flag,' "meza" 'Mass,' and "gerra" 'war' should
abound in *Santa Kruz apaiza*, a novel about a priest and guerrilla fighter

of the name Manuel Santa Kruz. In a similar vein, *Moreio*, a Provencal literary text about thwarted love, is only expected to contain words such as "nexka" 'girl,' "nigar" 'tear' (or 'cry'), "ark" 'chest,' "lore" 'flower,' and "izar" 'star,' while a Spanish picaresque novel about a young boy serving various masters in a hypocritical and corrupt society is bound to be filled with nouns such as "gura" 'desire' (or 'will'), "dirua" 'money,' "goseak" 'hunger,' "yazo" 'place,' and "gizagaxua" 'unfortunate man.' Yet even among the nouns, some indications of borrowings can be observed. Place names, for example, give away the language of the original work. In *Mireio*, the proper name "Mireio" is one of the preferred words ("Mireille" in French), as is "provenzako" 'Provencal.' Both hint at the origins and the setting of the story. Similarly, in *Santa Kruz apaiza*, the typical Basque last name Lizarraga appears with greater frequency, as does "euskaldunak" 'Basque speakers.' Finally, in Orixe's translation of *Lazarillo de Tormes* many words appear which seem to be closely connected to Spanish, such as "aiz" ('axe'; 'hacha' in Spanish), "oa" ('wave'; 'ola' in Spanish), "aen" ('breath,' 'air'; 'aliento' in Spanish), "abade" ('abbot'; 'abad' in Spanish), "yazo" ('place'; in Spanish, 'yacer' means to lie in a resting place, especially in the context of a burial). While these could be examples of Spanish influence, it is more likely that some of them entered Basque from Latin during Roman colonization (Trask, *History* 189, 259–61).

However, translationese often manifests itself in function words and syntax, areas prone to unnoticed errors because, as James Pennebaker notes, our brain is not wired to notice those "small, stealthy words" that account for less than 1‰ of our vocabulary while making up almost 60% of the words we use (Pennebaker ix). Morphology and syntax also seem more susceptible to translationese, given that they are more deeply ingrained than lexis and more closely tied to early cognitive development.

For example, in his translation of *Lazarillo de Tormes*, Orixe seems to overuse the past form of the verb *eduki* 'to have,' such as *eban*, *neban*, *ebala*, *ebazan*, and *ebanean*. This may be due to the fact that Spanish has a tendency to express narratives in the past tense. Meanwhile, Basque has multiple ways of expressing the past, and a literal translation from a language like Spanish may lead to an over-reliance on this tense.

Another example are possessive forms such as *nire* 'my,' *neure* 'my own,' and *niri* 'to me.' These tend to be more explicitly stated in Spanish than in Basque, as the latter often conveys possession by using a combination of verb forms and suffixes attached to nouns. Orixe's translation overuses these forms, no doubt because the author allows himself to be ruled by Spanish morphology and syntax.

As we can see, Oppose allows us to direct our attention to places where translationese might occur. While it does not answer the question of what translationese is, it provides a glimpse into the fabric of the language and suggests where the tears might be. The question which could be asked at the end of Tests IV and V is where Orixe's works would be positioned in a corpus comprising original Basque works and translations into Basque: would they align with the translations, assuming these form a distinct cluster in relation to original Basque texts?

Test VI

To examine the positioning of translated works within a mixed corpus, Basque novels were analyzed alongside literature translated into Basque.

Corpus

The literature was procured from Armiarma, spanning from Shakespeare's *Hamlet* to Steinbeck's short story cycle, *The Pastures of Heaven*. 49 works were added into the corpus from English (18), French (12), Spanish (6), Russian (5), German (3), Czech (1), Danish (1), Italian (1), Norwegian (1), and Portuguese (1). Prominent translators included Anton Garikano (who translated two works from English and two from German) and Jose Morales Belda (translator of Russian works).

Method

A series of cluster analyses ranging from 100 to 1,000 MFWs and employing Cosine Delta was visualized as a bootstrap consensus tree (Figure 6). The analysis was done without culling.

Figure 6: Stylometric analysis of Basque literature and translated literature (no culling).

Results

The bootstrap consensus tree revealed a division between original Basque prose and prose translated into Basque. Original Basque works predominantly clustered by author, showing minimal chronological progression, except for the distinct grouping of twenty-first-century texts. Notably, contemporary works by Borde, Larrazabal, and Olabarri clustered with translated literature. Azkue's *Ardi galdua* aligned with Erkiaga's *Araibal zalduna* among the Basque originals, as did Orixe's translations.

Test VII

Corpus

For the final analysis, a bootstrap consensus tree was conducted using the corpus from Test VI, encompassing original Basque novels and Basque translations from English, Spanish, French, Russian, German, Czech, Danish, Italian, Norwegian, and Portuguese. The attribution analysis was executed multiple times using varying vectors of MFWs and culling.

Method

The Delta measure was applied starting at 100 MFWs and increasing by increments of 100 up to 2,000 MFWs. Culling was also progressively applied starting from 0% to 100% in increments of 20. The data obtained was entered into Gephi and a map of literature in Basque was created using network analysis (see Figure 7).
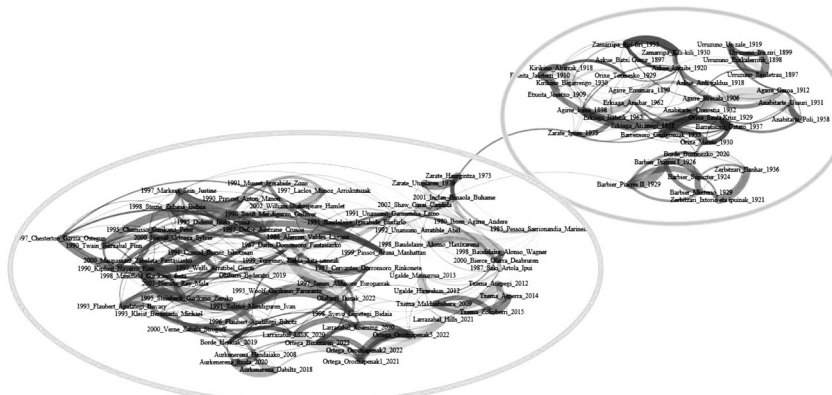
Figure 7: A network analysis of literature in Basque (with culling).

Results

The map obtained via network analysis illustrates the distribution of Basque novels. Two primary clusters are evident: the smaller, rightmost cluster seems to primarily contain original Basque works, while

the leftmost cluster comprises mostly translated works. Some newer Basque novels (penned in the twenty-first century) group with the translated works.

A modest chronological pattern can be noted among recent novels, but no clear temporal sequence is evident in the cluster of original Basque works. This may reflect Lasagabaster's view of the Basque novel's developmental discontinuity and lack of logical progression. Yet at the same time, it is important to acknowledge that the Basque corpus may simply be influenced by stylistic elements that transcend simple chronological development. By these I mean the various dialects spoken within the Basque Country. More research is needed to determine that; it is worth noticing, though, that, compared to works predating the standardization of Basque, the recent novels written in unified or standardized Basque show a tendency toward chronological clustering.

Conclusions

Firstly, let me begin by making the obvious, yet nevertheless very important observation: stylometric methods such as cluster analysis effectively identify authorship in Basque literature, despite the language's complex inflectional and agglutinative structure. Although authorship attribution accuracy was initially not as high for the corpus as that for English texts, culling has helped to increase it to almost 93%.

Secondly, culling has shown to be useful for discriminating between texts in different dialects of the same language. By allowing users to exclude features which do not meet a specified occurrence threshold, culling leads to more accurate clustering.

Thirdly, in agreement with Moretti, it cannot be denied that visual representations such as maps, graphs, and trees reveal deeper patterns not discernible through text alone. For example, cluster analysis helped us notice that some of Orixe's works were in fact translations. A bootstrap consensus tree not only corroborated these findings but also pointed to strong dialectal influence in the case of the novels *Ardi galdua*, *Araibar zalduna*, and *Garoa*. Furthermore, the dialectical signal was found to be stronger than the authorial one, as some works clustered according to dialect and apart from other works by the same author.

Fourthly, stylometric analysis of Orixe's works has revealed instances of translationese (from French and Spanish). The word lists obtained using the Oppose function give insights into Orixe's language use, showing that the frequency of his words changes when translating as

opposed to writing. While it does not provide ready answers, the analysis suggests there may be identifiable patterns which distinguish Orixe's writing from his translations and points to various paths which future work in translation and comparative literature research could take.

Finally, the map of Basque works and translated literature has shown a unique character of pre-twenty-first-century Basque novel, which for the most part forms a separate cluster on the map. Twenty-first-century Basque works have been shown to cluster with translated literature. Euskara batua and its increasing popularity as well as ongoing globalization may be the underlying reasons for the loss of distinction between newer novels penned in Basque and the translations. This may be a consequence of what literary historians have viewed as a developmental discontinuity in the history of the Basque novel, or it may be an effect of dialect impact on the corpus, or perhaps of both.

To conclude, besides providing answers to a series of questions, the research presented in this article points toward a vast field of the unknown and invites further stylometric research of Basque literature. What exactly is the impact of dialect on the linguistic style of Basque authors? To what extent does it influence the stylometric results and to what extent are these results a reflection of the Basque novel's unique history? How strong is the potential linguistic and sociocultural contamination of Basque literature by French and Spanish? To what degree can this influence be evidenced by a stylometric comparison of Basque novels with translations from French and Spanish? How prominent is euskara batua in contemporary Basque literature? And last but not least, what would a map of all Basque works look like? For so-called small literatures, there exists not only the temptation but the very real possibility of mapping the entire terrain. Such a map could shed new light on both traditional literary studies and stylometric research itself.

WORKS CITED

Bastian, Mathieu, et al. "Gephi: An Open Source Software for Exploring and Manipulating Networks." *Proceedings of the International AAAI Conference on Weblogs and Social Media*, vol. 3, no. 1, 2009, pp. 361–362.

Benzine, Vittoria. "The Earliest Words Ever Written in Basque Have Been Found Engraved on a Revelatory 2,100-Year-Old Bronze Relic." *Artnet News*, 15 November 2022, news.artnet.com/art-world/hand-of-irulegi-earliest-basque-script-navarre-spain-2210963. Accessed 9 January 2024.

Brooks, Cleanth. *William Faulkner: The Yoknapatawpha Country*. LSU P, 1989.

Bulson, Eric. *Novels, Maps, Modernity: The Spatial Imagination, 1850–2000*. Routledge, 2017.

Collins, Roger. *The Basques*. Basil Blackwell, 1986.

Eder, Maciej. "Visualization in Stylometry: Cluster Analysis using Networks." *Digital Scholarship in the Humanities*, vol. 32, no. 1, 2017, pp. 50–64, https://academic.oup.com/dsh/article/32/1/50/2957386. Accessed 9 January 2024.

Eder, Maciej, and Jan Rybicki. "Do Birds of a Feather Really Flock Together, or How to Choose Training Samples for Authorship Attribution." *Literary and Linguistic Computing*, vol. 28, no. 2, 2013, pp. 229–236.

Eder, Maciej, et al. "Stylometry with R: A Package for Computational Text Analysis." *The R Journal*, vol. 8, no. 1, 2016, pp. 107–121, https://journal.r-project.org/archive/2016/RJ-2016-007/index.html. Accessed 9 January 2024.

Egurtzegi, Ander. "Metathesis of Aspiration as the Source of Anticipatory Voicelessness in Basque." *Journal of French Language Studies*, vol. 29, no. 2, 2019, pp. 265–279.

Evert, Stefan, et al. "Towards a Better Understanding of Burrows's Delta in Literary Authorship Attribution." *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*, edited by Anna Feldman et al., Kerrville (TX), The Association for Computational Linguistics, 2015, pp. 79–88, https://aclanthology.org/W15-0709/. Accessed 9 January 2024.

Jansen, Wim. *Beginner's Basque*. Hippocrene Books, 2007.

Juvan, Marko. *Worlding a Peripheral Literature*. Palgrave Macmillan, 2019.

Kurlansky, Mark. *The Basque History of the World*. Penguin, 2001.

Lasagabaster, Jesús María. "Introduction: Basque Literary History." *Basque Literary History*, edited by Mari Jose Olaziregi, translated by Amaia Gabantxo, Center for Basque Studies, University of Nevada, 2012, pp. 13–21.

Larramendi, Manuel de. *De la antiguedad y universalidad del bascuenze en España: de sus perfecciones y ventajas sobre otras muchas lenguas*. Vol. 1, Por Eugenio Garcia de Honorato, 1728.

Michelena, Luis. "Lengua común y dialectos vascos. Berrarg." *Obras completas*, vol. 7, edited by Joseba A. Lakarra and Íñigo Ruiz Arzalluz, Universidad del País Vasco, 1981, pp. 291–313.

Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, 2005.

Mukarovsky, Hans G. "Outline of a Lexicostatistical Study of Basque and the Mande Languages, with a Note on Fula." *Euskalarien nazioarteko jardunaldiak*, Blbao, Euskaltzaindia, 1981, pp. 199–212, https://www.euskaltzaindia.eus/dok/ikerbilduma/6777.pdf. Accessed 9 January 2024.

Olaziregi, Mari Jose. "Worlds of Fiction: An Introduction to Basque Narrative." *Basque Literary History*, edited by Mari Jose Olaziregi, translated by Amaia Gabantxo, Center for Basque Studies, University of Nevada, 2012, pp. 137–200.

Pennebaker, James W. *The Secret Life of Pronouns*. Bloomsbury, 2011.

Rask, Rasmus. *Investigation of the Origin of the Old Norse or Icelandic Language*. Translated by Niels Ege, John Benjamins, 2013.

Rybicki, Jan. "A Second Glance at the Stylometric Map of Polish Literature." *Forum Poetyki*, no. 10, 2017, pp. 6–21, http://fp.amu.edu.pl/a-second-glance-at-a-stylometric-map-of-polish-literature/#sdfootnote1sym. Accessed 9 January 2024.

Rybicki, Jan. "Pierwszy rzut oka na stylometryczną mapę literatury polskiej." *Teksty drugie*, no. 2, 2014, pp. 106–128.

Rybicki, Jan. "Stylometric Translator Attribution." *The Translator and the Computer*, edited by Tadeusz Piotrowski and Łukasz Grabowski, Wydawnictwo Wyższej Szkoły Filologicznej we Wrocławiu, 2013, pp. 193–204.

Trask, R. L. *The History of Basque*. Routledge, 1997.

Trask, R. L. "Origins and Relatives of the Basque Language: Review of the Evidence." *Towards a History of the Basque Language*, edited by José Ignacio Hualde et al., John Benjamins, 1995, pp. 65–100.

Vennemann, Theo. *Europa Vasconica—Europa Semitica*. Edited by Patrizia Noel Aziz Hanna, de Gruyter, 2003.

Zuazo, Koldo. *Euskalkiak, herriaren lekukoak*. Donostia, Elkar, 2003.

Stilometrični pogled na baskovski roman

Čeprav velja baskovščina za verjetno najstarejši jezik na evropski celini, se v pisni obliki pojavi šele v šestnajstem stoletju. Prvi baskovski roman je izšel šele dobrih 300 let pozneje in žanr do danes ni bil izčrpno literarnovedno raziskan. Članek si za cilj zastavi stilometrično analizo izbranih baskovskih romanov 20. in 21. stoletja, pridobljenih s spletnih platform Armiarma in Booktegi. Ti so analizirani na podlagi pogostosti najpogostejših besed s klastrsko analizo (ang. *cluster analysis*), sledi primerjava s tujimi romani, prevedenimi v baskovščino. Rezultati kažejo, da se baskovski izvirniki jasno razlikujejo od prevedenih del, kar dokazuje edinstven jezikovni značaj baskovskega romana. Predstavljenih je nekaj jezikovnih vzorcev, ki bi lahko bili razlog za to razlikovanje. Vizualizacije rezultatov razkrivajo kronološki razvoj baskovskega romana in njegov prispevek k širši literarni krajini.

Analiza 90 slovenskih romanov in opusa Ivana Cankarja z računalniško stilometrijo

Ivana Zajc

Univerza v Novi Gorici, Raziskovalni center za humanistiko, Vipavska 13, 5000 Nova Gorica, Slovenija
https://orcid.org/0000-0002-5502-6772
ivana.zajc@ung.si

Članek predstavi oddaljeno branje slovenske književnosti iz obdobij slovenskega literarnega realizma in moderne. Prvi del članka vsebuje prikaz rezultatov računalniške stilometrične analize korpusa 90 slovenskih romanov iz omenjenih obdobij, izvedene v programskem okolju R s paketom Stylometry with R. Analiza je pokazala, da je ključni signal, po katerem se vključeni romani medsebojno razlikujejo, avtorski, pri čemer najbolj izstopajo romani Pavline Pajk in Ivana Cankarja. Ker je izstopanje najbolj izrazito pri Pavlini Pajk, članek predstavi tudi stilometrično raziskavo besedišča, ki je za njene romane najznačilnejše. Na tej podlagi članek ponudi sklep, da k specifičnosti njenih romanov med drugim prispeva žanr sentimentalnega ljubezenskega romana, ki zaznamuje njen opus. V drugem delu članka je z isto metodo analiziran Cankarjev literarni opus, v korpus pa so poleg romanov vključena tudi njegova druga dela iz različnih ustvarjalnih obdobij. Iz rezultatov analize sta razvidna obdobni in žanrski razvoj Cankarjevega literarnega sloga.

Ključne besede: slovenski roman / Cankar, Ivan / oddaljeno branje / računalniška stilometrija / digitalna humanistika

Digitalne metode v literarni vedi

Kadar k literarnim delom pristopamo s pomočjo oddaljenega branja, jih raziskujemo s kvantitativnimi metodami, v sodobnem času z uporabo računalniške tehnologije.[1] Pristop oddaljenega branja je leta 2000 predlagal Franco Moretti, pozneje pa je o tovrstnem raziskovanju literature Matthew L. Jockers vplivno govoril kot o makroanalizi, Stephen

---

113

Ramsay pa kot o algoritemski kritiki. Prvi poskusi oddaljenega branja so bili izvedeni preštevno, brez uporabe računalniške tehnologije,[2] danes pa kvantitativne raziskave praviloma analizirajo obsežnejše literarne zbirke ali druge podatke s pomočjo digitalnih orodij in statističnih pristopov. Pri sodobnem oddaljenemu branju ne gre le za analizo besedišča v književnih delih, ampak za razumevanje pisanja kot kompleksnega polja odnosov, ki ga je mogoče kvantitativno modelirati (Underwood, »Distant Reading«). Tovrstne raziskave se lahko nanašajo na različne elemente, na primer na korpus literarnih besedil v celoti, na dele literarnega besedila (npr. na dialoge ali karakterizacijo likov), na posamezno besedilo ali na metapodatke o književnosti in literarnem sistemu. Raziskave se lahko ukvarjajo tudi z ustaljenimi predpostavkami literarne zgodovine, ki jih lahko s pomočjo digitalnih tehnologij prevprašujemo, na primer s časovnim potekom literarnozgodovinskih obdobij in smeri ter z njihovimi lastnostmi. S tem področje literarne vede obogatimo z novim tipom kvantitativnih podatkov (Eve 153), ki jih interpretiramo s kvalitativnimi metodami (Underwood, »Distant Reading«). Kvantitativni podatki so se v literarni vedi uporabljali že prej, vendar so se z vzponom digitalnih informacijskih tehnologij možnosti za raziskave razširile. Massimo Salgaro izpostavi naslednje razlike med kvantitativnim in kvalitativnim raziskovanjem v literarni vedi: drugačen odnos med teorijo in raziskovanjem: kvantitativne raziskave so deduktivne in strukturirane po stopnjah, medtem ko je kvalitativno raziskovanje bolj odprto; različne tipe konceptov: kvantitativno raziskovanje predvideva operativne koncepte, medtem ko so koncepti pri kvalitativnem raziskovanju odprti in se nenehno razvijajo; različne tipe psihološke interakcije med raziskovalcem in opazovanim fenomenom: kvantitativno raziskovanje je znanstveno in nevtralno, medtem ko kvalitativno raziskovanje predvideva empatično identifikacijo s predmetom raziskovanja oziroma prevzemanje njegove perspektive; razlike v reprezentativnosti podatkov: kvantitativno raziskovanje teži k statistično reprezentativnim vzorcem, medtem ko kvalitativno raziskovanje izhaja iz posamičnih primerov, ki niso statistično pomembni; razlike v odnosu do matematičnih in statističnih tehnik: v nasprotju s kvalitativnim raziskovanjem kvantitativno raziskovanje te tehnike intenzivno uporablja; različne implikacije rezultatov: pri kvantitativnem raziskovanju so rezultati posplošeni, pri kvalitativnem pa specifični (Salgaro 51).

---

[2] Za genealogijo oddaljenega branja od starejših analognih tehnik do sodobnih računalniških metod gl. Underwood, »A Genealogy«.

Bemma Adwetewa-Badu opozarja, da digitalna orodja ne morejo opravljati kritičnega in analitičnega dela, ki ga opravljajo literarni raziskovalci, ki pa po drugi strani prav tako ne morejo opravljati funkcij digitalnih orodij (Adwetewa-Badu). Digitalne metode v literarni vedi s tega vidika ne prelamljajo s tradicionalnim bližnjim branjem in kvalitativnimi literarnovednimi metodami, ampak ponujajo dodatne podatke, ki lahko te raziskave poglobijo. Med kritikami digitalnih raziskav književnosti so na primer opozorila, da vzorci, ki jih v obsežnih zbirkah literarnih del najdevajo statistični algoritmi, niso pomenljivi (Nan 605), poleg tega digitalni humanistiki t. i. analogna humanistika očita »teoretsko podhranjenost, opiranje na zastarela pojmovanja teksta, naivni realizem ter trivialnost, nezadostnost ali celo zgrešenost računalniških rezultatov, na prvi pogled neovrgljivih« (Juvan, Šorli in Žejn 55). S strani literarne vede se pojavlja očitek, da se digitalnohumanistične raziskave ne naslanjajo na literarnovedne ugotovitve, pač pa zgolj prikazujejo podatke (Nan 605), oziroma da lahko raziskovalec med številnimi metodami računalniške analize besedil izbere tisto, ki najbolj podpre njegovo tezo (Juola, »The Rowling Case« 102). Bližnje branje nam omogoča vpogled v omejeno število literarnih del, ki jih lahko posameznik prebere, velika količina besedil v literarni zgodovini pa je ob tem spregledana kot del tega, čemur Margaret Cohen in za njo Moretti pravita »veliko neprebrano« (Moretti 8). Literarne raziskave se tako osredotočajo na besedila, ki so izpostavljena v kanonu, vendar so po obsegu omejena, njihova izbira pa je pristranska (Eve 9). A niti oddaljeno branje samo po sebi ne opravi te težave, saj ne vključuje vseh obstoječih literarnih besedil, ampak samo tista, ki so digitalizirana, kar so običajno predvsem kanonizirana besedila. Poleg tega ni konsenza o merilih za določanje reprezentativnosti korpusa. Slovenska književnost je v primerjavi z literaturami v jezikih, ki jih govori veliko več govorcev, v tem pogledu izjema, saj je digitaliziran že velik delež starejše književnosti.

Računalniška stilometrija

Sodobna stilometrija omogoča računalniško analizo literarnega stila (Eder idr. 107). S to metodo pogosto določamo avtorstvo literarnih del; med znanimi primeri so prepoznavanje avtorice knjig o Harryju Potterju J. K. Rowling, ki je po letu 2012 začela pisati pod psevdonimom Robert Galbraith (prim. Juola, »The Rowling Case«; Juola, »Rowling«) in preverjanje avtorstva del Shakespearea in Marlowa (Fox

idr.). Stilometrična analiza upošteva in primerja pogostost pojavljanja posameznih besed v vseh delih, vključenih v korpus. Individualen avtorski slog sicer pomembno določajo funkcijske besede, ki pogosto nimajo leksikalnega pomena (Kestemont). V dosedanjih raziskavah slovenske književnosti je Andrejka Žejn (Žejn, »Računalniško podprta«) z računalniško stilometrijo primerjalno analizirala literarna dela Janeza Ciglerja in Christopha Schmida, obravnavala pa je tudi pripovedno prozo od sredine 17. do sredine 19. stoletja (Žejn, *Izhodišča*). Stilometrična raziskava je podprla tudi analizo dramatike avtorice Simone Semenič (glej Zajc), z metodo »rolling stylometry« pa je bil obravnavan roman *Rokovnjači* Josipa Jurčiča in Janka Kersnika (Mandić in Zajc).

Tako kot omenjene raziskave je bila tudi raziskava, predstavljena v tem članku, izvedena v paketu Stylo v programskem jeziku R, ki ga je razvila poljsko-belgijska Skupina za računalniško stilistiko (Eder idr.) za analizo literarnih besedil na podlagi statističnih izračunov pogostosti besed ali zlogov.[3] Z meritvijo uporabe najpogostejših besed v besedilih iz korpusa so besedila razporejena glede na slogovne podobnosti; v vizualnem prikazu so tako dela, ki so si glede na uporabo besed podobna, bližje drugo drugemu. Z uporabo klastrske analize so podatki o romanih razvrščeni glede na njihove podobnosti ali razlike. Analiza upošteva in primerja frekvenco pojavljanja besed enakovredno v vseh delih korpusa. Članek prikazuje klastrsko analizo dveh korpusov slovenskih literarnih del, kjer so bili podatki o romanih razvrščeni glede na njihove podobnosti ali razlike. Analiza je temeljila na relativni frekvenci od 100 do 1000 najpogostejših besed v posameznem korpusu. Uporabljena je bila statistična metoda za merjenje razdalje Delta (prim. Burrows; Evert idr.). Rezultati raziskav so predstavljeni v obliki dendrograma in v obliki dvodimenzionalne vizualizacije bližine oziroma oddaljenosti vključenih literarnih besedil. Stilometrična analiza samodejno razporedi besedila glede na t. i. signale, ki izstopajo s tem, da določajo razvrščanje besedil glede na njihovo oddaljenost oziroma bližino z vidika stila.

Korpus 90 slovenskih romanov

V prvem delu raziskave sem oblikovala korpus 90 slovenskih romanov iz 19. in začetka 20. stoletja: romane sem shranila v besedilne datoteke (txt), urejene na način, ki je dostopen paketu Stylo in omogoča tudi

---

[3] Ta del raziskave, predstavljene v članku, je rezultat sodelovanja v akciji COST Distant Reading for European Literary History (CA16204) pod mentorstvom dr. Joanne Byszuk z Inštituta za poljski jezik Poljske akademije znanosti v Krakovu.

vizualizacijo v programski opremi Gephi. Besedila sem za analizo pripravila tako, da sem korpus besedil pregledala in uredila, iz njih izključila paratekst ter jih shranila v format txt in specifično urejene mape, do katerih program lahko dostopa. Besedila v korpusu sem črpala iz različnih virov: iz slovenskega dela Evropske zbirke literarnih besedil (ELTeC-slv)[4] in iz literarnih besedil, ki so prosto dostopna na slovenskem Wikiviru.[5] Iz tega osnovnega korpusa sem ustvarila druge sklope besedilnih korpusov za izvajanje specifičnih analiz. Korpus 90 romanov vsebuje dela slovenskih avtoric in avtorjev iz obdobij realizma in moderne.

Signal predstavlja merljiv podatek, ki ga lahko razberemo tudi iz vizualizacij, na primer kot jasne razdalje med določenimi skupinami besedil, kar nakazuje na njihovo slogovno raznolikost (prim. Evans). Kadar se literarna dela razlikujejo predvsem glede na njihovo avtorstvo, gre za avtorski signal, kadar pa se razlikujejo na primer glede na žanr ali spol, gre za žanrski oziroma spolni signal (prim. Jockers). Poleg tega se lahko pokaže izrazit vzorec kronološkega razvoja literarnega sloga posameznega avtorja (Evans). Koncept po drugi strani določen signal pojasnjuje, na primer z ugotovitvijo, da besedila, med katerimi so očitne razlike, spadajo v različne literarne žanre. V digitalni humanistiki tvorimo argumente na osnovi konceptov, medtem ko lahko računalniki izmerijo le signale (Heuser in Le-Khac).

S stilometrično analizo 90 slovenskih romanov odgovarjam na naslednji raziskovalni vprašanji: ali najdemo signale, ki niso povezani zgolj z avtorstvom, na primer spolne signale? in ali kateri od avtorjev in avtoric iz korpusa odstopa glede na svoj slog pisanja?

Stilometrična analiza 90 slovenskih romanov: rezultati in razprava

Slika 1 prikazuje rezultate raziskave: posamezna točka označuje določeno literarno besedilo, ki vsebuje pripis avtorja in naslova, črte pa označujejo povezave med besedili glede na splošno slogovno podobnost. S temnejšim barvnim odtenkom so označena dela avtoric, s svetlejšim pa dela avtorjev.

---

[4] Korpus vsebuje 100 daljših proznih del v slovenskem jeziku, ki so izšla med letoma 1836 in 1921.

[5] Wikivir kot projekt Mirana Hladnika je »pomemben za razumevanje branja v digitalnem okolju, ker si zadaja cilj ustvarjanja edinstvene spletne knjižnice slovenske književnosti, uporabnicam in uporabnikom pa omogoča, da se srečajo z besedili, ki niso znana ali kanonizirana« (Ilin 195).
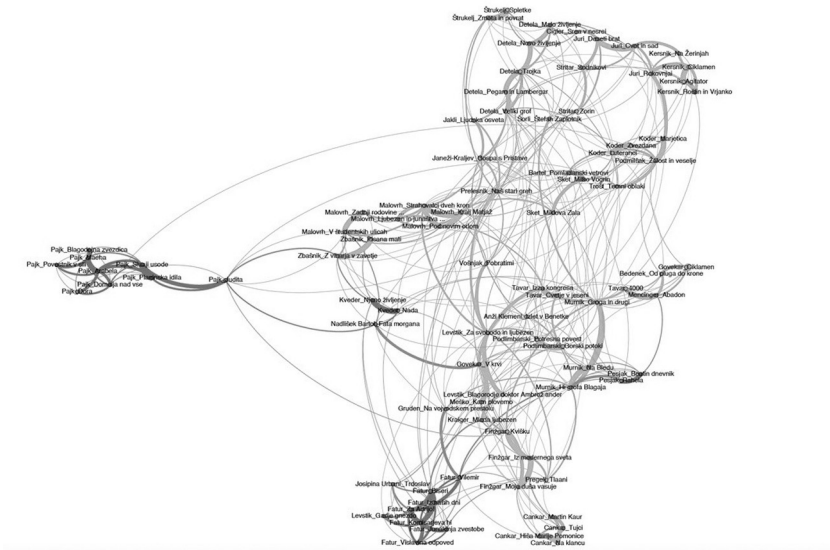
Slika 1: Stilometrična analiza 90 slovenskih romanov.

V zgornjem delu prikaza vidimo skupino avtorjev slovenskega realizma, od katerih sta si po stilu na primer blizu Josip Jurčič in Janko Kersnik. Njun roman *Rokovnjači*, ki ga je začel pisati Jurčič in dokončal Kersnik, je razvrščen med Kersnikova dela.[6] Deloma se kaže signal spola, ki pa ni izrazit, saj se skupaj razvrščajo le nekatere avtorice, med njimi pa so tudi avtorji, tako da se na primer delo Frana Levstika *Gadje gnezdo* razvrsti v bližini romanov Lee Fatur in Josipine Turnograjske. Med drugim so dela Zofke Kveder slogovno blizu delu Marice Nadlišek Bartol, dela Miroslava Malovrha so blizu delom Frana Zbašnika, dela Frana Saleškega Finžgarja pa romanu Ivana Preglja *Tlačani*. Kot je razvidno iz vizualizacije, spol ni ključen signal za razvrščanje teh del v skupine, temveč je najizrazitejši signal njihovo avtorstvo: na primer besedila Ivana Cankarja ali Lee Fatur so se razvrstila v skupini v spodnjem delu prikaza. Skrajno levo so se v ločeno skupino razvrstila dela Pavline Pajk, iz česar lahko razberemo, da so izrazito stilno drugačna od ostalih besedil v tem korpusu. Sklepamo, da je razlogov za to več: Pavlina Pajk se je slovenskega jezika, v katerem je pozneje pisala romane, naučila šele pri 16 letih, njen prvi jezik pa je bila italijanščina; otroštvo je preživela v Milanu in Trstu, po smrti staršev je živela pri stricu v Solkanu in nato

---

[6] Za natančnejšo stilometrično analizo Jurčič-Kersnikovega romana *Rokovnjači* gl. Mandić in Zajc 2020.

pri bratu, hodila je v uršulinsko šolo v Gorici; zaradi specifične jezikovne situacije v njenem otroštvu in mladosti sklepamo, da je bila tudi njena raba slovenščine specifična; poleg tega je pisala žanrsko drugačna besedila kot drugi avtorji in avtorice v korpusu, in sicer sentimentalne ljubezenske romane. Da bi pridobila še dodatne informacije o rabi jezika pri Pavlini Pajk, sem izvedla dodatno raziskavo, pri kateri sem uporabila funkcijo Oppose iz paketa Stylo, ki omogoča, da ugotavljamo razlike med najpogosteje uporabljenimi besedami v posamezni skupini literarnih besedil oziroma vidimo, katere besede se v posamezni skupini besedil pojavljajo bistveno pogosteje kot v drugi skupini. Ustvarila sem korpus besedil Pavline Pajk na eni strani (t. i. primarni sklop) in enako količino drugih literarnih besedil različnih njenih sodobnikov na drugi strani (t. i. sekundarni sklop), da bi izvedla primerjalno analizo besedišča del Pavline Pajk v primerjavi z drugimi reprezentativnimi romani tega obdobja slovenskega realizma.[7] Ugotavljala sem, katere besede so najznačilnejše za dela Pavline Pajk, tj. primarnega sklopa, kar pomeni, da so v besedilih sekundarnega sklopa redke, pa tudi, katere besede so najznačilnejše za sekundarni sklop in torej redke v besedilih primarnega sklopa. Naslednja slika prikazuje seznam najpogostejših besed, ki se jim primarni sklop izogiba (na levi strani), in seznam besed, ki so v primarnem sklopu pogostejše v primerjavi s sekundarnim sklopom (na desni strani).

---

[7] Janez Cigler, *Sreča v nesreči*; Fran Detela, *Malo življenje*; Fran Govekar, *V krvi*; Josip Jurčič, *Deseti brat*; Janko Kersnik, *Agitator*; Anton Koder, *Luteranci*; Fran Levstik, *Gadje gnezdo*; Josip Stritar, *Sodnikovi*.
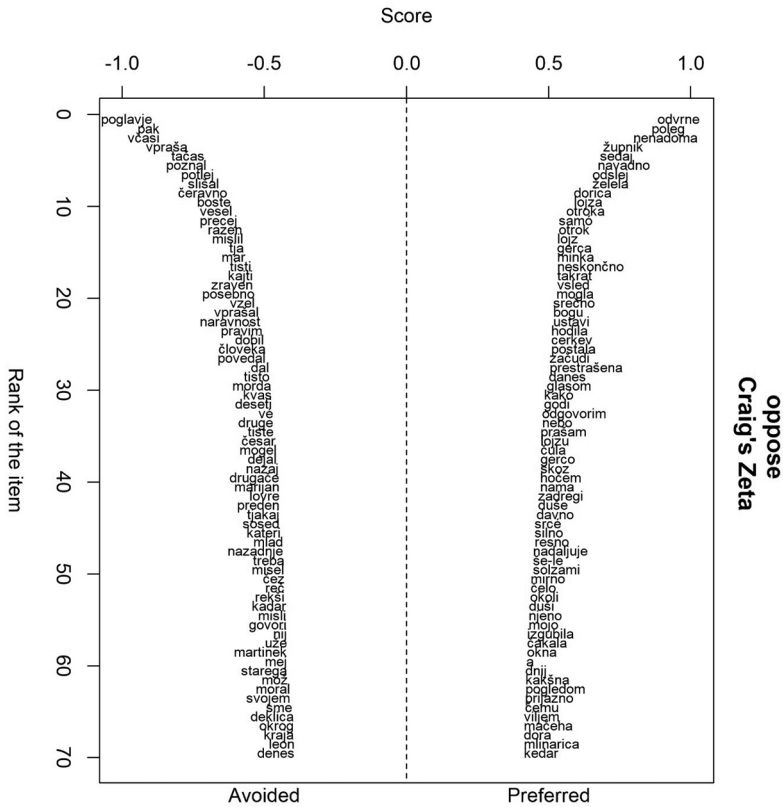
Slika 2: Besedišče del Pavline Pajk v primerjavi z drugimi besedili njenega časa.

Kot lahko vidimo na Sliki 2, med najbolj priljubljenimi besedami v korpusu del Pavline Pajk najdemo besede, povezane z ženskimi liki – omembe ženskih literarnih oseb (»mačeha«, »Dora«, »njeno« ipd.) ter glagolske oblike in pridevnike, ki se nanašajo na ženske like (»prestrašena«, »čakala«, »hodila«, »postala«, »želela« ipd.) –, oziroma besede, ki so značilne za žanr sentimentalnega romana (»srce«, »duše«, »solzami« ipd.), omembe župnika, cerkve, otrok ipd. Avtoričina specifična raba jezika sicer zahteva obsežnejše raziskave z bližnjim branjem, ki presegajo namen te študije.

Raziskava korpusa besedil Ivana Cankarja

Čeprav se računalniške stilometrične raziskave najpogosteje nanašajo na dela z različnim (pogosto tudi neznanim) avtorstvom, se v drugem delu raziskave osredotočam na žanrsko raznolik opus posameznega avtorja, in sicer Ivana Cankarja. Tudi ta eksperiment sem izvedla v paketu Stylo v programskem jeziku R.

Cankar velja za enega glavnih predstavnikov slovenske literarne moderne. Njegov opus šteje približno 30 samostojnih knjižnih publikacij, številna dela pa je objavljal tudi v tedanji periodiki. Teme njegovih del so socialne, moralne in nacionalne, pisal je poezijo, dramatiko in prozo, uveljavil je žanr simbolistične skice oziroma črtice.

Dekadenca in simbolizem sta se v slovenski literaturi pojavila po letu 1897, ko je Cankar ustvaril prve pesmi in črtice s tovrstnimi vplivi. Do sredine devetdesetih let 19. stoletja je bilo v ospredju njegove ustvarjalnosti pesništvo: »/s/ pripovedno prozo se do leta 1896 ni intenzivneje ukvarjal in jo je pisal bolj ali manj iz neliterarnih nagibov ter po tradicionalnem realistično-naturalističnem vzorcu, vendar ga je že naslednje leto z bolj ali manj deepiziranimi vinjetami […] začel tudi razkrajati.« (Čeh 45) Te črtice vnašajo prelom v Cankarjevo ustvarjalnost in so zametek poznejše zbirke *Vinjete* (Kocijan 36). Cankarjeve romane po drugi strani zaznamujeta zvrstni sinkretizem, tj. neklasična epska struktura, in žanrski sinkretizem, tj. preplet različnih žanrov, ki so bili v slovenskem prostoru pogosto inovativni (glej Zupan Sosič). Avtorja je močno zaznamovalo bivanje na Dunaju, kjer je s presledki preživel deset let svojega življenja (1898–1909), ob tem pa je »sprejel periferijo Dunaja kot svoj dom, kot samotarski, kontemplativni avtor se ni poskusil uveljaviti v dunajski kulturi, čeprav je prav prek nje sprejemal različne filozofske in literarne tokove« (Jensterle-Doležal 68).

Kot je pokazal prvi del raziskave, se izbor iz Cankarjevega opusa (prozna besedila *Martin Kačur*, *Tujci*, *Hiša Marije Pomočnice* in *Na Klancu*) razlikuje od opusov njegovih sodobnikov, saj se v vizualizaciji stilometrične analize razporedi v svoj sklop (gl. skrajno desno spodaj na Sliki 1), ki je še najbližje Finžgarjevemu delu *Moja duša vasuje* in Pregljevemu delu *Tlačani*. Drugi del raziskave se posveča izključno Cankarjevemu opusu in preverja, ali je stilno homogen ali heterogen. Za namene te stilometrične analize sem oblikovala korpus 47 žanrsko raznolikih Cankarjevih del na podlagi besedil, zbranih v Wikiviru, ki vsebuje mnoga Cankarjeva besedila. Med vključenimi proznimi deli najdemo tako krajše črtice kot romane in druga daljša besedila. Da bi zadostila kriteriju primerljivosti vključenih besedil glede na dolžino,

sem v korpus poleg proznih povesti in romanov vključila zbirke črtic in ne posameznih črtic, enako pa velja tudi za poezijo. Prvo izhodišče raziskave je raznolikost Cankarjevega opusa, ki vsebuje vse od proznih del prek poezije do dram. Poleg tega raziskava upošteva dejstvo, da je literarna zgodovina Cankarjev opus delila na različna ustvarjalna obdobja, in sicer sem izbrala tisto periodizacijo njegovega pisanja, ki je del šolskih kurikulov v slovenskem prostoru, saj gre za uveljavljeno periodizacijo (prim. Bernik, *Obzorja slovenske*). Ta Cankarjevo delo deli na zgodnje obdobje med letoma 1897 in 1900, srednje obdobje med letoma 1901 in 1909 ter pozno obdobje, ki traja od leta 1910 do avtorjeve smrti.[8] Cilj drugega dela raziskave je odgovoriti na naslednji raziskovalni vprašanji: ali lahko Cankarjeva dela uvrstimo v različna ustvarjalna obdobja? In ali lahko Cankarjeva dela uvrstimo v različne literarne vrste oziroma zvrsti?

Korpus del Ivana Cankarja: rezultati in razprava
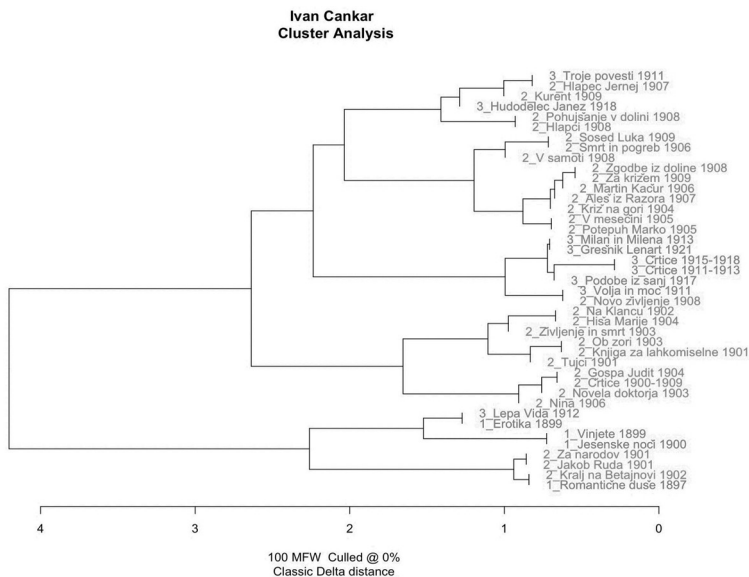
Slika 3: Klastrska analiza del Ivana Cankarja.

S klastrsko analizo sem identificirala skupine literarnih besedil v korpusu, ki so si stilno podobna (Eder 51). Slika 3 prikazuje rezultate na podlagi analize 100 najpogostejših besed; rezultati tovrstne analize so se ujemali tudi, kadar sem uporabila izhodišče 500 najpogosteje uporabljenih besed v posameznih besedilih korpusa. Kot je razvidno iz vizualizacije, se Cankarjev opus deli na dva dela: dela, ki so na sliki skrajno spodaj, spadajo v prvo obdobje njegovega ustvarjanja in se najbolj razlikujejo od ostalih, višje razvrščenih Cankarjevih besedil. Tudi ta se sicer nadalje delijo v dve skupini, ki se kronološko ujemata. V prvo skupino, kjer najdemo predvsem Cankarjeva zgodnja dela, se ob pesniško zbirko *Erotika* (1899) izjemoma razvrsti tudi avtorjeva poznejša drama *Lepa Vida* (1912), kar si razlagam kot avtorjevo vrnitev k poetičnemu slogu, ta drama pa je sicer nastajala dlje časa. Bernik (*Ivan Cankar* 375) opozarja, da se je Cankar z *Lepo Vido* vrnil k poeziji. Poleg tega navaja, da se ta drama vsebinsko povezuje z avtorjevo prvo »dramatično sliko« *Romantične duše*, kar potrjuje tudi stilometrična analiza, ki obe deli razvrsti v bližino (Bernik, *Ivan Cankar* 375). Cankar je v začetnem obdobju poleg poezije ustvarjal tudi dramatiko. Njegovi zgornji drami *Jakob Ruda* (1901) in *Romantične duše* (1897) se skupaj s poezijo iz prvega obdobja razvrstita v bližini drame *Kralj na Betajnovi* (1902), ki očitno stilno spada še v zgodnji del opusa, kar pomeni, da nad žanrskim signalom prevladuje časovni oziroma da se dela stilno razlikujejo glede na čas nastanka. To potrjuje tudi dejstvo, da se tudi drami *Pohujšanje v dolini šentflorjanski* in *Hlapci* razporedita sicer skupaj, kar pomeni, da sta si stilno podobni, vendar se umestita med pripovedna besedila. Ko govorimo o dramatiki, v Cankarjevem opusu torej prevlada časovni signal, saj se dela razvrstijo predvsem glede na obdobje njihovega nastanka, avtorjev dramski opus pa je torej stilno heterogen. Na sredino in v skrajno zgornji del prikaza se razporedijo dela, ki so nastala v srednjem obdobju, kar pomeni, da je to obdobje stilno heterogeno. Izrazito vejo, ki se stilno razlikuje od drugih, predstavljajo črtice, ki so nastale v poznem obdobju, tj. zbirki črtic iz obdobij 1911–1913 in 1915–1918 ter *Podobe iz sanj* (1917), medtem ko se črtice iz obdobja 1900–1909 od drugih črtic razlikujejo in stilno razporedijo med ostala dela iz drugega obdobja Cankarjevega ustvarjanja.

Sklep

Raziskava, predstavljena v prispevku, je bila izvedena v paketu Stylo v programskem jeziku R in se deli na dva eksperimenta: na stilometrično analizo korpusa 90 slovenskih romanov in stilometrično analizo opusa del Ivana Cankarja. Analiza je pokazala, da je ključen signal, po katerem se ta literarna dela razlikujejo, njihovo avtorstvo. Posebej izstopata Pavlina Pajk in Ivan Cankar, saj so njuna besedila v vizualizaciji razvrščena izrazito ločeno od drugih del, vključenih v korpus. Poleg tega se deloma kaže signal spola, ki pa ni izrazit, saj se skupaj razvrščajo le nekatere avtorice, med njimi pa najdemo tudi avtorje: na primer Levstikovo delo *Gadje gnezdo* se razvrsti v bližini romanov Lee Fatur in Josipine Turnograjske. Stilometrična analiza Cankarjevega opusa je pokazala na razpoznavna časovna obdobja njegove ustvarjalnosti. Stilno homogeno in izrazito drugačno od drugih delov Cankarjevega opusa je njegovo prvo ustvarjalno obdobje, v katero se je sicer razvrstila tudi drama iz leta 1902 *Kralj na Betajnovi*. Drugo obdobje je stilno bolj heterogeno, v tretjem obdobju pa kot stilno specifičen del Cankarjevega opusa izstopajo njegove poznejše črtice.

LITERATURA

Adwetewa-Badu, Bemma. »Poetry from Afar: Distant Reading, Global Poetics, and the Digital Humanities«. *Modernism/Modernity Print+*, let. 5, št. 1, 2020, https://modernismmodernity.org/forums/posts/adwetewa-badu-poetry. Dostop 12. 4. 2024.

Avsenik Nabergoj, Irena. »Ivan Cankar med domovino in tujino«. *Dve domovini*, št. 20, 2004, str. 95–111.

Bernik, France. *Obzorja slovenske književnosti*. Slovenska matica, 1999.

Bernik, France. *Ivan Cankar: Monografija*. Litera, 2006.

Burrows, John. »'Delta' – A Measure of Stylistic Difference and a Guide to Likely Authorship«. *Literary and Linguistic Computing*, let. 17, št. 3, 2002, str. 267–287.

Da, Nan, Z. »The Computational Case against Computational Literary Studies«. *Critical Inquiry*, let. 45, št. 3, 2019, str. 601–639.

Čeh, Jožica. »Cankarjeva metafora v vinjetnem obdobju«. *Slavistična revija*, let. 45, št. 1–2, 1997, str. 45–58.

Eder, Maciej. »Visualization in Stylometry: Cluster Analysis Using Networks«. *Digital Scholarship in the Humanities*, let. 32, št. 1, 2017, str. 50–64.

Eder, Maciej, Jan Rybicki in Mike Kestemont. »Stylometry with R: A Package for Computational Text Analysis«. *The R Journal*, let. 8, št. 1, 2016, str. 107–121.

Evans, Mel. »Style and Chronology: A Stylometric Investigation of Aphra Behnʹs Dramatic Style and the Dating of *The Young King*«. *Language and Literature*, let. 27, št. 2, 2018, str. 1–49.

Eve, Martin Paul. *The Digital Humanities and Literary Studies*. Oxford University Press, 2022.

Evert, Stefan, idr. »Understanding and Explaining Delta Measures for Authorship Attribution«. *Digital Scholarship in the Humanities*, let. 32, št. 2, 2017, str. 4–16.

Fox, Neal P., Omran Ehmoda in Eugene Charniak. »Statistical Stylometrics and the Marlowe-Shakespeare Authorship Debate«. *Proceedings of the Georgetown University on Language and Linguistics*, Georgetown University, 2012, https://cs.brown.edu/research/pubs/theses/masters/2012/ehmoda.pdf. Dostop 12. 4. 2024.

Heuser, Ryan, in Long Le-Khac. »A Quantitative Literary History of 2,958 Nineteenth-century British Novels«. *Stanford Literary Lab*, 2012, https://litlab.stanford.edu/assets/pdf/LiteraryLabPamphlet4.pdf. Dostop 12. 4. 2024.

Ilin, Darko. »Wikisource platforma kao biblioteka slovenačke književnosti i kulture«. *Slovenika: časopis za kulturu, nauku i obrazovanje*, št. 9, 2023, str. 195–211.

Jensterle-Doležal, Alenka. »Fenomen mesta v opusu Ivana Cankarja in Zofke Kveder«. *Jezik in slovstvo*, let. 55, št. 5–6, 2010, str. 57–70.

Jockers, Patrick. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013.

Juola, Patrick. »Rowling and 'Galbraith': An Authorial Analysis«. *Language Log*, 16. 7. 2013, http://languagelog.ldc.upenn.edu/nll/?p=5315. Dostop 12. 4. 2024.

Juola, Patrick. »The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions«. *Digital Scholarship in the Humanities*, let. 30, št. 1, 2015, str. 100–113.

Juvan, Marko, Mojca Šorli in Andrejka Žejn. »Interpretiranje literature v zmanjšanem merilu: 'oddaljeno branje' korpusa 'dolgega leta 1968'«. *Jezik in slovstvo*, let. 66, št. 4, 2021, str. 55–76.

Kestemont, Mike. »Function Words in Authorship Attribution: From Black Magic to Theory?«. *Proceedings of the 3rd Workshop on Computational Linguistics for Literature*, ur. Anna Feldman, Anna Kazantseva in Stan Szpakowicz, The Association for Computational Linguistics, 2014, str. 59–66.

Kocijan, Gregor. *Kratka pripovedna proza v obdobju moderne*. Znanstveni inštitut Filozofske fakultete, 1996.

Mandić, Lucija, in Ivana Zajc. »Stilometrična analiza avtorskega sloga Jurčič-Kersnikovega romana *Rokovnjači*«. *Slovansko jezikovno in literarno povezovanje ter zgodovinski kontekst*, ur. Ina Poteko, Študentska organizacija Filozofske fakultete, 2020, str. 7–13.

Moretti, Franco. »Domneve o svetovni literaturi«. *Grafi, zemljevidi, drevesa in drugi spisi o svetovni literaturi*, ur. in prev. Jernej Habjan, Studia humanitatis, 2011, str. 5–25.

Salgaro, Massimo. »The Digital Humanities as a Toolkit for Literary Theory: Three Case Studies of the Pperationalization of the Concepts of 'Late Style,' 'Authorship Attribution,' and 'Literary Movement'«. *Iperstoria*, let. 12, št. 2, 2018, str. 50–60.

Underwood, Ted. »Distant Reading and Recent Intellectual History«. *Debates in the Digital Humanities* 2016, ur. Matthew K. Gold in Lauren F. Klein, University of Minnesota Press, https://dhdebates.gc.cuny.edu/read/untitled/section/3b96956c-aab2-4037-9894-dc4ff9aa1ec5. Dostop: 12. 4. 2024.

Underwood, Ted. »A Genealogy of Distant Reading«. *Digital Humanities Quarterly*, let. 11, št. 2, 2017, http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html. Dostop 12. 4. 2024.

Zajc, Ivana. »Elementi monodrame in avtobiografskosti v besedilih Simone Semenič«. *Amfiteater*, let. 7, št. 2, 2019, str. 80–98.

Zupan Sosič, Alojzija. »Deset romanov Ivana Cankarja in sodobna definicija romana«. *Slavia Centralis*, let. 13, št. 1, 2020, str. 136–152.

Žejn, Andrejka. »Računalniško podprta stilometricčna analiza pripovedne literature Janeza Ciglerja in Christopha Schmida v slovenščini«. *Fluminensia*, let. 32, št. 2, 2020, str. 137–158.

Žejn, Andrejka. *Izhodišča slovenske pripovedne proze. Dvestoletna tradicija slovenske pripovedne proze: od sredine 17. do sredine 19. stoletja.* ZRC SAZU, Inštitut za slovensko literaturo in literarne vede, 2021, https://ispp.zrc-sazu.si/. Dostop 12. 4. 2024.

Analyzing 90 Slovenian Novels and the Oeuvre of Ivan Cankar Using Computational Stylometry

The article presents a distant reading of Slovenian literature from the periods of realism and early modernism (*moderna* in Slovenian). The first part of the article displays the results of a computational stylometric analysis of a corpus of 90 Slovenian novels from the two periods, conducted in the programming environment R with the Stylometry with R package. The results show that the primary signal differentiating the included novels is the authorial one, with Pavlina Pajk and Ivan Cankar standing out the most. Since Pajk is the most distinct author in the corpus, the article also includes a stylometric investigation into the most typical vocabulary of her works. On this basis, the article concludes that her writing differs from others due to the genre of sentimental romance novels which characterizes her work. In the second part of the article, Cankar's literary oeuvre is analyzed using the same methodology, and the corpus includes not only his novels but also other works from different periods of his writing career. The results of the analysis reveal the development of Cankar's literary style in terms of genre and periodization.

Primerjava Josipa Jurčiča in Ivana Cankarja z računalniškimi metodami za zaznavanje semantičnih premikov

Andrejka Žejn, Marko Pranjić, Senja Pollak

Inštitut za slovensko literaturo in literarne vede ZRC SAZU, Novi trg 2, 1000 Ljubljana, Slovenija
https://orcid.org/0000-0001-8028-0193
andrejka.zejn@zrc-sazu.si

Institut »Jožef Stefan«, Jamova cesta 39, 1000 Ljubljana, Slovenija
https://orcid.org/0000-0002-8645-9714
marko.pranjic@ijs.si

Institut »Jožef Stefan«, Jamova cesta 39, 1000 Ljubljana, Slovenija
https://orcid.org/0000-0002-4380-0863
senja.pollak@ijs.si

V prispevku uvodoma predstavimo heterogenost in interdisciplinarnost digitalne humanistike kot dva osrednja in medsebojno povezana koncepta. V osrednjem delu vpeljemo uporabo metode zaznavanja semantičnih premikov z uporabo kontekstualnih besednih vložitev pri analizi literarnih del. Potencial metode prikažemo na primerjalni analizi pripovednih opusov Josipa Jurčiča in Ivana Cankarja, kanoničnih slovenskih avtorjev, ki sta ustvarjala v različnih obdobjih, in sicer z avtomatskim prepoznavanjem besed, katerih pomeni se med avtorjema najbolj razlikujejo. Nadaljnja interpretacija temelji na kvalitativni analizi avtomatsko pridobljenih rezultatov in na uvrščanju besed, prepoznanih kot relevantnih za razlike med Jurčičevim in Cankarjevim slogom, v pomenska polja. Pokazali smo, da je pristop, ki temelji na kontekstualnih besednih vložitvah, mogoče uporabiti za analize literature z zadovoljivimi rezultati. S tem z vidika literarne zgodovine ponujamo nov vpogled v Cankarjevo in Jurčičevo pripovedništvo, saj pokažemo, da razlika med Jurčičevim (romantičnim) realizmom in Cankarjevo moderno sloni tudi na semantiki besed, povezani z gibanjem ter družbenimi in psihološkimi dejanji in procesi.

Ključne besede: slovenska književnost / Jurčič, Josip / Cankar, Ivan / obdelava naravnega jezika / semantična analiza / digitalna literarna veda / digitalna humanistika

127

Uvod

Digitalna humanistika, ki se kot široko polje raziskovalnih praks razvija od šestdesetih let 20. stoletja naprej, ima zaradi svoje heterogenosti množico definicij, kar hkrati pomeni, da obče veljavne definicije področja ni mogoče podati (Alvaro 50; Kuhn in Callahan 291; Rodríguez Ortega 2; Terras idr.).[1] Polje digitalne humanistike je namreč nenehen proces (Vanhoutte) oziroma amorfno, fluidno in fleksibilno področje raziskav.[2] Zato tudi vprašanje, kaj je digitalna humanistika, ne anticipira odgovora, ampak stalno raziskovanje in redefiniranje (McCarty 1233; Earhart 1, 117–119). Kolektivno ime *digitalna humanistika* (Svensson, »Humanities« 42) in njena posplošena definicija ne pokrijeta niti različnih digitalnih pristopov niti tega, kako digitalne prakse spreminjajo posamezna polja humanistike, zato je po Amy Earhart smiselna njena segmentacija (Earhart 119). Eden od segmentov digitalne humanistike je digitalna literarna veda (1–10),[3] ki je obenem področje v okviru literarnih ved (Zajc in Purg) in njihova kontinuiteta (Ganascia 3). Ker je bilo vse od začetkov digitalne humanistike temeljni predmet preučevanja besedilo, je bila digitalna literarna veda dolgo časa eno njenih najplodnejših področij (Murray § 10; Ganascia 1). V digitalni literarni vedi je mogoče vsaj v grobem ločevati med praksami, posvečenimi ohranjanju, dokumentaciji, prezentaciji in diseminaciji literarnih besedil (najbolj široko lahko te prakse zajamemo s pojmom digitalne izdaje), in kvantitativnimi analizami

---

[1] Članek je nastal v okviru raziskovalnega programa »Tehnologije znanja« (P2-0103), raziskovalnega programa »Literarnozgodovinske, literarnoteoretične in metodološke raziskave« (P6-0024) in raziskovalnega projekta »Računalniško podprta večjezična analiza novičarskega diskurza s kontekstualnimi besednimi vložitvami« (J6-2581), ki jih financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije.

[2] Nedefiniranost kot posledica obsežnosti digitalne humanistike je vključena tudi v diskurz o digitalni humanistiki kot »krovnem pojmu« (Svensson, »Beyond«; Presner 195–196; Alvarado) in njegovih ironičnih izpeljankah, kakršna je na primer ugotovitev, da »šotor ni dovolj velik« za pokritje celotnega spektra digitalne humanistike (Terras idr.). Ironiziranje metafore je zaznavno tudi pri Stephenu Ramsayju, ki ekstenzivno naštevanje, kaj vse je mogoče uvrstiti v digitalno humanistiko, sklene z besedno zvezo »prostorno platno« (»capacious canvas«; Ramsay 240).

[3] Angleški izraz za digitalno literarno vedo, *digital literary studies*, se je v literarni vedi »razplamtel« (Hoover idr.) v prvi polovici drugega desetletja 21. stoletja. Glede na Googlov n-gram za obdobje do leta 2019 se je ta besedna zveza na internetu – ki je ne nazadnje pogost ali celo najpogostejši medij digitalnohumanističnih objav – začela pojavljati sredi prvega desetletja 21. stoletja; njena pojavnost je izrazito naraščala do leta 2017, ko se je začel izrazit upad.

literature z najrazličnejšimi metodami t. i. oddaljenega branja ter interpretativne vizualizacije raznovrstnih podatkov iz literature, pridobljenih z računalniškimi metodami (sem sodita na primer analiza omrežij in literarna geografija). Področji se ne izključujeta med seboj in se medsebojno prepletata. Tako kot digitalno humanistiko tudi digitalno literarno vedo zaznamujeta odprtost in razvoj, povezan predvsem z razvojem računalništva in novih metodologij.

Poleg heterogenosti in z njo povezane predefiniranosti in/ali nedefiniranosti ima digitalna humanistika močan pečat kontroverznosti. Ta se zdi najbolj izrazit v digitalni literarni vedi, saj je k izzivalnosti v največji meri prispeval teoretik oddaljenega branja, ki je sicer metaforična in provokativna (Juvan idr. 57; Jannidis in Lauer 30) izraza *natančno branje* in *oddaljeno branje* predstavil kot opozicijo med branjem in nebranjem literarnih besedil (Moretti 11) in s tem sprožil ostre razprave med zagovorniki kvantitativnega oddaljenega branja in zagovorniki tradicionalnega natančnega branja, ki so se razplamtele v prvem desetletju 21. stoletja in so segle tudi zunaj ozkega akademskega prostora. Kljub mnenju, da so te debate danes postane in neplodne (Murray § 3), se digitalna literarna veda še vedno spopada z očitki, da so njene raziskave neuporabne (ne povedo ničesar, česar še ne vemo), trivialne (omejene so na štetje besed), celo neoliberalne (Eve 1). Ahistorični pogled na digitalno literarno vedo bi morda navedenim očitkom celo pritrdil, s historičnega vidika pa so te prve raziskave, kot ugotavlja Simon Mahony (Mahony 372), postavljale temelje discipline in sprožile nadaljnje preusmerjanje pozornosti od tehnologije kot nekakšne služabnice humanistike k interdisciplinarnosti kot nujni specifiki, na katero se opirajo opredelitve digitalne humanistike in njenih podpodročij (McCarthy 1225; Presner 195–196).

Jean-Gabriel Ganascia pri presojanju interdisciplinarnosti digitalne humanistike ugotavlja, da digitalna humanistika pripada tako t. i. kulturnim znanostim kakor t. i. naravnim znanostim, saj s stališča predmeta obravnave sodi med prve, po metodološki plati in po tem, da uporablja obsežne podatkovne zbirke, ki so avtomatsko procesirane, pa je bliže drugim (Ganascia 2–3).[4] V digitalni literarni vedi ni antagonizma med logiko znanosti kulture in razvijanjem orodij, ki pomagajo interpretirati ogromne količine podatkov z upoštevanjem obstoječih teorij.[5] Victoria Kuhn in Vicki Callahan pa v zvezi z

---

[4] Pri tem izhaja iz postavk nemških filozofov Heinricha Rickerta in njegovega učenca Ernsta Cassirerja, ki sta v prvi polovici 20. stoletja utemeljila razlikovanje med kulturnimi in naravnimi znanostmi.

[5] Za nasprotne poglede gl. Ganascia 4.

interdisciplinarnostjo digitalne humanistike ugotavljata, da doba digitalne kulture premešča meje med znanstvenimi disciplinami, ki so bile zarisane med vzponom kulture tiska (Kuhn in Callahan 292). V primerjavi s horizontalno interdisciplinarnostjo dveh disciplin, kakršni sta na primer zgodovinopisje in literarna veda, kjer dodajanje ali nadomeščanje elementov poteka brez bistvene spremembe v strukturi ali logiki udeleženih disciplin, je radikalnost digitalne humanistike v njenem potencialu, da vertikalno razširi interdisciplinarnost. Vertikalna razširitev postavlja najrazličnejše discipline pred nove izzive ter spreminja njihovo strukturo in logiko. Računalniške metode niso zgolj aplicirane na literaturo, ampak spremenijo sam način raziskovanja literature, s tem ko zastavljajo nova vprašanja. Digitalna humanistika ponuja predvsem miselno obzorje, ki omogoča nov pogled na analiziranje in interpretiranje kulturne produkcije (Bode 1, Piper 2). Inherentno prekrivanje računalništva in humanistike, v njenem okviru pa tudi literarne vede, in mrežne povezave med njima vzpostavljajo prostor združevanja vednosti, ki presega enostavno ločevanje in v katerem običajno sodelujejo strokovnjaki tako iz naravoslovja kot iz humanistike. Potreba po harmonizaciji vodi v razvijanje nove hermenevtike, v kateri potenciale kvantitativnih in matematičnih formulacij kombiniramo s humanističnimi (literarnovednimi) vidiki. V skladu z vsebinsko odprto naravo digitalne humanistike na novo vzpostavljeni prostor epistemološke produkcije še vedno potrebuje razvoj, uskladitev in utrditev (Rodríguez Ortega 4), kar je ne nazadnje imanentna lastnost vsake metode, o čemer priča nenehen razvoj metodoloških pristopov tako znotraj kot zunaj digitalne literarne vede, in sicer največkrat kot odziv na pomanjkljivosti že uveljavljenih metod. S tem povezan je tudi metodološki pluralizem, ki je še posebej izrazit v digitalni literarni vedi, kjer gre že v izhodišču za kombiniranje oddaljenega in natančnega branja oziroma kvantitativnega in kvalitativnega pristopa.

Čeprav glavni argumenti zagovornikov oddaljenega branja poudarjajo možnost sočasnega pogleda na obsežne količine literarnih del in vključitev literarnih del, ki jih tradicionalna literarna veda z metodami natančnega interpretativnega branja ni vključila v kanon in s tem tudi ne v svoje raziskave,[6] se digitalna literarna veda pogosto ukvarja s

---

[6] Čeprav ta argument močno odmeva, odkar ga je zapisal Moretti (Moretti 8), velja poudariti, da je že Burrows več kot desetletje prej zapisal, da literarna veda ne more več zanemarjati potenciala »elektronskega medija« in nadaljevati z literarno zgodovino, ki temelji na majhnem vzorcu in ne upošteva »tretjine, dveh petin, polovice« gradiva (Burrows 1).

kanoniziranimi avtorji,[7] ne pa nemara z obsežnimi korpusi, kakršnih posameznik ne bi uspel niti prebrati. Vzrok za osredotočenost na kanon bi lahko iskali v tem, da so ti avtorji in njihova dela najbolj raziskani, tako da je mogoče iz njih lažje izpeljati nova raziskovalna vprašanja, na katera je mogoče odgovoriti v okviru digitalne literarne vede. Ne glede na obseg korpusa je, kot ugotavlja Thomas Rommel, s kvantitativnimi metodami mogoče med drugim določati značilnosti stila, ki jih pri natančnem branju besedila ni mogoče prepoznati ali izluščiti, toda vseeno vplivajo na splošni vtis besedila (Rommel 90). Računalnik lahko učinkovito prepozna distribucijske vzorce, ki pomagajo razumeti učinek besedila. Za pomenljive rezultate je poleg literarnega koncepta, iz katerega izhaja računalniška kvantitativna analiza, in samega računalniškega orodja nujna ustrezna literarnovedna interpretacija.

Prepoznavanje, katere jezikovne značilnosti imajo v besedilu funkcijo markerjev stila in katere so stilistično nevtralne, temelji na njihovi pojavnosti v različnih, a povezanih kontekstih (Enkvist 34–35). V različna, a povezana konteksta spadata tudi pripovedna opusa kanoniziranih slovenskih literatov Josipa Jurčiča (1844–1881) in Ivana Cankarja (1876–1918), ki sta v literarnozgodovinskem središču prispevka. Jurčič in Cankar sta pisala žanrsko raznovrstno pripovedno prozo, Jurčič v obdobju realizma oziroma romantičnega realizma, Cankar v obdobju moderne, predvsem pa sta oba postavila mejnike v razvoju slovenskega romanopisja in pripovedne proze t. i. dolgega 19. stoletja: Jurčič kot avtor prvega slovenskega romana in s tem začetnik klasičnega slovenskega romana (Kos, »Cankar«), Cankar pa kot avtor, po zaslugi katerega je slovenski roman v obdobju moderne postal pravi, umetniški roman (Pirjevec 73). V nadaljevanju predstavljena računalniško podprta kvantitativna raziskava, ki je rezultat interdisciplinarnega sodelovanja med literarno vedo in informatiko, je bila usmerjena v razkrivanje razlik v diskurzu teh dveh avtorjev na podlagi pomenske primerjave z uporabo kontekstualnih besednih vložitev. Ta računalniška metoda za zaznavanje semantičnih premikov je usmerjena na pomen besede, ki izhaja iz njene tipične besedne okolice.

Za računalniško analizo besedil je ključnega pomena način predstavitve besed in dokumentov. Velik preskok na področju obdelave naravnega jezika so omogočile globoke nevronske mreže za učenje gostih vektorskih vložitev besed in dostopnost velikih prednaučenih jezikovnih modelov (BERT: Bidirectional Encoder Representations

---

[7] Značilno je, da je v stilometričnih in drugih računalniško podprtih kvantitativnih raziskavah med najpogosteje obravnavanimi literati William Shakespeare.

from Transformers; Devlin idr.). Medtem ko so bile prve goste vložitve statične oziroma globalne, tako da je bil eni besedi pripisan en vektor glede na vse kontekste, v kateri se dana beseda pojavlja, so novejše predstavitve, ki jih uporabljamo tudi v tem prispevku, kontekstualne. Pri teh ima vsaka raba besede (oziroma žetona, ki lahko predstavlja tudi del besede) svojo predstavitev, ki se spreminja glede na lokalni kontekst. Tovrstne vložitve bolje zajamejo polisemijo in zaznavajo razlike v pomenu določenega leksema v različnih kontekstih. Druga pomembna novost novejših metod pa je v dostopnosti prednaučenih modelov: modele, naučene na velikih jezikovnih korpusih, lahko donaučimo za specifični korpus ali nalogo; s tem obdržimo splošno jezikovno znanje in razumevanje odnosa med besedami, ki pa ga prilagodimo specifičnosti posamezne uporabe.

Metode za zaznavanje semantičnih premikov na podlagi vektorskih vložitev so bile uporabljene za vrsto nalog. Različne raziskave (Kutuzov idr., »Tracing«; Tang; Tahmasebi idr.) zaznavajo pomenske premike posameznih besed s pomočjo kontekstualnih vložitev. Sorodne metode pa so bile uspešno uporabljene tudi na primer za analizo stališča (Azarbonyad idr.; Martinc idr., »EMBEDDIA«), zaznavanja dogodkov (Kutuzov idr., »Diachronic«), premikov v diskurzu (Schlechtweg idr.) in diahrone razsežnosti novic (Martinc idr., »Leveraging«), vključno z novicami o covidu-19 (Montariol idr.).

Glavna inovacija v tem prispevku je uporaba metode zaznavanja semantičnih premikov na področju digitalne humanistike, in sicer pri primerjalni analizi literarnih del. Izkaže se, da je metoda, ki temelji na primerjavi distribucij gruč (*clusters*) kontekstualnih besednih vložitev (Montariol idr.), primerna za primerjavo avtorskih literarnih opusov in za nova spoznanja o literaturi. Z vidika literarne vede prispevek prinaša nov vpogled v diskurz dveh kanoničnih slovenskih avtorjev in razgrinja enega od sestavnih delov stila kot skupka številnih formalnih značilnosti, večplastnega celovitega sistema, ki ga lahko preučujemo kvalitativno ali kvantitativno (Herrmann idr. 44–45). Analiza je namreč pokazala, da razlika med Jurčičevim realizmom in Cankarjevo moderno med drugim sloni na semantiki besed, ki jih je mogoče uvrstiti v določena pomenska polja.

V nadaljevanju so predstavljeni izhodiščni literarni koncepti, metodologija, proces analize, interpretacija in sklepne ugotovitve, pri tem pa se izmenjujeta in medsebojno dopolnjujeta kvantitativni in kvalitativni pristop oziroma oddaljeno in natančno branje.

Slovenska pripovedna proza od Jurčiča do Cankarja

Ko se je v šestdesetih letih 19. stoletja začel vzpon slovenskega pripovedništva, so bile literarne težnje povezane z vznikom slovenskega meščanstva, v njih pa je oporo našlo tudi slovensko nacionalno gibanje (Kmecl, »Problematika« 164–165). Josip Jurčič, eden vodilnih predstavnikov tega obdobja, je svoje pripovedništvo začel razvijati leta 1861 z objavami krajših in daljših povesti, leta 1866 pa je postal zastavonoša slovenskega romanopisja z romanom *Deseti brat*. Modele za roman je prevzel po literarnem programu Frana Levstika in zgodovinskih romanih Walterja Scotta (Kos, *Pregled* 160). Še isto leto, ko je izšel njegov prvi roman, je nadaljeval s pisanjem povesti, romanov in krajših proznih del (Kmecl, *Josip*). Njegova začetna dela so izraziteje navezana na ljudsko izročilo, tudi z elementi pravljičnosti, kasneje pa je vedno bolj ubesedoval zgodovinsko in meščansko tematiko (Kmecl, *Josip* 18–19, 37). Jurčič je svojo ustvarjalnost pojmoval kot realizem, medtem ko ga je literarna veda uvrstila v romantični realizem, na prehod med romantiko in realizmom, ali celo bližje romantiki (Kmecl, *Josip* 137; Kos, *Pregled* 131).

Na začetku osemdesetih let 19. stoletja se v slovenski pripovedni prozi začne razmah realizma (Kos, *Pregled* 131), ki proti koncu stoletja preraste v naturalizem, od preloma stoletja do konca 1. svetovne vojne pa pripovedništvo zaznamuje slovenska moderna kot specifično križanje različnih literarnih smeri. Vodilni predstavnik tega obdobja je Ivan Cankar, čigar pripovedni in dramski opus še danes veljata za vrhunec slovenske literature. Žanrsko v njegovi pripovedni prozi že na začetku prevladuje kratka proza, ki proti koncu njegovega ustvarjanja postane celo edina zvrst (Čeh Steger 89). Od daljše proze je po letu 1900 objavil devet romanov in z njimi v slovensko književnost uvedel nove tipe romana. Njegov pripovedni slog velja za novost v primerjavi s prejšnjimi obdobji. Ko je na začetku devetdesetih let 19. stoletja objavil svoje prve spise, se je sicer še zgledoval po domači realistično--naturalistični tradiciji, ki jo je močno zaznamoval Jurčič, ob koncu stoletja pa so na njegovo kratko prozo pa tudi romane močno vplivali moderni evropski tokovi, zlasti dekadenca, impresionizem in simbolizem. Od tod subjektivistični pogled na svet, usmerjenost dogajanja v čutno-čustveni svet in uvajanje pesniških postopkov (Kos, *Pregled* 239; Zupan Sosič 232; Čeh Steger 89–90).

Korpus besedil

Jurčičeva in Cankarjeva pripovedna besedila so v elektronski obliki javno dostopna v spletni digitalni knjižnici *Wikivir: Slovenska leposlovna klasika*, ki je nastala in se dopolnjuje v okviru projekta digitalizacije slovenske leposlovne klasike v javni domeni. Tako je bilo mogoče razmeroma hitro pridobiti pripovedna besedila obeh avtorjev za računalniško analizo. Korpus, ki zajema dobra dva milijona žetonov (*tokens*), tj. jezikovnih enot, ki ustrezajo besedi ali delu besede, obsega 68 literarnih besedil; podrobnejši podatki o velikosti korpusa so prikazani v Tabeli 1.

	Josip Jurčič	Ivan Cankar
Število dokumentov	36	32
Število stavkov	31646	49217
Število žetonov	842604	1223164

Tabela 1: Pregled obsega analiziranega korpusa, ki ga sestavljajo ročno zbrana, javno dostopna dela Josipa Jurčiča in Ivana Cankarja; tokenizacija z modelom SloBERTa besedilo razdeli na žetone.

Metodologija

Za odkrivanje razlik v besedni rabi med avtorjema je bilo treba metodologijo za odkrivanje semantičnih premikov, ki so jo opisali Syrielle Montariol, Matej Martinc in Lidia Pivovarova (Montariol idr.), prilagoditi s prostodostopno kodo (dostopno na naslovu https://github.com/RSDO-DS3/SloSemanticShiftDetection). Medtem ko je bila izvirna metoda namenjena analizi besednih premikov skozi čas, je tako prilagojena metoda omogočila primerjavo avtorjev.

Kontekstualne besedne vektorske vložitve smo pridobili z modelom SloBERTa (Ulčar in Robnik Šikonja): prednaučeni model smo z maskiranjem besed prilagodili uporabljenemu literarnemu korpusu (učenje poteka pet epoh). Za gradnjo kontekstualnih besedilnih vložitev smo dokumente najprej razdelili na stavke, vsak stavek pa obrezali na 256 žetonov. Od konteksta odvisne vložitve za vsak žeton so bile ustvarjene s seštevanjem zadnjih plasti izhoda kodirnika modela. Zaradi tokenizacije, ki v modelih tipa BERT preslika eno vhodno besedo v več žetonov, so bile besede, sestavljene iz več žetonov, predstavljene kot povprečje vseh vektorjev žetonov, ki besedo sestavljajo. Za vsako besedo smo nato

kontekstualne vektorske vložitve združili v gruče podobnih rab, kjer smo uporabili algoritem k-means (in nastavitev za pet gruč po zgledu Montariol idr.).

V naslednjem koraku smo metodo uporabili za primerjavo distribucij gruč. V nasprotju s pristopom Syrielle Montariol idr., kjer so avtorji primerjali rabe v različnih obdobjih, smo komponento za delitev korpusa uporabili za delitev po imenu avtorjev. Besede smo tako rangirali glede na razliko v distribuciji gruč besednih rab med avtorjema, pri čemer smo se naslonili na metriko Jensen-Shannon divergence oziroma JSD (Lin). Prav tako smo lahko za vsako besedo vizualizirali in interpretirali gruče rab, in sicer preko besed, ki gručo opisujejo (z uteževanjem tf-idf), in vizualno primerjali distribucijo rab med avtorjema.

Rezultati in izbor besed za analizo

Z uporabo opisane metodologije smo dobili rangiran seznam besed z največjo zaznano pomensko spremembo med dvema avtorjema. V nadaljevanju smo rezultate, natančneje prvih 400 besed z največjo zaznano pomensko spremembo, kvalitativno (ročno) preverili. Pri tem smo presojali, katere besede na seznamu so najbolj povedne kot diferencialni markerji stila med Jurčičem in Cankarjem, in sicer s pomočjo grafičnih prikazov razporeditve gruč besednih rab pri posameznem avtorju in kontekstualnih izpisov posameznih besed,[8] ki so bili generirani ob analizi. Z ročnim pregledom smo dobili kvalitativni vpogled v uspešnost zaznavanja pomenskih razlik z vidika uporabe metode pri literarni analizi razlik v rabi besed v besedilih različnih avtorjev. Izmed 400 besed z največjo razliko v rabi jih je bilo 57 izbranih kot ustreznih.

Natančnost, kot je opredeljena v kontekstu iskanja informacij, je del ustreznih rezultatov med vsemi vrnjenimi rezultati. Sistem za zaznavanje semantičnih premikov vrne vse besede iz korpusa, ki so bile najdene, ne glede na njihovo semantično spremembo. Zato ocenjujemo natančnost pri rangu oziroma natančnost@k, ki je opredeljena kot delež ustreznih rezultatov med najvišjimi vrnjenimi k rezultati. To metriko, ki pokaže ročno ocenjen odstotek relevantnih besed za analizo v odnosu do pozicije na rangiranem seznamu, predstavljamo na Sliki 1. Med zgornjimi 50 besedami je tako 38 % besed zanimivih za analizo, na vseh 400 besedah pa je ta odstotek 14,25.

---

[8] Ob kontekstualnem primeru je podan tudi podatek o avtorju in besedilu, iz katerega je primer.
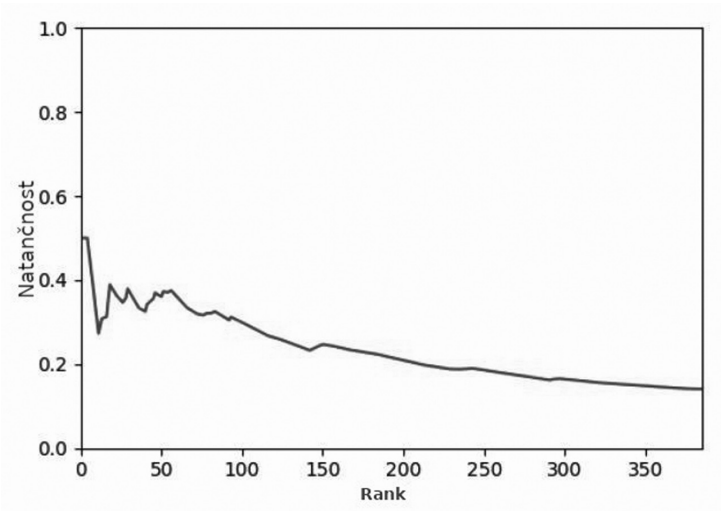
Slika 1: Natančnost glede na rangiranje (natančnost@k), ki prikazuje odstotek besed, ki so zaznane z metodo za analizo semantičnih premikov in so zanimive tudi z vidika literarne vede.

Metodo za zaznavanje pomenskih razlik s kontekstualnimi vložitvami ocenjujemo kot dobro izhodišče za primerjavo avtorjev literarnih besedil. Je pa pri interpretaciji odstotka nerelevantnih besed za analizo potrebna previdnost. V primerih nerelevantnih besed ne gre nujno le za omejitev pristopa k odkrivanju pomenskih sprememb, ampak tudi za to, da so lahko nekatere besede, kjer pride do različne rabe, vsaj na prvi pogled težje prepoznavne kot relevantne, mogoče pa je tudi, da je del besed nepomemben z vidika ciljne raziskave na področju literarne analize.

Natančna primerjava avtorjev

V iskanju skupnih točk in razlik med besedami, ki so bile med rezultati kvantitativne analize izbrane za podrobnejšo presojo, smo izbrane besede v naslednji fazi[9] združevali v pomenska polja. Pri uvrščanju v pomenska polja načeloma združujemo besede, ki imajo skupno semantično komponento in ki jih uporabljamo, ko govorimo o istem pojavu ali isti topiki (Heuser in Le-Khac 4), v posameznem pomenskem polju pa so besede organizirane glede na medsebojna pomenska razmerja, in

---

[9] Poudariti velja, da pri presojanju, kateri pomenski premiki se zdijo najrelevantnejši, še nismo razmišljali o uvrščanju v pomenska polja, tako da je izbor nastal neodvisno od tu prikazanega grupiranja besed.

sicer je relevantna ne le sopomenskost, ampak tudi protipomenskost ter nad- in podpomenskost (Stubbs 36).

Ker splošna taksonomija pomenskih polj za slovenščino ni izdelana, smo se pri razvrščanju oziroma združevanju besed oprli na semantično taksonomijo USAS (UCREL[10] Semantic Analysis System). Ta vsebuje 21 preddefiniranih osnovnih semantičnih kategorij ali konceptov, ki so nadalje diferencirani z različnim številom podkategorij, tako da je skupno število podkategorij 232. Vsak leksem je mogoče uvrstiti v več pomenskih polj, a smo zaradi preglednosti izbrali vsakič le eno polje in zaradi razmeroma majhnega števila besed upoštevali le osnovno taksonomijo z 21 razredi. Razporeditev besed v pomenska polja[11] je prikazana v Tabeli 2, v kateri si pomenska polja sledijo od tistih z največ leksemi navzdol.

Pomensko polje	Uvrščene besede
»gibanje in lokacija«	kod, spustiti, prevzeti, seči, nazaj, odpeljati, naravnost, zapreti, stopiti, poslati, jug, pot, hoditi, popotnik, najti
»družbena dejanja in stanja«	skupaj, služabnik, tuj, neznan, nebesa, tujec, slovenski, poljub, stisniti
»psihološka dejanja in stanja«	slišati, vzbuditi, čuti, premišljevati, zagledati, spati, želeti, želja
»govorna dejanja«	zaklicati, zasmejati, molčati, potihoma, izgovorjen, glasen, praviti
»predmeti, snovi in lastnosti«	hlad, mraz, ogenj, kaplja, luč, mrzel
»telo in posameznik«	lice, bolečina, smrt
»čustvena dejanja, stanja in procesi«	veselje, sreča, težko
»čas«	čas, tema, mrak
»arhitektura, zgradbe, hiša/dom«	hiša
»uprava in javne domene«	vojska
»številke in mere«	tisoč

Tabela 2: Razvrstitev izbranih besed v pomenska polja.

[10] Sistem je bil razvit v Univerzitetnem središču za računalniške korpusne raziskave jezika (University Centre for Computer Corpus Research on Language) na Univerzi Lancaster v Veliki Britaniji. Taksonomija t. i. semantičnih domen je bila sicer zasnovana za avtomatsko analizo pomenskih polj oziroma semantičnih domen v korpusni stilistiki, vpeljana v številnih kvantitativnih korpusnih raziskavah literature in prevedena v več jezikov.

[11] Prepoznanih je bilo 11 semantičnih polj, torej polovica vseh, ki so v taksonomiji USAS. Nobena beseda ni bila uvrščena v sledeča pomenska polja: »splošno in abstraktno«; »umetnost in obrt«; »hrana in kmetijstvo«; »denar in trgovanje«; »zabava, šport in igre«; »življenje in živa bitja«; »izobraževanje«; »svet in okolje«; »znanost in tehnologija«; »imena in slovnične besede«.

Iz te razporeditve je mogoče razbrati, da večina besed, v katerih je bila zaznana pomembna pomenska sprememba glede na sobesedilni kontekst v opusih izbranih avtorjev, spada v semantično polje »gibanje in lokacija«. Razmeroma obsežni sta tudi pomenski polji »družbena dejanja in stanja« (to polje v najširšem smislu opisuje družbene in druge medčloveške odnose) in »psihološka dejanja in stanja« (besedišče iz tega pomenskega polja opisuje človekov notranji svet). Tem po obsegu tesno sledi pomensko polje »govorna dejanja«, temu pa polje »predmeti in snovi in lastnosti«, pri čemer – kot bomo na enem od primerov prikazali v nadaljevanju – gre pri pomenskem premiku predvsem za razmerje med osnovnim in simbolnim pomenom. Bistveno manj obsežna so pomenska polja »telo in posameznik«, »čustvena dejanja, stanja in procesi« ter »čas«, ki vsebujejo po tri lekseme. V zadnje tri kategorije (tj. »arhitektura, zgradbe, hiša/dom«, »uprava in javne domene« ter »številke in mere«) je bil uvrščen le po en leksem, zato je njihova vloga razmeroma zanemarljiva.

Iz kategorizacije v pomenska polja je mogoče izpeljati ugotovitev, da je pomembna diskurzivna razlika med avtorjema vgrajena v kronotop njunih pripovednih del; tega predstavlja »gibanje in lokacija«, najobsežnejše pomensko polje, v kombinaciji s pomenskim poljem »čas«, ki je sicer med manj obsežnimi. Za ilustracijo na Sliki 2 podrobneje predstavljamo grafični prikaz in sobesedilne primere za leksem *kod* iz najobsežneje zastopanega pomenskega polja.
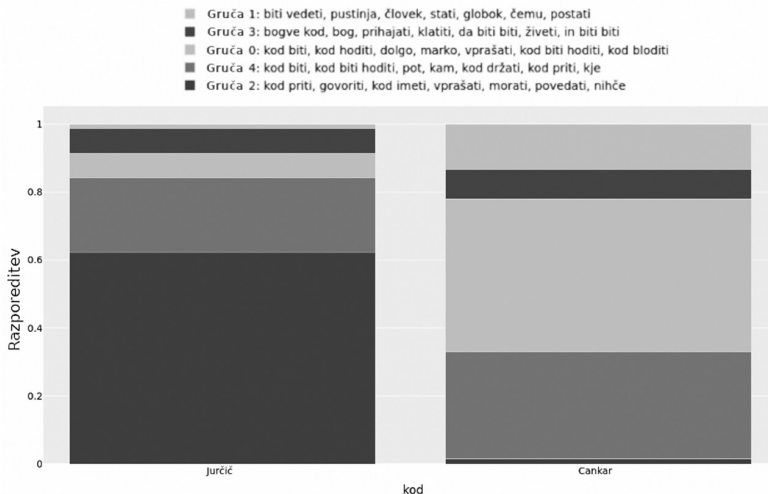
Slika 2: Grafični prikaz pomenskih razlik v rabi leksema *kod*, zaznanih z metodo za zaznavanje semantičnih premikov; posamezna gruča predstavlja tipično besedno okolico izbranega leksema.

Leksem *kod* je četrti na seznamu vseh leksemov, ki izkazujejo visoko stopnjo pomenske spremembe (JSD 0,291). Primerjava pokaže, da v Jurčičevem opusu pri *kod* izrazito prevladuje 2. gruča – *kod priti*, *govoriti*, *kod imeti*, *vprašati*, *morati*, *povedati* in *nihče* – medtem ko pri Cankarju prevladuje gruča 0 – *kod biti*, *kod biti hoditi*, *dolgo*, *Marko*,[12] *vprašati*, *kod biti hoditi* in *kod bloditi*. Ta gruča je tudi pojmovno povezana s 1. gručo, ki je pri Cankarju prav tako pogosta, namreč *biti vedeti*, *pustinja*, *človek*, *stati*, *globok*, *čemu* in *postati*. Razliko med tipičnim sobesedilom lahko prepoznamo kot razmerje med Jurčičevim ciljnim gibanjem likov v konkretnem literarnem prostoru in času, medtem ko se pri Cankarju leksem *kod* navezuje na prostorsko nedoločljivost in brezciljno tavanje. Spodnji kontekstualni primeri ponazarjajo, kako je ta semantični premik tipično izražen v besedilih:

> Vedel in videl je, kod se pot vije in kje zopet v hosti izgine. (Josip Jurčič, *Sin kmečkega cesarja*)

> Leonu pak je še posebej pokazal, kod se pride na stranski kor, kjer se dobro k oltarju in na prižnico vidi in kamor navadno hodijo civilni ljudjé njegove fare. (Josip Jurčič, *Cvet in sad*)

> Kod blodim tri dni, ali morda že teden dni? (Ivan Cankar, *Troje povesti*)

> Kod bega mati, da bi prinesla mesá in morda celó jabolk in orehov? (Ivan Cankar, *Hiša Marije Pomočnice*)

Drugi izbrani primer, ki ga podrobneje interpretiramo, je leksem *ogenj* iz pomenskega polja »predmeti, snovi in lastnosti« (gl. Sliko 3). Gruča 2 – *oči*,[13] *obraz*, *lice*, *ustnica*, *sijati*, *vesel*, *svet ogenj* –, ki pri Cankarju prevladuje, izraža povezave ognja z obrazom, kar nakazuje simbolnost, medtem ko je pri Jurčiču zaznana izrazita povezava s konkretnimi dejanji in samim obstojem ognja, na primer *zakuriti*, *vreči*, *prijeti*, *biti* (gl. gručo 4).

---

[12] Lastna imena so sicer posebna, za nas nepomembna kategorija. Če je lastno ime zaznano kot beseda, ki se spreminja, kaže na primer na isto ime dveh literarnih likov. Če se pojavlja pri opisu gruče, sicer kaže na sopojavljanje z besedo, vendar nam o sami rabi in analizi razlik ne pove veliko, tako da pri interpretaciji njihova posamezna pojavnost na celoto rezultatov tako rekoč ne vpliva.

[13] V tem kontekstu velja opozoriti na kvantitativno raziskavo besed in pomenskih polj v pripovedni prozi in dramatiki Ivana Cankarja, ki ugotavlja, da je najpogostejša polnopomenska beseda pri Cankarju *oči*. Raziskava izpostavlja tudi pogostost besed za označevanje delov telesa, zlasti podrobnosti obraza (Mikolič 23, 32).
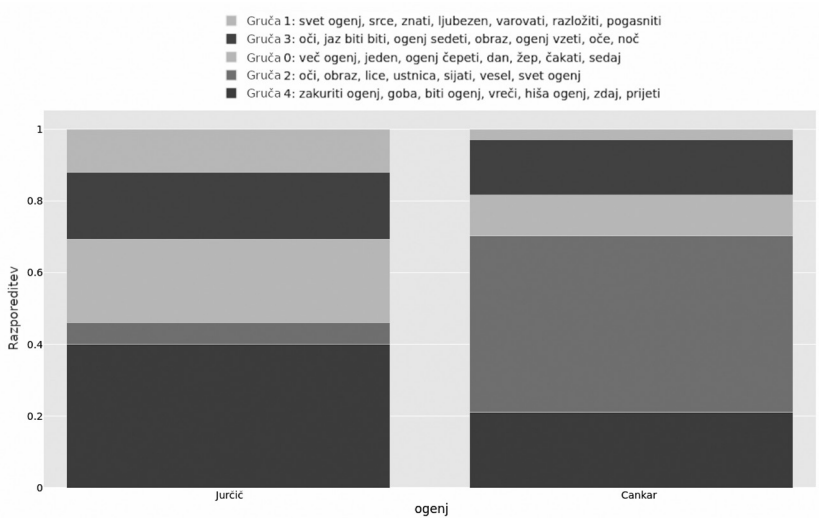
Slika 3: Grafični prikaz pomenskih razlik v rabi leksema *ogenj*, zaznanih z metodo za zaznavanje semantičnih premikov.

Poleg tega primerjava kontekstualne rabe pokaže, da je pri Cankarju leksem *ogenj* v povezavi z deli obraza, v katerih se zrcali notranjost človeka, najobičajneje simbol strasti, življenjske energije:

> Nobenega ognja ni bilo v črnih, motnih očeh. (Ivan Cankar, *Vinjete*)

> Njen obraz je bil, kakor jih je ljubil: svež in vesel, ves poln ognja in mladosti, njene temnosive oči so gledale prešerno in odkrito[.] (Ivan Cankar, *Tujci*)

> Na potepuhovih kolenih je sedela ženska in ga je objemala z golo roko okoli vratu; nanjo je sijala vsa luč, kakor v ognju so bila lica, ampak iz vsega ognja so žarele velike, pregrešne ciganske oči. (Ivan Cankar, *Zgodbe iz doline Šentflorjanske*)

Drugače je pri Jurčiču, kjer leksem *ogenj* najpogosteje poimenuje dejansko substanco:

> Nazaj prišedši, je celo v kuhinjo stopil, iz velikega ognja, pri katerem je Franica kosilo kuhala, žareč ogel za pipo vzel ter molče zopet odšel[.] (Jurčič, *Sosedov sin*)

Kot je prav tako mogoče razbrati iz grafičnih prikazov, ima *ogenj* tudi pri Jurčiču lahko simbolni pomen, a v bistveno manjši meri kakor pri

Cankarju, večinoma pa tudi v drugačnem kontekstu. Navedimo značilen primer:

> Nikakor ne morem tajiti, da me srce ne vleče do hčere mojega gospodarja, da, to moram izpovedati se tebi, predragi moj, če je eno bitje na božjem daljnem svetu, za katero bi se jaz z vsem mladostnim ognjem vnel, za katero bi norel in gorel, kakor je kdaj kak človek mogel, bila bi to enaka deklica, kakor je ta, katere sem ti v zadnjem pismu omenil. (Josip Jurčič, *Deseti brat*)

Sklepne ugotovitve

Zgornja primera vsebinske interpretacije grafičnih prikazov rezultatov kvantitativne analize in delitev besed, za katere je bil prepoznan pomenljiv semantični premik, v pomenska polja kažejo, da metoda omogoča globlji primerjalni uvid v sam način, na katerega avtorja na semantični ravnini gradita svoj literarni stil, in v to, katera pomenska polja, ki so obenem tudi skupna pomenska polja pripovedne proze izbrane dvojice avtorjev, so nosilci prehoda od Jurčičevega (romantičnega) realizma k Cankarjevi simbolistični moderni. Analiza in interpretacija sta pokazali, da sprememba v pomenski plati ni toliko vezana na neposredno izražanje čutno-čustvene komponente, kot bi nemara pričakovali, kolikor je izražena v vsaj na videz splošnejših pomenskih poljih, kot so – poleg »gibanja in lokacije« – »družbena dejanja in stanja«, »psihološka dejanja in stanja« in »govorna dejanja«.

V tej raziskavi so bile besede oziroma pomenski premiki, ki naj bi bili najbolj pomenljivi z vidika razmerja slogov izbranih avtorjev, prepoznani z (ročno ali neavtomatsko) kvalitativno analizo, nato pa so bile izbrane besede razporejene v pomenska polja. Za primerjavo med ročno izbranimi besedami in prvimi 100 besedami na seznamu smo v pomenska polja naknadno razporedili vseh prvih 100 besed na avtomatsko rangiranem seznamu. Izkazalo se je, da to na obseg posameznih pomenskih polj ni imelo bistvenega vpliva,[14] pa tudi nova pomenska polja se niso oblikovala, iz česar je mogoče sklepati, da bi do enako relevantnih rezultatov kakor s kvalitativno analizo rezultatov avtomatske analize prišli tudi z razporeditvijo vseh prvih 100, 150[15] ali celo več besed v pomenska polja, kar bi zahtevalo bistveno manj ročnega dela.

---

[14] Nekoliko obsežnejše je bilo le polje »govorna dejanja«.

[15] Prim. Sliko 1, iz katere je razvidno, da natančnost, ki je izračunana iz primerjave med kvantitativno in kvalitativno (ročno) analizo, po približno 150 prvih besedah na rangiranem seznamu izraziteje upade.

Rezultati nakazujejo tudi možnosti nadaljnjih raziskav, saj bi bilo mogoče pomenski premik opazovati glede na obdobja ustvarjanja teh dveh avtorjev (zlasti ker za oba velja, da sta iz zgodnje faze, navezane na tradicijo, ter preko vrhunca ustvarjalnosti prešla v svoje zrelo obdobje) ali glede na žanrsko delitev njune pripovedne proze. Vsekakor je raziskavo mogoče razširiti tudi na druge avtorje in obdobja pripovedne proze dolgega 19. stoletja in širše. Poleg tega bi bila zelo zanimiva medjezična analiza literarnih del, ki bi temeljila na večjezičnih jezikovnih modelih, kot sta mBERT (Devlin idr.) in XLM-R (Conneau idr.), saj bi omogočila preučevanje evolucije literarnega sloga ne le med žanri, ampak tudi med različnimi jeziki in kulturami.

Z računalniškega vidika se zdi, da pristop za odkrivanje semantičnih sprememb, ki so ga predlagali S. Montariol, Martinc in L. Pivovarova, dobro ustreza literarni analizi in preučevanju razlik v rabi besed med avtorji. Možna pomanjkljivost pri uporabi metod, ki temeljijo na velikih prednaučenih jezikovnih modelih, je v koraku prilagoditve domene za uporabo v literarni vedi. Čeprav model v fazi dotreniranja prilagodimo novi porazdelitvi besed, je velikost korpusa v naši študiji majhna. Zato bi ga bilo smiselno najprej prilagoditi na velikem korpusu vseh dostopnih literarnih besedil in ga šele nato prilagoditi posebni študiji primera. Razlike med rabo besed v literarni umetnosti so lahko kompleksnejše od korpusov, na katerih je bil usposobljen model, ki poleg literature zajema tudi veliko neliterarnih besedil. Poleg tega je predpogoj za računalniško primerjavo besedne semantike med dvema korpusoma tudi število pojavitev iste besede, to pa je v nasprotju s človeško sposobnostjo sklepanja o pomembnih razlikah v semantiki z enim samim primerom.

V prispevku smo se dotaknili primerjalne analize v kontekstu literarne umetnosti z uporabo velikih jezikovnih modelov. Ta vrsta analize z uporabo zaznavanja semantičnih sprememb je sama po sebi omejena na semantične razlike besed. Toda umetniškega sloga seveda ne opredeljujeta le semantika in sposobnost avtorja, da izčrpno izrazi in celo preobremeni semantiko besede, ampak tudi druge razsežnosti pisanja.

LITERATURA

Alvarado, Rafael C. »The Digital Humanities Situation«. *Debates in the Digital Humanities*, ur. Matthew K. Gold, University of Minnesota Press, 2012, str. 50–55.

Azarbonyad, Hosein, idr. »Words are Malleable: Computing Semantic Shifts in Political and Media Discourse«. *CIKM '17: Proceedings of the 2017 ACM Conference on Information and Knowledge Management*, Association for Computing Machinery, 2017, str. 1509–1518.

Bode, Katherine. *Reading by Numbers: Recalibrating the Literary Field*. Anthem Press, 2012.

Burrows, John F. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Clarendon Press, 1987.

Conneau, Alexis, idr. »Unsupervised Cross-lingual Representation Learning at Scale«. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ur. Dan Jurafsky idr., The Association for Computational Linguistics, 2020, str. 8440–8451, https://aclanthology.org/2020.acl-main.747.pdf. Dostop 5. 2. 2024.

Čeh Steger, Jožica. »Kratka proza«. *Ivan Cankar: literarni revolucionar*, ur. Aljoša Harlamov, Cankarjeva založba, 2018, str. 88–141.

Devlin, Jacob, idr. »BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding«. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 1. zv., ur. Jill Burstein idr., The Association for Computational Linguistics, 2019, str. 4171–4186, https://acl-anthology.org/N19-1423.pdf. Dostop 5. 2. 2024.

Earhart, Amy E. *Traces of the Old, Uses of the New: The Emergence of Digital Literary Studies*. University of Michigan Press, 2015.

Enkvist, Nils Erik. »On Defining Style«. *Linguistics and Style*, ur. Nils Erik Enkvist idr., Oxford University Press, 1964, str. 1–56.

Eve, Martin Paul. *The Digital Humanities and Literary Studies*. Oxford University Press, 2022.

Ganascia, Jean-Gabriel. »The Logic of the Big Data Turn in Digital Literary Studies«. *Frontiers in Digital Humanities*, št. 2, 2015, https://doi.org/10.3389/fdigh.2015.00007. Dostop 5. 2. 2024.

Herrmann, J. Berenike, idr. »Revisiting Style, a Key Concept in Literary Studies«. *Journal of Literary Theory*, let. 9, št. 1, 2015, str. 25–52.

Heuser, Ryan, in Lang Le-Khac. »A Quantitative Literary History of 2,958 Nineteenth-century British Novels«. *Stanford Literary Lab*, 2012, https://litlab.stanford.edu/assets/pdf/LiteraryLabPamphlet4.pdf. Dostop 5. 2. 2024.

Hoover, David L., idr. *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*. Routledge, 2014.

Jannidis, Fotis, in Gerhard Lauer. »Burrows's Delta and Its Use in German Literary History«. *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*, ur. Matt Erlin in Lynne Tatlock, Camden House, 2014, str. 29–54.

Juvan, Marko, idr. »Interpretiranje literature v zmanjšanem merilu: 'oddaljeno branje' korpusa 'dolgega leta 1968'«. *Jezik in slovstvo*, let. 66, št. 4, 2021, str. 55–76.

Kmecl, Matjaž. *Josip Jurčič: pripovednik in dramatik*. Zavod Republike Slovenije za šolstvo, 2009.

Kmecl, Matjaž. »Problematika slovenske proze 19. stoletja«. *Zbornik predavanj / XV. seminar slovenskega jezika, literature in kulture, 2.–14. julija 1979*, ur. Breda Pogorelec

in Ljubica Črnivec, Filozofska fakulteta, Pedagoško-znanstvena enota za slovanske jezike in književnosti, 1979, str. 157–182.

Kos, Janko. »Cankar in problem slovenskega romana«. *Sodobnost*, let. 24, št. 5, 1976, str. 413–423.

Kos, Janko. *Pregled slovenskega slovstva*. DZS, 2010.

Kuhn, Virginia, in Vicki Callahan. »Nomadic Archives: Remix and the Drift to Praxis«. *Digital Humanities Pedagogy*, ur. Brett D. Hirsch, Open Book Publishers, 2012, str. 291–308, https://books.openbookpublishers.com/10.11647/obp.0024. pdf. Dostop 5. 2. 2024.

Kutuzov, Andrey, idr. »Diachronic Word Embeddings and Semantic Shifts: A Survey«. *Proceedings of the 27th International Conference on Computational Linguistics*, ur. Emily M. Bender idr., The Association for Computational Linguistics, 2018, str. 1384–1397, https://aclanthology.org/C18-1117. Dostop 5. 2. 2024.

Kutuzov, Andrey, idr. »Tracing Armed Conflicts with Diachronic Word Embedding Models«. *Proceedings of the Events and Stories in the News Workshop*, ur. Tommaso Caselli idr., The Association for Computational Linguistics, 2017, str. 31–36, https://aclanthology.org/W17-2705.pdf. Dostop 5. 2. 2024.

Lin, Jianhua. »Divergence Measures Based on the Shannon Entropy«. *IEEE Transactions on Information Theory*, let. 37, št. 1, 1991, str. 145–151.

Mahony, Simon. »Cultural Diversity and the Digital Humanities«. *Fudan Journal of the Humanities and Social Sciences*, št. 11, 2018, str. 371–388.

Martinc, Matej, idr. »EMBEDDIA Hackathon Report: Automatic Sentiment and Viewpoint Analysis of Slovenian News Corpus on the Topic of LGBTIQ+«. *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*, ur. Hannu Toivonen in Michele Boggia, The Association for Computational Linguistics, 2021, str. 121–126, https://aclanthology.org/2021. hackashop-1.17. Dostop 5. 2. 2024.

Martinc, Matej, idr. »Leveraging Contextual Embeddings for Detecting Diachronic Semantic Shift«. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, ur. Nicoletta Calzolari idr., European Language Resources Association, 2020, str. 4811–4819, https://aclanthology.org/2020.lrec-1.592. Dostop 5. 2. 2024.

McCarty, Willard. »Humanities Computing«. *Encyclopedia of Library and Information Sciences*, ur. Miriam Drake, Marcel Dekker, 2003, str. 1224–1235.

Mikolič, Vesna. *Ali bereš Cankarja?*. Slovenska matica, 2021.

Montariol, Syrielle, idr. »Scalable and Interpretable Semantic Change Detection«. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, ur. Kristina Toutanova idr., The Association for Computational Linguistics, 2021, str. 4642–4652, https://aclanthology.org/2021.naacl-main.369. Dostop 5. 2. 2024.

Moretti, Franco. »Domneve o svetovni literaturi«. *Grafi, zemljevidi, drevesa in drugi spisi o svetovni literaturi*, ur. in prev. Jernej Habjan, Studia humanitatis, 2011, str. 5–25.

Murray, Simone. »Varieties of Digital Literary Studies: Micro, Macro, Meso«. *DHQ: Digital Humanities Quarterly*, let. 16, št. 2, 2022, http://www.digitalhumanities. org/dhq/vol/16/2/000616/000616.html. Dostop 5. 2. 2024.

Piper, Andrew. »There Will Be Numbers«. *Journal of Cultural Analytics*, let. 1, št. 1, 2016, https://culturalanalytics.org/article/11062-there-will-be-numbers. Dostop 5. 2. 2024.

Pirjevec, Dušan. »Problem slovenskega romana«. *Literatura*, let. 9, št. 67–68, 1997, str. 63–75.

Presner, Todd. »Comparative Literature in the Age of Digital Humanities«. *A Companion to Comparative Literature*, ur. Ali Behdad in Dominic Thomas, Blackwell, 2011, str. 193–208.

Ramsay, Stephen. »Who's In and Who's Out«. *Defining Digital Humanities: A Reader*, ur. Melissa Terras idr., Ashgate, 2013, str. 239–241.

Rodríguez Ortega, Nuria. »Five Central Concepts to Think of Digital Humanities as a New Digital Humanism Project«. *Artnodes*, št. 22, 2018, https://doi.org/10.7238/a. v0i22.3263. Dostop 5. 2. 2024.

Rommel, Thomas. »Literary Studies«. *A Companion to Digital Humanities*, ur. Susan Schriebman idr., Blackwell, 2004, str. 88–96.

Schlechtweg, Dominik, idr. »SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection«. *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, International Committee for Computational Linguistics, 2020, https://aclantho-logy.org/2020.semeval-1.1. Dostop 5. 2. 2024.

Stubbs, Michael. *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell, 2005.

Svensson, Patrik. »Beyond the Big Tent«. *Debates in the Digital Humanities*, ur. Matthew K. Gold, University of Minnesota Press, 2012, str. 36–72.

Svensson, Patrik. »Humanities Computing as Digital Humanities«. *Digital Huma-nities Quarterly*, let. 3, št. 3, 2009, http://www.digitalhumanities.org/dhq/ vol/3/3/000065/000065.html. Dostop 5. 2. 2024.

Tahmasebi, Nina, idr. »Survey of Computational Approaches to Lexical Semantic Change Detection«. *Computational Approaches to Semantic Change*, ur. Nina Tahmasebi idr., Language Science Press, 2021, str. 1–91.

Tang, Xuri. »A State-of-the-art of Semantic Change Computation«. *Natural Language Engineering*, let. 24, št. 5, 2018, str. 649–676.

Terras, Melissa, idr. »Selected Definitions From the Day of Digital Humanities: 2009–2012«. *Defining Digital Humanities: A Reader*, ur. Melissa Terras idr., Ashgate, 2013, str. 279–287.

Ulčar, Matej, in Marko Robnik Šikonja. *Slovenian RoBERTa Contextual Embeddings Model: SloBERTa 2.0*. Fakulteta za računalništvo in informatiko, 2021, http://hdl. handle.net/11356/1397. Dostop 5. 2. 2024.

Vanhoutte, Edward. »The Gates of Hell: History and Definition of Digital | Humanities | Computing«. *Defining Digital Humanities: A Reader*, ur. Melissa Terras idr., Ashgate, 2013, str. 119–156.

Zajc, Ivana, in Peter Purg. *Digitalna humanistika in literatura*. Založba Univerze, 2023, https://www.ung.si/sl/zalozba/43/digitalna-humanistika-in-literatura. Dostop 5. 2. 2024.

Zupan Sosič, Alojzija. »Romani«. *Ivan Cankar: literarni revolucionar*, ur. Aljoša Har-lamov, Cankarjeva založba, 2018, str. 200–233.

Comparing Josip Jurčič and Ivan Cankar Using Computational Semantic Change Detection Methods

The article begins with a presentation of the heterogeneity and interdisciplinarity of the digital humanities as two central and interrelated concepts. In the main part of the article, the method of detecting semantic changes based on contextual word embeddings for the analysis of literary works is introduced. The method's potential is demonstrated through a comparative analysis of the narrative works of two canonical Slovenian authors belonging to two distinct literary periods, Josip Jurčič and Ivan Cankar, in particular through the automatic recognition of words whose meanings differ between the authors. The differences in literary style are further interpreted via a qualitative analysis of the automatically obtained results, followed by a manual categorization into semantic fields of the words that were qualitatively identified as informative of stylistic differences between Jurčič and Cankar. The article shows that the approach based on contextual word embeddings can be used for literary analysis with satisfactory results. This enables narratology to gain new insight into the oeuvres of Cankar and Jurčič, as the article shows that the difference between Jurčič's (romantic) realism and Cankar's kind of modernism (moderna) is also based on the semantics of discourses related to movement and social and psychological actions and processes.

Kvantitativna analiza razmerij med semantičnimi polji v slovenski pripovedni prozi dolgega 19. stoletja

Lucija Mandić

Inštitut za slovensko literaturo in literarne vede ZRC SAZU, Novi trg 2, 1000 Ljubljana, Slovenija
https://orcid.org/0009-0000-7858-6513
lucija.mandic@zrc-sazu.si

V prispevku je uporabljena metoda vektorskih vložitev besed za analizo odnosov med semantičnimi polji v slovenski pripovedni prozi t. i. dolgega 19. stoletja. Z uporabo tehnologije Word2Vec v programskem jeziku Python je bila opravljena analiza Korpusa daljše slovenske pripovedne proze (KDSP 1.0). Za potrebe analize so bila konstruirana semantična polja za štiri družbene institucije: ekonomijo, politiko, kulturo in gospodinjstvo. Nabor besed za posamezno semantično polje je bil pridobljen z identifikacijo 50 besed z največjo kosinusno bližino vektorski predstavitvi vsake institucije. Nabor tako pridobljenih vektorjev je ponudil kvantitativno osnovo za raziskovanje odnosov med temi družbenimi institucijami, upodobljenimi v literarnih delih, zajetih v korpusu. Ugotovitve kažejo na pomembno prekrivanje med diskurzivnima poljema politike in kulture, s tem pa omogočajo kvantitativni pristop k pojavu, ki ga tradicionalna, kvalitativna literarna veda obravnava s konceptoma prešernovske strukture in slovenskega kulturnega sindroma.

Ključne besede: digitalna literarna veda / slovenska pripovedna proza / prešernovska struktura / slovenski kulturni sindrom / semantična analiza / vektorske vložitve besed

Računalniški pristopi k semantični analizi literarnih besedil

Pospešen tehnološki razvoj na področju tehnik strojnega učenja je sprožil metodološke premike tudi na področju humanistike.[1] Nadzorovano in nenadzorovano strojno učenje sta se izkazala za učinkoviti metodi za analize velikih količin besedilnih podatkov; predvsem nenadzorovane tehnike strojnega učenja so uporabne za luščenje informacij iz nestrukturiranega besedila. Vektorske vložitve besed (ang. *word embeddings*)

---

[1] Članek je nastal v okviru raziskovalnega programa »Literarnozgodovinske, literarnoteoretične in metodološke raziskave« (P6-0024), ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije.

147

so tehnika procesiranja naravnega jezika, ki uporablja strojno učenje za določanje konteksta posameznih besed v besedilu. V digitalni literarni vedi spada ta metoda med najbolj priljubljene pristope k semantičnim analizam (Hatzel idr. 203). Uporablja se za lociranje literarnih postopkov (Schneider idr.), za prepoznavanje avtorskega sloga (Eder in Šeļa), za prostorske analize (Herrmann, Byszuk in Grisot) in analize sentimenta (Brottrager idr.). V slovenski digitalni literarni vedi z izjemo semantične analize del Zofke Kveder (Pollak, Martinc in Mihurko) možnosti, ki jih ponujajo vektorske vložitve besed, še niso bile podrobneje raziskane.

Pri vektorskih vložitvah besed je semantična bližina besed izražena z numeričnimi koordinatami. Pri učenju jezikovnega modela so glede na kontekst vsaki besedi dodeljene koordinate večdimenzionalnih vektorjev, ki glede na določitve parametrov obsegajo med $n = 100$ in $n = 500$ dimenzij. Te določijo težo vektorja in njegovo numerično reprezentacijo, ki je za vsak vektor edinstvena. Pomensko sorodne besede, ki se pojavljajo v podobnih kontekstih, imajo posledično podobne vektorje in so si na ta način blizu v vektorskem prostoru. Razdalja med vektorji je običajno izračunana s pomočjo manhattanske, evklidske ali kosinusne razdalje.

Tehnologija Word2Vec (Mikolov idr.) v Pythonovi knjižnici Gensim, ki je uporabljena v analizi, predstavljeni v članku, je dvoslojno nevronsko omrežje, ki pretvori besedilo v nabor edinstvenih vektorjev, razporejenih v visokodimenzionalnem prostoru. Analize semantičnih relacij, ki jih zazna Word2Vec, so pokazale, da vektorske vložitve predvidijo položaje besed, ki presegajo preprosto semantično bližino; s seštevanjem in odštevanjem vektorskih reprezentacij besed model zazna tudi analogne odnose med pomeni besed. Avtorji orodja so z naslednjim izračunom vektorjev: »kralj« – »moški« + »ženska« prišli do vektorja, ki je najbližje vektorski reprezentaciji besede »kraljica«. Po podobnem načelu lahko jezikovni model predvidi na primer relacije med državami in njihovimi glavnimi mesti. Če vemo, kaj je Pariz v relaciji do Francije, model pravilno predvidi, da gre v primeru Berlina in Nemčije za primerljiv odnos. S tega vidika gre pri vektorskih vložitvah besed za kompleksnejšo reprezentacijo pomenskih odnosov med besedami kakor pri preprostih analizah kolokacij.

Word2Vec je nenadzorovana tehnika strojnega učenja, ki pa pri učenju jezikovnega modela omogoča nastavitve naslednjih parametrov: velikosti okna (ang. *window size*), s katero določimo obseg konteksta besede, ki bo uporabljen za učenje: na primeru povedi »Prvi slovenski roman je napisal Josip Jurčič.« bi bili v primeru velikosti okna *window*

= 2 za učenje uporabljeni dve besedi na vsaki strani ciljne besede – če je ciljna beseda »roman«, bo levi del konteksta »prvi« in »slovenski«, desni pa »je« in »napisal«; velikosti vektorja (ang. *vector size*), s katero določimo obseg skritega nivoja nevronske mreže, od te pa je odvisna že omenjena dimenzija vektorskega prostora; najnižje frekvence pojavitev besed v korpusu: če na primer uporabimo nastavitev *min_count = 100*, bodo pri učenju jezikovnega modela upoštevane le besede, ki se v korpusu pojavijo najmanj stokrat.

Word2Vec omogoča aplikacijo dveh različnih arhitektur prediktivnega modela za učenje: zvezne vreče besed (ang. *Continuous Bag-of-Words* oziroma CBOW) in preskočnega n-grama (ang. *Skip-gram* ali SG). CBOW poskuša prek konteksta priti do ciljne besede (za učenje besede »roman« bi kot vhod v nevronsko mrežo uporabili besede »prvi«, »slovenski«, »je« in »napisal« ter pri izhodu pričakovali besedo »roman«). Kontekst je na ta način predstavljen kot »vreča besed«, zajetih v oknu fiksne velikosti okoli ciljne besede. Arhitektura SG deluje v obratni smeri: kot vhod v nevronsko mrežo je uporabljena ciljna beseda (v tem primeru »roman«), na podlagi katere se model nauči njenega konteksta (pri izhodu iz nevronske mreže torej pričakujemo besede »prvi«, »slovenski«, »je« in »napisal«).

Izbira arhitekture je v veliki meri odvisna od velikosti korpusa. Medtem ko je arhitektura CBOW bistveno hitrejša, arhitektura SG natančneje predstavi redke besede in je zato primernejša za majhne nabore podatkov (Mikolov idr.), med katere z 11.454.627 besedami sodi tudi Korpus daljše slovenske pripovedne proze.

Korpus in predpriprava besedila

Za analizo je bil uporabljen Korpus daljše pripovedne proze KDSP 1.0 (Mandić in Erjavec). Korpus vsebuje 262 literarnih besedil, daljših od 20.000 besed, ki so izšla med letoma 1836 in 1918. Besedila so bila zajeta iz treh virov: iz Digitalne knjižnice Slovenije (dLib), projekta Slovenska leposlovna klasika na spletu, v okviru katerega so besedila v odprtem dostopu objavljena na portalu Wikivir, in korpusa ELTeC-slv (Erjavec idr.). Optično prepoznavanje znakov (OCR) vseh besedil je bilo ročno pregledano in popravljeno, korpus pa je bil nato stavčno segmentiran, lematiziran in tokeniziran. V korpusu so označene tudi imenske entitete ter morfosintaktične oznake MULTEXT-East in Universal Dependencies. Korpus je bil avtomatsko anotiran s programom CLASSLA (Ljubešić in Dobrovoljc).

Ta verzija korpusa je dostopna v formatu XML-TEI na repozitoriju CLARIN.SI.

Korpus je opremljen tudi z bibliografskimi in biografskimi metapodatki, kjer so bili ti na voljo. Preostali metapodatki vsebujejo tudi informacije o desetletju izida, literarni vrsti, literarni podvrsti, tematiki in stopnji kanoničnosti, ki je bila posebej relevantna za to analizo. Stopnja kanoničnosti je bila določena na podlagi vključenosti v slovenska šolska berila po letu 1980 oziroma v knjižno zbirko Zbrana dela slovenskih pesnikov in pisateljev. Dela, ki so vključena v omenjena izbora, so označena z visoko stopnjo kanoničnosti (teh je v celotnem korpusu 80), preostala pa z nizko stopnjo kanoničnosti (182).

Konstrukcija semantičnih polj

Analiza se v veliki meri naslanja na članek Laure Nelson, ki v svoji raziskavi konstruiranja družbenih institucij na ameriškem jugu t. i. dolgega 19. stoletja izhaja iz Marxove, Webrove ter feministične in kritične teorije družbe (Nelson 4–5). Osredotoča se na politiko, ekonomijo, kulturo in gospodinjstvo kot na osrednje družbene instance v obdobju industrijske revolucije v ZDA, ko politika, ekonomija in kultura izpodrivajo gospodinjstvo v zasebno sfero, podrejeno javni (4–5). Ker obravnava pol-periferno okolje v obdobju industrijske revolucije, so lahko družbene institucije, na katere se osredotoča, tudi izhodišče za semantične analize korpusa KDSP, ki zajema literarna dela, objavljena od druge četrtine 19. stoletja do konca 1. svetovne vojne. Po zgledu študije Laure Nelson bo kot ključen del analize reprezentacija teh družbenih institucij v korpusu prikazana s pomočjo konstrukcije semantičnih polj, ki se nanašajo nanje. Za razliko od omenjene študije, ki z vektorskimi vložitvami besed analizira korpus dnevniških, pisemskih, spominskih in drugih avtobiografskih spisov navadnih ljudi, bo v tem primeru analiziran korpus literarnih besedil s poudarkom na semantičnih poljih politike in kulture, namesto odnosa med intersekcionalnimi družbenimi identitetami in omenjenimi družbenih institucijami, pa bodo v ospredju odnosi med samimi institucijami.

Za te štiri institucije so bila skonstruirana semantična polja s pomočjo že omenjenega orodja Word2Vec v knjižnici Gensim v programskem jeziku Python. Jezikovni model je bil naučen na celotnem korpusu KDSP 1.0, ki pa je bil za potrebe učenja primerno razčlenjen in filtriran. Pri učenju so bile upoštevane samo leme glagolov in samostalnikov, izločene pa so bile imenske entitete. Na ta način so bile

ohranjene le polnopomenske besede z izjemo pridevnikov. Parametri so bili nastavljeni na *sg = 1* (uporabljena je bila arhitektura SG), *window size = 3*, *vector_size = 100* in *min_count = 10*. Z nastavitvijo spodnje meje frekvence besed, ki so bile uporabljene za učenje, je bil v veliki meri izključen šum (npr. napačno označene besede), ki se običajno pojavi ob avtomatskem označevanju starejših besedil. V prvem delu analize je bil za učenje modela Word2Vec porabljen celoten korpus besedil, v drugem delu analize pa sta uporabljena še podkorpus z nizko stopnjo kanoničnosti in podkorpus z visoko stopnjo kanoničnosti. Obsegi vseh treh korpusov po filtriranju nepolnopomenskih besed, pridevnikov in imenskih entitet so razvidni v Tabeli 1.

	Celoten korpus	Visoka stopnja kanoničnosti	Nizka stopnja kanoničnosti
Število dokumentov	262	80	182
Število pojavnic (ang. *token*)	3.790.873	1.197.035	2.593.838
Število različnic (ang. *type*)	62.392	36.690	48.506

Tabela 1: Velikosti korpusa in podkorpusov, ki so bili uporabljeni za učenje jezikovnih modelov Word2Vec.

Word2Vec vsebuje funkcijo *most_similar*, ki najde vektorje, ki so naj-bližje vektorski reprezentaciji besede, ki jo iščemo. Funkcija omogoča tudi, da s seštevanjem vektorjev iščemo semantično sorodne besede dveh ali več besed hkrati oziroma da semantično polje reduciramo z njihovim odštevanjem. To je uporabno predvsem pri večpomenskih besedah, ki jim na ta način lahko natančneje določimo kontekst. Semantično polje za sfero kulture je bilo pridobljeno z vektorji, ki so najbližje vektorjema »umetnost« + »literatura«, za sfero politike sta bila uporabljena vektorja »narod« + »država«, za ekonomijo »denar« + »kapital«, za gospodinjstvo pa vektor »družina« + »otrok«. Za vsako semantično polje je bilo izlu-ščenih 50 vektorjev z najbližjo kosinusno razdaljo[2] (Tabela 2).

---

[2] Kosinusna razdalja meri kosinus kota med dvema vektorjema in se giblje od -1 do 1, pri čemer sta pri vrednosti 1 vektorja identična, pri -1 nasprotna, pri 0 pa ležita pravokotno drug na drugega. Odštevanje kosinusne podobnosti od 1 daje mero razda-lje, kjer višje vrednosti implicirajo večjo podobnost med vektorji.

Izhodiščni vektorji	Semantično polje
umetnost + literatura	slikarstvo, ukus, slovstvo, kompozicija, proizvod, klasik, umotvor, stroka, realist, znanost, učenjak, matematika, časnikar, politik, duhovitost, impresionizem, naturalist, literat, filozofija, tehnika, predavati, proučevati, teorija, proučiti, izobraženost, problem, književnost, filozof, estetika, impresionist, izobraženec, razvoj, beležiti, študij, umeteljnost, veleum, razumnost, izobraženje, primes, zmisel, založnik, simbol, zanimevati, individualnost, jezikoslovec, umetnik, slikanje, izobrazba, poznanje, uporabljati
narod + država	ustava, boritelj, cesarstvo, revolucija, kerščanstvo, samostalnost, sloga, blaginja, enakopravnost, duhovenstvo, prekucija, geslo, gospodstvo, narodnost, vladanje, grofija, izobraženec, razširjanje, neodvisnost, enakost, živelj, napredek, blagovestnik, podjarmiti, državnik, duhovščina, oblastnik, oznanjevalec, prosveta, razširjevanje, organizirati, piratje, vernost, omika, suženjstvo, veleum, prekucuh, književnost, izročilo, katoličanstvo, individualnost, krščanstvo, slovanstvo, dinastija, republika, luteranstvo, državljan, vodilo, monarhija, očak
denar + kapital	svota, kredit, glavnica, gotovina, hranilnica, banka, izposojevati, štediti, posojevati, potrošiti, glda, odstotek, novec, prihranek, prištediti, novčič, belič, obresti, kupnina, varčevati, izplačevati, najemščina, posojati, znesek, trošek, procent, obrestovati, potrositi, vsota, primanjkljaj, tisočak, forint, svotica, honorar, zaračuniti, skopariti, plačevanje, zvišati, borza, odškodnina, srečka, poboljšek, vknjižiti, obogateti, izposojati, doklada, posojilo, zavarovalnica, pristojbina, zapravljati
družina + otrok	deca, dec, dvojček, otročič, sinček, preživiti, graščakinja, pestunja, punčika, pestovati, oskrbnica, edinica, prvorojenka, bolehnost, olikati, porod, posestnica, izrejati, najdenka, družinica, odrasti, tilda, polbrat, tašča, stariš, primožiti, edinka, dojka, suknarija, pomožiti, rednica, umerti, sinaha, vnuček, ranar, razvajati, očim, tovarišija, jedinec, rejenka, deco, načelnica, malček, pomreti, rednik, sneha, ptujec, soseščina, pomočnica, naslednica

Tabela 2: Semantična polja, pridobljena z iskanjem najbližjih vektorskih reprezentacij besed v celotnem korpusu.

Vizualizacija odnosov med diskurzivnimi polji

Vektorske predstavitve besed, ki so bile pridobljene z učenjem modela Word2Vec, obstajajo v stodimenzijskem vektorskem prostoru. Da bi bližino med vektorji lažje razumeli, moramo vektorski prostor zreducirati na tri ali celo dve dimenziji. Za jasnejšo predstavitev prekrivanja semantičnih polj je bila v tem primeru uporabljena dvodimenzionalna vizualizacija. Za redukcijo vektorskih dimenzij je bila uporabljena metoda glavnih komponent (ang. *principal component analysis* ali PCA), katere cilj je identifikacija ključnih komponent v naboru podatkov, ki v čim večji meri povzamejo variabilnost podatkov. S pomočjo PCA je tako mogoče zmanjšati dimenzionalnost podatkov in hkrati predstaviti informacije s čim manjšo izgubo natančnosti (Nelson 6).

Z redukcijo vseh 200 pridobljenih vektorskih predstavitev besed na dve dimenziji je bil ustvarjen diagram raztrosa, na katerem so razvidni odnosi med štirimi obravnavanimi semantičnimi polji (Slika 2). Ker so na diagramu predstavljeni le približki vektorskih koordinat, iz vizualizacije ne moremo interpretirati razdalj med posameznimi oznakami, še vedno pa lahko sklepamo, da so si gruče vektorjev, ki se prekrivajo, bolj podobne kakor tiste, ki so jasno ločljive. Z diagrama je razvidno, da so si vsa semantična polja blizu, pri čemer je gruča vektorjev, ki označujejo semantično polje ekonomije, najbližje gruči, ki predstavlja semantično polje gospodinjstva. To polje se dotika vektorskih gruč, ki predstavljata semantični polji kulture in politike in ki se prekrivata. Predvsem prekrivanje semantičnih polj politike in kulture nakazuje, da se v literarnih delih, ki so zajeta v korpus, temi kulture in politike pojavljata v skupnih kontekstih.
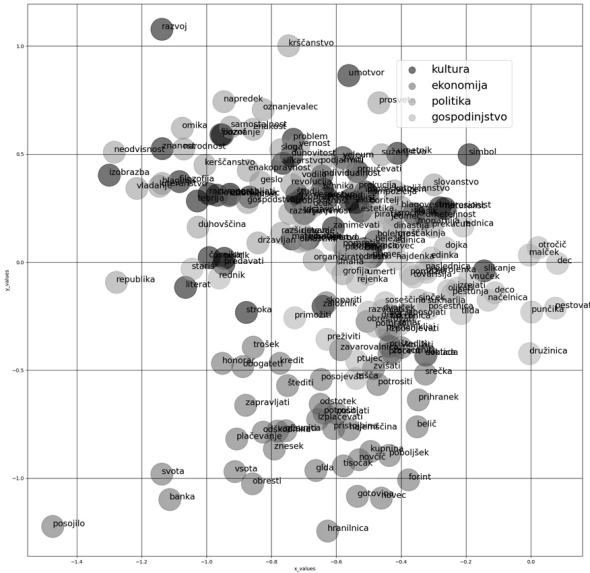
Graf 1: Vektorske vložitve besed na primeru jezikovnega modela, naučenega na celotnem korpusu.

Ker je bil namen analize preveriti, v kolikšni meri so odnosi med temi semantičnimi polji pogojeni s kanoničnostjo literarnih del, je bila analiza ponovljena še na podkorpusih del z visoko (Graf 2) in nizko (Graf 3) stopnjo kanoničnosti. Na grafih raztrosa je najočitnejša razlika v spremembi razmerja med gospodinjstvom in ekonomijo, pri čemer sta ti polji pri kanoničnih besedilih bolj prekrivni kakor pri nekanoničnih. Semantični polji kulture in nacionalne politike ostajata prekrivni v obeh podkorpusih, pri čemer je prekrivnost pri kanoničnih besedilih nekoliko očitnejša, saj je več besed v obeh semantičnih poljih identičnih (npr. »načelo«, »program«, »vrlina«, »znanost«, »navduševati«).
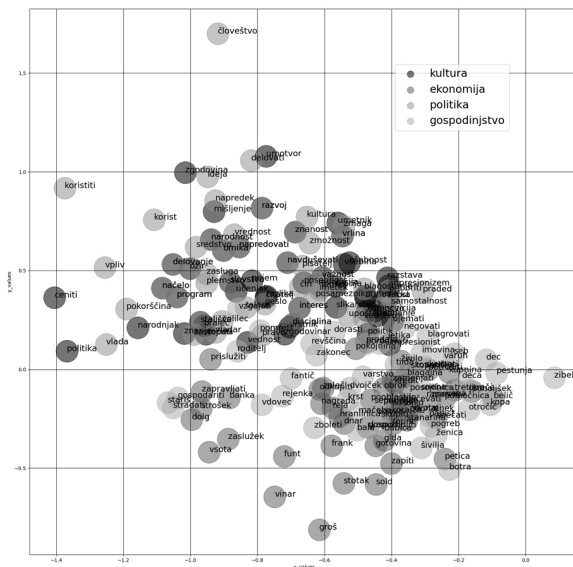
Graf 2: Vektorske vložitve besed na primeru jezikovnega modela, naučenega na podkorpusu besedil z visoko stopnjo kanoničnosti.
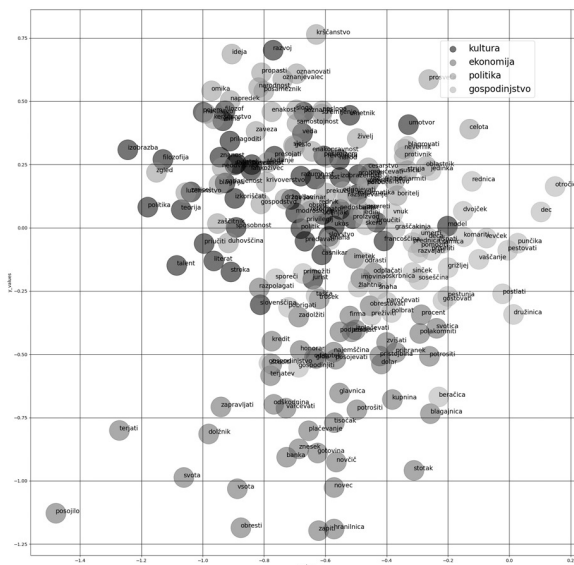
Graf 3: Vektorske vložitve besed na primeru jezikovnega modela, naučenega na podkorpusu besedil z nizko stopnjo kanoničnosti.

Evalvacija s pomočjo hierarhičnega gručenja

V kakšnih odnosih so posamezna semantična polja, lahko preverimo tudi s postopkom hierarhičnega gručenja (prim. Kljun, Teršek in Žitnik 8–9). Ker nas zanima razdalja med semantičnimi polji in ne posameznimi vektorji, so najprej izračunane povprečne vrednosti vseh vektorjev v posameznem semantičnem polju. Na ta način so pridobljeni štirje povprečni vektorji, ki imajo kakor izhodiščni vektorji 100 dimenzij. Razdalja med povprečnimi vektorji je izmerjena kot 1 simcos, kjer je simcos kosinusna podobnost med dvema povprečnima vektorjema. Za določanje razdalje med povprečnimi vektorji je bila izbrana Wardova metoda, ki je namenjena minimiziranju variance znotraj gruč. Rezultat so trije dendrogrami, na katerih je tako kakor na diagramih raztrosa razvidno, da sta si semantični polji kulture in politike bližji v primerjavi s semantičnima poljema ekonomije in gospodinjstva (Grafi 4–6). Večja ko je podobnost med gručami, nižje na osi y se stikata veji, ki povezujeta dve gruči.
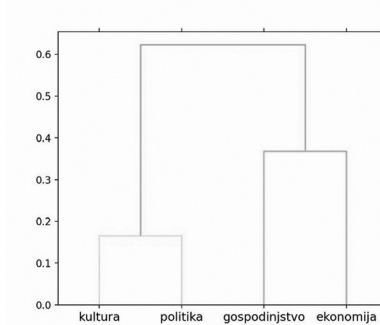
Graf 4: Hierarhično gručenje vektorskih reprezentacij besed na primeru celotnega korpusa.
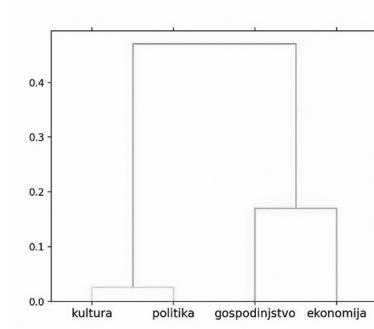
Graf 5: Hierarhično gručenje vektorskih reprezentacij besed na primeru podkorpusov besedil z visoko stopnjo kanoničnosti.
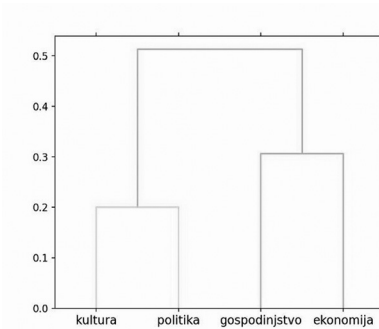
Graf 6: Hierarhično gručenje vektorskih reprezentacij besed na primeru podkorpusov besedil z nizko stopnjo kanoničnosti.

Kot gaže Graf 5, je za podkorpus besedil z visoko stopnjo kanoničnosti značilna bližina med semantičnima poljema kulture in politike, pa tudi med poljema ekonomije in gospodinjstva. Bližina semantičnih polj je (zlasti v primeru gospodinjstva in ekonomije) manj izrazita v korpusu del z nižjo stopnjo kanoničnosti.

Prešernovska struktura

Ker so rezultati analize pokazali najočitnejše prekrivanje semantičnih polj politike in kulture, se bo razlaga rezultatov podrobneje posvetila literarni tematizaciji teh dveh institucij v slovenskem prostoru v t. i. dolgem 19. stoletju. V zvezi s prekrivanjem semantičnih polj kulture in nacionalne politike se zdita relevantna dva koncepta slovenske preddigitalne literarne vede: Pirjevčev koncept prešernovske strukture iz leta 1969 in Ruplov koncept slovenskega kulturnega sindroma iz leta 1976. Oba koncepta temeljita na natančnem branju slovenskega literarnega kanona 19. stoletja. Pirjevec je s prešernovsko strukturo konceptualiziral delo in recepcijo Franceta Prešerna; koncept je (tam še pod imenom »Prešernova struktura«) vpeljal takole:

> [P]esnik je »voljan biti« pesnik in s tem tudi žrtev le, dokler mu je »zagotovljeno«, da je edini izvoljeni organ nebeške poezije in rajske lepote. Vse to mora biti »zagotovljeno«, sicer ostane narod ne le brez resnice, marveč tudi brez potrdil in brez mitologije. In vse dokler je narod blokirano gibanje, so takšna zagotovila tudi zares dana, iz česar hkrati sledi, da je Prešernova struktura zares plodna lahko le v določeni splošno zgodovinski strukturi. (Pirjevec 78)

Ruplov koncept slovenskega nacionalnega sindroma povzema to osnovno strukturo in jo s Prešerna razširi še na kanonične prozaiste 19. stoletja: Levstika, Stritarja, Jurčiča, Kersnika, Tavčarja in nazadnje Cankarja. Ruplova izhodiščna predpostavka je, da literatura »ni le preprost sektor družbene produkcije ali npr. družbene vrhnje stavbe, ampak skuša nadomeščati vse (oz. večino) funkcij, ki jih v razvitih družbah vršijo še (oz. predvsem) drugi sektorji vrhnje stavbe (pravno- -politični, izobraževalni, znanstveni … sektor)« (Rupel 424), njegova sklepna ugotovitev pa, da slovenska književnost 19. stoletja »nima tiste kvalitete, ki jo ima umetnost v razvitih družbah, kjer pretežno izraža oz. odraža družbene probleme in recimo osebne občutke oz. čustva«, saj po Ruplu »[l]iteratura pri Slovencih ni le odraz in izraz, temveč – ker nadomešča celo vrsto subsistemov – vtis, vzorec, naročilo, ukaz, ki ima določeno akcijsko, spodbujevalno funkcijo« (426–427).

157

Pirjevčevo fenomenološko študijo, ki je nastala kot del polemike o neoavantgardni poeziji poznih šestdesetih let, in Ruplovo kulturno-sociološko analizo, ki si prizadeva za strožji družboslovni metodološki aparat, druži stališče, da slovenska literatura estetsko stagnira zaradi svoje vloge pri konstruiranju nacionalne identitete, in sicer naj bi bila literatura nadomestek za politično, deloma pa tudi gospodarsko uveljavljanje nedržavotvornega naroda, podrejenega tuji oblasti (Juvan 298).

Oba koncepta sta bila deležna številnih kritik, v zadnjih letih predvsem na račun dejstva, da slovenski književnosti že v 19. stoletju v resnici ni pripadal monopol pri konstruiranju nacionalne zavesti, da se estetske stagnacije ne da pripisati le narodotvorni vlogi literature, da tovrstna vloga literature ni izključno slovenski fenomen ter da se velik del trivialne literature s svojim pomenom za narodni preporod ni ukvarjal (Juvan 315–317; Dović 288–290).

Za razlago rezultatov zgornje analize je relevantno predvsem dejstvo, da omenjeni teoretiki večinoma izpeljujejo koncepta prešernovske strukture in slovenskega kulturnega sindroma predvsem iz Prešernove avtotematske poezije, kjer naj bi bila literatura kot sredstvo legitimacije nacionalnega gibanja tudi najjasneje tematizirana.[3]

Rupel sicer razširi nabor besedil tudi na daljšo prozo in publicistiko 19. stoletja, a je uspešnejši pri branju neliterarnih spisov, saj išče zgolj artikulacijo političnih nazorov, ne pa specifične literarne tematizacije nacionalne vloge literature, ki jo pri Prešernu iščejo ostali trije teoretiki.[4] To prekrivanje literarne in politične tematike, ki ga Pirjevec, Juvan in Dović odkrivajo predvsem pri Prešernu, kvantitativna semantična analiza odkriva v daljši prozi, v kakršni je Rupel iskal politično tematiko brez ozira na literarno. Rezultati analize napeljujejo na tezo, da gre pri tej semantični prekrivnosti za reprodukcijo širše uveljavljenega diskurza, ki polji približuje drugo k drugemu in ki se ne pojavlja le v besedilih, ki eksplicitno tematizirajo sočasno politično (ali literarno) dogajanje (kakor na primer Kersnikova romana *Ciklamen* in *Agitator*, ki jima Rupel namenja veliko pozornosti), temveč posredno učinkuje v celotnem korpusu.

Tako Pirjevec, Juvan in Dović kakor Rupel prešernovsko strukturo oziroma slovenski kulturni sindrom obravnavajo izključno v kontekstu

---

[3] Po Pirjevcu in Ruplu z natančnim branjem Prešerna tudi Marko Juvan pride do zaključka, da je »nastavke za oblikovanje podobe in vloge 'nacionalnega pesnika', ki vsekakor sodi v območje prešernovske strukture in slovenskega kulturnega sindroma, […] oblikoval že Prešeren sam v svoji avtotematski poeziji« (Juvan 313).

[4] Eksplicitno izražanje političnega stališča pri analizi kanoničnih besedil obravnavanih avtorjev najde pri Prešernu, Tavčarju, Kersniku in Cankarju, medtem ko pri Levstiku, Jurčiču in Stritarju odkriva le implicitno političnost (Rupel 431).

slovenskega literarnega kanona. Pirjevec ugotavlja, da je literarni prostor, v katerem je nastajala (kanonična) književnost, »reprodukcija temeljnih razsežnost naroda« in pripada tistemu delu ljudstva, ki je bilo aktivno vpleteno v oblikovanje slovenske nacionalne politike, torej narodnozavednemu meščanstvu (Pirjevec 77). Obstajal pa naj bi tudi del ljudstva, na katerega nacionalna politika ni vplivala, v skladu s tem pa tudi vzporeden sistem »literature za ljudstvo«. Drugi pol slovenskega literarnega prostora naj bi bila po Pirjevcu torej literatura, ki se s svojo nacionalno vlogo ni obremenjevala in je ostala tudi zunaj nacionalnega kanona. Poznejše kritike prešernovske strukture in slovenskega kulturnega sindroma te Pirjevčeve teze niso problematizirale; prej jim je njegova redukcija prešernovske strukture na problem, ki zadeva le literarni kanon, služila kot argument za njeno nedoslednost (Juvan 316). Kvantitativni pristop nam omogoča, da poleg kanoničnih avtorjev ti konceptualizaciji slovenskega literarnega kanona apliciramo tudi na ostalo slovensko literaturo 19. stoletja. V nasprotju z uveljavljenim prepričanjem je primerjava med kanoničnimi in nekanoničnimi avtorji namreč pokazala, da semantična bližina literature in nacionalne politike za nekanoničen del korpusa ni nič manj značilna kot za kanoničnega.

LITERATURA

Brottrager Judith, idr. »Modeling and Predicting Literary Reception. A Data-Rich Approach to Literary Historical Reception«. *Journal of Computational Literary Studies*, let. 1, št. 1, 2022, https://jcls.io/article/id/95/. Dostop 12. 4. 2024.

Dović, Marijan. *Prešeren po Prešernu: kanonizacija nacionalnega pesnika in kulturnega svetnika*. LUD Literatura, 2017.

Eder, Maciej, in Artjoms Šeļa. »One Word to Rule Them All: Understanding Word Embeddings for Authorship Attribution«. *Digital Humanities 2022 Combined Abstracts*, Univerza v Tokiu, 2022, str. 199–202, https://dh2022.adho.org/. Dostop 12. 4. 2024.

Erjavec, Tomaž idr. *Slovenian Novel Corpus (ELTeC-slv): April 2021 release (v2.0.0)*. Zenodo, 2021, https://doi.org/10.5281/zenodo.4662600. Dostop 12. 4. 2024.

Hatzel, Hans Ole, idr. »Machine Learning in Computational Literary Studies«. *it – Information Technology*, let. 65, št. 4–5, 2023, str. 200–217.

Herrmann, J. Berenike, Joanna Byszuk in Giulia Grisot. »Using Word Embeddings for Validation and Enhancement of Spatial Entity Lists«. *Digital Humanities 2022 Combined Abstracts*, Univerza v Tokiu, 2022, str. 239–241.

Juvan, Marko. *Prešernovska struktura in slovenski kulturni sindrom*. LUD Literatura, 2012.

Kljun, Maša, Matija Teršek, in Slavko Žitnik. »Pomenska analiza kategorij sovražnega govora v obstoječih označenih korpusih«. *Uporabna informatika*, let. 30, št. 1, 2021, str. 3–18.

Ljubešić, Nikola, in Kaja Dobrovoljc. »What Does Neural Bring? Analysing Improvements in Morphosyntactic Annotation and Lemmatisation of Slovenian, Croatian and Serbian«. *Proceedings of the 7ᵗʰ Workshop on Balto-Slavic Natural Language Processing*, ur. Tomaž Erjavec idr., Association for Computational Linguistics, 2019, str. 29–34, https://aclanthology.org/W19-3704. Dostop 12. 4. 2024.

Mandić, Lucija, in Tomaž Erjavec. *Corpus of Longer Narrative Slovenian Prose KDSP 1.0.* ZRC SAZU, 2023, http://hdl.handle.net/11356/1823. Dostop 12. 4. 2024.

Mikolov, Tomáš, idr. »*Efficient Estimation of Word Representations in Vector Space*«. *International Conference on Learning Representations*, 2013, https://arxiv.org/abs/1301.3781. Dostop 12. 4. 2024.

Nelson, Laura K. »Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South«. *Poetics,* št. 88, https://doi.org/10.1016/j.poetic.2021.101539. Dostop 12. 4. 2024.

Pirjevec, Dušan. *Vprašanje o poeziji. Vprašanje naroda*. Obzorja, 1978.

Pollak, Senja, Matej Martinc, in Katja Mihurko. »Natural Language Processing for Literary Text Analysis: Word-Embeddings-Based Analysis of Zofka Kveder's Work«. *Proceedings of the Workshop on Digital Humanities and Natural Language Processing (DHandNLP 2020)*, ur. Maria José Finatto idr., CEUR-WS, Aachen, 2020, http://ceur-ws.org/Vol-2607/paper4.pdf. Dostop 12. 4. 2024.

Rupel, Dimitrij. *Svobodne besede od Prešerna do Cankarja*. Lipa, 1976.

Schneider, Felix idr. »Data-Driven Detection of General Chiasmi Using Lexical and Semantic Features«. *Proceedings of the 5ᵗʰ Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, ur. Stefania Degaetano-Ortlieb idr., Association for Computational Linguistics, 2021, str. 96–100, doi.org/10.18653/v1/2021.latechclfl-1.11. Dostop 12. 4. 2024.

A Quantitative Analysis of Relations between Semantic Fields in the Slovenian Narrative Prose of the Long Nineteenth Century

The article analyzes the relationships between semantic fields in Slovenian narrative prose of the long nineteenth century using the method of word embeddings. The corpus of longer Slovenian narrative prose (KDSP 1.0) was analyzed using the Word2Vec technology in the Python programming language. For the purposes of the analysis, semantic fields were constructed for four social institutions: economy, politics, culture, and the household. A set of words for each semantic field was obtained by identifying the 50 words with the greatest cosine proximity to the vector representation of each institution. The set of vectors obtained in this way became the quantitative basis of an investigation into the relations between these social institutions as they are narrated by the literary texts included in the corpus. The findings reveal a significant overlap between the semantic fields of politics and culture, thus offering a quantitative approach to a phenomenon that traditional literary scholarship tends to conceptualize as the Prešernian structure or the Slovenian cultural syndrome.

Razprave / *Articles*
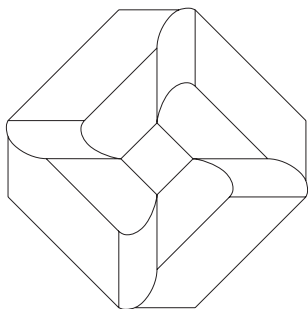
Ronsardova in Cowleyjeva pindarjevska oda

Vid Snoj

Univerza v Ljubljani, Filozofska fakulteta, Oddelek za primerjalno književnost in literarno teorijo, Aškerčeva cesta 2, 1000 Ljubljana
https://orcid.org/0000-0001-9128-9919
vid.snoj@ff.uni-lj.si

Pindar je po vrnitvi na Zahod, ki ga je po koncu antike od grštva ločila jezikovna pregrada, spet dobil status »prvaka lirikov«, kakršnega je imel pri Grkih. V pozni renesansi je najvišja zvrst lirike postala oda, kot so epinikij tudi imenovali že aleksandrijski učenjaki, in njen najimenitnejši zgled Pindarjev epinikij s triadično kitično zgradbo. Ob tem je na recepcijo Pindarjevega pesništva merodajno vplivala pesem, v kateri je Horacij Pindarja upodobil kot divjo gorsko reko in mu tako dal podobo metričnega divjaka. V ljudskem jeziku je pindarjevsko odo s triadično zgradbo uveljavil Pierre de Ronsard, s čimer si je prislužil trajno mesto v francoski in evropski literarni zgodovini. Prijel se ga je sloves francoskega Pindarja. Pindarjevske ode so začeli pisati po Ronsardovem zgledu, ne da bi Pindarja poznali iz prve roke, in Ronsardova oda je segla tudi v Anglijo. Vendar mejnik v angleškem pindariziranju pomenijo »Pindarjevske ode« Abrahama Cowleyja, za katere sta značilni zrahljana metrika in svoboda preskakujoče misli. S prevajanjem Pindarja in pesnjenjem po njegovem »entuziastičnem načinu« je Cowley med svojimi rojaki sprožil modo iregularnih pindarjevskih od, to modo pa je čez čas zajezil William Congreve s popolnoma nasprotno podobo vélikega grškega lirika. V njej se Pindar kaže kot povsem regularen pesnik, čigar metrika se sklada s sovisnostjo njegovih misli.

Ključne besede: starogrška poezija / Pindar / literarna recepcija / pindarjevska oda / Horacij / Ronsard, Pierre de / Cowley, Abraham

Na Zahod so rokopise Pindarjevih epinikijev iz Bizanca še pred njegovim padcem leta 1453 prinesli italijanski humanisti, kot eden prvih učenjak in pesnik Francesco Fidelfo.[1] Pindarja so začeli posnemati pesniki med njimi, sprva v grščini in latinščini, čeprav je vse, kar na primer v Fidelfovih latinskih pesmih spominja na Pindarja, mogoče najti tudi pri Horaciju (prim. Schmitz 24). Posnemanje Pindarja je

---

[1] Razprava je nastala v okviru raziskovalnega programa P6-0239, ki ga je sofinancirala Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije.

163

v Italiji sčasoma, v dvajsetih in tridesetih letih 16. stoletja, prešlo v ljudski jezik. Sredi tega stoletja je v vrh zunaj klasičnih jezikov po odmevnosti, če že ne po globini, pognalo s pindarjevsko odo Pierra de Ronsarda v Franciji.

Dobro je znano, da je v srednjeveški Evropi nastalo drugačno pesništvo od antičnega. V ljudskih jezikih romanskih dežel je mere z enakim trajanjem kot gradnik zamenjalo enako število zlogov v verzni vrstici, ki jo je pod vplivom poantičnega latinskega pesništva s preprosto enozložno rimo na koncu krepila čedalje večzložnejša asonanca in za njo nato rima (prim. Gasparov 101). Posledica humanističnega vračanja k antiki na področju lirike pa je bila, da je v renesančno pesništvo zmagoslavno stopila oda. »Oda« je, ko je šlo za tematsko visoko pesem bodisi s preprostejšo horacijevsko bodisi z zapletenejšo pindarjevsko kitično zgradbo, v Evropi postala ime za imenitnejšo in v Franciji sploh za najvišjo lirsko zvrst, čeprav te besede za svoje pesmi nista rabila ne Horacij ne Pindar. Za prevedek *eîdosa*, s katerim so Pindarjev slavospev v čast zmagovalcu na športnih igrah označevali aleksandrijski učenjaki v alternaciji z »epinikijem« (prim. Revard 122), so jo izbrali renesančni preučevalci pesništva in pesniki sami. Čeprav *eîdos* v aleksandrijski rabi ni imel pomena same po sebi bivajoče resničnosti ali večnega lika oziroma vzorca stvari kakor pri Platonu, so jeziki renesanse izgovarjali »odo« kot izborno, odlično in navsezadnje tako rekoč kot vzorčno vrsto lirike. Po antičnih zgledih ubrana oda je v ljudskih jezikih Evrope dosegla prevlado nad dediščino njihovih spevov in pesmi iz srednjega veka.

Pindarjevo vračanje na Zahod se je godilo prek Horacija. V pozni renesansi je Horacij predal prvenstvo spet Pindarju, čeprav so Horacijeve pesmi tudi pozneje pogosteje izdajali in komentirali kakor Pindarjeve ter jih več prevajali in posnemali, obravnavali v šoli in verjetno zunaj nje prav tako več brali. Mnoge izmed izdaj Pindarjevih epinikijev v 16. in še v 17. stoletju so na naslovnici imele Kvintilijanovo oznako *princeps lyricorum*[2] in na začetku skrajšano Horacijevo pesem v Pindarjevo čast, njenih prvih sedem kitic, ki so izrecna hvalnica Pindarju. Pred očmi bralca je, preden je prišel do epinikijev, vstala Horacijeva podoba Pindarja.

Prva kitica omenjene pesmi, druge iz četrte knjige Horacijevih *Pesmi*, je nikakor ne preprosta zatrditev Pindarjeve nenadkriljivosti. Horacij Pindarjevega posnemovalca, se pravi tega, ki gre z vnemo za Pindarjem v tekmo z njim, primerja z Ikarjem, letalcem iz grškega mita:

---

[2] Quintilian 466; *Šola govorništva* 8,6,71. Prim. slovenski prevod, ki Pindarja splošneje, a hkrati presežno preoznači iz »prvaka lirikov« v »prvaka pesnikov« (617).

dvigne se »z dajdalsko močjo voščenih kril«, a bo strmoglavil in dal ime »steklenemu morju« (4,2,1–4).[3] Ker je Pindarjev posnemovalec kakor Ikar, ki je poletel v nebo na umetnih krilih iz peres, katera je z voskom zlepil njegov oče Dajdal, Pindarju z dajdalsko veščino nikakor ne bo kos. Njegov delež bo slava, vendar bo to slava posnemovalca. Njegovo ime bo preživelo. Morje, v katero bo strmoglavil, bo sicer poimenovano po njem, posnemovalcu, toda to bo morje ogledalo, na katerem se bo za prihodnje rodove zibal odsev nedosegljivega Pindarja.

Horacij predvsem ne krene za Pindarjem v posnemanje njegove ritmično razgibane metrike. Namesto tega ubira svojo pesem v metrično pravilnih, enakomernih, ritmično enoličnih sapfičnih kiticah. O Pindarju oziroma njegovem spevu, ki ga je Pindar sam imenoval »deroč val« (Ol. 10,10),[4] nato sledi (4,2,5–8):

> monte decurrens velut amnis,
> imbresquem super notas aluere ripas,
> fervet immensusque ruit profundo
>   Pindarus ore ...

> Kot reka, ki z gore dol teče,
> od déžja narasla čez znane bregove,
> Pindar besni in iz globokih ust strmoglavlja
>   neizmeren …[5]

Horacij v figuri od dežja narasle, besneče, v dolino strmoglavljajoče gorske reke pesni Pindarjevo pesniško veličino. V tej figuri se pevec zliva s petjem: Pindar je reka. Pindarja reko zajema in nosi metrično besnenje. Horacij tudi pravi, da je Pindar *numerisque fertur / lege solutis* (11–12), »od mer nošen, / osvobojenih zakona«, v svojih ditirambih, a hkrati sugerira »paradigmatičnost 'ditirambičnega' načina za *vse* Pindarjevo pesništvo« (Freis 33; kurziva v izvirniku) in kot tisto obliko božanskega navdiha, ki ga žene v metrično besnenje, bakhično obsedenost. S to svojo pesniško označitvijo Pindarjevega pesništva je za stoletja opredelil njegovo podobo na Zahodu. Ker pa se iz antike ni ohranilo nobeno takšno delo o liriki, kot je Aristotelova *Poetika*, ki

---

[3] Horace, *Odes* 220. Najbolj znana antična upesnitev mitske zgodbe o Ikarjevem letu je Ovidijeva iz *Metamorfoz* (8,183–215); za slovenski prevod glej *Metamorfoze* 53–55.

[4] Snell–Maehler 42, zv. 1; prim. Senegačnik 85. Pindarja po Snellovi in Maehlerjevi izdaji prevajam sam, v opombah pa napotujem tudi na slovenski prevod, delo Braneta Senegačnika.

[5] Moj prevod; enako velja tudi za prevode v nadaljevanju.

je postala temelj za novoveško teorijo epike in dramatike, se je misel o liriki na splošno celo neposredno vrtela okrog Pindarja in Horacija. Kot pravi Hans-Henrik Krummacher: »Humanistična teorija lirike je zvečine eksegeza Pindarjevih in Horacijevih pesmi, ki jo usmerjata dve antični poroštvi, to, kar je v svojem katalogu vzorčnih avtorjev o obeh povedal Kvintilijan, in počastitev Pindarja na začetku druge ode v četrti knjigi Horacijevih od.« (Krummacher 14)

Po drugi strani je Horacij vplival na preobrazbo epinikija v enkomij oziroma, kar najbolj splošno rečeno, pesem v človekovo čast. Pindarjev epinikij je, z njegovima besedama, »spev« ali »hvalnica« za zmago na igrah, a slavljenje v njem nikdar ni namenjeno samo zmagovalcu, ampak *tudi in predvsem* bogu. Začetek »Drugega olimpijskega slavospeva« se glasi: »Hvalnice, fórming gospodarice, / komu bogov, komu junakov, komu naj mož zapojemo slavo?« (Snell–Maehler 8, zv. 1; prim. Senegačnik 26) Pindar človeka s tem, da ga postavi na zadnje mesto, opominja na njegov položaj; prvi med ljudmi, čigar zmaga si zasluži slavljenje, je za mitskim herojem in za bogom. Horacij pa v svojih *Pesmih*, na začetku dvanajste iz prve knjige, to zaporedje obrne: »Koga mož ali junakov boš, Klio, za slavljenje / z liro ali rezko piščaljo izbrala? / Koga bogov?« (Horace, *Odes* 44) Izbira tu pade na *principsa*, na Avgusta.

A ne le to. V renesansi seveda že zdavnaj ni bilo več starodavne ustanove panhelenskih iger, od katerih so najstarejše, olimpijske, trajale skoraj toliko kot grška antika, od leta 776 pr. Kr. do poznega 4. stoletja, ko je rimski cesar Teodozij I. z odlokom zaprl templje in odpravil poganske obrede ter z njimi vred tudi igre. Vendar Horacij v dveh pesmih iz četrte knjige, četrti in štirinajsti, v katerih se sicer očitno zgleduje pri Pindarjevih epinikijih, namesto zmagovalca na igrah slavi tistega v vojni. Še več: čeprav je to v prvem primeru Druz in v drugem mladi Tiberij, pogled upira v vladarja, ki stoji za njima – »njegove oči so na Avgustu« (D'Angour 61). Ta pogled, to oko, ki se je na vladarja premaknilo že v helenističnem grškem pesništvu, je Ronsarda in druge poznorenesančne pesnike za njim napeljalo k temu, da so Pindarjev slavospev za zmago na igrah priredili v odo za vojaško zmago v kraljevo slavo. V poštev je kmalu začela prihajati tudi vrsta različnih priložnosti, ki so pindarjevski odi rabile kot povod za slavljenje namesto zmage, skupaj s slavljenci od kralja do kopice drugih pomembnih ljudi, ki so postajali njeni naslovljenci. Tako se je Pindarjev slavospev – *epiníkion*, se pravi »zmagospev«, kot so ga nekdaj poimenovali Aleksandrijci in s tem poimenovanjem zadeli njegovo osredinjenost na zmago (*níke*) – razpustil v enkomij. Pindar je postal »pojem enkomiastičnega pesnika« (Schmitz 98).

Z vzponom ode v hierarhiji lirskih zvrsti je recepcija Pindarja šla v drugo smer kot pri nemških reformiranih bralcih. Na podlagi dovzetnosti, ki so jo za cvet poganskega mišljenja in pesnjenja začeli izkazovati že krščanski apologeti v antiki, je zanimanje za Pindarjeve gnome, izreke z nravstveno vsebino, oživelo v renesansi, in sicer še zlasti potem, ko so po padcu Bizanca z učenjaki prišle na Zahod trume rokopisov, grško spisje s pesništvom vred. Na začetku 16. stoletja so se druga za drugo pojavile kar tri tiskane izdaje Pindarjevih epinikijev v grškem izvirniku, najprej dve v Italiji, prva leta 1513 v Benetkah in druga s sholijami dve leti pozneje v Rimu – ta je sploh najstarejša tam natisnjena knjiga –, ter nato še tretja leta 1526 v Baslu. Živo zanimanje za Pindarja je vzniknilo tako pri reformistih, krščanskih humanistih, ki so ostali v rimskokatoliški Cerkvi, kakor pri reformiranih, ki so izstopili iz nje. Humanisti so sledili klicu *ad fontes*, »k izvirom«, ter se ob odporu do zapletene učenosti poznosholastične logike in dialektike vračali ne samo k izročilu cerkvenih očetov, ampak tudi antične filozofije in pesništva.

Pindarja v *Adagia*, zbirki grških in latinskih pregovorov in modrostnih izrekov, prosto navaja eden največjih, če ne celo največji humanist Erazem Rotterdamski, čigar vračanje h grško-rimski antiki se je godilo skladno z zgodnjekrščanskim razumevanjem njene udeleženosti v semenskem logosu. Po drugi strani so med reformiranimi, ki so v svojem nasprotovanju Rimu in navezanosti katoliškega krščanstva na rimstvo še toliko bolj iskali stik z grštvom ter se pri tem navduševali nad Pindarjem, začele nastajati zbirke gnom iz njegovih epinikijev. Že bežen pogled pa pokaže, da je pesnike francoske Plejade namesto gnome v Pindarjevih epinikijih pritegoval mit. Vendar Plejadovci dovzetnosti za Pindarjevo pesništvo niso samo obrnili drugam, ampak so si prav s pindariziranjem, v katerem so se preizkušali v svojem pesništvu, izdolbli pomembno mesto v lapidariju francoske in evropske literarne zgodovine.

»Plejada« je krog mladih pesnikov okrog sebe poimenoval Ronsard, in sicer po skupini sedmih zvezd, katerih svetloba se zliva v eno samo žarenje, ter morda tudi z mislijo na skupino aleksandrijskih pesnikov iz 3. stoletja pr. Kr., ki je že nosila to ime. Ronsard ni prvi v Franciji rabil besede »oda« niti nemara ni prvi pisal od v francoščini, vsekakor pa je s svojim pesniškim nastopom zmagovito uveljavil pindarjevsko odo s triadično zgradbo.

Pindarjeve epinikije je v obliki, v kateri jih poznamo, vzpostavila aleksandrijska učenost. Prvotno so obstajali v glasbi in plesu, a se je od njih ohranilo samo besedilo. To je na podlagi rokopisov, ki so se v

aleksandrijsko knjižnico stekali iz vse Grčije, izoblikoval Aristofan iz Bizanca, ki je na začetku 2. stoletja pr. Kr. na čelu knjižnice nasledil Eratostena iz Kirene. Pindarjevi epinikiji so v verzni in kitični obliki, v kateri so prišli do nas, njegovo delo, pri čemer so mu bili morda v pomoč še zlasti rokopisi z glasbeno notacijo (prim. Irigoin 8). V Aristofanovi ureditvi imajo triadično kitično zgradbo. Sestavljajo jih strofe in antistrofe, ki jih sklepajo epode, vendar nimajo stalnih stopic oziroma metričnih enot; vsaka strofa, nasprotno, pozna svoje stopične kombinacije, ki jim ustrezajo tiste v parni antistrofi, in vsaka epoda spet svoje, s katerimi respondirajo stopične kombinacije v drugih epodah istega epinikija.

Tako je aleksandrijska učenost približno razbrala Pindarjeve mere, po njegovih epinikijih ubrane ode s triadično zgradbo pa je v ljudskem jeziku začel pisati Ronsard. Kot v obširni in prodorni knjigi o delovanju antičnega literarnega izročila v zahodni literaturi pravi Gilbert Highet, je Ronsard postal »utemeljitelj povzdignjenega lirskega pesništva po antičnih zgledih ne le za Francijo, ampak za vso moderno Evropo« (Highet 274).

Ronsard in njegovi so se s Pindarjem seznanjali prek Jeana Dorata, ki velja za pravega začetnika filološkega preučevanja grškega pesništva v Franciji. Dorat jim je pojasnjeval Pindarjev težki jezik in razkazoval lepote njegovega pesništva, ob tem da je tako kot italijanski humanisti tudi sam pisal pesmi v grščini in latinščini, med njimi latinske pindarjevske ode. Plejadovci, ki so v javnost stopili leta 1549 s spisom Joachima Du Bellayja *Obramba in poveličanje francoskega jezika*, pa so poudarek postavili na ljudski jezik, na pisanje v francoščini. V tem Plejadinem »manifestu« Du Bellay zavrača barbarstvo starega francoskega pesništva ter ob vpeljavi grških in rimskih pesniških zvrsti v francoščino, predvsem »od, še neznanih francoski Muzi« (Du Bellay 30), terja od pesnikov celo tvorjenje novih besed iz grških in latinskih. Kajti za to, da francoščina še ni zacvetela, ni kriva narava tega jezika, ampak kultura »tistih, ki so ga imeli v varstvu, a ga niso dovolj obdelali [*ne l'ont cultivée à suffisance*]« (8). Dobra zemlja jezika še ni bila deležna prihoda prave kulture in kultivacije.

Najodličnejša izpolnitev Du Bellayjevega programa je prišla s pindarjevsko odo že leto poznejе, ko so izšle Ronsardove *Prve štiri knjige od*. Čez dve leti je »prvim štirim knjigam« sledila še peta, Ronsard pa je imel prav poseben razlog za to, da je tehnični, paratekstualni izraz »knjige« iz notranjosti svoje prve zbirke, v kateri zarisuje njeno ureditev, postavil v naslov. »Knjige«, in sicer »štiri«, so se v naslovu te zbirke znašle zaradi pesniških del, s katerimi se kosa njegova oda – štirih knjig

Pindarjevih epinikijev in prav tako štirih knjig Horacijevih *carmina*. Čeprav Ronsardove ode razen iz Pindarja in Horacija zajemajo tudi iz Anakreonta in drugih antičnih virov, so najbolj stremljive med njimi zasnovane po Pindarjevo. Triadično kitično zgradbo jih ima dober ducat v prvi knjigi.

S štirimi knjigami od je Ronsard odgovoril na Du Bellayjev poziv k pisanju v Franciji dotlej neznane zvrsti lirike, poglavitnega obeta za *illustration*, »poveličanje« ali »poplemenitenje« francoščine. Kot sam pravi v predgovoru »Bralcu«, se je »napotil po neznani poti«, da bi pokazal, kako »slediti Pindarju in Horaciju« (Ronsard 11). V pričevanje o takšni napotitvi citira Horacijeve besede: »Prvi zasádil svobodne stopinje sem v prazno, / nisem s tujo nogo jih vtisnil« (Ep. 1,19,21–22; Horace, *Satires* 382). Pot v neznano je v jeziku svoje domovine sicer tlakoval s toposi; »neznana pot« sama je namreč postala topos, saj se že v ozadju navedenih Horacijevih besed svetlikajo Kalimahova »nezvožena póta« (prim. »Himno Apolonu«, v. 27–28; Callimachus 28) in za Kalimahovimi Pindarjeva »nezvožena pot«.[6] Kljub temu pa je Ronsard zatrjeval svoje prvenstvo. V odi iz svoje prve zbirke trdi: »Prvi sem / v Franciji pindariziral« (2,2,36–37), in to trditev v različnih oblikah večkrat ponovi, v eni izmed krajših in preprostejših poznejših od, ki so ostale zunaj obeh zbirk, na primer z besedami, da se prav v njegovih odah Pindarjeve »Tebe lahko prepoznajo / pofrancozene od nas« (34–35; Ronsard 135, 444).

Še več: Ronsard o sebi pravi, da se je na svoji »francoski poti« ločil tudi od Horacija, saj ta s tem, ko ni poskušal loviti Pindarjevih ritmov, svojih pesmi seveda tudi ni gradil s strofami, antistrofami in epodami ter medstrofičnimi in medepodičnimi responzijami. V četrti epodi ode, posvečeni Du Bellayju, navezujoč se na začetek Horacijeve pesmi v Pindarjevo čast pribije (1,11,165–168):

> Par une cheute subite
> Encor je n'ay fait nommer
> Du nom de Ronsard la mer
> Bien que Pindare j'imite. (Ronsard 103)

> Zaradi nenadnega padca
> nisem z imenom Ronsard
> poimenoval znova morjá,
> čeprav posnemam Pindárja.

---

[6] Prim. fr. 7b,10–12; Snell–Maehler 36, zv. 2.

Česar ni zmogel Horacij, je zdaj Ronsard: zavrnil je Horacijevo svarilo in šel za Pindarjem, vendar tako, da je oral francoščino, ki je bila dotlej v pomanjkljivem varstvu oziroma oskrbi, kot je zapisal Du Bellay. Izognil se je nevarnosti, o kateri je Horacij dejal, da grozi posnemovalcu Pindarja, in s pindarjevsko odo, ki ima triadično zgradbo, presegel Horacija. Kot Pindarjev posnemovalec ni strmoglavil in ni kakor Ikar dal morju svojega imena. Dvignil se je visoko v nebo kakor Pindar, a je hkrati, bi lahko dodali v duhu Du Bellayjevega »manifesta«, v francoskem jeziku našel zemljo, ki jo je obdelal do cveta in sadu.

Tako je o svojem dosežku pesnil Ronsard, takoj pa mu je pritegnil Du Bellay. V zadnjem sonetu iz *Oljke*, prve zbirke sonetov v francoščini, ki je izšla istega leta kot Ronsardova zbirka od, pravi (5–8):

> Quel Cygne encore' des Cygnes le plus beau
> > Te presta l'aile et quel vent jusqu'aux cieux
> > Te balança le vol audacieux
> > Sans que la mer te fust large tombeau? (Du Bellay 146)

> Kateri labod, med labodi povrhu najlepši,
> > ti let je pripravil in veter kateri let drzni
> > ti vse do neba zravnotéžil,
> > ne da širen grob morje postalo bi ti?

To retorično vprašanje, postavljeno spet v očitni navezavi na Horacijevo hvalnico Pindarju, nima neznank: ti, katerega nagovarja Du Bellay, je Ronsard in labod Pindar. Veter, ki nosi oba – Pindarja, ki je pripravil Ronsardov let, in Ronsarda, ki je Pindarju sledil tako, da ga je s svojim letom *nasledil* –, je isti veter, božanski navdih. Po njem se posnemovalec izenačuje s svojim zglednikom, po njem mu pripada visoko dostojanstvo njegovega naslednika. Temu Du Bellay v satirični pesnitvi »Muzagnomahija ali vojna podgan« dodaja, češ, »to Pindar je francoski, / ki s Tebami in z Apulijo / jezik bogati vendomski« (274–276; Du Bellay 154). Ronsard, ki se je od pindarjevske ode potem vrnil k Horaciju in lažjim kitičnim oblikam, je torej Pindarjeve Tebe in Horacijevo Apulijo pripeljal v Vendôme, deželo, kjer se je rodil. S posnemanjem antičnih zgledov pod božanskim navdihom je skladno s programom plemenitenja francoskega jezika obogatil svojo domovino.

Sodbo, da je Ronsard francoski Pindar, so sprejeli tudi drugi pisci v Franciji. V pisanju nekaterih izmed njih je francoski Pindar kot najvišji zgled včasih celo lahko zamenjal grškega. Pindarjevo pesništvo je pogosto delovalo drugotno, prek Ronsardovega; pindarjevske ode so

pisali po Ronsardovih, ne da bi Pindarja poznali iz prve roke. Kmalu po uveljavitvi Plejadinega pesništva so se sicer zlasti z reformirane, hugenotske strani začeli pojavljati napadi na njegovo posvetnost in poganstvo, in Ronsardovo pesništvo je bilo razglašeno za nravstveno manj vredno od Pindarjevega. Vendar tudi hugenotska polemika proti Ronsardu ni mogla mimo Horacija. Najostrejšo ost je dobila v jasnem namigu na Horacijevo figuro za Pindarjevega posnemovalca, namreč da je Ronsard padel kakor Ikar.

Ronsard v svojih pesmih pogosto govori o *fureur divine*, »božanski blaznosti«. V srednjem veku na Zahodu so za *furor divinus* kot izvir pesništva sicer vedeli od piscev poznoantičnega Rima; meniški skriptorij ga je »ohranil in črkoval po tem zgledu«, a je šele »ustvarjalni eros italijanske renesanse v teh črkah znova obudil duha« (Curtius 439). *Furor divinus* se je v pesništvo prvič po antiki vrnil tik pred iztekom *Quattrocenta* z latinsko »Himno soncu« italijanskega humanističnega učenjaka in pesnika grškega rodu Micheleja Marulla. Ronsard, ki je Marulla cenil in mu posvetil epitaf, je svoj furor prikazoval kot pindarjevski, in Plejadovci so božansko blaznost, kot priča vrsta njihovih poetoloških izjav, vzdignili na svoj prapor. Toda o tej blaznosti ne govori Pindarjevo pesništvo.

»Vedežuj, Muza, jaz pa bom prerokoval,« pravi Pindar[7] in sebe glede na to, da je za domovino Muz v grškem mitu veljala Pierija, imenuje tudi »pojoči prerok Pierid« (Paj. 6,6; Snell–Maehler 27, zv. 2; prim. Senegačnik 340). Samega sebe Pindar torej uzira »v enakem razmerju do Muz, kot je tisto, v katerem stoji prerok nasproti orakeljskemu bogu«, oziroma, še natančneje, sam »stoji nasproti Muzi, kot nasproti Pitiji stojijo svečeniki v Delfih« (Bowra 4, 8). Kakor je Pitija, visoka Apolonova svečenica v Delfih, neartikulirano govorila ali jecala v vedeževalskem blaznenju, to pa je ljudem, ki so po delfski orakelj prihajali z vseh koncev Grčije, artikulirano in včasih zajeto v heksametre posredoval svečenik oziroma *prophétes* kot tisti, ki govori naprej (iz gr. *prophemí*, »povem vnaprej«, »napovem« ali tudi »povem naprej«),[8] tako govori Pindar. Muza je kakor Pitija: ona vedežujoč blazni. Nasprotno je on, Pindar, kakor delfski profet ali prerok: njeno blaznenje umerja v človeku umljivo, vendar kljub temu govorico, ki se

---

[7] Fr. 140d; Snell–Maehler 119, zv. 2; prim. Senegačnik 363.

[8] Antični viri o tem, kako je vedeževala Pitija, si med sabo nasprotujejo. Nekateri govorijo o Pitijini obsedenosti od boga, ki spominja na ekstatični trans, drugi – kljub *maní* – o jasnosti njenega duha in govorice. Sodobno mnenje, ki se nagiba na drugo stran, povzema Fritz Graf, namreč da je Pitija »govorila s svojim glasom bodisi v heksametrih bodisi v prozi, ki so jo potem v verze spravili izvedenci« (Graf 56).

temni po svojem izviru. In kakor zmage na igrah ni brez človekovega napora, tudi zmagospev, ki prihaja od Muze, terja človeški delež. Ta ni v *téchne*, ampak v *sophíi*, »modrosti«, ki je nekaj takega kot vednost, vznikajoča »znotraj«: »Moder je, kdor mnogo ve po naravi« (Ol. 2,86; Snell–Maehler 12, zv. 1; prim. Senegačnik 31). Pindarjeva ni veščina ali umetnost, katere se je mogoče priučiti, ampak modrost nekoga, ki se mu *phyâ*, »po naravi«, se pravi po vznikanju v njem samem, daje, da misli pesniško – in ve pesniti.

O tem, da je pesnik, ki govori pod božanskim navdihom, iz uma in da je pravi govorec njegovih besed bog, da pesništvo tako kot vedeževalstvo (*mantikê*), ki si besedni koren deli z blaznostjo (*manía*), izvira iz blaznenja in da je stanje, v katerem je pesnikova duša med pesnjenjem, enako bakhičnemu navdušenju, skratka, da je pesništvo »blazneča umetnost« – o *furor divinus sive poeticus* je govoril Platon.[9] Ko so Ronsard in njegovi sledilci imenovali in opisovali svoj *furor poeticus*, niso govorili po Pindarjevo, ampak so izhajali iz Platona, po katerem je božansko blaznenje pesnikov »'dionizično' stanje« (Moffitt 61). Pri tem pa jih je spet usmerjala Horacijeva podoba »ditirambičnega« Pindarja, po kateri Pindarjevo metrično besnenje izdaja bakhično obsedenost.

Toda Ronsardov slog je kljub njegovemu sklicevanju na furor daleč od bogastva Pindarjevega sloga. Pri Pindarju so besede nabite s pomeni (in zato minsko polje za vsakega prevajalca). Prav tako pomenljiva je njihova skladnja. Pindarjeve besede, ki stojijo kakor stebri,[10] valovijo v pomenjanju. Njegovi stavki v silovitem razmahu prestopajo vrstice in kitice ali pa se začnejo in končajo, še preden se konča vrstica. Nasprotno se Ronsardovi stavki zravnajo. Ti stavki so premočrtni, tečejo gladko naravnost in se pogosto približajo rimani prozi. Tudi če se misel v krajših vrsticah nadaljuje iz ene v drugo, se skoraj zmeraj sklene z vrstico, ki jo kot zvočni obroč sklepa rima. V Ronsardovem besednjaku je veliko pomanjševalnic, a redko kaj podobnega Pindarjevim novotvorbam. Ronsardova govorica sicer deluje temno, zlasti kadar namiguje na malo znane mite, vendar se pomen, brž ko bralec prepozna njihov vir, iz tega vira tudi lepo zjasni. Ronsard v jeziku ne orje na globoko kakor Pindar.

---

[9] Prim. *Ion* 533d–534e in *Fajdros* 244a–245b (Platon 961–962, 545–546).

[10] V spisu *O sestavi besed* že Dionizij iz Halikarnasa prepoznava Pindarja za imenitnega predstavnika sloga, ki ga imenuje *austeré harmonía*, »rezko skladanje«. Za ta slog je značilno, da povedi, celo posamezne besede stojijo v stavku trdno in neodvisno od drugih kakor grobo izklesani, nezglajeni kamniti bloki. Prim. Dionysius of Halicarnassus 168.

Poleg tega je metrika francoske ode tuja Pindarjevi. Pindarjevske ode, ki so jih v renesansi pisali v grščini, se zvesto držijo »metrične sheme« tega ali onega Pindarjevega epinikija. Tiste v latinščini pogosto jemljejo verzne in kitične oblike, znane iz latinske lirike, in tako ene kot druge kombinirajo v nove sestave. Francoske *odes pindariques* pa imajo verze različnih dolžin, največkrat z osmimi, desetimi ali dvanajstimi zlogi in venomer sklepane z rimo, ter epode krajše od strof in antistrof, vendar v vseh strofičnih parih in vseh epodah zmeraj eno in isto verzno obliko (prim. Schmitz 133). Metrična zgradba vrstic se od prve do zadnje skoz vso pesem ne spreminja. Namesto večobličnosti Pindarjeve ritmične metrike v teh odah vlada stroga enoobličnost.

Prav tako tuje kot Pindarjevo ritmično metriziranje je bilo francoskim pindaristom njegovo pesniško mišljenje. Ronsardova oda francoski kraljici Katarini Medičejski, recimo, ki se začne s prikazom pesniškega furorja, ima to, kar mora imeti epinikij: ima mitsko pripoved o ustanovitvi Firenc, ki spominja na tisto o ustanovitvi Kirene iz »Devetega pitijskega slavospeva«, in prehod od slavljenja mesta, v katerem se je kraljica rodila, k slavljenju njenega »božanskega rodu« (1,3,34; Ronsard 44), družine Medičejcev. Vendar Ronsard nato namesto zmage slavi kraljičine kreposti. Njegove ode so na splošno enkomiji, sestavljeni zvečine s prerazporejanjem in raznoličenjem toposov iz epinikijskega in sploh antičnega literarnega izročila.

Čeprav Ronsard ni bil prvi, ki je pindariziral v Evropi, je s svojimi odami sprožil trend slavljenja vélikih ljudi, dogodkov iz njihovih življenj in njihovih kreposti. S pindarjevsko odo je vladal francoskemu pesništvu v drugi polovici 16. stoletja, prek njega pa je slava Pindarjevega pesništva segla tudi v Anglijo.

Prva pindarjevska oda v angleščini je »Pandora« Johna Southerna iz leta 1584. Southern se je tako kot Ronsard v Franciji hvalil, da prvi pindarizira v Angliji. Vendar je v peščici pesmi pindariziral po Ronsardovo. Njegovo pesnjenje je skromno, plagiatorski rop Ronsarda.

Prvi angleški pindarjevski odi, ki sta se posrečili, sta nastali precej pozneje, a v istem letu. Leta 1629 je Ben Jonson ob smrti mladega plemiča Henryja Morisona napisal odo v čast njemu in njegovemu prijatelju Luciusu Caryju, enemu svojih občudovalcev, ki so kot samorazglašeni »Benovi sinovi« njega samega, svojega literarnega očeta, šteli za najpomembnejšega angleškega kritika in pesnika tistega časa. V tej odi Morisona in Caryja, slaveč njuno nerazvezljivo prijateljstvo, primerja s Kastorjem in Poluksom, polbratoma dvojčkoma iz grškega mita, ki sta imela različna očeta, prvi špartanskega kralja Tindareja, drugi Zevsa, in od katerih je drugi, Poluks, nesmrten po svojem očetu,

173

ko je Kastor padel v boju, izbral drugo od možnosti, ki mu ju je dal na izbiro oče Zevs, namreč ne da bo sam ostal nesmrten in prebival z očetom na Olimpu, ampak da bo vzel nase polbratovo smrtnost in si z njim delil svojo nesmrtnost ter da bosta skupaj tako na Olimpu kakor v Hadu.[11] Jonsonova oda ima več *conceits*, zapletenih primerjav, ki s sopostavljanjem tistega, kar si je med sabo po mnenju večine ljudi nepodobno, dajejo misliti. Jonson na primer svojega »sina« in njegovega umrlega prijatelja z besednim prestopom imenuje *twi-lights*, »dvo- / luči, *Dioskura*« (92–93; Jonson 235, kurziva je v izvirniku): Cary in Morison sta kot *twins* s *twi-* razločena v dvojčični luči, v luč- -in-temo, a sta vendarle dvojna luč – oziroma s čezjezikovno besedno igro, ki skupno poimenovanje mitskih dvojčkov seje skoz sito latinščine, združena hkrati kot *Di-oscura*, Dvo-temneža (prim. Fitzgerald 5). Oda ima enako umerjene rimane verze neenakih dolžin in triadično kitično zgradbo.

Istega leta kot Jonsonova je nastala oda »Ob jutru Kristusovega rojstva« Johna Miltona, največjega angleškega baročnega pesnika. Milton to pesem, ki je delno sad njegovega mladostnega branja Pindarja, imenuje »ponižna oda« (24; Milton 17), a jo po štirih uvodnih kiticah z mednapisom preimenuje in uglasi v himno. Pesem ob preitju v himno dobi drugačno verzno in kitično obliko; dolgi enakomerni rimani verzi se začnejo izmenjavati s kratkimi in kitice se približajo kitičnemu vzorcu italijanske kancone, ne izgrajujejo pa se, tako kot že ne na začetku, v triade. Milton ob rojstvu Božjega Sina daje slišati vik in krik grških in rimskih, egipčanskih in kanaanskih bogov, ki zapuščajo preročišča, človeške domove in svetišča, ter jok, vzdihovanje in toženje božanstev, ki se poslavljajo od izvirov in gajev. Razlega se spev trume novih duhov, angelskega zbora, in zliva z glasbo sfer, ki je bila od Adamovega padca naprej neslišna pod nebesnim svetovjem, vendar jo je s spustom skoz sfere v smrtno človeško glino, tako da je te sfere odprl, naredil slišno Odrešenik in ji z basi zdaj spet, kakor ob stvarjenju, priteguje zemlja.

Proslavljanje odhoda poganskih bogov iz svetišč je imelo pomembno vlogo že v krščanskih himnah renesanse, predvsem tistih, ki slavijo božič. Toda Milton na začetku svoje ode, ki bo postala himna, samemu sebi zakliče: *O run, prevent them with thy humble ode*, »O, teci, jih s svojo ponižno odo prehiti«. S tem sebespodbujevalnim klicem se spusti v tekmo za prvenstvo v slavljenju pred tremi *mágoi* ali »modrimi« z Vzhoda, ki so

---

[11] O Poluksovi izbiri primerjaj Pindarjev »Deseti nemejski slavospev« (55–90); Senegačnik 275–277.

po evangeljski pripovedi krenili na pot za novovzšlo zvezdo in prvi prišli počastit Božjega otroka na kraj njegovega rojstva (Mt 2,1–12). Pri tem se v angleškem glagolu *prevent* namesto siceršnjega »odvrniti« ali »preprečiti« prebije naprej pomen latinskega *praevenire* – »priti prej«. Milton prihaja, da prvotne častilce prehiti z dostojnim slavljenjem Kristusovega rojstva. »Od zvezde vodeni čarodeji« (23; Milton 17), ljudje starega sveta, niso vedeli, koga so prišli počastit, ničesar od tega, kar zdaj razkošno slaveč razgrinja Miltonova oda-himna. Njegov spev v čast Bogu, ki se je učlovečil v svetu med ljudmi za njihovo odrešenje, a kot Stvarnik vsega »v resnici 'prihaja prej'« (Fitzgerald 33), premaguje in odvrača poganske dišave čarodejev.

Za mejnik v angleškem pindariziranju so obveljale *Pesmi* Abrahama Cowleyja, ki so izšle leta 1656 in kot poseben razdelek vsebujejo »Pindarjevske ode«. Cowley, eden tako imenovanih metafizičnih pesnikov, je Pindarja prevajal v angleščino in sam pesnil po njem. Pred svoje ode je postavil prevoda dveh Pindarjevih slavospevov, »Drugega olimpijskega« in »Prvega nemejskega«, ter jima dodal še imitacijo Horacijeve pesmi v Pindarjevo čast. V uvodu k »Pindarjevskim odam« pojasnjuje, zakaj se je odločil za nedobesedni prevod: »Če bi se kdo lotil prevajanja Pindarja besedo za besedo, bi si drugi mislili, da je en blaznež prevajal drugega« (Cowley 155). Zato je prevajal Pindarjev *način govora*, ki ga je prevzel tudi v svojem pesnjenju. V opombi k odi »Vstajenje« pravi: »Ta oda je resnično pindarjevska, s tem ko skladno z njegovim entuziastičnim načinom iz ene stvari vpada v drugo.« (183) Za poglavitno potezo Pindarjevega govora je prepoznal nenadne prehode z enega predmeta na drugega, čeprav se ti niti v njegovem prevajalskem pindariziranju ne vrstijo tako odsekano kot pri Pindarju.

Cowley si je pri prevajanju Pindarjevega »entuziastičnega načina« jemal veliko svobode. Pindarjeva epinikija imata v njegovem prevodu ne samo verze različnih dolžin, ampak tudi kitice z različnim številom verzov, v katerih se povrhu neenakomerno mešajo raznoliki metrični impulzi. Z iregularnostjo verzov in kitic se uveljavlja svoboda v metriki. A čeprav ta svoboda vsekakor daje vtis približka Pindarjevemu metričnemu besnenju, kakršnega je v prestopanju gorske reke čez bregove upodobil Horacij, prej kot posrednik besnenja deluje kot tamponski sloj in branik pred njim. Tudi to spet nemara izdaja Cowleyjevo lastno pesnjenje. V »Vstajenju« se Cowley namreč obrne na svojo Muzo, češ, naj »pindarjevskega svojega Pegaza krêpko obrzda, / ki besneti res začenja« (54–55; Cowley 183).

Z metrično svobodo pa se druži svoboda smisla. Cowleyjevo prevajanje po smislu je svobodno v tem, da dodaja vezne člene. V »Prvem

nemejskem slavospevu«, ki ga Pindar začne z nagovorom otoka Ortigije, Cowley na primer zamenja »posteljo Artemide« s »porodno posteljo / svetle Latone« (3–4; Cowley 170). Da bi vzpostavil jasnejšo povezavo med otokom in boginjo, »posteljo« namesto Artemidi pripiše njeni materi Leto (oziroma latinizirani Latoni) in z dodatnim pridevnikom »porodna« pokaže, da je bila Artemida rojena na Ortigiji. Čeprav pri zamenjavi Pindarjeve metafore s svojo izpusti Artemido, ki jo izrecno imenuje Pindar, *porodna* postelja Latone, ker je *rojstna* postelja Artemide, razjasni zvezo: Ortigija je rojstni kraj te (v prevodu neimenovane) boginje.

Še znatno večji je Cowleyjev poseg na začetku druge epode, kjer Pindar brez zveznega prehoda začne podajati mit o Heraklovem prvem junaškem dejanju, namreč kako je že v otroški zibki zadavil kači, ki ju je nadenj poslala Hera. Tu Cowley zgradi nič manj kot most med mitskim herojem in Hromijem, zmagovalcem na nemejskih igrah in slavljencem Pindarjevega slavospeva. Za to vstavi kar pet verzov o tem, da se noben drug izmed mladeničev ni tako zgodaj podal v tekmo za krepost kakor Hromij, in v zadnjem, petem, pribije, da na pot kreposti »nihče ni nikdar prej od njega krenil razen Alkida« (79; Cowley 173), kot Herakla imenuje po njegovem dedu. Herakles v prevodu šele potem, ko ga Cowley jasno postavi Hromiju v predhodništvo, zadavi kači.

Na Cowleyja je delovala Horacijeva podoba Pindarja. Kot prevajalec Pindarjevega »entuziastičnega načina« pa je izbral srednjo pot. Ni se naselil v figuri Ikarja, ampak njegovega očeta Dajdala, ki ni letel ne prenizko ne previsoko, preblizu. Čeprav je po svobodnosti svoje metrike postal znan in razvpit med sodobniki, je vendarle izbral zmernost. Pindarjev »entuziastični način« je lovil z njo, zrahljano metriko, in s prevajanjem po smislu, s posegi vanj: z regulacijskimi konektorji, mostogradnjo, protetiko smisla. Toda ker srednji poti pripada, da se drži stran od skrajnosti, nikdar ne vodi tja, kamor visoka. Cowley s svojim »ne preblizu« ne utira poti *k* izviru Pindarjeve govorice. Dobesedno prevajanje Pindarja je bilo zanj pot v blaznost. Svobodo, ki si jo je jemal pri prevajanju, je pri tem rabil kot zaščito za oba: »Cowleyjevo prevajanje hoče ohraniti pozneževo duševno zdravje [*sanity*],« ugotavlja John T. Hamilton, a hkrati »zunaj azila zadržati tudi svoj zgled, Pindarja samega« (Hamilton 177). Pesniško blaznost je vsekakor želel ločiti od klinične norosti ter pesnika in njegovega prevajalca sebe rešiti pred tem, da bi postala dvonorca.

Cowley je kot prevajalec in posnemovalec Pindarja sam pisal ode, ki seveda nimajo triadične zgradbe, ampak iregularen verz in kitico. Cowleyjeve »Pindarjevske ode«, prevedeni Pindarjevi in njegove lastne,

so s svojo oblikovno iregularnostjo postale mejnik zaradi vpliva, ki so ga imele ne samo na množice verzifikatorjev, ampak tudi na pesnike. Cowleyjevo prevajanje Pindarja je zagovarjal John Dryden, ki je leta 1668 postal prvi angleški *poeta laureatus*, češ, o Pindarju je »splošno znano, da je temen pisec, da mu manjka povezanosti [*Connexion*]«, zaradi česar je bilo za genija, »kakršen je tisti gospoda Cowleyja, pač nujno, da Pindarju da govoriti angleško« (Dryden 1013) – in zagovor je veljal tudi za Cowleyjevo pesnjenje. Poleg tega je Dryden, kar zadeva Pindarjeve ritme, stavil na zadržanost oziroma cowleyjevsko zmernost. Za svojo najboljšo pesem je imel »Aleksandrovo gostijo«, ki je pindarjevska oda iregularnega cowleyjevskega kova, enako pa je tudi stoletje pozneje visoko cenjeni Thomas Gray za svoji najbolj posrečeni pesmi štel pindarjevski odi »Napredek pesništva« in »Bard«, le da sta ti v nasprotju z Drydnovo grajeni strogo triadično.

Več pindarjevskih od brez triadične zgradbe je v angleščini morda sicer še pred Cowleyjem spesnil drug metafizični pesnik, Richard Crashaw. Odo je, recimo s slavljenjem Kristusovega rojstva in razglašenja, poskušal prilagoditi krščanski himniki, vendar je trume za sabo potegnil Cowley, ne Crashaw. In še: Cowley je odo s tem, da jo je pisal v čast vznesenosti, usodi, življenju in tako naprej, odcepil od take ali drugačne priložnosti, kakršnekoli že se je držala dotlej. V tem pesniškem izročilu so pozneje v angleščini nastale takšne mojstrovine, kot je na primer Wordsworthova oda nesmrtnosti.

Po Cowleyju se je znatno več od regularne pisala iregularna oda, dokler ni mode iregularnih pindarjevskih od zajezil William Congreve. Leta 1706 je napisal odo kraljici Ani in vzel nase »popravo 'napake' pindarjevske iregularnosti« (Wilson 161). Svoji odi je postavil na čelo »Govor o pindarjevski odi«, v katerem je zavrgel Horacijevo podobo Pindarja in ob tem, da je metrično svobodo cowleyjevskega kova povezal z divjo svobodo misli, zavrnil obe. Verzi takšne ode, pravi Congreve, niso nič drugega kot »povez [*Bundle*] blodnih nesovisnih misli, izraženih v povezu prav takšnega iregularnega kitičnega zavoja [*in a like Parcel of irregular Stanza's*]« (Congreve 144). Iregularno pindarjevsko odo pri tem loči od Pindarjeve:

> *Nasprotno ni nič bolj regularnega od Pindarjevih od*, kar zadeva tako njegovo natančno ohranjanje mer in števila kitic in verzov kakor tudi nepretrgano sovisje [*perpetual Coherence*] njegovih misli. Kajti čeprav so njegove zastranitve pogoste in prehodi nenadni, vendarle vselej obstaja neka skrivna povezava [*Connexion*], ki ji kljub temu, da se zmeraj ne kaže očesu, nikdar ne spodleti, da se priobči [*communicate itself*] bralčevemu razumevanju. (Congreve 144, kurziva je moja)

Congrevu se je prikazala podoba, nasprotna Horacijevi podobi metričnega divjaka: Pindar je *popolnoma regularen* pesnik, njegova regularna, sama v sebi skladna metrika se ujema s prav tako regularnim »sovisjem njegovih misli«. Povezava med temi mislimi sicer nikakor ni očitna, a se kljub temu zmeraj naznani bralčevemu razumevanju – še več, in to ostaja neizrečeno: razumevanju prinaša obet razjasnjenja temne Pindarjeve govorice ali celo razjasnjenje samo. Povezava obstaja tako med mislimi kakor tudi med mislimi in merami, med miselno in metrično zgradbo Pindarjevih od. To pomeni, da je Pindarjeva ritmična metrika Congrevu v povezavi oziroma skladnosti z miselnim sovisjem pokazala *apolinični* obraz. Po njegovi kritiki se je pindarjevska oda pri marsikaterem pesniku, na primer pri Grayu, vrnila v regularno triadično obliko.

Ena od spodbud za Congrevovo presojo Pindarjeve ode nemara izhaja iz razločevanja med njegovimi epinikiji in ditirambi. Že Antonio Sebastiano Minturno namreč v spisu *O pesniku* zatrjuje strogo regularnost epinikijske triade in v nasprotju z zgovorno Horacijevo sugestijo o metričnem besnenju v vseh Pindarjevih pesmih postavi domnevo, da Pindarja »nosijo iregularni ritmi« (Minturnus 397) *samo* v ditirambih. Druga spodbuda za to presojo pa je vsekakor prišla od wittenberške izdaje Pindarjevih epinikijev. To knjigo, ki je izšla leta 1616, dobrih dvesto let po prvih dveh v Italiji, je z grškimi besedili, latinskimi prevodi in obširnim komentarjem ter z uvodi, metrično analizo in tabelno sinopso retorične zgradbe za vsak epinikij posebej pripravil Erasmus Schmid. Na začetku knjige je med drugim natisnjena pesem Emericiusa Thurza, rektorja univerze v Wittenbergu, v čast izdajatelju. Tretja kitica te pesmi se glasi (9–12):

> Monte decurrens velut amnis, imbres
> Qvem ſuper notas aluêre ripas,
> Eluit ſordes, Lyriciq; mendas
> SCHMIDIUS aufert. (Schmid 15)

> Kot reka, ki z gore dol teče, od déžja
> narasla čez znane bregove,
> nesnago izpira, stran Liriku
> SCHMID odnaša napake.

Prva dva verza sta čisti citat prvih dveh verzov Horacijeve pesmi v Pindarjevo čast. Toda druga dva spremenita njen scenarij: prej je pesnik, ki je poskušal posnemati Pindarja, strmoglavil, zdaj pa skupaj s Pindarjem, »Lirikom« z veliko začetnico, kakor gorska reka, ne kakor Ikar, strmoglavlja njegov izdajatelj Schmid, in sicer tako, da besedilo

Pindarjevih epinikijev čisti in v njem odpravlja napake, ki so se vanj prikradle v tekstni predaji.

Schmid je hkrati kakor reka, ki jo je Herakles speljal skoz Avgijev hlev. Njegovih emendacij, narejenih ob primerjavi vrste rokopisov, je več kot šeststo. Poleg tega v pismu »Naklonjenemu bralcu«, ki ga je objavil v poskusni delni izdaji nekaj let prej in dal skoraj nespremenjenega ponatisniti v integralni izdaji Pindarjevih epinikijev, izraža prepričanje, da bo ta izdaja njemu, bralcu, naredila Pindarja ne le razumljivega, ampak v njem tudi ne bo več videl nobene »temnosti, ki so si jo sami pri sebi umišljali nekateri« (Schmid 4), ker jim je njegovo pesništvo zatemnilo slabo stanje besedila. Iz Schmidovega prepričanja, da odpravljanje napak izbrisuje temnost Pindarjeve govorice, veje silovit hermenevtični optimizem, ki izhajajoč iz predpostavke, da je prvotni tekst jasen, izganja iz njega temo in bralčevemu razumevanju obeta jasnost brez temnosti. Ta optimizem prihaja na spregled tudi v Congrevovem poudarjanju oblikovne in vsebinske regularnosti Pindarjevega pesništva in postavitvi Pindarjeve ode v popolno nasprotje z odo njegovih posnemovalcev.

Tako kot Congreve je Cowleyjevo pindariziranje in pesnjenje po njem pindarizirajoče množice nekaj pozneje grajal Samuel Johnson, morda najuglednejši književnik v angleški zgodovini sploh, čigar delo obsega vse od pesmi do kritiških spisov in celo slovarja angleškega jezika. »Dr. Johnson«, kot so ga pogosto imenovali, v obširnih *Življenjih pesnikov*, ki jih je napisal proti koncu življenja, pravi: »Ta ohlapna in brezzakonska verzifikacija […] je takoj preplavila naše pesniške knjige; vse fante in dekleta je zgrabila prijetna moda, in tisti, ki niso bili zmožni početi ničesar drugega, so lahko pisali kakor Pindar.« (Johnson 40) Johnson se je sicer strinjal z mnenjem mnogih sodobnikov, oprtim na Horacijevo avtoriteto, namreč da je Pindarjev verz metrično svoboden. Vendar ima ta svoboda po njegovi sodbi svojo mero. Verzi se v Pindarjevih epinikijih vračajo z enakim številom zlogov in dosežejo enakomernost na ravni kitice. Svobodno oblikujoči se metrični vzorci v prvih treh kiticah vzpostavljajo pravilo za vse druge triade, in kar se izvzorči v prvi triadi, dobi natančne metrične responzije v vseh, kolikor jih še sledi. Kar je sprva slišati rezko, navsezadnje zveni ubrano kakor na Apolonovi liri.

Johnson seveda še zdaleč ni imel zadnje besede v zgodbi, ki jo piše recepcija Pindarja na Zahodu, verjetno pa je eden redkih piscev, ki so imeli pred očmi oba Pindarjeva obraza, dionizičnega in apoliničnega, čeprav mu je prvega prerasel Pindarjev drugi obraz. Horacijeva podoba Pindarja je delovala še naprej in na obe strani. Vzbujala je prav tako številne »za« kot »proti«.

LITERATURA

Bowra, Cecil M. *Pindar*. Clarendon Press, 1964.

Callimachus. *Hecale / Hymns / Epigrams*. Loeb 129, grški izvirnik z angleškim prevodom, Harvard University Press, 2022.

Congreve, William. »A Discourse on the Pindarique Ode«. *Complete Works*, Delphi, 2018, str. 144–148.

Cowley, Abraham. *Poems: Miscellanies, The Mistress, Pindarique Odes, Davideis, Verses Written on Several Occasions*, izd. A. R. Waller, Cambridge University Press, 1905.

Curtius, Ernst Robert. *Evropska literatura in latinski srednji vek*. Prevod Tomo Virk, LUD Literatura, 2002.

D'Angour, Armand. »Horace's 'Victory Odes': Artifices of Praise«. *Bulletin of the Institute of Classical Studies. Supplement*, let. 112, 2012, str. 57–71.

Dionysius of Halicarnassus. *Critical Essays I: On Literary Composition*. Loeb 466, grški izvirnik z angleškim prevodom, Harvard University Press, 2007.

Du Bellay, Joachim. *Œuvres complètes: La défense & illustration de la langue françoise suivi de l'Olive et quelques autres œuvres poétiques*. Zv. 1, izd. Léon Séché, Revue de la Renaissance, 1903.

Dryden, John. »Preface Concerning Ovid's Epistles«. *Complete Works*, Delphi, 2013, str. 1006–1016.

Fitzgerald, William. *Agonistic Poetry: The Pindaric Mode in Pindar, Horace, Hölderlin, and the English Ode*. University of California Press, 1987.

Freis, Richard. »The Catalogue of Pindaric Genres in Horace 'Ode' 4.2«. *Classical Antiquity*, let. 2, št. 1, 1983, str. 27–36.

Gasparov, Mihail Leonovič. *Očerk istorii evropejskogo stiha*. 2. izd., Fortuna Limited, 2003.

Graf, Fritz. *Apollo*. Routledge, 2008.

Hamilton, John T. *Soliciting Darkness: Pindar, Obscurity, and the Classical Tradition*. Harvard University Press, 2003.

Highet, Gilbert. *The Classical Tradition: Greek and Roman Influences on Western Literature*. Oxford University Press, 2015.

Horace. *Odes and Epodes*. Loeb 33, latinski izvirnik z angleškim prevodom, Harvard University Press, 2004.

Horace. *Satires. Epistles. The Art of Poetry*. Loeb 194, latinski izvirnik z angleškim prevodom, Harvard University Press, 1989.

Irigoin, Jean. *Histoire du texte de Pindar*. C. Klincksieck, 1952.

Johnson, Samuel. *The Lives of the English Poets*. Zv. 1, izd. Roger Lonsdale, Oxford University Press, 2006.

Jonson, Ben. *The Workes of Benjamin Jonson*. Zv. 2, Meighen 1640.

Krummacher, Hans-Henrik. *Lyra. Studien zur Theorie und Geschichte der Lyrik vom 16. bis zum 19. Jahrhundert*. Walter de Gruyter, 2013.

Milton, John. *Complete Shorter Poems*. Izd. Stella P. Revard, Willey-Blackwell, 2009.

Minturnus, Antonius Sebastianus. *De poeta, ad Hectorem Pignatellum, Vibonensium Ducem, Libri sex*. Apud Franciscum Rampazetum, 1559.

Moffitt, John F. *Inspiration: Bacchus and the Cultural History of a Creation Myth*. Brill, 2005.

Ovidij Naso, Publij. *Metamorfoze. Izbor*. Prevod Kajetan Gantar, Mladinska knjiga, 1977.

Quintilian. *The Orator's Education*. Loeb 126, latinski izvirnik z angleškim prevodom, Harvard University Press, 2001. (Slovenski prevod: *Šola govorništva*. Prevod Matjaž Babič, Šola retorike Zupančič&Zupančič, 2015.)

Platon. *Zbrana dela*. Zv. 1, prevod Gorazd Kocijančič, Mohorjeva družba, 2004.

Revard, Stella P. *Pindar and the Renaissance Hymn-Ode: 1450–1700*. Arizona Center for Medieval and Renaissance Studies, 2001.

Ronsard, Pierre de. Œu*vres complètes*. Zv. 2, izd. Prosper Blanchemain, P. Jannet, 1857.

Schmid, Erasmus, izd. *Pindarou Periodos, hoc est Pindari Olympionikai, Pythionikai, Nemeonikai, Isthmionikai*. Wittenberg, 1616.

Schmitz, Thomas. *Pindar in der französischen Renaissance. Studien zu seiner Rezeption in Philologie, Dichtungsstheorie und Dichtung*. Vandenhoeck & Ruprecht, 1993.

Senegačnik, Brane, prev. *Slavospevi in izbrani fragmenti*, Pindar, Družina, 2013.

Snell, Bruno in Herwig Maehler, izd. *Pindari carmina cum fragmentis. Epinicia*. Zv. 1, B. G. Teubner, 1980.

Snell, Bruno in Herwig Maehler, izd. *Pindari carmina cum fragmentis. Fragmenta. Indices*. Zv. 2, B. G. Teubner, 1975

Wilson, Penelope. »Pindar and English Eighteenth-Century Poetry«. *Bulletin of the Institute of Classical Studies, Supplement*, let. 112, 2012, str. 157–168.

Ronsard's and Cowley's Pindaric Ode

After his return to the West, which had been separated from Greek culture by the language barrier after the end of antiquity, Pindar regained the status of "champion of the lyricists" that he had enjoyed with the Greeks. In the late Renaissance, the highest genre of lyric became the ode, as the epinician was also already called by Alexandrian scholars, and Pindar's epinician with its triadic stanzaic structure was established as its prime example. The poem in which Horace depicted Pindar as a wild mountain river and gave him the image of a metrical wild man also had a decisive influence on the reception of Pindar's poetry. In the vulgar language, Pierre de Ronsard made prominent the Pindaric ode with a triadic structure and thus earned himself a lasting place in French and European literary history. He became known as the French Pindar. Pindaric odes began to be written after the Ronsard's model, without first-hand knowledge of Pindar, and Ronsard's ode reached England. The milestone in English pindarising, however, is Abraham Cowley's "Pindarique Odes", which are characterized by loose metrics and the freedom of leaping thought.

By translating Pindar and poeticizing after Pindar's "enthusiastic manner", Cowley triggered a fashion for irregular Pindaric odes among his compatriots, a fashion that was curbed after a while by William Congreve's diametrically opposed image of the great Greek lyricist. In it, Pindar appears as a purely regular poet whose metrics correspond to the coherence of his thoughts.

The "Cura pastoralis" Fragment from the Archives of the Republic of Slovenia: A Completely Different Story

Anke Lenssens

Ghent University, Faculty of Arts and Philosophy, Blandijnberg 2, 9000 Gent, Belgium
https://orcid.org/0000-0003-3756-3336
Anke.Lenssens@UGent.be

The fragment "Cura pastoralis" is the oldest manuscript fragment in Slovenia and is kept in the Archives of the Republic of Slovenia under the file number AS 1073, II–1r. It consists of one bifolium. One side of it is very badly damaged, as the fragment was once used for the binding of a book that served as the marriage register of the town of Trbovlje between 1669 and 1704. The register is still kept in the Diocesan Archives of Maribor. The bifolium is cataloged as a fragment of a ninth-century manuscript containing the "Cura pastoralis" of Gregory the Great. After a thorough examination, however, it turned out to be part of a work by Paterius of Brescia, Gregory the Great's first secretary, which may have seen the light of day in the Freising scriptorium. There are some physical and content-related similarities between the fragment and the oldest complete manuscript in Slovenia, the Ecloga of Lathcen, which was written in the same period, namely at the end of the first half of the ninth century.

Content and physical appearance

The so-called "Cura Pastoralis" fragment, known under signature AS 1073, II-1r in the Archives of the Republic of Slovenia, has a dimension of 26,9 by 35,7 centimeters and the text is divided over 30 lines.[1] The fragment goes back to the second quarter of the ninth century and was most likely copied in a Southwest-German writing school (Golob,

183

Srednjeveški rokopisi 182). The complete text seems to be written by the same hand. There is little reason to doubt the dating of the *bifolium*. Therefore, its Carolingian origin is also very straightforward.
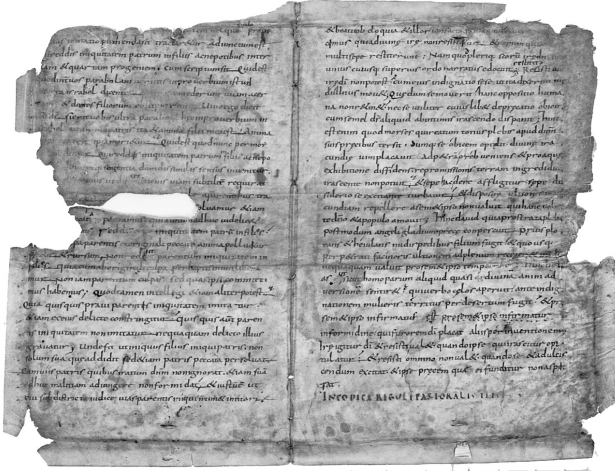

Figure 1: Inner side of the bifolium AS 1073, II-1r.

Figure 2: Outer side of the bifolium AS 1073, II-1r.

About 20 years ago, Nataša Golob defined the fragment as a part of a manuscript containing the *Cura* or *Regula Pastoralis* by Gregory the Great (Golob, "Karolinški fragment" 277–281). At the bottom of the

right page of the bifolium on the inner side we can indeed read the following title in bold letters: IN CODICE REGULE PASTORALIS III. After taking a look at the exact content of the inner side of the bifolium, it became clear that it contains parts of another famous work by Gregory the Great, namely the *Moralia in Iob*. The left page on the inner side contains a part of chapter 51 of book 15. The right page holds part of chapter 16 of book 9 (Gregorius I and Adriaen, *Moralia in Iob: libri I–X* 473–476; *libri XI–XXII* 784–786). This explanation was also added in a more recent description of the manuscript fragment by the same author (Golob, *Srednjeveški rokopisi* 183). The outer side of the bifolium is severely damaged and almost impossible to read. However, on the right side of the outer part of the bifolium, it is possible to distinguish a title written in bold letters followed by a *capitalis* D on the next line. By putting in a little effort and using a light source, we are able to distinguish the ink from the parchment, especially because the scribe was rather generous with the use of ink while writing headings. The heading reads: IN EXPOSITIONE BEATI IOB LIBRO XXXV. The presence of this title reassures us again that it is indeed a part of the *Moralia in Iob* we are dealing with. Contrary to what was always assumed and again pointed out in the most recent work on the fragment, Golob assumes that the left page of the damaged side of the bifolium must contain a part of the third book of the *Regula Pastoralis*, because this was mentioned in the title below the excerpt of book 9 chapter 16 of the *Moralia in Iob* (Golob, *Srednjeveški rokopisi* 183).

However, it is not as straightforward as it seems. The damaged side of the bifolium was never actually examined with full attention. One could say it is simply impossible to read, but with a little effort, it is still possible to distinguish some words[2] on the left page of the damaged side. At first sight, we can distinguish a part of chapter 25 from book 3 of the *Regula Pastoralis* (Migne, *Patrologiae Volume 77* 97–98). This would make sense considering the title IN CODICE REGULE PASTORALIS III written on the previous page. However, it would be very strange to add parts of the *Regula Pastoralis* when there are still parts of the *Moralia in Iob* to come, among which is chapter 51 of book 15. According to our bifolium, the last sentence of book 9 chapter 16 is: "Irae igitur Dei et resisti valet, quando ipse qui irascitur, opitulatur, et resisti omnino non valet, quando se ad ulciscendum excitat, et ipse

---

[2] Some examples of words that are still visible, are: *de porta*, *necesse est*, *tanta aequalitate*, *culpas*, *occidat vir*, *amicum*, *profecto esse*, *universa plebs*, *papilionis* and *populus de egyptia*.

precem quae ei funditur non aspirat." This is not the last sentence in the original version of the *Moralia in Iob* (Gregorius I and Adriaen, *Moralia in Iob: libri I–X* 473–476). When we further examine the left page of the damaged side, we see that there is more than just a part of *Regula Pastoralis*. There is an excerpt of the first homily of the second part of the *Homilae in Hiezechielem*, another work by Gregory the Great (Migne, *Patrologiae Volume 75* 935–948). This excerpt directly follows the one from the *Regula Pastoralis*. Consequently, we have excerpts, not complete chapters or texts, from Book 9 chapter 16 of the *Moralia in Iob*, chapter 25 of Book 3 of the *Regula Pastoralis* and the first homily of the second Book of the Homilies on Ezekiel. This exact order of excerpts is to be found in a work called *De Expositione Veteris ac Novi Testamenti liber de diversis S. Gregorii Magni libris concinnatus* by Paterius of Brescia, more precisely in the part on Exodus (Migne, *Patrologiae Volume 79* 747–749). The titles we can distinguish in the fragment, are therefore subtitles that identify the works from which the excerpts are taken. In principle, we should be able to find four of them throughout the fragment. On the right page of the damaged bifolium, we can distinguish two of them. As was already mentioned, the first one is still more or less visible: IN EXPOSITIONE BEATI IOB LIBRO XXXV. We can find another one at the bottom of the same page that is very poorly visible and can be read as: IN EXPOSITIONE BEATI IOB LIBRO XV (Migne, *Patrologiae Volume 79* 751). Still very well readable is the subtitle referring to the excerpt from the *Regula Pastoralis*: IN CODICE REGULE PASTORALIS III. We should be able to find the subtitle introducing the excerpt from the first homily of the second part of the *Homilae in Hiezechielem*. Unfortunately, the left page of the damaged side of the bifolium is in a very bad condition in the place where this subtitle should be. On line 17 of the page in question, we can vaguely distinguish the use of capital letters. This is likely the place where the excerpt of the homily was introduced. Considering all this, we see that the text on the right page of the intact side is continued on the left page of the damaged side. The same can be said about the right page of the damaged side and the left page of the intact side. Content wise, there is only a small gap between the end of the text on the left page and the beginning of the part on the right page of the damaged side. This means there was not more than one bifolium present inside of this one when it was still inside the manuscript.

The work of Paterius where AS 1073, II–1r was originally part of, is sometimes also referred to as *Liber Testimoniorum*. It still survives in 123 works, both in complete versions and fragments (Martello 431,

435). Now we can add the 124th to the tradition. Paterius was the notary and later on *secundicerius*[3] under Gregory the Great (Étaix 78). As a contemporary and close collaborator of the famous pope, Paterius likely had the original versions at his disposal. Without doubt, his work certainly is characterized by its high quality. According to the author himself, the anthology was divided into three parts: two on the Old Testament and one on the New Testament. He wrote this in the prologue of his work, but eventually the work has not survived the test of time in its entirety. We do possess the parts from Genesis till the Song of Songs. The last two parts, Proverbs and the Song of Songs, appear to be in a much rougher stage and less accurate (Martello 431–432). It was therefore argued by Étaix that these parts were not originally written by Paterius (Étaix 66–68). More recent research on the *Liber Testimoniorum* was carried out by Castaldi and Martello (Castaldi and Martello 23–107). They assume that the revision process was never completed for all the parts of the *Liber Testimoniorum* and that the revised parts therefore got lost because they were probably written down on inferior material. Only the last two parts of the 14 parts that have been copied through the centuries have survived in their unrevised form. Curiously, the work only started to gain fame in the eighth century. The oldest surviving fragments and "complete" versions go back to the same century (Martello 431–433). It is rather remarkable that Gregory the Great was often cited through the work of Paterius by a number of very well-known Medieval writers and theologians, like Bede and Rabanus Maurus (Étaix 67). Apart from that, the work of Paterius is still very valuable for the study of the works of Gregory the Great, as the *Liber Testimonorium* refers to unpublished fragments of the pope. This can, of course, be explained by the function Paterius held in Rome as the notary and later *secundicerius* of Pope Gregory. He would have been able to use unedited versions of the works of his pope (Étaix 75–78). Without doubt, the work of Paterius was considered as valuable and often used as a reference work for the oeuvre of Gregory the Great.

When we take a closer look at the fragment itself, we can notice something striking. On the right page of the readable side of the bifolium, we notice some underlining on line 24. It concerns the following part of the sentence: "pro semetipso infirmatur." One could wonder why these words in particular were emphasized. The whole sentence

---

[3] A *secundicerius* is the second in rank at the pontifical chancery (Boudinhon 122–123).

goes as follows: "Et pro semetipso infirmatus est pro semetipso infirmatur in formidine qui furorem Dei placat aliis per interventionem."[4] As can be seen, the same message, "pro semetipso infirmatus est pro semetipso infirmatur," is given twice by two different grammatical constructions. The underlining is indeed very peculiar, as it does not seem to have a clear purpose. However, there is a manuscript from the same period with very similar content that holds the same type of underlining. The manuscript in question contains the *Ecloga de moralibus Job*, written by a certain Irish monk Lathcen or Laidcend who died in 661 in the monastery of Clonfertmulloe (Lathcen and Adriaen v). It is kept in the National and University Library in Ljubljana (NUK) under the registration mark MS 6. The *Ecloga* is actually a very condensed version or summary of the Moralia in Job (Lathcen and Adriaen v). In this manuscript, the underlining was often used to point out mistakes in the text. Here, the correct words were mostly written above the erroneous part of the sentence. We do find many manuscripts where the corrections were added in the margins, but the practice of expunctuation existed as well. This meant that the incorrect part of the sentence was underlined with dots in order to warn the reader to ignore it (Rudy 59–60). The underlined words in our fragment are exactly the ones that the reader should ignore, so the underling is definitely a case of expunctuation.
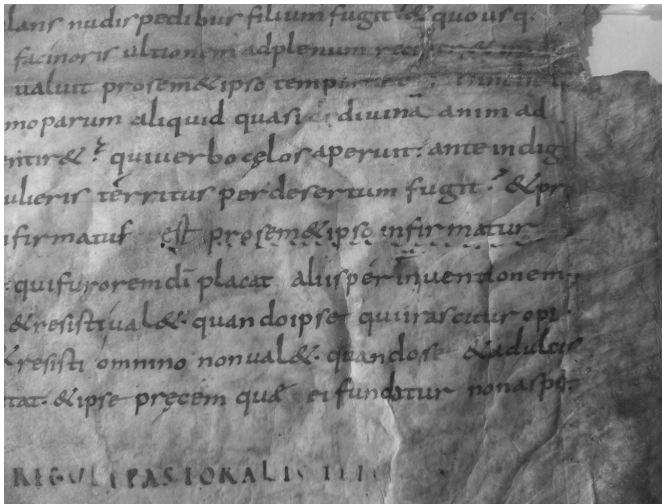

Figure 3: Expunctuation is clearly visible on AS 1073, II–1r.

---

[4] "And he who shows weak for himself in fear, who appeases the Wrath of God for others through intervention." All translations in this article are done by the author.

Origin and value of the fragment

Where the manuscript that contained our fragment is originally from, is difficult to tell with certainty. We know it was used as binding material for the marriage register of the parish of St. Martin spanning the period 1669–1704. In one of Nataša Golob's works on manuscript fragments in Slovenia, she remarks that the discarded manuscripts that came to be recycled in the bindings of books often came from monasteries or other institutions near the place where the book was assembled (Golob, "Srednjeveški pergamentni fragmenti" 103). In our case, the book is a blank notebook, but we should be able to apply the same theory. In order to do so, we must first try to find out where the notebook was assembled. Fortunately, the marriage register holds some watermarks with a dimension of 3,5 by 4 centimeters[5]. We can clearly distinguish a watermark in the form of a rather plain shield with a cross beam holding a simple curlicue. However, the origin of the watermark has proven to be completely untraceable. This is certainly a setback, but there is still another clue that can give us some more insight into the origin of the fragment. Apparently, there were no standard forms for the registration of marriages before 1784, so the parishes were not instructed from above on where they should buy their notebooks and what quality they should have. Standardization was only introduced due to the reforms introduced by the Holy Roman Emperor Joseph II (Štih, Simoniti and Vodopivec 241). It would be rather odd if a small parish like St. Martin bought its notebooks at a place far removed from its own location. I therefore suggest the marriage register must have been assembled by a bookbinder that was relatively close by. What is also clearly visible, is the grid that was added to the paper in the same way as is done with the watermark. In order to add lines to the paper, metal strings were put inside the paper scoops. This technique came about in 1745 in Nürnberg. This type of gridded paper was normally of a higher standard and meant for chanceries and more generally for writing (Weiss and Weiss 170–174).

---

[5] Many thanks to Igor Filipič, archival advisor at the Diocesan archives in Maribor, for sending the photographs of the watermarks and giving the information on the marriage register and former structure of the diocese.

Figure 4: Watermark and grid on the paper of the marriage register (1669–1704) of St. Martin from the Diocesan Archives of Maribor, Parish of Trbovlje—Sv. Martin, sig. 0271, marriage book 01 1669–1704.

In short, we can be fairly sure that the parish of St. Martin acquired a notebook made with rather qualitative paper that was likely assembled nearby. Considering that the notebook was made not too far away from the Trbovlje region, the manuscript fragment must have been taken from a codex that was discarded from a location that was relatively close by. Considering that the handwriting was determined as South-East German and was possibly even of Freising origin (Golob, *Srednjeveški rokopisi* 181–182), it is possible that the work of Paterius travelled to the broader area around Trbovlje due to the presence of the bishopric of Freising in what was then Carantania. However, it is very unlikely that Freising was already active south of the river Drava (Drau) in the ninth century. Their first possession in this area was the Loka dominion in 973 (Sickel 56–57) and this is still relatively far away from the area of Trbovlje. Another possession that was in the hands of Freising and at a more or less equal distance from Trbovlje as is Škofja Loka, was the area around Klevevž. The first mentioned property of Freising in this area was Vinji Vrh. This was in 1074, just over one hundred years after the acquisition of the Loka dominion (Blaznik 5). This does not mean Freising could not have been active in this area a little sooner, but it seems unlikely they would have been active around Klevevž in

the ninth or even tenth century. If this manuscript somehow ended up in St. Martin through interference from the bishopric of Freising, it was most likely due to its presence in the area of Škofja Loka. The manuscript could have only been brought from Freising if this was done more than 150 years after it was copied. This is indeed possible, considering the theory of Golob that peripheral areas of bishoprics, or in this case a remote possession, were often supplied with manuscripts that were discarded because newer copies were already in use at the center of the see (Golob, "Karolinški fragment" 280). Shortly put, the location of discovery of the fragment of Paterius' work and the watermarks on the paper of the marriage register cannot give us a definitive answer about the origin of the manuscript where the fragment once belonged to. We cannot rule out the Freising scriptorium as a possible candidate, but neither can we confirm it was copied there based solely on what has been discussed earlier.

However, there are other clues that still hint towards a Freising origin. The works by Gregory the Great alongside the oeuvre of other patristic fathers were copied eagerly by the Freising scriptorium. One could definitely speak of a tradition starting off during the episcopacy of Arbeo (764–783) and dwindling a bit by the time of Anno (854–875). It knew its height during the episcopacy of Hitto (810/12–835), which overlaps with the period our fragment could be dated to approximately. One could assume the codex containing the compilation work of Paterius saw the light during the heydays of patristic literature in Freising. The patristic writers that were held in especially high regard were Ambrose, Augustine and Gregory. The middle of the ninth century was known for all its copying activities regarding the works of the Church fathers. This was all due to the Carolingian renaissance. It was important to possess these esteemed patristic works, especially for Cathedral schools (Mass 190–191). It is beyond any doubt that the work of Gregory the Great had an educational purpose. It is therefore most likely that the manuscript where our fragment was part of, belonged to a certain center of knowledge, as an anthology of the work of Gregory the Great must have come in very handy. Interestingly, we can also detect the popularity of the works of Gregory the Great at the monastery of St. Gall. We know that the oldest *abbreviatio* of the *Moralia in Iob* by Lathcen was recommended to Salomon, the future bishop of Konstanz, by his mentor Notker from St. Gall in his *De interpretibus divinarum scripturarum*. Notker suggested this work because it is much more condensed than the original work of Gregory the Great (Castaldi 374–375, Aris 362; Migne, *Patrologiae Volume 131*

996–997). In a way, this type of adaptation made it easier to process the material of very extensive works.

What seems to be often overlooked, is that Notker also recommends the *Liber Testimoniorum* of Paterius in the first chapter of the same book. Chapter one recommends works that mainly discuss the Pentateuch (Migne, *Patrologiae Volume 131* 993–1004). About Paterius' work, he literally states: "Quod si excerptum Paterii, quod de libris beati Gregorii per ordinem singulorum librorum deflorando confecit, unquam reperire potueris, illud tibi ad omnimodam sufficiet sapientiam." (Migne, *Patrologiae Volume 131* 995)[6] Paterius' anthology was definitely highly valued by Notker and apparently not so easy to get hold of. The work provides the reader with the possibility to look up what Gregory wrote about a specific passage in the Old Testament without having to plough his way through multiple works of considerable size, as it is composed of quotations from various writings of Gregory the Great.

We could argue that the *Liber Testimoniorum* and the *Ecloga* could have fulfilled a similar function. Both works can be seen as compact versions of much larger works. This definitely comes in handy for the training of new clergy and for looking up the answers to specific questions on the themes they discuss. At the same time that Salomon was trained, his brother by blood Waldo and future bishop of Freising, received his education in St. Gall as well (Mass 24). Both brothers could have been influenced by the recommendations of Notker. However, our fragment of the *Liber Testimoniorum* and the copy of the *Ecloga* from the NUK (MS 6) were presumably both created a bit earlier than 850, so before the episcopates of the two brothers. Waldo only became bishop of Freising in 884 and Salomon took over the see of Konstanz in 890 (Mass 72–73, 84). Both works were written in the same geographical area and have a very similar appearance. As was already mentioned, the paleographical examination of the fragment shows us that it was most likely part of a codex that was written in Southwest Germany, more specifically Freising, or the Eastern part of what is now Switzerland (Golob, *Srednjeveški rokopisi* 181–182). MS 6 from the NUK containing the *Ecloga* has also been attributed to a Southern or Southwestern German writing school. In the ninth century the work of Lathcen was mainly copied in Murbach, Reichenau, Salzburg, Konstanz and other

------

[6] "But if you could ever find the excerpt of Paterius, which he produced by selecting the books of the blessed Gregory through the order of the individual books, it will be sufficient for you in all matters of wisdom."

centers in the area (Kos 300–301). It is possible Notker's recommendations might have had something to do with this. There is a discrepancy in time between the recommendations and the writing of the *Ecloga* and the *Liber Testimoniorum*, so these works were not copied under Notker's influence. This does not mean, however, that the introduction of these works is solely related to Notker. It is very likely that he simply propagated the usage of the works by Lathcen and Paterius because they were commonly seen as valuable and useful by the monks at St. Gall. The tradition of using these works could certainly date from before the *De interpretibus divinarum scripturarum* was written, which was definitely before Salomon became the bishop of Konstanz in 890 (Migne, *Patrologiae Volume 131* 993–994). These recommendations from St. Gall could have easily found their way into Freising, as there is earlier proof of close ties between the monastery and the Freising scriptorium. In fact, all the different writing schools in the Southwestern part of the Carolingian Empire were well connected (Golob, *Srednjeveški rokopisi* 182). It is therefore plausible to assume that Freising started copying the works of Paterius and Lathcen under the influence of the monastery of St. Gall. Therefore, it seems possible that the South-West German writing school we are looking for, is indeed the one of Freising. Due to a lot of similarities between the writing style of MS 6 and our fragment and the fact that the work of Lathcen was also propagated from St. Gall, we could carefully think about a Freising origin of the *Ecloga* (MS 6) as well. If we assume both manuscripts found their way into what is now Slovenia, it most likely happened after 973, so more than one hundred years after the manuscripts were created. The only possible explanation for this, could be the fact that older versions of important works were sent to the newer churches and religious centers in the peripheral areas of a bishopric (Golob, "Karolinški fragment" 280). This is what could have happened to both the manuscript of the *De interpretibus divinarum scripturarum* where fragment AS 1073, II–1r was part of, as the *Ecloga* (MS 6).

Of course, a definitive answer to the origin of the fragment shall probably never be given, but a Freising provenance seems to be the most plausible option so far. It is possible both works could have been used for the training of new local Slavic clergy or perhaps German clergy that proceeded their further education on the spot. This last possibility springs to mind because of a canon from the Council of Reims in 813. It says that the clergy should study to be able to better understand its duties. The works that are fit for study, are explicitly mentioned: the Bible, the canons, the Rule of Benedict, the *Regula Pastoralis* of

Gregory the Great and other writings of the Church Fathers. In this study through readings there should also be a particular focus on Mass, baptism, penance and the eight cardinal sins (von Hefele 758–759). However, where exactly the manuscripts were used is difficult to say, as it is very much possible that certain sites of religious training got lost through the ages.

Conclusion

Now that we know the true content of the fragment and have an idea of its provenance, it can be seen as much more than just the oldest manuscript fragment on Slovenian soil. It is a new addition to the manuscript tradition of *De interpretibus divinarum scripturarum* of Paterius of Brescia and hints towards a Freising provenance. Because of this more thorough examination of the fragment, it was possible to link it content wise and especially on a visual and paleographical level to MS 6 from the NUK that contains the *Ecloga* by Lathcen.

WORKS CITED

Aris, Marc-Aeilko. "Notker Balbulus." *Neue Deutsche Biographie*, 19th ed., edited by Franz Menges, Duncker & Humblot, 1999, p. 362.
Blaznik, Pavle. *Zemljiška gospostva v območju freisinške dolenjske posesti*. Slovenska akademija znanosti in umetnosti, 1958.
Boudinhon, Auguste. "Notaries." *The Catholic Encyclopedia: An International Work of Reference on the Constitution, Doctrine, Discipline, and History of the Catholic Church*, vol. 11, edited by Charles George Herbermann, New York, The Encyclopedia Press, pp. 222–223.
Castaldi, Lucia. "Lathcen." *La trasmissione dei testi latini del Medioevo*, edited by Lucia Castaldi and Paolo Chiesa, Sismel, 2012, pp. 374–387.
Castaldi, Lucia, and Fabrizio Martello. "Tempera quasi aurum: origine, redazione e diffusione del 'Liber testimoniorum' di Paterio." *Filologia mediolatina. Rivista della Fondazione Ezio Franceschini*, vol. 18, 2011, pp. 23–107.
Étaix, Raymond. "Le Liber testimoniorum de Paterius." *Revue des sciences religieuses*, vol. 32, no. 1, 1958, pp. 66–78.
Golob, Nataša. "Karolinški fragment 'Cura pastoralis'." *Arhivi*, vol. 25, no. 1, 2002, pp. 277–281.
Golob, Nataša. "Srednjeveški pergamentni fragmenti v knjižnih vezavah." *Zgodovinski časopis*, vol. 71, 2017, pp. 70–104.
Golob, Nataša. *Srednjeveški rokopisi in rokopisni fragmenti: Arhiv Republike Slovenije*. Arhiv Republike Slovenije, 2018.
Gregorius I and Marc Adriaen. *[S. Gregorii Magni] Moralia in Iob: libri I–X / Cura et studio Marci Adriaen.* Brepols, 1979.

Gregorius I and Marc Adriaen. *[S. Gregorii Magni] Moralia in Iob: libri XI–XXII / Cura et studio Marci Adriaen.* Brepols, 1979.

von Hefele, Karl Joseph. *Conciliengeschichte. Dritter Band (zweiter Auflage).* Freiburg im Breisgau, Herder'sche Verlagshandlung, 1877.

Kos, Milko. "Ljubljanski rokopis Lathcenove 'Ecloga de moralibus in Job'." *Razprave Znanstvenega društva za humanistične vede*, vol. 2, 1925, pp. 289–302.

Lathcen and Marc Adriaen. *Egloga, quam scripsit Lathcen filius Baith de Moralibus Iob quas Gregorius fecit. Corpus Christianorum Series Latina vol. 145.* Brepols, 1969.

Martello, Fabrizio. "Paterius." *La trasmissione dei testi latini del Medioevo*, edited by Lucia Castaldi and Paolo Chiesa, Sismel, 2012, pp. 431–446.

Mass, Josef. *Das Bistum Freising in der späten Karolingerzeit.* Seitz und Höfling, 1969.

Migne, Jacques-Paul. *Patrologiae cursus completus ...: Series latina, Volume 131.* Paris, Garnier fratres, 1857.

Migne, Jacques-Paul. *Patrologiae cursus completus ...: Series latina, Volume 75.* Paris, Garnier fratres, 1862.

Migne, Jacques-Paul. *Patrologiae cursus completus ...: Series latina, Volume 77.* Paris, Garnier fratres, 1862.

Migne, Jacques-Paul. *Patrologiae cursus completus ...: Series latina, Volume 79.* Paris, Garnier fratres, 1862.

Rudy, Kathryn. *Piety in Pieces: How Medieval Readers Customized their Manuscripts.* Open Book Publishers, 2016.

Sickel, Theodor. *Diplomata regum et imperatorum Germaniae. Bd. 2, Tl. 1. Die Urkunden Otto des II. (Ottonis II. Diplomata).* Hannover, Hahn, 1888.

Štih, Peter, Vasko Simoniti and Peter Vodopivec. *Slovenska zgodovina: družba—politika—kultura.* Inštitut za novejšo zgodovino, 2008.

Weiss, Karl Theodor, and Wisso Weiss. *Handbuch der Wasserzeichenkunde.* VEB Fachbuchverlag, 1962.

Fragment »Cura pastoralis« iz Arhiva Republike Slovenije: povsem drugačna zgodba

Ključne besede: krščanska književnost / srednjeveški rokopisi / 9. stol. / fragmenti / Gregor Veliki / Paterij iz Brescie / Lathcen: *Ecloga* / Trbovlje / Freising

Fragment »Cura pastoralis« je najstarejši rokopisni odlomek na Slovenskem in se hrani v Arhivu Republike Slovenije pod številko AS 1073, II–1r. Sestavljen je iz enega bifolija. Ena stran je zelo poškodovana, saj je bil fragment nekoč uporabljen za vezavo knjige, ki je med letoma 1669 in 1704 služila kot poročna knjiga mesta Trbovlje. Matično knjigo še vedno hranijo v Nadškofijskem arhivu v Mariboru. Bifolium je kataloziziran kot fragment rokopisa iz 9. stoletja, ki vsebuje »Cura pastoralis« Gregorja Velikega. Vendar pa se je po temeljiti raziskavi izkazalo, da gre v resnici za del dela Paterija iz Brescie,

prvega tajnika Gregorja Velikega, ki je morda luč sveta ugledalo v freisinškem skriptoriju. Med odlomkom in najstarejšim celotnim rokopisom na Slovenskem, Latchenovo *Eclogo*, ki je nastala v istem obdobju, in sicer ob koncu prve polovice 9. stoletja, je nekaj fizičnih in vsebinskih podobnosti.

John Donne in Slovenian and the "Reflexive" Translation Theory of Antoine Berman

John Stubbs

University of Ljubljana, Philosophy of Arts, Department of English and American Studies, Aškerčeva cesta 2, 1000 Ljubljana
https://orcid.org/0009-0005-3765-7334
johnanthony.stubbs@ff.uni-lj.si

In his last book, Antoine Berman developed his theory of "translation criticism" by considering the poetry of John Donne in French translation. The present article applies that theory to Slovenian translations of Donne by Janez Menart and Marjan Strojan. It makes particular use of Berman's concept of "marked" textual features to consider the survival of Donne's "far-fetched" stylistic devices in Menart's and Strojan's translations. The article suggests that translation critics outside the Slovenian milieu should be aware of the preconceptions fostered by the hierarchy of "high" and "low" impact cultures. Silvia Kadiu, commenting on Berman, stresses the importance of "reflexivity" on the part of literary translators. When, as translation critics, we consider these two translators' place within Slovenian culture and investigate the critical erudition they display towards Donne's historical milieu, we find that Strojan and Menart pay exceptional respect to the principle of reflexivity.

Keywords: literary translation / English poetry / Donne, John / Slovene translations / Menart, Janez / Strojan, Marjan / reflexivity

The theorist Antoine Berman attempted to resolve the enduringly tricky question of how we should assess literary translation. Berman synthesized his ideas in his last book, *Towards a Translation Criticism: John Donne*, and debated the success of a number of French translations of the infamously challenging English poet.[1] My remarks here will apply Berman's methods to another case study: translations of Donne included in the Slovenian language anthology of English poetry edited by Marjan Strojan, which has become a standard reading text and reference work in Slovenia since it first appeared in 1997. The discussion

---

[1] For synopses of the book and summary of its place in Berman's oeuvre, see Davis 2013 and Berman's translator's introduction (Berman vii-xvii).

197

will center on Berman's conception of the textual features that translator and critic should determine as "marked." The hierarchy of "high" as opposed to "low" impact cultures will prove to have a bearing here, one that Berman's theory, while not explicitly acknowledging it, effectively encompasses and counteracts. I will end by drawing on the work of Silvia Kadiu to support Berman's suggestion that in certain circumstances we should understand translation itself as a "reflexive" act of criticism. The Slovenian language treatments of Donne by Strojan and Janez Menart, I will argue, fall under this category.

<div align="center">***</div>

The criticism of literary translation complicates the challenge posed in assessing any form of translation. The question is never merely "Does the translation get the message across?" but "are the literary merits of the translation adequate to convey the message?" and, of course, "what actually *is* the message?"

For Antoine Berman, a distinguished specialist in the field, a conflict between schools of translation criticism exacerbates the problem. Some critics are loyal above all to the historical particularity of the original text as they perceive it. Berman is less than polite towards them. "Engagé" analyses, as he calls them, are all about denunciation; "denouncing," moreover, "with precision." They involve the "meticulous tracking of the incoherencies, poor systematicity, and biases of the translators" (Berman 32–33). In contradistinction if not outright opposition to such readers are translation critics of what Berman calls the "Socio-critical" or "Tel Aviv" School. Such critics stress the "norms" translators follow and the cultural discourses in which they participate. Berman explains the "Tel Aviv" rationale as follows: "To analyze a translation without going back to the system of norms that shaped it, then to judge it on this basis [i.e. without considering the position and background of the translator], is thus absurd and unjust, since the translation could not be otherwise, and since, as an act of translation, it only had meaning insofar as it was an operation subjected to these norms." (39) Berman as such is kinder, on the whole, to Tel Aviv critics of translation than their "engagé" counterparts, but he still finds their methods deficient. The "Tel Aviv" school would implicitly accept as translation anything that presents itself as translation. Berman sees this as an overture to bowdlerisation: translations can all too easily sacrifice challenging content (or passages that are merely "hard" to translate) to "the literary norms of the receptor culture" (43).

The conflict between "engagé" and "socio-critical" analyses is a painful one for Berman because he deeply wishes to defend the integrity of translation, which is to say the literature of translation, as literary art in its own right. He stresses the "creative and autonomous role of translation in Western history" (Berman 39): it is for him a noble endeavor. He is unapologetic (and a tad provocative) in insisting on the Latinate and indeed essentially "Roman" nature of the discipline of translation (9–11); he claims that there would be no "space for *colinguism*" in "the West" without translated literature.

Berman develops the engagé / socio-critical distinction very much on his own terms, but parallels and precedents in the work of other theorists bear him out. Skopos theory offers a less idiosyncratic description of the competing traditions that can guide critics of translation. Some seek the freedom of the modes Katherine Reiß called "communicative" or "creative" translation; some will place a literary translator's efforts against the standards required by "philological" translation (Reiß and Vermeer 124–125). Once all due distinctions are made, those categories fall in rather neatly with Berman's schools of criticism: the engagé critics are natural allies of philologists while socio-critical "Tel Avivians" are likely to be patrons of creative-communicators.

These were rifts and torsions that Berman hoped to overcome. His own theory seeks to synthesize and to some extent reconcile the two approaches. The translation critic must be an elucidator as much as an evaluator; he (Berman and his translator use the masculine pronoun) must both study meticulously the key features and historical setting of the original (Berman 54–58), and consider the "position" and "product" of the translator himself (58–62). With great care, the critic may then proceed to set up a "confrontation" of translation and original. In doing so, he must remain aware of factors that together comprise the translator's "horizon." The translator's hermeneutic horizon, for Berman, who cites the "post-Heideggerian" tradition which developed the metaphor, is made up of "the set of linguistic, literary, cultural and historical parameters that "determine the ways of feeling, acting, and thinking of the translator" (63).

Berman resists what he calls the "infeudation" of literary translation to any particular school of theory that might "tell translation 'what to do'": he is against "traductologies" *per se* (Berman 53). Nevertheless, a "philological" or "engagé" priority, for him, does bind critic and translator alike: namely a strong working knowledge of the critical traditions that pertain to a given text. For Berman, marked features of a literary text are "those stylistic characteristics, whatever they may be, that

individuate the writing and the language of the original" (51); and yet, broad as their remit might seem, neither critic nor translator have complete autonomy in deciding what the "marked" or salient features of the original work might be. Those features have already been "marked" by other hands, including, in some cases, the author's. Berman approvingly cites Chateaubriand resolving to surround himself "with all the disquisitions of the scholiasts" in order to translate *Paradise Lost* (Berman 52).

Still greater rigor is required, for Berman, on the part of the translation critic. The care he urges in this regard, Silvia Kadiu argues, stems from his sense that "reflexivity" is essential to good translation, a belief "that it is no longer possible to practice translation without reflecting on it" (Kadiu 97). Translation, for him, is a critical activity in itself. "Translation criticism" is as such an ambiguous term. It can denote either the criticism of literary translation or, in specific conditions, the criticism performed by literary translation. Since, in the former sense, it is in effect the "criticism of criticism" (98), Berman insists that good translation critics will always know more and see further than good translators. The critic's readings should be "more connected, more systematic, than those of the translator" (Berman 54). With little hope of reaching that standard, I shall at least try to apply it to two Slovenian translators' treatments of two poems by Donne.

<center>***</center>

"Marked" stylistic traits, for Berman, are by no means the most elegant or beautiful in a writer's repertoire: they are the most "necessary" (Berman 55). Although Berman devotes almost half his book on translation criticism to the discussion of a single poem by Donne (Elegie XIX), he does not say a great deal about the history of reception that has enfolded and transmitted Donne's poetry. By common consent, nevertheless, Donne's poetic style is marked by a fascination for striking, yet often logically contrived and outlandish figures and constructions. It arguably marks one of his chief contributions to the later seventeenth-century tradition. As one recent authority observes: of other English "metaphysical" poets, often labelled members of a "school of Donne," "it is traditionally said that the group was united by the use of far-fetched comparisons, or "conceits," that drew attention to their own ingenuity—although this is more evidently a feature of Donne's work than that of other members of the group" (Burrow). Partly on account of this predisposition (and partly due to Donne's notorious

arcaneness), his friend Ben Jonson warned that Donne's writing "for not being understood, would perish" (Jonson, "Conversations" 466, [187–188]). Modernist criticism of course demanded that we cherish that preoccupation with what Jonson disapprovingly termed "far-fetched descriptions" (Jonson, "Explorata" 427, [2197–2198]). In either case, regardless of the critical view taken, "far-fetchedness" qualifies as a "marked" feature of a distinctive Donnean text.

The basic premise of Donne's well known "Song" ("Goe, and catch a falling star") is that it is easier to experience miracles than for honesty to prosper or a beautiful woman to be faithful. Here is a signal moment:

> If thou beest borne to see strange sights,
> Things invisible to see,
> Ride ten thousand daies and nights,
> Till age snow white haires on thee. (Hayward 4)

There is broad agreement that Donne wrote for an extremely well read, rhetorically literate coterie readership, which encountered his secular poems in manuscript. Some if not all of his first readers will have identified the idea of age "snowing" white hair "on thee" with a Horatian metaphor, "the snows of the head" ("capitis nives"; Horace, *Carmina*, 4.13.12). Classical rhetoricians disapproved of such contrivances. Marcus Quintilian, the eminent Roman authority, singled out Horace's trope as an example of questionable practice. It demonstrated that metaphors could at times be "harsh" when "derived from distant resemblances" (Quintilian 435, note 22, vol. 3). Ben Jonson, who greatly admired Quintilian, echoed him directly when he declared that "metaphors farfet [far-fetched] hinder to be understood" (Jonson, "Explorata" 431, [2359–2360]). As I mentioned above, Jonson seems to have had constructs of this sort in mind when he worried that Donne's work would perish "for not being understood."

Donne, however, as so often, made this particular trope work: in his poem, unlike Horace's, one does not think of dandruff. The odd tenor of the image fits well among the fabulous sights and outlandish events his listeners will need to experience before they meet a) "an honest mind" or b) a "woman true and fair," even though "the snows of the head" are in themselves a commonplace phenomenon. Donne's use of the idea moreover serves a further purpose, one that he may well have expected a rhetorically educated reader to spot. Some with Donne's learning may have recalled the specific place where Quintilian gently chides Horace; some will merely have recognized that Donne's use of

the trope went against ancient authority. This was the sort of thing one should not do, in poems or in letters or sermons or, for that matter, state documents: the blizzard of white hair advertises the writer's affinity for, and willingness to use, a class of devices that classically inspired teachers of rhetoric vigorously discouraged. From the viewpoint of intellectual history, it lends Donne's writing a rebellious, almost disdainfully learned quality, and further accentuates the wearily contemptuous air of the "Song." It is a tiny, yet telling indicator of his outlook, and possibly his status, among the Elizabethan and Jacobean avant-garde. It is a feature "marked," moreover, by Ben Jonson's unmistakable disapproval. [2]

Modern readers of Donne will, unless they have had some prior contact with Horace's odes, remain entirely oblivious to this slight contextual tremor until and unless a footnote or some other expository guide informs them of it. The notion of age *snowing* white hairs accordingly poses the translator a challenge. It would be all too easy to opt for a bland if clearer alternative such as "till old age makes your hair go white." Here is Marjan Strojan's translation of the passage:

> Če čudés ti vid je dan,
> glej nevidene reči,
> jezdi tisoč dni in dan,
> da lase Čas zasneži. (Strojan, *Antologija* 127–128)

> (If sight of miracles is given to you,
> See things never seen before,
> Ride a thousand days and a day,
> So that Time covers hairs with snow/
> So that Time starts to snow hairs.)[3]

---

[2] Marotti drew on prior manuscript scholarship to refine the coterie context and inspired many further studies; Pebworth is an early example. Ettenhuber offers a recent view of how knowledge of rhetorical convention guided both Donne and his readers (or listeners) in negotiating "far-fetched comparisons" (Ettenhuber 401). On the wider classical and early modern scepticism towards such devices, see Berry; Cooke 222–225; Skinner 88. My doctoral topic some twenty years ago was the bearing of the far-fetched on Shakespeare; I remain grateful to my supervisors, Gavin Alexander and Colin Burrow, for their thoughts in this area as in others. I returned to the subject in publications on later figures and periods (e.g. Stubbs, *Jonathan Swift* 92–93).

[3] Paul de Man warned mockingly of the dangers of translating translation (de Man 35; commenting on Benjamin 81). I offer my back translations in a duly cautioned spirit of pragmatism, for readers unfamiliar with Slovenian.

Putting the translation alongside the original reveals a number of disparities; not least a typological distinction between "Age" in Donne's poem, and "Time" in Strojan's reading. The "harsh" metaphor of age's snow is notably present, however. Rather strikingly, Strojan uses the resources of Slovenian both to preserve the far-fetchedness of Donne's image, and to soften it. As my (far from definitive) back translation indicates, the line concerned may be heard, and translated, two ways. The first sense of the verb *zasnežiti* recorded by *SSKJ* (the standard Slovenian dictionary), "to start to snow," gives a line very close indeed to Donne's.[4] The second standard sense of the verb means to cover an object or surface with snow; if this is the active sense, we see the rider's hair turning white under the snows of time. There is no Slovenian counterpart to "white" or "on thee" in Strojan's reading. Donne's image has as such acquired lapidary focus, a precision which some might feel in fact alters or intensifies it. It has also taken on a spiritual pathos that Donne's text does not confer. The "snows of time" are invisible, figurative and defined only in their effects; a snow of white hair is, by comparison, wackily imaginable yet more than faintly absurd. Both images are present in Strojan's working of the line. Those who read Slovenian will also note that Strojan's line also has a quite different phonic concentration and evenness to Donne's; a smoothness maintained throughout the whole translation. I might (and I beg pardon for the play on words) seem to be splitting hairs; but the auditory and illocutionary qualities of a poem in verse translation are naturally a vital aspect of its meaning—and I will return to that side of things towards the end.

Berman is in no doubt at all that such minutiae are crucial. He exhorts us to discover as much as we can about the translator in our attempts to explain such details. "We want to know if he has written articles, studies, dissertations, monographs about the works he has translated and, finally, if he has written about his own practice as a translator, about the principles that guide it, about his translations and translation in general" (Berman 58). Most Slovenian readers of these pages will know of Marjan Strojan's verse translations of *Paradise Lost* and *The Canterbury Tales*, and his credentials as an interpreter of the later medieval and early modern canon. Strojan (1949–) is a distinguished and prolific poet in his own right, and has had considerable

---

[4] I am reliably informed, moreover, that the *SSKJ* entry on *zasnežiti* does not entirely capture actual usage: most Slovenians would hear in this context a perfective verb lending an emphasis which perhaps only the future perfect can render in English: "Ride a thousand days and a day / So that time will have snowed grey hairs."

influence on Slovenian literary culture beyond the sphere of translation alone. He has moreover spoken of translation as a "unique poetic act," by which working poets in effect place their skills at the service of their text, but invariably transmit something of themselves to it (Bratož). He might seem, as such, very eligible for recruitment to the "creative" school that socio-critical Tel Avivians admire. With regard to Donne, however, engagé philologists should note that not long before his *Antologija angleške poezije* first appeared, Strojan published a scholarly yet very accessible short account of Donne and his work. Strojan's article shows keen awareness of the erudition of Elizabethan-Jacobean literary circles and their familiarity with demanding forms such as the "artificial conceit" (Strojan, "John Donne" 7).

It seems reasonable to claim that, when translating Donne's "Song," Strojan decided that the line about snowy hair (or snowing hair, to be more precise) certainly mattered. The extent to which he was aware of the exact Horatian source is probably indeterminable; the notes to his translation of the "Song" ignore the line (Strojan, *Antologija* 685). Yet Strojan's phrasing creates a double effect. Time starting to snow white hairs is a Donnean thought; time covering hairs with snow abides by a principle closer to Quintilian's or Ben Jonson's. In Strojan's translation the image with which Donne transgressed the classical rule is present, and yet the rule itself is tacitly re-imposed.

Strojan's decision to retain the image of snow is all the more striking if one considers details in the original he decided to exclude. The original poem's opening sardonically urges the listener to "Goe, and catch a falling star." Strojan's Poet instead says:

Daj, ujemi zvezdo v dlan (Strojan, *Antologija* 127)

(Go on, catch a star in the palm of your hand)

Strojan's opening imperative, *daj*, is untranslatable (literally, it means "give"), yet entirely idiomatic. An entirely Slovenian expressive choice has as such displaced Donne's very characteristic, more than faintly sardonic "Go" (a literal rendering of which in Slovenian, *pojdi*, would sound most discouraging). More importantly still, is that the glancing and uncatchable flash of Donne's falling star disappears altogether. Instead, and it would seem unaccountably, Strojan introduces something of his own—the palm ("dlan") of the star-catcher's hand. Here Strojan has surely breached Berman's rule about preserving salient features of the original.

That said, a larger, not quite literal parity between Donne's line and Strojan's is worth considering. Strojan knew, as Donne almost certainly did not, that although one might not catch a shooting star at any considerable altitude, it was quite possible to pick up and pocket a shard of meteorite.[5] Plucking a fixed star out of the sky and holding it in one's palm is obviously inconceivable, even though the thought of doing so is both vivid and compelling. Accordingly, the core principle of Donne's line, namely an impossibility related to some direct intervention in a matter of astronomy, remains intact.

Here one might say that Strojan follows a maxim laid down by another theorist of translation, indeed another eminent Slovenian practitioner with whom he worked closely. Recalling his first efforts at translating Chaucer at the age of eighteen, Strojan has spoken of the tutelage of Janez Menart (Crnović). Menart (1929–2004) was later Strojan's fellow translator of Donne (among other English poets) for the 1997 anthology. Among Menart's maxims for the translator of poetry was the following:

> The translator is obliged to mediate between the poet and readers of the second language. In doing so he will wish by all means to ensure that the reader of the translation will experience to the greatest possible extent the "same" feeling he would derive from the original, if he spoke the language in which the poem was first written (Menart, "O prevajanju" 666).

The philological / engagé traductologist would protest: yet the *skopos* for translation set out here is undoubtedly relevant to Strojan's management of Donne's "Song." Although he cancels Donne's falling star, Strojan surely preserves the dominant "feeling" of Donne's poem—its incredulity about virtue, its challenge that the reader or listener perform the impossible. Indeed, there is possibly an overlap here between Berman's theorising, "post-Heideggerean" approach to "marked" textual features and Menart's lighter, aphoristic notion of "feeling." My next concern will be the extent to which one of Menart's anthology contributions manifests the same stylistic traits in Donne preserved by Strojan, as prescribed by Berman, and whether it follows his own self-declared principle of translation.

<div align="center">***</div>

---

[5] The astronomical nature of shooting or falling stars was established by Giovanni Schiaparelli in 1862.

If we consider Menart's treatment of a poem to which Berman devotes a great deal of attention, Donne's nineteenth "Elegie," "To his Mistress Going to Bed," few would find any loss of the original's overwhelming, indeed provocative and controversial erotic charge. Consider the following lines, first in Donne:

> Licence my roving hands, and let them go,
> Before, behind, between, above, below.
> O my America! my new-found-land,
> My kingdome, safeliest when with one man man'd. (Hayward 96)

And in Menart's version:

> Naj blodijo mi roke gori, doli
> in spredaj, še vmes dovoli!
> O ti, Amerika! Naj te spoznam,
> kraljestvo, varno le, če vladam sam!

> (May my hands wander up, down,
> and in front; allow them to go between, as well!
> O you, America! May I know you,
> a kingdom, only safe if I rule it alone!)

The combination of carnal and imperialist imperatives—and the uncertain irony that hovers over the latter—is still present. Significantly, however, Menart modulates the imperative voice heard through Donne's imperious "Licence," by using the third person with the particle *naj* to express a wish or "understated command" ("omiljen ukaz"; see *SSKJ* 1, a). Menart adds, and he takes away. The Slovenian verb "bloditi" (wander, stray) takes on a comic vagrancy here that Donne's simple "go" doesn't have. The geographical and suggestive particularity of Donne's "Newfoundland" meanwhile, disappears.

There is much throughout the translation to satisfy *les engagés*, and much to vex them after second thoughts. Menart had reservations about the capacity of the Slovenian language to furnish cognates for archaic or recondite diction in pre-modern texts (Menart, "O prevajanju" 668); yet he is never short of an equivalent, by necessity somewhat antiquated term for each of the various garments and accessories the poem's "mistress" casts off. There are also moments when he succeeds in achieving a near-perfect literal rendering of an English line that is also wonderfully euphonious in Slovenian. "Tu, kjer je prst, pritisnil bom svoj žig" ("Here, where my finger is, I shall set my stamp")

is Menart's treatment of Donne's "Then where my hand is set, my seal shall be" (Strojan, *Antologija* 125; Hayward 97). The task of the translation critic involves weighing up the significance of Menart's replacing Donne's hand with just a finger, or making the first person "I" rather than "my seal" the active subject of the second clause.

Despite the closeness of translation, in letter and spirit, to the original, the shifts noted above indicate that the "feeling" of Menart's translation differs quite strongly from Donne's. A single finger of course suggests contact of a kind quite distinct from that which the entire, controlling hand of Donne's Poet establishes. The formal "you" occurs throughout when Donne's speaker addresses his "mistress" (as "madam," moreover, in l.1); Menart's Poet uses the informal Slovenian *ti* pronoun. Donne's avoidance of his customary informal "thou" is distinctive if not exceptional amid his "love poetry," and contributes to a disquieting quality in the original that Menart tries on the whole to subdue. There is little in Menart's text of the provocative humor that emerges in Donne's opening couplet:

> Come, Madam, come, all rest my powers defie,
> Until I labour, I in labour lie. (Hayward 96)

The speaker compares his frustration in waiting (for the "labour" of lovemaking) to the pangs of childbirth: a claim that would stir objections in any age. This is the first of a number of figurative gender exchanges that conclude with the Poet eventually comparing himself to a midwife. The figure doubtless qualifies as a "farfet metaphor" of the kind both Quintilian and Ben Jonson disliked (few women experiencing parturition would compare the sensation to impatient sexual arousal). Menart, for his part, drops the idea of the male speaker's imagined womb dilating.

> Daj, pridi že, moč ne trpi čakanja,
> napor nedela k delu me priganja. (Strojan, *Antologija* 124)

> (Go on, come here, [my] strength can't bear waiting,
> the strain of not working is pestering me to start work.)

Berman's theory of translation criticism urges us to inquire why the translator might cancel such a salient point in the original. Given the control of idiom and undertone demonstrated elsewhere in the same translation, it is inconceivable that Menart missed the meaning of the English phrase "in labour." He has substituted Donne's anatomical

impossibility with an altogether much fainter paradox, "napor nede-la"—the strain (or effort) of inactivity. He was possibly uncomfortable with the drastic if entirely figurative sexual transformation the line imagines, or indeed its incredible sexism. Much of the original, in lexis, structure and spirit is manifestly there in the Slovenian. The inflective echo of "labour" in Donne's line survives in the recurrence of "delo" in Menart's, although the precise figurative form of the original disappears.[6] Similarly, Menart's "Daj, pridi že," carries a remarkably exact prosodic echo of Donne's "Come, Madam, come." Notwithstanding such fidelity, the outlandishness—indeed, the outrageousness—of the original opening has disappeared. Perhaps Menart felt that the poem could do without it; perhaps he felt it was in poor taste. Certainly, he rejected any option that might have compromised the idiomatic clarity of his own Slovenian line or the smooth rhyme of his couplet. For whatever reason, he has applied a Jonsonian principle and cancelled a far-fetched figure: in effect, he has edited Donne's poem. In doing so, unlike Strojan in his handling of the "falling star," Menart broke his own rule. He manifestly did not preserve the "feeling" of the original.

In fact, as we have seen, Menart has made somewhat free with his text overall; he has softened its pungent, radical humour, introduced an air of genuine intimacy that the original arguably lacks, and made it resemble the "poem of love, joy and nudity" that Berman, for one, believed it to be (Berman 13).[7] The question is, should he be applauded or reprimanded for doing so; or, to put it in less "engagé" terms, how should we—we trainee translation critics—explain the changes Menart chose or was led to make?

Berman's ideas would to some extent rescue Menart's translation from a blunt verdict that the translator "got things wrong". Yet there is something else present, on the hermeneutical horizon, as it were, that even Berman does not consider: namely the possibility that Menart did not feel he was getting things wrong but that he was putting certain things right, and that he felt entitled to do so.

Another factor affecting discussion of the success or otherwise of such a translation is the sort of translation culture to which it belongs. Theorists define national or regional literary cultures as "low" or

---

[6] Donne's device is an example of antanaclasis, the repetition of a word (here, "labour") with a change of meaning in the second instance.

[7] I first encountered Menart's version of the elegie when the late Uroš Mozetič read it aloud at a literary event in Ljubljana: I was immediately struck by it as an undoubtedly brilliant poem—but one quite different to the Donnean text. Menart's performance is a poem of love; Donne's is not.

"high" impact: as Martina Ožbot explains in a recent collection of papers, Slovenian works routinely fall under the latter category, and the imbalance is equally felt in the sphere of translation. Addressing trends in that sphere, Ožbot works with the longstanding linguistic distinction between "foreignization" and "domestication." She considers, for instance, an Italian version of Miroslav Košuta's poem "Daljava" ("Distance") (Ožbot 34) as an example of how Italian translation very often "domesticates" Slovenian-language structures and idioms with forms that better suit Italian. The English-Slovenian context obviously offers no direct parallel to the cultural locus of Western Slovenia and North-East Italy that Ožbot approaches. A number of principles she summarizes do however apply: there is a basic tendency for "central cultures," notably, "to be more self-focused, while peripheral cultures may be more open to the Other" (86). Among the trends such openness displays in the practice of translation is a greater willingness within "peripheral" cultures to sacrifice the target language's idiomatic norms for the sake of preserving characteristics of the source text. As Ožbot puts it: "In Slovenian target texts, foreignization is often encountered, especially at the level of lexis, and appropriation of source elements is a common characteristic of literary texts translated into Slovenian." (87)

Foreignizing their language in this way was an accommodation that both Menart and Strojan were patently unwilling to make. At significant moments, indeed, they pointedly resist the prevailing trend of central-peripheral relationships. Ožbot, guided by the work of Itamar Even-Zohar, observes that such practice, in emphasizing "acceptability" (to the target culture) is more typical of translators working within central cultures. Peripheral cultures typically place more stress on the "adequacy" or accuracy of translation (Ožbot 86). Menart and Strojan appear to have challenged this unwritten arrangement. How and why they did so merits further scrutiny.

<p style="text-align:center">***</p>

Berman, to his credit, would urge us to go "in search of the translator" (Berman 57) regardless of the language group to which "he" belongs, peripheral or central (or median): and when we do so we learn that Menart, like his younger colleague Strojan, was a distinguished and prolific translator of medieval and early modern literature. On further investigation it also transpires that Menart, a figure proverbially "born for success" (Glavan), was and remains highly regarded in Slovenia for

his own poetry as well as his translations. He was a major figure within the country's postwar milieu.[8]

Menart's intermittently "editorial" approach to Elegie XIX can be better understood when we consider the sort of poet he was. Unlike Donne, Menart's diction and figurative language exhibit a preference for the fecund minimum. Like Jonson, whom he clearly admired, he seems to have felt that a composition "hath blood and juice, when the words are proper and apt, their sound sweet, and the phrase neat and picked" (Jonson, "Explorata" 436, [2570–2573]). Menart's Slovenian diction could be extremely rich, yet he obeyed a strikingly strict Elizabethan rule of aptness: his poetry displays none of Donne's taste for cryptic extravaganzas. It could, nevertheless, be distinctly "metaphysical," as in one of his most famous short lyrics, "Jaz" ("I" or "Self") (first published in 1953):

> Pred ogledalom nem stojim
> in tujca pred seboj strmim.
>
> Kot da zrem prvič ta obraz,
> vprašujem ga: Si ti res jaz?
>
> Zamišljeno me zro oči
> in vprašajo: Sem jaz res ti?
>
> In trezno pravi mu moj jaz:
> Jaz nisem ti, ti nisi jaz;
>
> jaz sem le jaz, ki se mi zdi,
> in ti si jaz le za ljudi;
>
> a pravi jaz je dan za dnem
> uganka meni in ljudem. (Menart, *Stihi* 8)
>
> (I stand mute before the mirror
> And stare at the stranger in front of me.
>
> As if beholding that face for the first time,
> I ask it: Are you really I?
>
> Thoughtfully its eyes gaze on me,
> And ask me, am I really you?

---

[8] For testimony to Menart's standing, see e.g. Mejak 317–319; Jenuš; Zlobec; Žerdin.

And, thinking straight,[9] my self tells it,
I am not you, you are not I,

I am just I as I seem to myself,
And you are I only for others.

But the real I is, day after day,
A riddle to myself and other people.)

My back translation can of course capture nothing of Menart's versatile simplicity, nor of the rich shifts of sense he creates from the repetition of small, everyday words, nor his carefree exploitation of the morphological resources of the Slovenian language in his discovery of rhyme. At moments, equally, the rhyme belongs in the nursery, consisting of no more than a straightforward joining of nominative nouns. *Jaz* in Slovenian is both the first person subject pronoun and the word for the self, and the unity of the self as such is split between those senses as well as between the poles of gazer and reflection; belying the complexity it stirs, the word rhymes crisply with *obraz*, the word for face. Conceptually, the poem strands its speaker in a perpetual feedback loop of perception and reflection; yet, simultaneously, it provides gnomic closure to satisfy any child committing it to memory. For it is a child's poem: a striking and accessible Slovenian treatment of the mirror phase published long before Slavoj Žižek got anywhere near that subject. To a student of Menart's translation of Shakespeare, the poem also offers a magical antidote to the pain of Sonnet 62, to which its vocabulary presents compelling parallels. Within the parameters of seventeenth century stylistics, its "neat and picked" use of form and diction are much more reminiscent of George Herbert, that avid student of Jonson, than of Donne.

The poetic mode displayed here is equally evident in Menart's practice as translator. His writing manifests what one could call a Jonsonian or, more strictly speaking, a neo-classical poetic; he applied the same principles to Donne's text—for its own good, as Jonson would have said. Those principles emerge throughout Menart's treatment of the elegie: in his redacting the strange idea of a man being in labor, or replacing Donne's rough possessive hand with a single, more sensitized finger. One might well disagree with Menart's choice, in cancelling the

---

[9] "Collectedly" might be a better translation here of "trezno" (literally, "soberly"). Amid the post-war existential agonies over human identity, Menart's "self" keeps a cool head—even while professing itself to be a fragmentary enigma.

distinctive, zany moment of Donnean inaptness which the opening couplet of Elegie XIX comprises, but it would be unwise to discount it or class it as an error of translation. The harshest words, moreover, are likely to arise from within the translator's own milieu: a scholarly and principled yet misguided critique, for example, reprimanded Menart for apparent inaccuracies in his early verse translation of Shakespeare's *Sonnets* (see Grosman).[10] I suggest that we refer the matter to another standard invoked by Antoine Berman.

<p style="text-align:center">***</p>

The larger point to be made with regard to "translation criticism" hinges on one of Berman's central tenets: translation should be a "reflexive" exercise. For Berman, translation is reflexive when it constitutes an act of criticism in its own right. In Silvia Kadiu's summary, he believed that "the critical nature of translational reflexivity is realized in the process of translation itself" (Kadiu 96).

If translators are to be credited as critics, a question of authority arises. The status of translators depends on their status within and the scale of their contribution to their own culture; it also depends on the status of that culture itself within an international hierarchy. Some languages may "domesticate" others, and others may not. Apparent divergences from the original in a translation by Seamus Heaney or Robert Lowell are likely to gain recognition as creative, interpretive or critical acts. Within the Slovenian milieu, clued-in readers are able to accept Menart's translations in the same spirit. They know and, generally speaking, revere his style, and they can recognize his hand (or finger) in the text of his translation. This is to say that they will see that Menart's treatment of a poem by Shakespeare or Donne represents a synthesis of his learning and his own capacities and instincts as a poet. His profile and back catalogue in fact reinforce awareness of the cultural contingency of his translation. To those unacquainted with Menart's "low impact" cultural setting, his apparent amendments of Donne are more likely to seem errors.

Kadiu warns us that the principle of reflexivity is always at risk of being lost in such situations. "Reflexivity in research," she tells us, "is built on an acknowledgement of the historical and ideological pressures forming researcher and researched alike" (Kadiu 14). Within the parameters of Berman's theory of translation criticism, this means that the critic

---

[10] I discuss the critical debate between Grosman and Menart in Stubbs, "Shakespeare's *Sonnets*."

must be open about "his" hermeneutical horizons while attempting to delineate the horizons of author and translator. Reflexivity should be more than a mere profession of faith in and adherence to the principle itself, or a superficially self-relativizing statement of intent: "no amount of highlighting, pointing or self-awareness can ensure an ethical practice of translation" (145). Kadiu voices deconstructive pessimism about language's ability to perform what it preaches, or find oneness with the Other it purports to signify. True reflexivity, she suggests, lies within the stylistic "play" of the translation itself and any critical appraisal of it; and even then, manifests itself as an absence rather than a presence, "at moments of hesitancy, ambiguity and opacity" (145).

For Kadiu, reflexivity is primarily a safeguard against the distortion of "minor" by "dominant" cultures. Without reflexivity, the outcome is invariably translation Kadiu finds unethical. "The main idea underlying this line of thought," she explains, "is that, in order to highlight manipulation (such as a stereotypical representation of the source culture), research must itself be reflexive and think about the conditions of its own emergence." (Kadiu 14)

The loss of reflexivity is surely a bad thing, at the level of practice or theory, with respect to "minor" and "dominant" cultures alike. How, then, are we to understand the loss in translation—Menart's apparent deletion—of a crucial feature such as Donne's idea of being "in labour" at the outset of Elegie XIX? Is Strojan's cancellation of Donne's image of a falling star, in the famous "Song," another unwarranted intervention? Their apparent un-reflexiveness would be all the more grievous, at such moments, since a resounding quality in their manner of translation is its confidence—a notable lack of the "hesitancy" Kadiu sees as a hallmark of reflexivity. Instead of hesitancy, Menart and Strojan offer critical erudition. We (as translation critics) should approach their work as Tel Avivians but with the philological equipment of the engagé scholar. If we consider Menart's version of Elegie XIX in the light of a discussion of his own poem, "Jaz," the stylistic harmony between the two is self-evident; and the relationship is a reflexive one. The poetic that guides Menart in handling Donne's verse is plain to see. The shared affinity for difficult speculation, yet with an aversion to spiky, far-fetched conceptualization and a tacit dislike of Donne's sardonic, often rather unkind manner is manifest in the rhyming tetrameters, dating back to the 1950s, which I considered above. Our assessment of Menart's achievement as translator in short depends on whether translation critics bother to investigate his own creative literary achievement. This is by no means tantamount to licensing a

bowdlerised or sketchy translation: no one can seriously suggest that Menart has offered a slapdash or approximate realization of the elegie. Nor did he approach the exercise by imposing an utterly alien aesthetic on his text. He did not, for instance, turn Donne's stanzas and couplets into unrhymed free verse, and thereby pre-emptively silence engagé quibbles with an implied "take it or leave it." Both Strojan and Menart's choice of Donne, and their management of his thought, his diction and his rhythms, reflect a profound involvement with the stylistic values that shaped his work. Those values influenced their own practice as poets; their translation of Donne is a profession of that involvement. As poets, nevertheless, both translators are closer to Jonson than to Donne himself. They admit as such—reflexively rather than directly—in their treatment of Donne's approach to form.

Jonson, in his signature fashion, declared "that Donne, for not keeping of accent [i.e., for his disregard of prosodic rules], deserved hanging" (Jonson, "Conversations" 462, [42]). Both Strojan and Menart rejected this "marked" stylistic trait, and opted to present a distinctly mellifluous, metrical Donne to their Slovenian readership. A glance back at the stanza quoted earlier from Donne's "Song" will illustrate the point. The jagged cadence of Donne's line, "Till age snow white haires on thee," is unrecognizable in the sibilant, suggestive smoothness of Strojan's rendering: "Da lase Čas zasneži." The shade of Jonson would commend Strojan here; as would that of Pope, who "versified" (that is, rewrote) Donne's second and fourth satires. Eliot, who praised the rhythms of real speech he found in Donne's metrical irregularity, might disapprove, though I cannot see him faulting Strojan's craft.

The translator of Donne's poetry would be in a terrible position were it not that the arguments over Donne's treatment of "accent," insofar as translation is concerned, effectively cancel one another out. To reproduce intentional discordancy in the original when translating poetry is, if not impossible, then at least a risky policy to pursue (except in obvious instances of, for instance, burlesque); the target language readership will almost certainly interpret prosodic irregularity as a failure on the part of the translator. In the case of Donne, however, the translator is spared: attempting to replicate the liberties Donne took with "accent" would be redundant and self-defeating. For by a paradox almost worthy of the poet himself the polarity between the critics and defenders of his prosody proves to be an artificial one. Ultimately, like East and West, the two schools meet: the early twentieth-century champions of the Donnean mode argue that he actually did "keep accent." Eliot or F.R. Leavis would merely maintain that he did so by

observing principles of euphony to which Ben Jonson and the neoclassical tradition could never subscribe.[11] Instead of striving to emulate Donne's hanging offence, Menart and Strojan chose the wise course of using their own strengths as versifiers in serving Donne's text. Where their translation seems to edit and amend that text, it is reflexive, which is to say open, about the operative system of values that guides such amendments. Their insistence on euphony, on "keeping of accent" in Slovenian, is a sustained expression of that system.

Accordingly, Menart and Strojan succeeded in "domesticating" Donne; they made his original uncannily present in their translation, yet captured it within an idiom that is unmistakably Slovenian. They did so, nevertheless, by applying a literary standard that was present in, and characteristic of, the culture Donne addressed. In the Slovenian translations considered here, Donne has been judged and, in places, reformed—but according to laws that his ghost might recognize from his own time, and his own land. An "ethical" act of translation criticism, by Antoine Berman's standards, has been performed.[12]

WORKS CITED

Benjamin, Walter. "The Task of the Translator." *Illuminations*, translated by Harry Zohn, Schocken, 2007, pp. 69–82.

Berman, Antoine. *Toward a Translation Criticism: John Donne.* Translated by Françoise Massardier-Kenney, The Kent State University Press, 2009.

Berry, Philippa. "Tudor Laughter on the Verge of Metamorphosis: Ben Jonson and the Comedy of the Far-fet." *Tudor Theatre*, *6: For Laughs?*, Peter Lang, 2002, pp. 249–260.

Bratož, Igor. "'Vsako prevodno dejanje je svojevrsten pesniški akt': s Sovrètovim nagrajencem." *Delo*, 3 June 1993, p. 5.

Burrow, Colin. "Metaphysical Poets." *Oxford Dictionary of National Biography*, 2007. https://doi.org/10.1093/ref:odnb/95605. Accessed 27 May 2024.

---

[11] The example of this defence that has always stayed with me is of Leavis singling out Donne's use of "Did," violating the iamb at the beginning of the second line of "The Good Morrow": the moment is representative of Donne's "consummate control of intonation, gesture, movement and larger rhythm" (Leavis 11–12). Leavis's praise was notoriously hard earned.

[12] Many friends and colleagues have helped me in my work on Slovenian translations of John Donne and other Elizabethan writers. In particular I would like to thank Matjaž Berger, director of the Anton Podbevšek Teater in Novo Mesto, who invited me to give a talk on Donne—accompanied by readings given by Barbara Ribnikar—in April 2022. Looking further back, my great mentor on this poet is David Colclough, author of the entry on Donne for the Oxford Dictionary of National Biography, and I renew my thanks to him here.

Cooke, Elizabeth J. "Thomas Hobbes and the 'Far-fetched'." *Journal of the Warburg and Courtauld Institutes*, vol. 44, 1981, pp. 222–232. https://doi.org/10.2307/751074

Crnović, Deja. "Prešernovi nagrajenci 2015: Marjan Strojan." *Siol.net*, 2015, https://zgodbe.siol.net/presernovi-nagrajenci/marjan-strojan/. Accessed 27 May 2024.

Davis, Nathaniel. "Toward a Translation Crticism: John Donne by Antoine Berman." *French Forum*, vol. 38, no. 3, 2013, pp. 170–173.

De Man, Paul. "'Conclusions.' Walter Benjamin's 'The Task of the Translator': Messenger Lecture, Cornell University, March 3 1983." *Yale French Studies No. 69: The Lesson of Paul de Man*, Yale University Press, 1986, pp. 25–46.

Ettenhuber, Katrin. "'Comparisons are Odious?' Revisiting the Metaphysical Conceit in Donne." *Review of English Studies*, vol. 62, no. 255, 2011, pp. 393–413.

Glavan, Mihael. *Pesnik se predstavi: rojen za uspeh: dokumentarna razstava ob 80. obletnici rojstva Janeza Menarta*. Exhibition, 3 Sept. 2009—4. Oct. 2009, Razstavna dvorana NUK, Narodna in univerzitetna knjžinica, Ljubljana.

Grosman, Meta. "Shakespearjevi Soneti v slovenščini." *Slavistična revija*, vol. 35, 1987, no. 3, pp. 303–320.

Hayward, John, editor. *The Complete Poetry and Selected Prose of John Donne*. Nonesuch Press, 1932.

Jenuš, Nataša. "Dobra poezija je zelo osebna stvar: pogovor z Janezom Menartom—ob avtorjevi sedemdesetletnici." *Večer*, 30 Oct, 1999, p. 42.

Jonson, Ben. "Explorata, or Discoveries." *The Complete Poems*, edited by George Parfitt, Penguin Books, 1975 [1996], pp. 375–458.

Jonson, Ben. "Conversations with William Drummond." *The Complete Poems*, edited by George Parfitt, Penguin Books, 1975 [1996], pp. 461–480.

Kadiu, Silvia. *Reflexive Translation Studies: Translation as Critical Reflection*. University of Chicago Press, 2019.

Leavis, F.R. *Revaluation: Tradition and Development in English Poetry*. George Stewart, 1947.

Marotti, Arthur F. *John Donne, Coterie Poet*. University of Wisconsin Press, 1986.

Mejak, Mitja. "Marginalije ob liričnem jubileju." *Sodobnost*, vol. 21, no. 4, 1973, pp. 313–327.

Menart, Janez. "O prevajanju poezije." *Sodobnost*, vol. 13, no. 8/9, 1965, pp. 665–680.

Menart, Janez. *Stihi mojih dni: Izbrano delo.* Prešernova družba, 1989.

Menart, Janez. "O prevajalskem ptičem mleku." *Prevajanje srednjeveških in renesančnih besedil / The Translation of Medieval and Renaissance Texts*, edited by Martina Ožbot, Društvo slovenskih književnih prevajalcev, 2002, pp. 61–91.

Ožbot, Martina, *Translation and Multilingualism: A Dynamic Interaction*. Znanstvena založba Filozofske fakultete, 2021.

Pebworth, Ted-Larry. "John Donne, Coterie Poetry, and the Text as Performance." *The English Renaissance*, vol. 29, no. 1, 1989, pp. 61–75.

Quintilian, Marcus Fabius. *Institutio oratoria / The Orator's Education*. 4 vols., translated by D. Russell, Harvard University Press, 2001.

Reiß, Katharina, and Hans Vermeer. *Towards a General Theory of Translational Action: Skopos Theory Explained*, translated by Christine Nord, Routledge, 2014.

Skinner, Quentin. *Reason and Rhetoric in the Philosophy of Hobbes*. Cambridge University Press, 1996.

Strojan, Marjan. "John Donne, pesmi ene knjige." *Tretji dan*, vol. 24, no. 12, 1995, pp. 7–9.

Strojan, Marjan, editor. *Antologija angleške poezije: Od začetkov do konca 20. stoletja*. Cankarjeva založba, 1997.

Stubbs, John. *Jonathan Swift: The Reluctant Rebel.* Penguin and Norton, 2017.

Stubbs, John. "Shakespeare's *Sonnets* in the Slovenian translation of Janez Menart." *Elope*, vol. 20, no. 2, 2023, pp. 105–120.

Zlobec, Ciril. "V našem času in svetu je videl hkrati svetlobo in sence: Janez Menart 1929–2004." *Delo*, 23 Jan. 2004, p. 11.

Žerdin, Ali. "Janez Menart: 1929–2004." *Mladina,* 26 Jan. 2004, p. 55.

John Donne v slovenščini in »refleksivna« prevodna teorija Antoina Bermana

Antoine Berman je v svoji zadnji knjigi predstavil teorijo »prevodne kritike« prek obravnavanja poezije Johna Donna v francoskem prevodu. Ta članek omenjeno teorijo aplicira na slovenske prevode Donna, ki sta jih napravila Janez Menart in Marjan Strojan. Še zlasti se opira na Bermanov koncept »zaznamovanih« besedilnih značilnosti in z njim obravnava ohranjanje Donnovih »daljnosežnih« slogovnih sredstev v Menartovem in Strojanovem prevodu. Razprava kaže, da bi se morali prevajalski kritiki zunaj slovenskega okolja zavedati predsodkov, ki jih spodbuja hierarhija med »visokimi« in »nizkimi« kulturami vpliva. Silvia Kadiu v komentarju k Bermanovi teoriji poudarja pomen »refleksivnosti« literarnih prevajalcev. Ko kot prevajalski kritiki razmišljamo o mestu teh dveh prevajalcev v slovenski kulturi in preiskujemo kritično erudicijo, ki jo izkazujeta do Donnovega zgodovinskega okolja, lahko ugotovimo, da Strojan in Menart izjemno spoštujeta načelo refleksivnosti.

The Antinomies of Latvian Literary Realism

Benedikts Kalnačs

University of Latvia, Institute of Literature, Folklore and Art, Mūkusalas iela 3, Riga, LV-1423, Latvia
https://orcid.org/0000-0001-6435-325x
benedikts.kalnacs@lulfmi.lv

This article traces the social and literary contexts of Latvian literary realism. During its rise in the nineteenth century, Latvian realism was seen as developing along the general lines of European culture that helped to foster aesthetic diversity and innovative potential. These processes are the focus of the first part of the article. The subsequent reshaping and distortion of the concept of realism during the period of Soviet occupation are then scrutinized in some detail. The investigation follows in the footsteps of the current rise of interest in realism, considering its importance in a broad spectrum of contexts, and claims that realism is an innovative aesthetic practice of great historical and contemporary relevance. However, it has also frequently been misunderstood or consciously misinterpreted by representatives of other aesthetic trends as well as misused by political regimes. The article points toward the necessity of restoring a sound interpretation of realist aspirations and their social and aesthetic contributions.

Keywords: Latvian literature / literary criticism / realism / socialist realism / the Soviet regime

This article sets out to explore the social and aesthetic contexts of the rise of realism in nineteenth-century Latvian literary culture and its initial as well as subsequent reception.[1] Considering the recent revitalization of research on realism, this is a topical scholarly issue on an international scale. Significantly, the history of realism is partly a history of response to any kind of imposed restrictions; on the other hand, realism has also been instrumentalized as a tool of imposing and legitimizing such restrictions. The investigation of realism's historical misuses may help to understand the mechanisms employed by twenty-first-century totalitarian and populist ideologies. In this article I intend to trace controversies regarding the term and its application to

---

219

literary practices in two different periods of Latvian realism: the early debates of the 1880s and 1890s that focused on social and moral issues; and the highly biased interpretation of realism under Soviet ideological rule (in particular, in the 1950s and 1960s). In the conclusion, I link some of the research questions to the current revival of interest in literary realism.

There are many factors that caused the steady decline of interest and the subsequent recovery of the concept of realism in literary scholarship. One of the main aspects of criticism has been the predominance of formalist approaches in the humanities in the second half of the twentieth century. As Pam Morris indicates, the aesthetic legacy of realism was challenged and significantly compromised by certain trends in Structuralism, Postmodernism and Deconstruction (Morris 24–44). With this kind of theoretical interventions in mind, even scholars who considered broader social contexts had at times been working along similar lines, which explains the continuous exclusion of realist practices from their primary focus. However, the scrupulous research done by numerous scholars interested in social and historical contextualization of aesthetic trends has in the long run directly or indirectly stimulated the reappearance of studies of literary realism. If we trace just one such line of development, Naomi Schor, in her 1993 book, *George Sand and Idealism*, powerfully argued the importance of idealism within the rich texture of nineteenth-century aesthetic debates. While the relationship of Sand to her contemporaries such as Balzac was retained, the emphasis on idealism shifted the power balance away from realism. Explicitly referring to Schor's thorough investigation, Toril Moi in her 2006 monograph on the Norwegian playwright Henrik Ibsen, often associated with nineteenth-century realist and naturalist interventions, argued that it is possible to propose a new interpretation of literary history where the main tension in the second half of the nineteenth century is that between idealism and modernism, thus making the presence of realism superfluous. In his turn, Fredric Jameson, who in the 2010s continuously carried out studies of the poetics of social form, also refers to Schor's monograph; however, his 2013 book *The Antinomies of Realism* (a title on which this article also capitalizes) already provides a striking attempt to re-address the complexity of realism as a literary and art form in nineteenth-century contexts and beyond. Jameson not only emphasizes the possibility and need to evaluate the full spectrum of classical realism with its narrative power and the ability to showcase the affects of human behavior (for him, the most important of realism's productive antinomies is that

between telling and showing) but also its potential continuation and impact in later periods of literary production.

Jameson's book is one important example among many that signal the interest in realism in the first quarter of the twenty-first century being on the rise. Having for a long time been looked upon with great suspicion, literary realism recently experiences a new wave of attempts in opening its innovations and challenges that reach well beyond the classical nineteenth-century period. These trends have even been described as a "new realist turn" (Esty and Lye 276). Significantly enough, both contextual and textual arguments are important here considering the opinion that realism responds to, investigates, questions, and reconstructs the complex ways human lives are lived in specific places and different conditions. In his introduction to the volume on literary realism, significantly titled *Reclaiming Realism*, Matthew Beaumont states that it is of vital importance for national as well as international scholarship to follow realism's different trajectories and contexts tracing its "cognitive as well as imaginative access to a material, historical reality" (Beaumont 2). Postcolonial Studies, Gender Studies, Trauma Studies, and New Historicism are among scholarly approaches to the concept of realism that contextualize its importance (Balaev). At the same time, textually oriented research provides an insight into realism's aesthetic complexities.

An important scholarly milestone is the two-volume investigation *Landscapes of Realism* recently published in the series of comparative literary histories in European languages. The first volume, *Mapping Realism*, already goes beyond the previously established limits and sets out for rereading literary realism "from both a European-language and a global comparative perspective" (Göttsche, Mucignat and Weninger 2). This is further explained in more detail:

> Rethinking realism as a transformative, multi-phased and multi-stranded literary-historical dynamic that started well before the programmatic realist movements of the mid-nineteenth century reveals that nineteenth-century realism is not the result of one defining invention or one discreet origin, but rather a cumulative effect of a range of interlinking developments in which inspired authors developed new ways of writing and, step by step, also a vocabulary for describing these innovations in terms that prepared ground for later realist theory. (104)

Importantly, this multi-layered development indicates a potential discrepancy between realist theory and practice: "It is obvious that realist literary practice was epistemologically more advanced than the emerging theory of realism with its baggage of philosophical idealism." (165)

221

Landscapes of Realism focuses on the multiplicity and diversity of realist efforts inviting further studies of various social and cultural contexts while at the same time suggesting a useful and widely conceived interpretative pattern.

Another important trend in contemporary research is provided by the concept of worlding realism (Goodlad), with an emphasis on realism's willingness to explore the relations between individual and the world, and looking for a particular aesthetic approach. This leads to the recognition that the world is in crisis, and studies of realism might be considered as helping to understand literature's potential interventions.

In recent years, Latvian scholarship also started to contribute to the interpretation of realism in local and comparative contexts. Among other attempts there is a new history of nineteenth-century Latvian literary culture (Daija and Kalnačs, *A New History*), studies of *fin-de-siècle* literary and cultural trends (Kalnačs et al., *Fin de siècle*), research on the nineteenth-century Latvian novel (Kalnačs, "The Genesis"), re-evaluation of the literary oeuvre of the turn-of-the-century Latvian author Rūdolfs Blaumanis (Blaumanis; Füllmann; Füllmann et al.; Kalnačs, *Pavērsiens*) as well as a comparative analysis of realism in nineteenth-century Baltic literatures (Kalnačs, "The Poliphony"). In this article, I build on these investigations to trace the role and interpretation of Latvian literary realism in two periods that for different reasons can be considered turning points in the reception history and have broader implications within international research: late nineteenth century as well as twentieth-century Soviet ideological rule.

The appearance and initial reception of nineteenth-century Latvian realism

The term 'realism,' first used by the philosopher Immanuel Kant in the late eighteenth century (Dierkes-Thrun 709), had during the first half of the nineteenth century already been well established in French literary scholarship. German authors and scholars made their own use of the term, conceiving it in a somewhat modified form of so-called poetic realism that preserved the ties with idealist aspirations. In turn, English and Russian theorists also had their take. These interlinked but at the same time distinctive traditions played a significant role in the understanding of the term in nineteenth-century Latvian culture.

Realism was first mentioned in Latvian literary space in an 1882 article written by Jānis Puriņš-Zvīgulis who at the time studied mathematics

at St. Petersburg University. The author argued that there is a crucial difference between sentimental and romantic poetics, on the one, and realist literature, delving into the representation of the everyday, on the other hand (Knope 107). However, this and other articles with their emphasis on realism's critical potential and demands for a direct involvement with topical social issues cover only a fraction of nineteenth-century realist aspirations that were significantly broader in scope. It is important to recognize that, contrary to what the late arrival of the term might suggest, realism in Latvian literary practice is of much earlier origin.

The nineteenth century was significant in the emergence of Latvian romantic nationalism that, relying on similar ideas and developments in other parts of Europe, brought with it ideas of the possibility of creating an elite culture (Ijabs). While opposing the Baltic German political hegemony, the Latvians at the same time were indebted to the main intellectual trends developed in the Baltic German cultural environment. From this perspective, the most important nineteenth-century move was marked by turning away from the so-called popular enlightenment with its aim to guide and advise the ethnically different and lower social class of Latvian peasants (Grudule; Daija, *Literary*). Instead, a conscious appropriation of the ideas of the most highly valued representatives of German Enlightenment as well as important later poets became a declared aim of the New Latvian movement that "created the contents for the concept of national identity, alongside language and literature, incorporating also an interpretation of national history and a positive self-image of the Latvian nation" (Mintaurs 107).

In an addition to these intellectually charged efforts, another significant trend was the realist turn toward the inclusion of everyday situations and personal experience. The formation of realism was a logical step in the development of nineteenth-century Latvian literature. Early traces of this process can be observed in the emergence of the Latvian language newspapers leading to the creation of the public space and the rise of secular literature (Šemeta). The gradual growth of self-consciousness, stimulated by the spread of national romanticism across different parts of Europe, encouraged ethnic Latvians to aspire to their agency, and the logical consequence of this development was a more careful attention to the living conditions of the population. Nineteenth-century Latvian realism relied on the experience received from various trends in German literature, including popular culture and so-called *Heimatliteratur*, roughly translatable

as homeland literature, as well as poetical realism. Those different strands were fused together by Latvian authors consciously searching to attract broader audiences while at the same time remaining faithful to their poetic goals. Realist practices were already well developed before the term realism was first used in criticism in the 1880s (Kalnačs, "Walking"). Even if putting emphasis on shorter forms such as novellas instead of the traditional impact of nineteenth-century novels (Buzard; Moretti; Watt), the efforts of Latvian authors matched other European literatures striving towards a complex representation of ordinary living conditions, with those not only providing mere background, but creating "a source of action in its own right and a main factor that causes, defines and helps" to form individual characters and to organize the plot (Pavel and Tihanov 46). The innovative potential of realist aesthetics also becomes evident in the tensions between idealism and realism in the Latvian press (Kalnačs, "Beauty"). This juxtaposition was somewhat later followed by a productive dialogue between realism and early modernism. Late nineteenth-century Latvian realism thus responded to the development of both society and culture, and manifested itself in constant interaction with other literary and art trends.

The main objections to realism from the early 1880s on came from two sides, those arguing for heightened requests regarding literary quality, on the one hand, as well as from critics with a staunch idealist persuasion, on the other. The latter ones had already made their point when objecting to the first novel by ethnic Latvian authors, the brothers Reinis and Matīss Kaudzīte, "Mērnieku laiki" ("The Surveyors' Times"), in 1879. As formulated by the first reviewers, their negative response was mainly due to missing ideals, the authors being accused of getting too sceptical in the depiction of their Latvian compatriots. One of the reviewers, whose opinion was closely echoed by other critics, resolutely stated that "the author of a novel always has an obligation to idealise life, to show it more beautiful and chaste than it in fact is" (Haralds 87). Another important topic of this kind of moral criticism was the representation of women that were still considered by many as mainly capable to carry out only domestic activities.

From the perspective of the quality demands, the contributions of Janis Jansons from the early 1890s are characteristic and important (Jansons). However, the critical stance of the supporters of realism was also characterized by considerable simplifications. One aspect of this was the link to the tradition of Russian literature as juxtaposed to the German impact as well as that of other European literatures. The early

theorists of Latvian realism were keen to turn away from the so-called poetical realism of German literature, bordering on aesthetic idealism, and look for Russian examples, in their eyes creating a true testimony of the existential hardships of the lower social classes. This early perspective with a particular emphasis on the importance of Russian literature turned out to be symptomatic for the later reception of realism during the Soviet era.

Altogether, however, in the late nineteenth century the decisive factors in Latvian literary culture were creative liveliness and aesthetic diversity. The vitality of late nineteenth-century realism benefited from ongoing discussions in literary circles. It was due to this fruitful exchange that realism was able to respond to the challenges of modernity and retained its role alongside modernist innovations. One testimony to this were the first novels dealing with the experience of the lower social classes in the city also emerging toward the end of the century, prominent among them Augusts Deglavs's *Zeltenīte* in 1896. Other authors such as Apsīšu Jēkabs, Rūdolfs Blaumanis, and Andrievs Niedra were scrupulous in their investigations of daily conditions in the countryside. Under such conditions, even the highly subjective and somewhat restricted views of Andrejs Upīts, one of the most ardent and polemical supporters of realism, were relevant as valuable contributions to literary debates (Upīts, *Latviešu*). The representation of individual experience and existential reflection were among the characteristic features of literature in the turn-of-the-century period.

This trend continued into the interwar decades following the establishment of an independent state in 1918, with prose writers Jānis Jaunsudrabiņš, Kārlis Zariņš, and Pāvils Rozītis being among those who insightfully reflected on the war experience and postwar conditions in Latvian society. One of the most important modernist poets of the 1920s and 1930s, Aleksandrs Čaks, opined that the uniqueness of an artwork can only be achieved by observing two main principles, the use of appropriate poetic means and close observation of the conditions of everyday life (Kalniņa 186). Realist trends thus continued to live alongside modernist aspirations in an atmosphere of constant innovation and aesthetic change.

During the subsequent period of Soviet communist rule, the climate of free exchange of opinions changed completely. By imposing strict ideological rules, realism was canonized, thus leading to an ideological distortion of its meaning. For decades to come, literary realism under the Soviet regime lost a great deal of its aesthetic appeal.

Literary realism within the constraints of Soviet ideology

The initial phase of Soviet rule in the aftermath of the 1940 occupation of Latvia was marked by an almost total neglect of earlier traditions. Gradual changes in the overall atmosphere only started to emerge after the death of the Soviet Communist party leader Iosif Stalin. In 1956, the first volume of a collection of articles of Latvian literary criticism comprising the period between 1874 and 1904 appeared. However, despite providing more complex insights into Latvian cultural history in comparison to the 1940s and early 1950s, the edition mirrored the ideological trends of the Soviet regime.

A crucial role in the implementation of strict ideological guidelines was played by the application of the term 'socialist realism' that referred to the ideological demands imposed by communist ideology. To grasp the process of implementation of the ideas of socialist realism we need to take a step back because the main ideological battles had already been fought in the early decades of the history of the Soviet Union. In the turbulent 1920s, avant-garde artists still had a considerable impact on different branches of culture, especially in cinema, theatre, and visual arts. However, the Soviet ideological conversion of the early 1930s with the declaration of socialist realism as the main principle of representation turned the initial climate of vital experimentation into its direct opposite, starting long-term dominance of ideologically pre-scribed art.

The rise of European realism was linked to the Enlightenment, when "for the first time the authority of metaphysical and divine came under challenge from a secular form of knowledge that claimed to reveal the truth of the material physical world" (Morris 3). The idea of socialist realism was a direct opposite suggesting an imaginary kind of revolutionary development of society. It fostered the opinion that it is "not enough to represent life as it is; it is necessary to show where it is going, and that is toward the inevitable future of the communist society. In short, what is introduced here is a quasi-philosophical, quasi-religious theological doctrine" (Becker 21–22).

Such an approach was not reality grounded, and there was never any clarity as to what exactly socialist realism means. This was a convenient situation for the authorities as they could turn against any artistic creation on the simple grounds of it not being adjusted to 'socialist realist' principles.

The ways the transfer of knowledge was supposed to work is exem-

plified by the First Congress of Soviet Latvian writers that took place in Riga in June, 1941, a year after the Soviet occupation, and a week before the intervention of Nazi Germany. Instead of earlier ways of acquiring knowledge through individual experience, the Soviet authors were requested to follow objectives declared by the guest delegates representing different republics with the whole process being orchestrated from Moscow, or the fourth Rome in Soviet leaders' self-perception (Clark).

In order to promote Soviet ideals, some well-known Latvian authors were mobilized and accepted the ascribed tasks. The most prominent example was that of Andrejs Upīts, born in 1877, who had been one of the most prolific Latvian writers and a staunch supporter of realism and naturalism throughout his literary career. In the introduction to his book *Reālisms literatūrā* (*Realism in Literature*) in 1951, Upīts declared that in the Soviet state realism had become the one and only possible aesthetic method (Upīts, *Reālisms* 3). However, a truthful depiction of different milieus, situations and characters, that was at the core of realist aesthetics, turned out to be incompatible with the ideological prescriptions given to the term 'socialist realism' by its founding fathers.

The historical rootedness in ideology was already characteristic of nineteenth-century Russian criticism that took literary realism as a weapon to promote social struggle (Göttsche, Mucignat and Weninger 151). Largely for this reason the tsarist rule considered realism as threatening the very foundations of class society. Paradoxically enough, while following along the same lines and indicating the representation of the battle of social classes as the primary target of realism, the Soviet regime overemphasized and instrumentalized its impact to a considerable extent stripping realism of its critical potential and aesthetic diversity.

In the orthodox take on the concept of socialist realism, it implied and even promoted the representation of everyday conditions, while at the same time this representation was never expected to be open to criticism and discussion. The long search for regaining art's potential of poetic complexity was twofold; it either meant turning away from 'realism' altogether, or implied an attempt to get rid of the label 'social-ist' voluntarily attached to it. In the post-Stalinist period, there were two principal avenues for an authentic expression of the artist. Either authors tried to look for alternatives to 'realism,' a trend blurred in their eyes with the socialist connotation; or they started to rub off the 'socialist' marker from realism and tried to come closer to the descrip-

tion of everyday experience. These processes are extremely rewarding to observe while literary practices gradually overcame ideological restrictions. The first post-Stalinist years brought some debates into the literary field which, however, remained rather wage (Gūtmane 223–262). The courage and ability to express bolder criticism was lacking as nobody knew exactly how far one would be allowed to go. From the second half of the 1950s onward, however, there was a slow but unstoppable move towards a more open approach in literary texts as well as in critical debates. In Soviet Latvia, the first steps of re-vitalizing culture were linked to the return of a more detailed display of individual experience, especially by young authors such as Visvaldis Lāms, Zigmunds Skujiņš, and Regīna Ezera born in the 1920s who started to enter the literary scene in the late 1950s.

One of the most characteristic cases was that of the prose writer Ēvalds Vilks. His example shows that even writers initially not fully antithetic to the Soviet regime faced insurmountable difficulties as soon as they started to express critical opinions. Vilks, born in 1923, worked for the Soviet press, and started to publish his short stories in the late 1950s. His choice of genre already marked an attempt to turn away from the officially promoted Soviet monumental novel, exemplified in Latvian literature by the efforts of Andrejs Upīts and Vilis Lācis that glorified the socialist regime. Instead, Vilks focused on the description of minor details that were hardly compatible with the heroic success narrative promoted by the Soviet authorities. These early efforts were already met with sharp criticism. The most crucial case became that of Vilks's short story *Divpadsmit kilometri* (*Twelve kilometres*, 1963). While working on this text, in a private letter Vilks stated: "In this story, I try to find an answer, how had it been possible for the personality cult to occur. This important question cannot be avoided." (Vilks 430) The plot is saturated by a meeting of three former classmates who in their conversations reflect on the last twenty years, thus bringing together the period of the first Soviet occupation with the early 1960s. The role of each of the three characters in earlier events, their potential guilt as well as the disillusionment about Soviet realities are the main points provided by Vilks's narrative. To an extent, the author still respected official policies of the period that allowed certain criticism of individual actors as long as the very principles of the Soviet regime were not questioned. After rounding up the first version of his text, as the manuscript shows, he added some more explicitly critical remarks before the story was printed in the literary magazine *Karogs*. This publication was met with severe criticism of the author apparently having

doubted the 'socialist' foundations of society. Vilks's prose did not fit the demands of the time exactly, while he was moving closer to more sincere and truthful 'realist' observations. For the subsequent printing in a short story collection in 1968, the author was forced to make several cuts that diminished the power of the message. This second publication of *Twelve Kilometres* almost directly coincided with the Soviet imperialist intervention in Prague.

Despite the significant setbacks and harsh criticism, Vilks's career path shows in a nutshell how approaches to Soviet reality gradually started to change. First, there was an adaptation to the existing conditions which in terms of Postcolonial Studies can be labelled mimicry. Secondly, some of the efforts displayed the phenomenon of hybridity, with more authentic elements already coming to the fore. And finally, in the discussion of the Soviet occupation and the restrictions imposed by the regime, the authors moved towards transculturation to express their thoughts more directly and freely, and from their own point of view. This was an attempt to revitalize realist aesthetics developed by early generations of Latvian writers, especially during the late nineteenth and early twentieth century. However, Vilks's case is symptomatic for the Soviet period as even a relatively modest attempt of 'realist' representation could not escape the claws of rigorous 'socialist' critique with the author coming under direct ideological pressure.

Conclusion: The temporalities of Latvian literary realism

The above timeline indicates that there have been several turns in understanding and practicing realism in Latvian literature and criticism. Nineteenth-century realist aspirations were closely related to more general European trends. Realist aesthetics gradually established itself in Latvian literature, reaching considerable maturity by the 1890s. The polemics of the period were productive and helped authors to sharpen their ideas and poetic tools. There was certainly an appeal of realist representations to the reading public that was preserved throughout the independence period of the 1920s and 1930s, when realist art could freely develop alongside other aesthetic trends.

The Soviet period, even if on the surface promoting realist representations, in fact imposed considerable restrictions on the implementation of realist aesthetics, for which the term 'socialist realism' was invented. While the socialist component implied the utopian nature of the project, attempts at more detailed and truthful representation

of the everyday were strictly censored. This ideological heritage left its mark on the literary scene at the turn of the twenty-first century when radical experiments tended to prevail, while realism was considered old-fashioned and outdated even from the perspective of literary history. Subsequently, a gradual re-evaluation of the possibilities of realist aesthetics took place in literary practice. This was also the case with literary scholarship providing new readings of realist texts and reformulating theoretical issues.

The twenty-first-century reception of the history of Latvian realism is inextricably linked to the heightened awareness of realism's potential in international scholarship. Certainly, a significant aspect is the demand for reality effects (Morris 97–118). The historical importance of realist authors is also determined by their willingness and effort to communicate with readers touching on the most painful experiences of human lives. The diversity and openness of contemporary criticism allows for a return to productive discussions of the particularities of realist aesthetics with contrasting opinions still sharing a common ground of recognizing the importance of the reality quest in literature.

In his book, *Literary Studies in Reconstruction,* Marko Juvan points toward the need to contextualize and locate every scholarly opinion:

> Due to various histories of literary theorists' personal socialization and their placement within the social-ideological heteroglossia, their metalanguages of literary phenomena are 'contaminated' with unconscious motives, shadows of cultural prejudices, ideas of the goals and purposes of scholarly activity, imaginary identifications, political opposition and so on. (Juvan 27)

All of that is certainly true regarding the study of literary realisms of different periods and cultures. It is also obvious that there are some general trends that cross the borders, bringing areas of research with an international relevance to the fore. One such phenomenon is the current spotlight on realism in literature.

WORKS CITED

Balaev, Michelle, editor. *Contemporary Approaches in Literary Trauma Theory*. Palgrave Macmillan, 2014.

Blaumanis, Rūdolfs. *Frost im Frühling*. Edited by Benedikts Kalnačs and Rolf Füllmann, Aisthesis, 2017.

Beaumont, Matthew. "Introduction: Reclaiming Realism." *A Concise Companion to Realism*, edited by Matthew Beaumont, Wiley-Blackwell, 2012, pp. 1–12.

Becker, George J. "Introduction: Modern Realism as a Literary Movement." *Documents of Modern Literary Realism*, edited by George J. Becker, Princeton University Press, 1963, pp. 3–38.

Buzard, James. *Disorienting Fiction: The Autoethnographic Work of Nineteenth-Century British Novels*. Princeton University Press, 2005.

Clark, Katerina. *Moscow, the Fourth Rome: Stalinism, Cosmopolitanism, and the Evolution of Soviet Culture, 1931–1941*. Harvard University Press, 2011.

Daija, Pauls. *Literary History and Popular Enlightenment in Latvian Culture*. Cambridge Scholars Publishing, 2017.

Daija, Pauls, and Benedikts Kalnačs, editors. *A New History of Latvian Literature: The Long Nineteenth Century*. Peter Lang, 2022.

Dierkes-Thrun, Petra. "Realism." *The Fin-de-Siècle World*, edited by Michael Saler, Routledge, 2015, pp. 706–718.

Esty, Jed, and Colleen Lye. "Peripheral Realisms Now." *Modern Language Quarterly*, vol. 73, no. 3, 2012, pp. 269–288.

Füllmann, Rolf. "*Andrikson* (lett. *Andriksons*) von Rūdolfs Blaumanis: eine transkulturelle Dorfnovelle und ihre didaktische Relevanz. Der zweisprachige Nationalautor Blaumanis im transkulturellen Raum des Baltikums der Jahrhundertwende." *Ost-Geschichten. Das östliche Mitteleuropa als Ort und Gegenstand interkultureller literarischer Lernprozesse*, edited by Mark-Oliver Carl, Sieglinde Grimm and Nathalie Kónya-Jobs, Vandenhoeck und Ruprecht, 2021, pp. 83–101.

Füllmann, Rolf, Antje Johanning-Radžienė, Benedikts Kalnačs and Heinrich Kaulen, editors. *Rūdolfs Blaumanis (1863–1908). Lettische Moderne und deutschsprachige Literatur*. Aisthesis, 2019.

Goodlad, Lauren. "Introduction: Worlding Realisms Now." *NOVEL: A Forum on Fiction*, vol. 49, no. 2, 2016, pp. 183–201.

Göttsche, Dirk, Rosa Mucignat and Robert Weninger, editors. *Landscapes of Realism: Rethinking Literary Realism in Comparative Perspectives. Vol. I, Mapping Realism*. John Benjamins, 2021.

Grudule, Māra. "Zemnieku jeb tautas apgaismība Latvijā – jaunlatviešu kustības avots." *Meklējumi un atradumi*, edited by Inguna Daukste-Silasproģe, Zinātne, 2005.

Gūtmane, Zanda. *Totalitārisma traumu izpausmes Baltijas prozā*. LU Literatūras, folkloras un mākslas institūts, 2019.

Haralds [Vensku Edvarts]. "'Mērnieku laiki.' Romāns no Kaudzītes Matīsa un Reiņa." *Latviešu literatūras kritika*, vol. 1, edited by Arvīds Grigulis and Vilis Austrums, Latvijas Valsts izdevniecība, 1956, pp. 85–90.

Ijabs, Ivars. "Another Baltic Postcolonialism: Young Latvians, Baltic Germans, and the Emergence of Latvian National Movement." *Nationalities Papers: The Journal of Nationalism and Ethnicity*, vol. 42, no. 1, 2013, pp. 88–107.

Jameson, Fredric. *The Antinomies of Realism*. Verso, 2013.

Jansons, Janis. *Literatūra un laikmets*. Liesma, 1972.

Juvan, Marko. *Literary Studies in Reconstruction: An Introduction to Literature*. Peter Lang, 2011.

Kalnačs, Benedikts. "Beauty, Truth and Form. Idealism and its Others in Nineteenth-Century Latvian Literary Debates." *CompLit. Journal of European Literature, Arts and Society*, vol. 2, no. 4, 2022, pp. 15–35.

Kalnačs, Benedikts. "The Polyphony of Late Nineteenth-century Baltic Realism." *Landscapes of Realism: Rethinking Literary Realism in Comparative Perspectives. Vol. I, Mapping Realism*, edited by Dirk Göttsche, Rosa Mucignat and Robert Weninger, John Benjamins, 2021, pp. 577–595.

Kalnačs, Benedikts, Pauls Daija, Eva Eglāja-Kristsone and Kārlis Vērdiņš. *Fin-de-siècle literārā kultūra Latvijā*. LU Literatūras, folkloras un mākslas institūts, 2017.

Kalnačs, Benedikts. "The Genesis, Ideology and Poetics of the Nineteenth-Century Latvian Novel." *A New History of Latvian Literature: The Long Nineteenth Century*, edited by Pauls Daija and Benedikts Kalnačs, Peter Lang, 2022, pp. 149–169.

Kalnačs, Benedikts. *Pavērsiens. Rūdolfs Blaumanis latviešu un Eiropas literatūrā*. LU Literatūras, folkloras un mākslas institūts, 2022.

Kalnačs, Benedikts. "Walking Through the Text: The Representation of Mobility in Late Nineteenth-Century Latvian Fiction." *World Literature Studies*, vol. 15, no. 3, 2023, pp. 89–98.

Kalniņa, Ieva E. "Gastropoētiskie tēli Aleksandra Čaka dzejoļu krājumos." *Ēdiena reprezentācijas kultūrā*, edited by Ieva Kalniņa, LU Akadēmiskais apgāds, 2023, pp. 184–205.

Knope, Elza. *Latviešu literatūras kritika 19. gs. otrajā pusē*. Latvijas PSR Zinātņu akadēmija, 1962.

Mintaurs, Mārtiņš. "The New Latvians: Strategies for the Creation of a National Identity." *A New History of Latvian Literature: The Long Nineteenth Century*, edited by Pauls Daija and Benedikts Kalnačs, Peter Lang, 2022, pp. 105–122.

Moi, Toril. *Henrik Ibsen and the Birth of Modernism: Art, Theater, Philosophy*. Oxford University Press, 2006.

Moretti, Franco. *Atlas of the European Novel 1800–1900*. Verso, 1998.

Morris, Pam. *Realism*. Routledge, 2003.

Pavel, Thomas, and Galin Tihanov. "What is realism? Ideas and debates." *Landscapes of Realism: Rethinking Literary Realism in Comparative Perspectives. Vol. I, Mapping Realism*, edited by Dirk Göttsche, Rosa Mucignat and Robert Weninger, John Benjamins, 2021, pp. 29–64.

Schor, Naomi. *George Sand and Idealism*. Columbia University Press, 1993.

Šemeta, Aiga. "The Media History of the Early Periodical Press in Latvian." *A New History of Latvian Literature: The Long Nineteenth Century*, edited by Pauls Daija and Benedikts Kalnačs, Peter Lang, 2022, pp. 49–65.

Upīts, Andrejs. *Latviešu jaunākās rakstniecības vēsture*. Golts, 1911.

Upīts, Andrejs. *Reālisms literatūrā*. Latvijas Valsts izdevniecība, 1951.

Vilks, Ēvalds. *Kopoti raksti*. Vol. 2, edited by Harijs Hiršs, Liesma, 1982.

Watt, Ian. *The Rise of the Novel: Studies in Defoe, Richardson and Fielding*. University of California Press, 1957.

Antinomije latvijskega literarnega realizma

Ključne besede: latvijska književnost / literarna kritika / realizem / socialistični realizem / sovjetski režim

Razprava preučuje družbene in literarne kontekste latvijskega literarnega realizma. Latvijski realizem se je med svojim vzponom v 19. stoletju razvijal v skladu s splošnimi smernicami evropske kulture, ki so pripomogle k estetski raznolikosti in inovativnemu potencialu. Tem procesom je posvečen prvi del članka. Nato so podrobneje obravnavani poznejše preoblikovanje in izkrivljanje pojma realizma v obdobju sovjetske okupacije. Raziskava sledi trenutnemu porastu zanimanja za realizem, upošteva njegov pomen v širokem spektru kontekstov in pokaže, da je realizem inovativna estetska praksa, ki ima velik zgodovinski, pa tudi sodobni pomen. Vendar so ga predstavniki drugih estetskih smeri pogosto napačno razumeli ali zavestno napačno interpretirali, prav tako pa so ga zlorabljali politični režimi. Članek opozarja na nujnost ponovne zdrave interpretacije realističnih teženj ter njihovih družbenih in estetskih prispevkov.

NAVODILA ZA AVTORJE

Primerjalna književnost objavlja izvirne razprave s področij primerjalne književnosti, literarne teorije, metodologije literarne vede, literarne estetike in drugih strok, ki obravnavajo literaturo in njene kontekste. Zaželeni so tudi meddisciplinarni pristopi. Revija objavlja prispevke v slovenščini ali angleščini, izjemoma tudi v drugih jezikih. Vsi članki so recenzirani.

Prispevke pošiljajte na naslov: marijan.dovic@zrc-sazu.si.

Razprave, urejene v programu Word, naj **ne presegajo 50.000 znakov** (vključno s presledki, sinopsisom, ključnimi besedami, z opombami, bibliografijo in daljšim povzetkom). Besedilo naj bo v pisavi Times New Roman, 12 pik, enojni razmik. Drugi prispevki – poročila, recenzije ipd. – lahko obsegajo največ 20.000 znakov (vključno s presledki).

Naslovu razprave naj sledijo **ime in priimek, institucija, poštni naslov, država, ORCID iD** in **e-naslov** avtorja oziroma avtorice.

Razprave imajo slovenski **povzetek** (1.000–1.500 znakov) in **ključne besede** (5–8), oboje naj bo v *kurzivi* tik pred besedilom razprave. **Angleški prevod** povzetka (preveden naj bo tudi naslov razprave) in ključnih besed je postavljen na konec besedila (za bibliografijo).

Glavni tekst je obojestransko poravnan; lahko je razčlenjen na poglavja s podnaslovi (brez številčenja). Med odstavkoma ni prazne vrstice, prva beseda v novem odstavku pa je umaknjena v desno za 0,5 cm (razen na začetkih poglavij, za citati in za ilustracijami).

Sprotne opombe so oštevilčene tekoče (arabske številke so levostično za besedo ali ločilom). Količina in obseg posameznih opomb naj bosta smiselno omejena. Bibliografskih referenc ne navajamo v opombah, temveč v kazalkah v sobesedilu neposredno za citatom oziroma povzetkom bibliografske enote.

Kazalka, ki sledi citatu ali povzetku, v okroglih oklepajih prinaša avtorjev priimek in številko citirane ali povzete strani: (Juvan 42). Kadar avtorja citata navedemo že v sobesedilu, v oklepaju na koncu citata zapišemo samo številko citirane ali povzete strani (42). Če v članku navajamo več enot istega avtorja, vsako enoto po citatu oziroma povzetku v kazalki označimo s skrajšanim naslovom: (Juvan, *Literary* 42).

Citati v besedilu so označeni z dvojnimi narekovaji (» in «), citati v citatih pa z enojnimi (' in '); izpusti iz citatov in prilagoditve so označeni z oglatimi oklepaji. Daljši citati (štiri vrstice ali več) so izločeni v samostojne odstavke brez narekovajev; celoten citat je zamaknjen desno za 0,5 cm, njegova velikost je 10 pik (namesto 12), nad in pod njim pa je prazna vrstica. Vir citata je označen v oklepaju na koncu citata.

Ilustracije (slike, zemljevidi, tabele) so priložene v ločenih datotekah z minimalno resolucijo 300 dpi. Objavljene so v črno-beli tehniki. Položaj ilustracije naj bo označen v glavnem tekstu (Slika 1: [Podnapis 1]). Avtorji morajo urediti tudi avtorske pravice, če je to potrebno.

V **bibliografiji** na koncu članka so podatki izpisani po standardih MLA:

– članki v periodičnih publikacijah:
Kos, Janko. »Novi pogledi na tipologijo pripovedovalca«. *Primerjalna književnost*, let. 21, št. 1, 1998, str. 1–20.

– monografije:
Juvan, Marko. *Literary Studies in Reconstruction. An Introduction to Literature.* Peter Lang, 2011.

* Mesto izdaje se pred založnikom navaja zgolj, če je bila knjiga izdana pred letom 1900, če ima založnik sedež v večih državah, ali če založnik ni splošno znan.

– zborniki:
Leerssen, Joep, in Ann Rigney, ur. Commemorating Writers in Nineteenth-Century Europe. Palgrave Macmillan, 2014.

– poglavja v zbornikih:
Novak, Boris A. »Odmevi trubadurskega kulta ljubezni pri Prešernu«. *France Prešeren – kultura – Evropa*, ur. Jože Faganel in Darko Dolinar, Založba ZRC, 2002, str. 15–47.

– članek v spletni reviji:
Terian, Andrei. »National Literature, World Literatures, and Universality in Romanian Cultural Criticism 1867–1947«. CLCWeb: Comparative Literature and Culture, let. 15, št. 5, 2013, https://doi.org/10.7771/1481-4374.2344. Dostop 21. 5. 2015.

– drugi spletni viri:
McGann, Jerome. »The Rationale of HyperText«. http://www2.iath.virginia.edu/public/jjm2f/rationale.html. Dostop 24. 9. 2015.

Za vse ostale primere glej *MLA Handbook*, deveta izdaja.

GUIDELINES FOR AUTHORS

Primerjalna književnost (Comparative Literature) publishes original articles in comparative literature, literary theory, literary methodology, literary aesthetics, and other fields devoted to literature and its contexts. Multidisciplinary approaches are also welcome. The journal publishes articles in either Slovenian or (American) English, and occasionally also in other languages. All published papers are peer-reviewed.

Articles should be submitted via e-mail: marijan.dovic@zrc-sazu.si.

Articles should be written in Word for Windows, Times New Roman 12, single-spaced, and **not longer than 50,000 characters** (including spaces, abstract, keywords, and bibliography).

The full title of the paper is followed by the author's **name**, **institution**, **address**, **country**, **ORCID iD**, and **e-mail address**.

Articles must have an **abstract** (1,000–1,500 characters, in *italics*) and **keywords** (five to eight), both set directly before the main text.

The **main text** has full justified alignment (straight left and right margins) and may be divided into sections with unnumbered subheadings. There are no blank lines between paragraphs. Each paragraph begins with a first-line indent of 0.5 cm (except at the beginning of a section, after a block quotation, or after a figure).

Footnotes are numbered (Arabic numerals follow a word or a punctuation directly, without spacing). They should be used to a limited extent. Footnotes do not contain bibliographical references because all bibliographical references are given in the text directly after a citation or a mention of a given bibliographical unit.

Each **bibliographical reference** is composed of parentheses containing the author's surname and the number of the page cited: (Juvan 42). If the author is already mentioned in the accompanying text, the parenthetical reference contains only the page number (42). If the article refers to more than one text by a given author, each reference includes a shortened version of the cited text: (Juvan, *Literary* 42).

Quotations within the text are in double quotation marks (" and "); quotations within quotations are in single quotation marks (' and '). Omissions are marked with ellipses (. . .) with no brackets, and adaptations are in square brackets ([and]). Block quotations (four lines or longer) have a left and right indent of 0.5 cm, are set in Times New Roman 10 (not 12), and are preceded and followed by a blank line.

Illustrations (images, maps, tables, etc.) should be provided in separate files at a minimum resolution of 300 dpi. They are published in black and white. The preferred positioning of illustrations is marked in the main text (Figure 1: [Caption 1]).

The bibliography at the end of the article follows the MLA style guide:

– Journal articles:
Kos, Janko. "Novi pogledi na tipologijo pripovedovalca." *Primerjalna književnost*, vol. 21, no. 1, 1998, pp. 1–20.

– Books:
Juvan, Marko. *Literary Studies in Reconstruction: An Introduction to Literature*. Peter Lang, 2011.

* The City of Publication should be given before the Publisher only if the book was published before 1900, if the publisher has offices in more than one country, or if the publisher is generally unknown.

– Edited volumes:
Leerssen, Joep, and Ann Rigney, editors. *Commemorating Writers in Nineteenth-Century Europe*. Palgrave Macmillan, 2014.

– Chapters in edited volumes:
Novak, Boris A. "Odmevi trubadurskega kulta ljubezni pri Prešernu." *France Prešeren—kultura—Evropa*, edited by Jože Faganel and Darko Dolinar, Založba ZRC, 2002, pp. 15–47.

– Articles in e-journals:
Terian, Andrei. "National Literature, World Literatures, and Universality in Romanian Cultural Criticism 1867–1947." *CLCWeb: Comparative Literature and Culture, vol.* 15, no. 5, 2013, https://doi.org/10.7771/1481-4374.2344. Accessed 21 May 2015.

– Other digital sources:
McGann, Jerome. "The Rationale of HyperText." http://www2.iath.virginia.edu/public/jjm2f/rationale.html. Accessed 24 Sept. 2015.

For issues not covered here, please refer to the *MLA Handbook*, 9th ed.