

Креирање и анализа корпуса текстова југословенских рок песама од 1967-2003.¹

УДК 811.163.41'322

САЖЕТАК: У раду се са теоријског и практичног аспекта анализира процес образовања и обраде корпуса текстова југословенских рок песама од 1967-2003. За преузимање грађе и XML аотирање коришћене су библиотеке lyricsmaster и yattag у језику Python. Корпус је прошао фазу препроцесирања, а XSL трансформацијом генерисани су основни статистички подаци. У апликацијама Слово Мајстор и LeXimir спроведена је рестаурација дијакритика (а у другој апликацији и фреквенцијска анализа). Проналажење друштвено-политичких тема вршено је у софтверу Unitex, док су преовлађујуће теме визуализоване у TreeCloud апликацији.

КЉУЧНЕ РЕЧИ: корпусна лингвистика, југословенски рокенрол, гребање веба, обрада природних језика, копање по тексту.

РАД ПРИМЉЕН: 15. април 2019.

РАД ПРИХВАЋЕН: 19. јуни 2019.

Људмила Петковић

ljudmila.petkovic@gmail.com

Универзитет у Београду

Београд, Србија

1. Увод

1.1 Феномен југословенског рокенрола

Музички аналитичари, социолози и антрополози се једногласно слажу у оцени да је југословенски рокенрол (познатији и као *Уи рок*)

¹ Рад је проистекао из мастер тезе „Креирање и анализа корпуса текстова југословенских рок песама у периоду 1945-2003“ која је одбрањена на Универзитету у Београду дана 18.3.2019. Израдом тезе је руководила проф. др Ранка Станковић која је допринела формулацији теме, уз напомену да је 1945. година у овом научном раду замењена 1967. годином.

оставио дубоког трага на просторима бивше Југославије. У питању је музички стил који је зачет 1961. године, са појавом група Урагани, Бијеле Стријеле и Силуете, а током исте деценије основани су и састави Црвени Кораљи (1962), Златни Дечази (1962) и Корни Група (1968) (Janjatović, 1998). Ју рок се развијао упоредо са процватом британске *beat* сцене (Раковић, 2011), коју су прославили бендови The Beatles и The Rolling Stones (Cooper and Cooper, 1993). Крајем педесетих и почетком шездесетих година, рокенрол се у Југославији поистовећивао са рокенрол и твист плесом, а не са музичким жанром (Раковић, 2011).

Од 1963. године, рокенрол у СФРЈ стиче статус жанра, који се од тада изводи првенствено помоћу електричних гитара, електричних бас гитара и бубњева (Раковић, 2011). Класични рок и рокабили су представљали веома популарне музичке форме међу Југословенима, што потврђују и обраде песама Елвиса Преслија, Бадија Холија, Чака Берија и др. (Арсенијевић et al., 2016). Седамдесете године су донеле уплив хипи покрета, уз додатно жанровско раслојавање и појаву хард-рока, прогресивног рока, арт-рока и других поджанрова. Крај седамдесетих и почетак осамдесетих година обележили су панк и нови талас, по узору на англоамеричке изворне варијанте (Арсенијевић et al., 2016). Заправо, југо-рок је пронашао своје место у друштву које је са одушевљењем прихватило производе западне културе, како посредством домаћих часописа (нпр. *Rock* и *Džuboks*) и страних радио емисија емитованих на Радију Луксембург, војним и пиратским станицама, тако и у виду филмова из земаља Запада. Такође су се пратили и западњачки модни трендови који су диктирали, између осталог, ношење дуге косе код младића или мини-сукања код девојака (Раковић, 2018).

Ипак, оно по чему се Ју рок посебно истицао били су слободарски текстови, посебно карактеристични за панк и новоталасну музику. Наиме, садржај текстова је неретко био политички ангажован, ироничан или, пак, вулгаран, па је самим тим и третиран као неподобан за тадашње југословенске друштвене прилике (Гајић, 2018). Доказ да није реч о изолованој појави представља и систематизација случајева (ауто)цензуре у Ју року.² Са друге стране, бројне песме тога доба опевале су отаџбину, док су у неким биле присутне и партизанске или бригадирске конотације (Божиловић, 2016). Божиловић додаје да срж Ју рока чини бунт и борба рок извођача за демократију, која је преточена у неконвенционалне, једноставне, непосредне и

² Видети веб страну [Balkanrock](#).

импровизоване музичке форме и текстове. Међутим, Ју рок, попут свог западњачког еквивалента, изнедрио је и уметнике који су личне животне ставове сублимирали у текстове песама са лирским, филозофским и интроспективним тоном (нпр. групе Азра, Идоли или Екатарина Велика).

1.2 Сродна истраживања

Према сазнањима аутора, студије југословенског рокенрола су се до сада заснивале превасходно на теоретским разматрањима, без примене рачунарских технологија (неки од таквих радова наведени су у одељку 1.1). [Zörnig et al. \(2016\)](#) спровели су једино истраживање које се бавило квантитативном анализом текстова песама из ере југословенског рокенрола. Део корпуса у наведеном раду чине текстови песама састава Рибља Чорба и његовог фронтмена, Боре Ђорђевића. У истом раду изложен је поступак рачунања фреквенција речи и лексичког варијабилитета коришћењем првенствено метрика *relative repeat rate* и *h-point*, што је омогућило класификацију корпусних текстова помоћу мултиваријационе анализе.

Са друге стране, аутоматско преузимање текстова песама са веба, њихова електронска обрада и анализа добијају све више пажње, о чему сведочи и велики број пројеката и радова. Постоји велики број корпуса текстова страних песама, нпр. [The Million Song Dataset³ \(Bertin-Mahieux et al., 2011\)](#), који се могу користити у истраживањима рачунарске анализе текста. До сада је вршено генерисање листа најчесталијих речи у аотираним корпусима текстова поп ([Kreyer and Mukherjee, 2007](#)) и рок песама ([Falk, 2013](#)), а спроведено је и моделовање тема у либрету пекиншке опере ([Zhang et al., 2017](#)) и корпусу текстова нумера прикупљених са сајта [SongMeanings \(Lukic, s.d.\)](#).

Када је реч о текстовима песама из домена рока или сродних жанрова, рачунарска анализа њиховог садржаја представља перспективну и изузетно развијену научну праксу. Неки занимљиви резултати које су ове технике произвеле јесу екстракција структуре текстова песама групе [The Beatles \(Mahedero et al., 2005\)](#) или смештање текстова песама Пола Макартнија и Џона Ленона на график *емоционалног сата* (енгл. *emotion clock*, у раду ([Whissell, 1996](#))). Истраживање које су спровели [Petrie et al. \(2008\)](#) је још један пример

³ [The Million Song Dataset](#) (на вебу).

интересовања за рачунарску обраду текстова Битлса. Ово истраживање бави се променом доминантног расположења у текстовима њихових песама, а коришћењем стилometriјске анализе открива стилистичке сличности и разлике између Ленона, Макартнија и Џорџа Харисона, као текстописаца.

Falk (2013) доноси резултате дијахронијске анализе лингвистичких особености у корпусу текстова страних рок песама од 1950. закључно са 1999. годином, на основу најфреквентнијих речи коришћених у свакој деценији. Haslam (2017) прати еволуцију тематских мотива у песмама кантаутора Леонарда Коена, што је визуализовано у виду *облака речи* (енгл. word cloud). Са становишта корпусне лингвистике, Taina et al. (2014) истражују језичке чиниоце по којима се разликују текстови песама из различитих хеви-метал поджанрова. Алата који се користе у психометријским истраживањима ради рачунања повезаности текста (Coh-Metrix)⁴ и откривања емоционалних, когнитивних и структуралних особина присутних у тексту (Linguistic Inquiry and Word Count)⁵ примењени су у покушају упоредне анализе текстова у рок, фолк, кантри, панк и гранџ песмама (Lightman et al., 2007). Циљ тог истраживања јесте откривање разлика у стилу писања извођача који су извршили самоубиство, и оних који нису били суицидни. Тематски удаљенији, али методолошки близак овом истраживању је рад који се бави техникама аутоматског проналажења полилексемских јединица у текстовима древних религиозних хинду поема помоћу *локалних граматика* (енгл. local grammars) у софтверу Unitex⁶ (Stein, 2012).

По свему судећи, анализа текстова југословенских рок песама са становишта рачунарске лингвистике чини се као недовољно развијена област истраживања. Из тог разлога, овај рад настоји да сагледа Ју рок песме кроз призму интердисциплинарности, како би се постојећа тумачења жанра поткрепила резултатима произведеним помоћу рачунарских алата за обраду корпуса.

1.3 Теоријско-методолошки оквир рада

Једно од тежишта рада је на примени техника корпусне лингвистике. Корпусна лингвистика представља развијену научну методологију, док је многи стручњаци сматрају и дисциплином, теоријом, парадигмом и

⁴ Coh-Metrix (на вебу).

⁵ Linguistic Inquiry and Word Count (скр. LIWC) (на вебу).

⁶ Unitex/GramLab (на вебу).

алатом (Taylor, 2008). Овај методолошки приступ бави се руковањем структурираним, машински читљивим и наменски бираним текстовима, који представљају основу за анализу различитих аспеката језика и његове употребе. У општем контексту корпусне лингвистике, „наменски бирани текстови” подразумевају колекцију одређених текстуалних јединица које су узорковане ради остваривања репрезентативности корпуса. Учесталост коришћења одређене речи или фразе, њихово проналажење у контексту кроз генерисање *конкорданце* (енгл. concordance), или екстракција метаподатака (нпр. имена аутора текста, датума објављивања, језика на коме је текст написан и сл.) из анотираног корпуса само су неки од могућих задатака корпусне лингвистике (McEnery and Hardie, 2012).

Корпусна грађа текстова Yu рок песама прикупљена је применом технике *гребања веба* (енгл. web scraping).⁷ Алгоритми за аутоматско прикупљање, препроцесирање и аотирање корпуса у складу са XML синтаксом имплементирани су у програмском језику Python, коришћењем библиотека lyricsmaster, xml.sax.utils и yattag. XSLT трансформација XML документа у XHTML формат омогућава преглед статистичких података о корпусу.⁸ Аутоматска рестаурација дијакритика вршена је у апликацијама Слово Мајстор⁹ и LeXimir.¹⁰ Методе корпусне лингвистике и обраде природних језика примењене су ради фреквенцијске анализе токена и колокација коришћењем LeXimir-а, а у софтверу Unitex конструисан је *коначни аутомат* (енгл. Finite State Automation) за екстракцију друштвено-политичких и културолошких тема. Употребљен је и алат TreeCloud (Gambette and Véronis, 2009) за визуализацију преовлађујућих тема у корпусу, у оквиру технике *копања по тексту* (Кешел and Шипка, 2008) или *ископавања из текста*¹¹ (енгл. text mining), која се заснива на откривању текстуалних образаца у неструктурираним подацима.

⁷ Наведени енглески термин је стандардизован у иностраној литератури, за разлику од чланака и радова објављених на овим просторима, у којима се срећу бројни покушаји јединственог терминолошког одређивања (нпр. *налажење података на вебу*, *стругање података* или *гребање веба*, да наведемо само неке). Поменути термини потичу из веб чланака „*Nalaženje podataka na internetu*” и „*SEO optimizacija kroz rečnik najbitnijih termina.*”

⁸ Кодови су доступни на [GitHub репозиторијуму](#) аутора (на вебу).

⁹ Слово Мајстор (на вебу).

¹⁰ LeXimir (на вебу).

¹¹ Термин преузет из [презентације](#) проф. др Цветане Крстев (на вебу).

Рад се састоји од пет одељака. У уводном делу размотрене су основне одлике Yи рока, релевантни радови на тему рачунарске обраде текстова домаћих и страних песама, као и методе истраживања коришћене у овом раду. Одељак 2 посвећен је опису поступка аутоматског прикупљања грађе за корпус Yи рок песама. У одељку 3 представљен је поступак полуаутоматског¹² препроцесирања и аотирања корпуса. Конкретни резултати примене корпуса изложени су у одељку 4, док одељак 5 доноси закључне коментаре и смернице за будући рад.

2. Прикупљање грађе методом *гребања веба*

Уже подручје истраживања представљеног у овом раду јесте рачунарска израда и анализа корпуса текстова југословенских рок песама, насталих у доба двеју југословенских држава – Социјалистичке Федеративне Републике Југославије (1945-1992) и Савезне Републике Југославије (1992-2003). Што се тиче критеријума прикупљања грађе, корпус чине текстови нумера на некадашњем српскохрватском језику, који је стандардизован Бечким књижевним договором 1850. године, да би се са распадом Југославије 1991. разложио на српски, хрватски и босански језик (Hentschel, 2003). У складу са тим, претраживали смо текстове песама на поменутих трима језицима, док су текстови песама на другим језицима који су били у употреби у Југославији – македонском, словеначком и језицима мањина – изостављени.

Када је реч о методи гребања веба, она се односи на употребу одређене рачунарске алатке за директно преузимање дигитализованог садржаја. Суштина ове праксе може се описати као аутоматизација свеобухватног и релативно брзог екстраховања и складиштења података са веба. Гребање веба намеће се као далеко ефикасније решење у односу на методе ручног копирања и чувања сваке информационе јединице понаособ. Осим тога, наведена техника може представљати и једино могуће решење за екстракцију података у случају када су опције „copy/paste“ онемогућене у веб читачу.

¹² Под полуаутоматским поступком подразумева се комбинација техника аутоматског и ручног препроцесирања и аотирања корпуса.

2.1 Опис извора података: сајт LyricWiki

LyricWiki¹³ представља комерцијални музички веб сајт који складишти текстове песама домаћих и страних извођача. Сајт је претражив према имену извођача, албума, песама, жанрова и продуцентских кућа. Доступне су и разне листе, попут листа најпопуларнијих албума из одређене године према мишљењу уредника угледних музичких часописа, и листа албума са музиком из филмова, као и обиље других текстуалних садржаја у области музике. Како је наведено у опису овог сајта, LyricWiki је јавно доступан и има дозволу за објављивање поузданих текстова песама.

База података сајта складишти више од 2.054.289 текстова (податак од 15.06.2019).¹⁴ Будући да LyricWiki садржи релативно велики број песама на српском и некадашњем српскохрватском језику, у почетној фази истраживања (док је сајт још увек био отворен за уређивање) преузимани су текстови у извођењу одабраних музичара из бивше Југославије.¹⁵ Разлог више за одабир овог сајта за гребање текстуалних података јесте чињеница да се са њега могу преузимати текстови преко наменског API-ја, тј. модула LyricWiki библиотеке lyricsmaster у језику Python, о којој ће бити више речи у наредном одељку.

2.2 Функционалност библиотеке lyricsmaster у језику Python

Аутоматско преузимање текстова песама са интернета представља ефикасну и релативно често коришћену методу која претходи електронској анализи корпуса. Један од репрезентативних примера такве праксе јесте употреба програмске библиотеке lyricsmaster, која је доступна на веб страни PyPI репозиторијума¹⁶ програмских пакета за рад у језику Python. Помоћу наведеног API-ја могу се прикупити текстови песама са неких од најпопуларних музичких сајтова (у даљем тексту: *провајдера*), као што су AZLyrics, Genius и др. На истој страни демонстрирано је коришћење наведеног алата да би се директно преузели и сачували текстови песама америчког репера Тупака Шакура, који су доступни на сајту LyricWiki. Ради анонимизације IP адресе корисника, постоји опција гребања сајта преко Tor Proxy Server-а.¹⁷

¹³ LyricWiki (на вебу).

¹⁴ Statistics (на вебу).

¹⁵ Списак извођача почиње од веб стране Language/Serbian.

¹⁶ lyricsmaster 2.8.1 (на вебу).

¹⁷ Tor Project (на вебу).

Имплементација поменути библиотеке резултовала је аутоматским екстраховањем текстуалног садржаја са различитих страница сајта LyricWiki, при чему је свака од тих страна посвећена неком од популарних југословенских извођача чији текстови песама су преузимани. Наиме, дефинисаном функцијом за гребанње веба и коришћењем методе `get_lyrics()` класе `lyricsmaster.providers.LyricWiki(tor_controller=None)`¹⁸ проналажене су жељене странице, са којих су прикупљани једино текстови песама, а не и други садржаји на истим странама (нпр. подаци у заглављу или на бочној траци веб стране). Конструисан је алгоритам који је са тих страна преузимао садржаје текстова према именима извођача, и то на следећи начин:

1. Најпре је за провајдера текстова одабран сајт LyricWiki;
2. Затим је променљивом `izvodjaci` дефинисана листа од тридесет извођача чији се текстови песама прикупљају. Имена извођача референцирамо тачно онако како су наведена на сајту (нпр. уместо ‘Yu-Grupa’, ‘yu grupa’, и сл., једино је дозвољен унос ниске ‘YU Grupa’). Ово се односи и на случајеве у којима два извођача деле исти назив (нпр. српска и финска група Negative). Уредници LyricWiki-ја су направили разлику између наведених група тако што су српском саставу додали одговарајућу ознаку „RS” за земљу порекла, како би се знало да је реч о бенду из Републике Србије, па је у код за гребанње веба унет модификован назив `Negative (RS)`;¹⁹
3. Креирана је функција `korpus()` чији аргументи су елементи наведене листе извођача, и за сваког од њих је покушано преузимање текстова помоћу методе `get_lyrics()`;
4. Потом су формирани објекти `discography`, `album` и `song`, док су увођењем параметара `title` и `lyrics` добијени наслови албума (`album.title`), песама (`song.title`) и сами текстови песама (`song.lyrics`);
5. Након тога, преузети корпусни материјал чуван је на локалној меморији рачунара методом `save()`, која је примењена на објекте `discography`, `album` и `song`.²⁰ Подразумевана апсолутна путања јесте `{user}/Documents/LyricsMaster/`, а сами текстови су сачувани у директоријуму `/izvodjac/album/` у формату `pesma.txt`.

¹⁸ `get_lyrics` је главна метода наведене класе (из документације на вебу).

¹⁹ Финска група није добила никакву ознаку, већ само назив „Negative“.

²⁰ `save()` је метода трију класа: `lyricsmaster.models.Discography`, `lyricsmaster.models.Album` и `lyricsmaster.models.Song`.

На LyricWiki странама са текстовима песама одређених извођача (нпр. YU групе) већина наслова била је форматирана у виду хипервеза које упућују на стране са доступним текстовима. Ипак, за неколицину наслова уопште нису били креирани текстуални садржаји (нпр. за песму „Џовек и Marsovac“), упркос формалном постојању наслова, што је изазивало обустављање рада алгорита. Ради руковања изузецима, увели смо исказ `try` за детекцију грешке на оном месту где је уочен проблем. Исказу смо придодали одредбу `except` и исказ `continue` који отклањају грешке и омогућавају програму да настави са даљим радом. Наведене грешке сумиране су у табели 1:

Песма/Извођач	Грешка	Узрок	Решење
„Џовек и Marsovac“	<code>AttributeError</code>	Неактиван линк	<code>try-except-continue</code>
‘Yu-Grupa‘	<code>TypeError</code>	Погрешан назив	‘YU Grupa‘

Табела 1: Проблематични случајеви у току прикупљања текстова.

2.3 Ограничења lyricsmaster-a

За одређене саставе и соло извођаче (Рибљу Чорбу, Екатарину Велику, Азру, Филм, Џибонија, Ђорђа Балашевића и др.) није било могуће прикупити текстове помоћу lyricsmaster-a, било због онемогућеног приступа подацима или услед непостојања текстова песама за датог извођача на сајту LyricWiki. Такође, у фази прикупљања текстова дошло се до закључка да би корпус требало да садржи текстове песама женских извођача, али за одређене женске извођаче чији се музички стил може окарактерисати као „рок“ (Калиопи, Слађана Милошевић и сл.) није било могуће аутоматски преузети текстове са истог сајта. Са друге стране, неки извођачи се уопште нису налазили у бази података, а самим тим ни њихови текстови (као што је то случај са Мајом Оџаклијевском).

Како бисмо надоместили овај недостатак, применили смо алтернативне критеријуме за накнадни одабир извођача тако што смо проширили жанровски опсег који ће бити покривен нашом грађом. Прецизније, укључили смо у корпус и поп извођаче (нпр. Нину Бадрић и

Зану) у чијим музичким аранжманима се запажају утицаји рок музике. Поред тога, додали смо Мадам Пиано²¹ и Горана Карана.²² Корпус представљен у овом раду састоји се од песама извођача наведених у табели 2:

Бајага	Електрични Оргазам	Неверне Бебе
Бајага и Инструктори	Забрањено Пушење	Негатив
Беби Дол	Зана	Нина Бадрић
Бијело Дугме	Идоли	Октобар 1864
Ван Гог	Индекси	Партибрејкерс
Галија	Јосипа Лисац	Прљаво Казалиште
Горан Бреговић	YU Група	Рани Мраз
Горан Каран	Кербер	Смак
Дивље Јагоде	Мадам Пиано	Хари Мата Хари
Дино Мерлин	Мирзино Јато	Хаустор

Табела 2: Списак 30 извођача у корпусу.

Најстарији албум у корпусу јесте *Наше доба* (1967) групе Индекси; са друге стране, најскорије објављени албуми заступљени у истој грађи јесу албуми *Од неба до неба* групе Дивље Јагоде, односно *Collection* Нине Бадрић (оба издата 2003).

2.4 Формирање стабла директоријума за смештање грађе

Извршавањем кода за гребање веба формирана је хијерархијски организована структура података, односно стабло директоријума са својим кореном и гранама. Тако добијену дрволику организацију корпуса можемо представити у виду извода са командне линије

²¹ Стручњак за YU рок, Петар Јањатовић, уврстио је Мадам Пиано у своју енциклопедију југословенског рока ([Janjatović, 1998](#)), због значаја који је ова представница цез и етно-звуча имала за развој југословенске музичке сцене.

²² Горан Каран такође није репрезентативан пример рок извођача, али не одступа превише од задатог оквира. Иако је Каран стекао популарност извођењем поп песама „са далматинским призвуком,“ он је започео каријеру као рок певач (податак преузет из његове [биографије](#) на вебу).

оперативног система, након покретања кода у програмском језику Bash:²³

```
alias tree="find . -print | sed -e 's;[~/]*/;|____;g;s;____|; |g'"
```

На слици 1 дат је парцијални приказ наведене структуре који је преузет са апликације Terminal на оперативном систему macOS, након што је претходно извршена навигација до директоријума LyricsMaster, где је смештен корпус текстова:

```
Last login: Mon Feb  4 15:21:48 on ttys001
Ljudmilas-MacBook-Air:~ ljudmilapetkovic$ cd /Users/ljudmilapetkovic/Documents/LyricsMaster
Ljudmilas-MacBook-Air:LyricsMaster ljudmilapetkovic$ alias tree="find . -print | sed -e 's;[~/]*/;|____;g;s;____|; |g'"
Ljudmilas-MacBook-Air:LyricsMaster ljudmilapetkovic$ tree
.
├── Mirzino-Jato
│   ├── .DS_Store
│   ├── Šećer i med
│   └── Apsolutno-Tvoj.txt
├── YU-Grupa
│   ├── YU-Grupa
│   ├── Trka.txt
│   ├── Crni-Leptir.txt
│   ├── More.txt
│   ├── Čudna-Šuma.txt
│   ├── Noc-Je-Moja.txt
│   └── Devojko-Mala-Podigni-Glavu.txt
├── .DS_Store
├── Rim 1994
│   ├── Blok.txt
│   ├── .DS_Store
│   └── Odlazim.txt
```

Слика 1: Стабло директоријума „LyricsMaster“ са командне линије.

Као што се може закључити на основу слике 1, корени директоријум носи подразумевани назив „LyricsMaster”, унутар којег су смештени директоријуми са називима извођача (нпр. Мирзино Јато или YU Група). Затим се за сваког извођача наводе називи албума (*Шећер и мед*, *YU Група*, *Рим 1994*, итд.), у којима су садржани текстови песама у .txt формату, уједно и крајњи чворови у структури дрвета (нпр. „Апсолутно твој”, „Црни лептир” итд.). Међу наведеним подацима формално се јавља и датотека „.DS_Store”, која није ни од каквог значаја за даљу обраду и анализу корпуса, и која ће стога у каснијим фазама бити елиминисана. Након процедуре гребања сајта, прикупљени садржаји текстова у .txt формату, који су ускладиштени у засебним директоријумима, у наредној

²³ По узору на [Bash код](#) (на вебу).

етапи припреме корпуса обједињени су у јединствену XML датотеку и аотирали.

3. Препроцесирање корпуса

3.1 DTD спецификација

Пре аутоматског генерисања корпуса у XML формату, састављена је *декларација типа документа* (скр. DTD), која спецификује логичку структуру XML документа, на коју се он реферише, и у односу на коју се проверава његова валидност. Овај корак је такође важан јер представља предуслов за даљу обраду XML документа, попут генерисања основних статистичких података из њега у прегледном XHTML формату коришћењем XSLT процесора, о чему ће бити речи у одељку 4.1. У валидном XML документу корпуса, примера ради, DTD декларише да корени елемент `<exYuPesme>` може имати више аутора, али да је његова вредност атрибута фиксирана празна ниска.

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="tabela-css-classes.xsl"?>
<!DOCTYPE exYuPesme [
<!ELEMENT exYuPesme (autor)+>
<!ATTLIST exYuPesme
xmlns CDATA #FIXED ''> ... ]
```

Слика 1: Декларација типа документа – одломак.

3.2 Проналажење свих текстова коришћењем модула os

Први корак у процесу генерисања XML документа представља примена модула `os` са методом `listdir()`, из стандардне програмске библиотеке језика Python. Када се метода позове из кореног директоријума „LyricsMaster“, она враћа листу директоријума, тј. имена извођача. Да би се прикупили називи поддиректоријума (наслови албума датог извођача) и датотека (називи песама на тим албумима), било је потребно форматирати ниске са апсолутним путањама помоћу `format()` функције. Крајњи делови путања, тј. све оно што се налази иза последње косе црте – *основни назив* (енгл. base name), представља вредности тражених категорија. У том поступку коришћене су две функције `os.path` модула: помоћу функције `join()`, елементи листе директоријума и датотека спојени су са крајевима текућих путања

директоријума, како би се приступило тим елементима. Друга функција, `isfile()`, комбинована је са исказима контроле тока `if` и `if not` како би се исправно утврдило да ли се генерише листа директоријума или датотека. Илустрације ради, у наставку су дате путање на основу којих можемо екстраховати име групе Смак, наслов њиховог албума *Црна дама* и песму са тог албума – „Даире“:

```
../../LyricsMaster/Smak  
../../LyricsMaster/Smak/Crna-Dama  
../../LyricsMaster/Smak/Crna-Dama/Daire.txt
```

Затим су учитани сами текстови песама помоћу функција `open()` и `read()`, при чему је задржана стиховна структура текстова која садржи нове редове коришћењем функције `split('\n')[:-1]`. Суштина целокупне описане процедуре јесте могућност правилног распоређивања података приликом аотирања датотеке у XML језику.

3.3 Елиминација сувишног садржаја

У корпусу је запажен изванредан број текстова песама написаних искључиво на страним језицима. Будући да је тема овог рада обрада корпуса текстова на српском, односно српскохрватском језику, извршено је ручно уклањање наведених нумера или, пак, читавих албума из корпусних директоријума. У овом одељку наводимо неке карактеристичне примере пречишћавања корпуса. Конкретно, избачене су песме на грчком, македонском, ромском, енглеском, португалском и пољском језику. Специфичан случај представља присуство вишејезичности у текстовима песама Горана Бреговића као соло извођача, који је и аутор песама «Κέρνα μας», са албума *Alkohol: šljivovica & champagne*; „7/8 & 11/8“, „Ederlezi“, „TV screen“, „Ausência“ (албум *Ederlezi*) и „To nie ptak“ (*Kayah & Bregović*). На албуму са музиком из филма *Arizona Dream* све песме су на енглеском језику, тако да је и он уклоњен.

Из корпуса су такође искључени текстови песама који су грешком приписани неком аутору. Примера ради, међу текстовима нумера на LyricWiki страни Нине Бадрић налази се и песма „Ubila si del mene“, која припада словеначком *boy* бенду Game Over. Крагујевачкој групи Смак такође су погрешно додељене две песме (“Muistoja” и “Myrsky”) истоимене групе из Финске. Поред тога, на албуму *Zašto ne volim sneg* групе Смак појављује се и неколико инструменталних нумера. Нису у

обзир узете ни песме за које LyricWiki још увек нема садржај (уместо којег се јављају три тачке), као у случају нумере „Ne mogu da kariram“, бенда Партибрејкерс. Такође су се јавили и дупликати једног текста под различитим насловима. Примера ради, за групу Забрањено Пушење задржан је текст за песму „Rošasna salva“ са тим насловом, док су дупликати са лажним насловима „Manijak“, „Vuk“ и „Ujka Sem“ брисани.

Нарочиту пажњу је требало обратити на одабир песама са концертних или компилацијских албума. Наиме, првобитно је замишљено да се такви директоријуми одмах одстрањују, јер је било очекивано да ће се у њима налазити нумере које већ постоје у корпусу. Ипак, уочено је присуство одређеног броја необјављених нумера на концертним албумима (нпр. „Na vrhovima prstiju“ са албума *Neka svetmir čije nemir* групе Бајага и Инструктори). Са албума највећих хитова Нине Бадрић (*Collection*) задржан је мали број јединствених песама. Још једну врсту дупликата представљале су и неке обраде песама, као за песму „Тако ти је мала моја кад љуби Bosanac“ Бијелог Дугмета из 1975. године, коју је обрадила група Забрањено Пушење 1998. године.

3.4 Анотирање и препроцесирање корпуса у XML формату

Yattag²⁴ представља програмску библиотеку за рад у језику Python којом се аутоматски могу додавати HTML и XML етикете приликом структурирања докумената. Овај API аутоматски додаје отворене и затворене изломљене заграде, и сваки почетни таг праћен је затвореним тагом. На овај начин, анотатор може брже и лакше обележавати текстове, јер програм смањује могућност јављања синтакстичких грешака. Yattag кроз модул `indent` такође подржава и аутоматско увлачење према општој хијерархијској структури XML документа, при чему се величина увученог простора може модификовати. Будући да је поступак анотације требало спровести на великом корпусу, идеја је била искористити функционалности наведене библиотеке ради ефикаснијег обележавања у складу са дефинисаним правилима означавања садржаја текстова. Та правила подразумевала су укључивање XML етикета елемената и атрибута за описивање одговарајућих делова садржаја корпуса по следећем поступку:

- Корени елемент дефинисан је етикетом `<exYuPesme>`;

²⁴ Yattag (на вебу).

- Аутори, односно извођачи, дефинисани су елементом `<autor>`, чији су атрибути `ime`, `brojAlbuma`,²⁵ `pol`, `zanr`, `rodnoMesto` (текстописци се не бележе, јер се не наводе ни на сајту);
- Албуми су дефинисани елементом `<album>`, који има атрибуте `naziv`, `godina` и `izdavac`;
- Песме су дефинисане елементом `<pesma>`, који има атрибут `naslovPesme`, док је за стихове резервисана етикета елемента ``.

Пошто метод `tag()` креира XML етикете, њему су као аргументи прослеђени само жељени називи етикете за означавање елемената и њихових атрибута. Нпр. албум као елемент има атрибуте за назив, годину објављивања и продуцентску кућу, што је дефинисано на следећи начин: `with tag('album', naziv=album, godina="", izdavac=""):`. Са друге стране, метод `text()` генерише текстуални садржај који није назив етикете. У циљу скраћивања кода, коришћена је Doc инстанца класе `yattag.Doc` и здружена метода `tagtext()`, која тој инстанци додаје садржај који наведена метода производи (називе етикете, атрибута и обичан текст). Након дефинисања свих неопходних параметара, примењен је `getvalue()` метод, како би се целокупан садржај претворио у велику ниску карактера. Овим путем текстови су смештени у нову XML датотеку.

Након пробног генерисања XML корпуса, у њему је уочен специјални карактер амперсенд (`&`), који је, по правилу, замењен карактером излазне секвенце (`&`). То је учињено да се амперсенд не би третирао као почетак референце ентитета и како би се у потпуности испоштовали принципи правилног XML структурирања. Због тога је у код за аотирање корпуса уметнута функција `escape()` модула `xml.sax.saxutils`. Иако се у корпусу често јавља и симбол за апостроф (`'`), он у процесу аутоматске замене карактера није замењен својим XML еквивалентом `'`, што није ометало генерисање исправно формираног XML документа. Ради веће прегледности, аутоматски су уклоњене и цртице у називима песама и албума које су остале приликом гребања веба (нпр. промена од „Da-Sam-Pekar“ у „Da Sam Pekar“).

Што се тиче ручног обележавања корпуса, елементу `<autor>` додате су вредности атрибута `zanr`: поп, рок и *world music*.²⁶ Осим жанра, додате су и вредности атрибута `rodnoMesto` за родно место извођача и `izdavac` за продуцентску кућу која је издала албум. Пример

²⁵ Подр. број албума обухваћених текстовима песама укључених у корпус.

²⁶ Према класификацији жанрова доступних на веб страни [Discogs](#).

полуаутоматски обележеног текста поменуते песме са албума *Концерт код Хајдучке чесме* групе Бијело Дугме, може се видети на слици 2.

```
<autor ime="Bijelo Dugme" brojAlbuma="13" pol="Grupa" zanr="Rok" rodnoMesto="Sarajevo">
  <album naziv="Koncert kod hajdučke česme" godina="1977" izdovac="Jugoton">
    <pesma naslovPesme="Da Sam Pekar">
      <li>Da sam pekar, mala moja</li>
      <li>Ne znam bi l'0 me htjela</li>
      <li>Kad bi noću bila sama</li>
      <li>Zemčke bi jela</li>
```

Слика 2: Одломак анотираог корпуса у софтверу oXygen XML Editor.

3.5 Аутоматска рестаурација дијакритика – софтвер LeXimir

Нормализација корпуса и његова припрема за рачунарску анализу не представља нимало тривијалан задатак, а, по правилу, доприноси генерисању информативнијих резултата у поређењу са рачунарском анализом непрепроцесираниог текста. Прикупљени текстови песама у Yч корпусу били су прилично неуједначени по питању употребе писма: већина текстова била је записана на латиници, а није мали број оних у којима специјална латинична слова (č, ć, ž, đ, š) нису садржала потребне дијакритичке знаке. Са друге стране, мањи број песама био је првобитно забележен на ћирилици. Почетна замисао била је да се целокупан корпус преслови на ћирилицу; од те идеје се одустало јер се у текстовима на српском језику појављују и одломци на страним језицима (нпр. “la musique c’est fantastiqu / prepare la revolution / et la femme est tres jolie / tre jolie comme un bonbon”).²⁷ Стога је одлучено да у поступку транслитерације текстови написани деградираном латиницом и ћирилицом буду аутоматски претворени у стандардну српску латиницу са дијакритичким знацима. Пресловљени текст корпуса у .txt формату коришћен је у поступку екстракције лексичких јединица из одељка 4.2.

Да би се одабрао најпогоднији алат, обављена је рестаурација дијакритика у апликацијама Слово Мајстор и LeXimir. Други алат користи електронске морфолошке речнике (са валидним облицима речи), локалне граматике и Корпус савременог српског језика (*Krstev*

²⁷ Према изворном тексту, са штампарским и правописним грешкама.

et al., 2018). Након тестирања обеју метода, дошло се до закључка да Слово Мајстор солидно решава проблем претварања деградираних у стандардизовану латиницу, уз напомену да је неким речима погрешно додао дијакритичке знаке, што је нарушило семантику речи. Тако је од глаголског облика „isecka“ погрешно створена реч „isečka“. Добра страна ове апликације јесте могућност накнадне исправке погрешно транслитерираних речи.

Са друге стране, LeXimir, који користи функционалности Unix софтвера за обраду корпуса, исту реч је оставио у исправном облику. На слици 3 може се видети да су потенцијални кандидати за одабир адекватног облика реч „isečka“, што је флективни облик именице „исечак“ у генитиву јединице, и „isecka“, односно 3. л. јд. аориста глагола „исецкати“. Алгоритам за исправљање на основу левог и десног текстуалног контекста потенцијалне речи за исправљање врши *дезамбигуацију* (разрешавање вишезначности). Одабрана је друга опција, где заменица „га“ може да стоји уз глагол „исецка“, а не и уз флективни облик именице „исечка“, јер би друга конструкција била граматички неисправна. Ипак, ручна евалуација је неопходна и након примене описаног алата, јер у неким речима (као нпр. у речи „оци“ у контексту „Trljam oci sanjive“) нису додати дијакритички знаци тамо где је требало да постоје.

Potencijalni kandidati	Levi kontekst	Ispr.	Desni kontekst
*7(uže(72)_uze(61))	*****	uze	*****
*7(isečka(5)_isecka(1))	s i ujeo kuvara Kuvar uze, uze nož	Isecka	ga na dva, na tri komada Došli, do
*7(oci(818)_oci(24))	žurno Kafu pijem, nestajem Trljam	oci	sanjive Da mi ne bi zaspale Vrata

Слика 3: Рестаурација дијакритика у алату LeXimir.

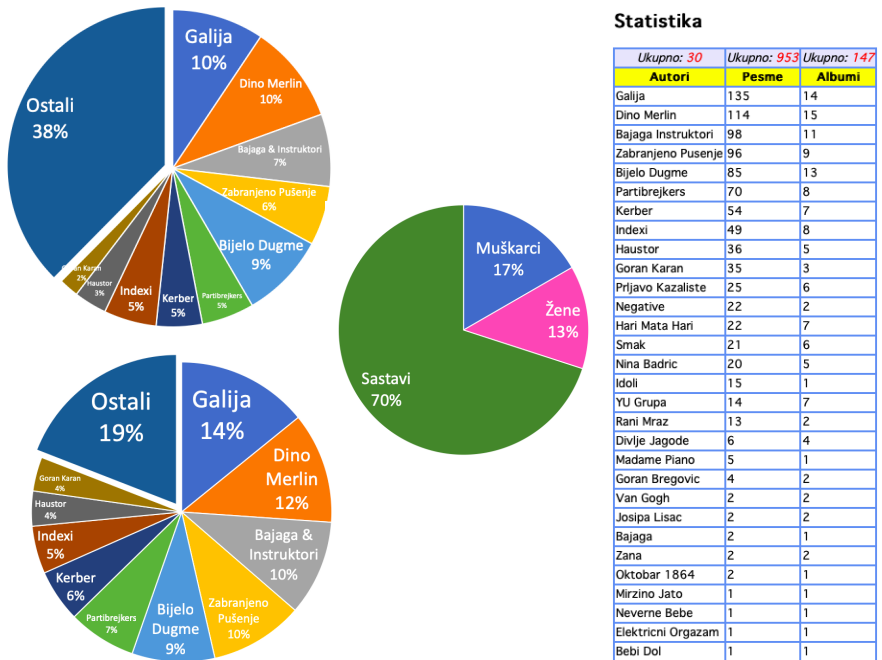
4. Рачунарска анализа корпуса

4.1 Статистички подаци о корпусу

DTD спецификација омогућила је постављање упита кроз навигацију над датим документом помоћу XSLT језика за трансформацију XML документа у друге формате. Нпр., следећи XSLT израз, за сваки елемент <autor> у документу, бира вредност атрибута *ime*, чиме се генерише листа свих извођача у корпусу:

```
<xsl:for-each select="//autor">
  <xsl:value-of select="@ime" />
</xsl:for-each>
```

Подаци из XML документа су трансформисани у XHTML табелу са слике 4, где се може видети укупан број аутора, песама и албума у корпусу. Кружни графикони са исте илустрације описују (у смеру казаљке на сагу, одозго надоле): заступљеност извођача у корпусу према броју албума, број мушких, женских и групних извођача, као и удео по броју песама.



Слика 4: Статистички подаци о YU корпусу.

У LeXimir-у су такође произведени основни статистички подаци о YU корпусу. Након рестаурације дијакритика, корпус садржи 248.807

токена, 16964 јединствене лексичке јединице, 116.972 просте форме (16.909 различитих) и 268 бројева (10 различитих). LeXimir нуди подршку и за приказ флективних облика токена, лематизованих облика, врста речи, фреквенција речи и колокација. Резултати фреквенцијске анализе могу се филтрирати према врсти речи. На основу извезене .xlsx датотеке можемо утврдити нпр. које су најфреквентније именице или колокације чији је главни члан именица. Слика 5 у наставку представља парцијални приказ резултата који сведочи о присуству најчесталијих именичких токена, међу којима се јављају речи *dan*, *noć*, *ljubav*, *srce* итд. Слика 6 доноси преглед одређених колокација у вези са ратом (*svetski rat*, *vojnu muziku*, *ratne filmove*).

Oblik	Lema	POS	Freq.
dan	dan	N	299
Al	Al	N	231
noć	noć	N	228
do	do	N	224
ljubav	ljubav	N	201
srce	srce	N	196
život	život	N	195
put	put	N	184
meni	mena	N	164
meni	meni	N	164
kraj	kraj	N	155
noći	noć	N	147
biti	bit	N	132
bila	bito	N	124
grad	grad	N	120

Слика 5: Токени.

svetski rat	svetski rat	N	3
novi svet	Novi svet	N	3
noćne ptice	noćna ptica	N	3
tam-tam	tam-tam	N	3
prošlog vremena	prošlo vreme	N	2
vojnu muziku	vojna muzika	N	2
železnička stanica	železnička stanica	N	2
crno grožđe	crno grožđe	N	2
sunčev zrak	sunčev zrak	N	2
malog medveda	Mali Medved	N	2
zlatne medalje	zlatna medalja	N	2
morske obale	morska obala	N	2
svetla budućnost	svetla budućnost	N	2
ratne filmove	ratni film	N	2

Слика 6: Колокације.

4.2 Екстракција тема из корпуса

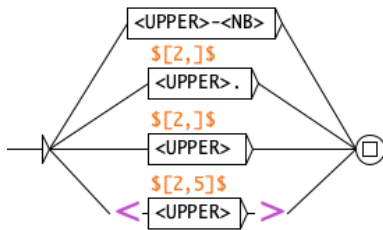
Познато је да су у текстовима песама југо-рока биле присутне друштвено-политичке и културолошке теме, те је у софтверу Upi-tex конструисан граф коначног аутомата за препознавање индикатора наведених тема. Помоћу регуларних израза и лексичких маски трагало се за лексичким обрасцима у виду скраћеница²⁸ састављених од:

- мајускуле, цртице и низа бројева (нпр. *B-52*);
- секвенце мајускула + тачка која се понавља најмање два пута (*K.P.*);

²⁸ Видети веб страну [Скраћенице и објашњења](#)

- секвенце мајускула + размак која се понавља најмање два пута (*ES EF ER JOT*);
- секвенце која садржи од две до пет мајускула (*CZ, CIA*).

Међу њима су се издвојили називи државних органа (нпр. *AVNOJ*), спортских клубова (*PFC*), назива телевизијских канала (*BBC*) итд. Граф са слике 7, осим скраћених назива представљених у резултату конкорданце на слици 8, такође препознаје и токене: *CZ, ES EF ER JOT, FK, TV, JRT, K.P., KGB, KK, MUP, O.K., PC, PFC, SUP* и *TAS*.



Слика 7: Граф за препознавање скраћеница.

Od istorijskog [AVNOJ](#)-a Do izbjegličkog
 Od istorijskog [AVNOJ](#)-a Do izbjegličkog
 Od istorijskog [AVNOJ](#)-a Do izbjegličkog
 Od istorijskog [AVNOJ](#)-a Do izbjegličkog
 nebu je pisalo [B-52](#) I veliki buketi od
 nebu je pisalo [B-52](#) I veliki buketi od
 šao Haše!{S} I [BBC](#) na mome radiju, Osl
 šao Haše!{S} I [BBC](#) na mome radiju, Osl
 niskim terenom [BMW](#)-a kešom plaćenog, o
 KGB zaspala je [CIA](#) odmara se naša mili
 , penjao se na [CK](#) i pevao pesnu protiv
 iti stan Preko [CNN](#)-a gledao sam Mufu S

Слика 8: Одломак из конкорданце.

Да би се утврдило да ли су друштвено ангажоване теме статистички значајне у корпусу, генерисан је визуални приказ преовлађујућих тема у виду *облака дрвета* (енгл. tree cloud) коришћењем алата TreeCloud, уз напомену да је анализа вршена на нелематизованим облицима. Овај програм комбинује графичко представљање података у виду облака речи са дрволиком структуром, тако да се, осим фреквенције речи, пружа увид и у начин груписања токена, према њиховом растојању у овој структури (видети слику 9). У апликацију је уграђена и листа *stop речи* (енгл. stop words) за српски језик, којој су, за потребе визуализације датог корпуса, придодате још неке речи. Креиран је визуални приказ у коме се издвајају кластери са значењем осећања, у виду речи *tuga, ljubav/-i, bol, srce/-a*. Од урбаних тема, присутне су речи *grad/-u, ulice*, делови тела су означени речима *ruke, lice, oči, srce/-a/-u*. Мотив времена је такође фреквентан у корпусу (*život, vr(ij)eme, godina/-e*). Према овом приказу, друштвено ангажоване теме нису у довољној мери заступљене у корпусу.

Литература

- Bertin-Mahieux, Thierry, Daniel PW Ellis, Brian Whitman and Paul Lamere. “The million song dataset”. In *Proceedings of the 12th International Conference on Music Information Retrieval*, 2011, 591–596. Датум приступа: 19. октобар 2019, <http://ismir2011.ismir.net/papers/OS6-1.pdf>.
- Cooper, Laura E. and B. L. Cooper. “The pendulum of cultural imperialism: Popular music interchanges between the United States and Britain, 1943-1967”. *Journal of Popular Culture* Vol. 27 no. 3(1993): 61. Датум приступа: 19. октобар 2019, <https://sci-hub.tw/10.1111/j.0022-3840.1993.00061.x#>.
- Falk, Johanna. “We will rock you : A diachronic corpus-based analysis of linguistic features in rock lyrics”. 2013. Датум приступа: 19. октобар 2019, <http://www.diva-portal.org/smash/get/diva2:605003/FULLTEXT02.pdf>.
- Gambette, Philippe and Jean Véronis. “Visualising a Text with a Tree Cloud”. *IFCS'09: International Federation of Classification Societies Conference* (2009): 561–569, 19. октобар 2019, <https://hal-lirmm.ccsd.cnrs.fr/lirmm-00373643/file/2009GambetteVeronis.pdf>
- Haslam, Thomas J. “Mapping the Great Divide in the Lyrics of Leonard Cohen”. *Rupkatha Journal on Interdisciplinary Studies in Humanities* Vol. 9, no. 1(2017): 1–10. Датум приступа: 19. октобар 2019, <http://rupkatha.com/V9/n1/v9n1s01.pdf>
- Hentschel, Elke. “The expression of gender in Serbian”. In *Gender across languages: The linguistic representation of women and men*, Vol. 3, 287–309. John Benjamins Publishing Company, 2003. Датум приступа: 19. октобар 2019, <https://epdf.tips/gender-across-languages-volume-iii-the-linguistic-representation-of-women-and-me.html>
- Janjatović, Petar. *Ilustrovana YU rock enciklopedija: 1960-1997*. Geopoetika, 1998. Датум приступа: 19. октобар 2019, https://monoskop.org/images/c/ca/Janjatovic_Petar_Ilustrovana_YU_Rock_Enciklopedija_1960-1997.pdf
- Kreyer, Rolf and Joybrato Mukherjee. “The style of pop song lyrics: A corpus-linguistic pilot study”. *Anglia-Zeitschrift für englische Philologie* Vol. 125, no. 1(2007): 31–58. Датум приступа: 19. октобар 2019, <https://doi.org/10.1515/ANGL.2007.31>
- Krstešević, Cvetana, Ranka Stanković and Duško Vitas. “Knowledge and Rule-Based Diacritic Restoration in Serbian”. *Proceedings*

- of the Third International Conference Computational Linguistics in Bulgaria, 41–51. 2018. Датум приступа: 19. октобар 2019, https://www.researchgate.net/publication/328416358_Knowledge_and_Rule-Based_Diacritic_Restoration_in_Serbian
- Lightman, Erin J., Philip M. McCarthy, David F. Dufty and Danielle S. McNamara. “Using computational text analysis tools to compare the lyrics of suicidal and non-suicidal songwriters”. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 29, 2007. Датум приступа: 19. октобар 2019, <https://cloudfront.escholarship.org/dist/prd/content/qt0dh4553j/qt0dh4553j.pdf>.
- Lukic, Alen. “A Comparison of Topic Modeling Approaches for a Comprehensive Corpus of Song Lyrics”, Tech report, Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, s.d. Датум приступа: 19. октобар 2019, http://alenlukic.com/assets/docs/lyric_topic_modeling.pdf
- Mahedero, Jose P.G., Álvaro Martínez, Pedro Cano, Markus Koppenberger and Fabien Gouyon. “Natural language processing of lyrics”. In *Proceedings of the 13th annual ACM international conference on Multimedia*, 475–478. 2005. Датум приступа: 19. октобар 2019, https://www.researchgate.net/profile/Pedro_Cano5/publication/221573745_Natural_language_processing_of_lyrics/links/00b7d52826f623edfb000000.pdf
- McEnery, Tony and Andrew Hardie. *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press, 2012. Датум приступа: 19. октобар 2019, <http://gen.lib.rus.ec/book/index.php?md5=33c1c5b6d73ea816dfb2a034f73bb176>
- Petrie, Keith J., James W. Pennebaker and Borge Sivertsen, “Things We Said Today: A Linguistic Analysis of the Beatles”, *Psychology of Aesthetics, Creativity, and the Arts* Vol. 2, no. 4(2008): 197. Датум приступа: 19. октобар 2019, <https://www.uvm.edu/pdodds/files/papers/others/2008/petrie2008a.pdf>.
- Stein, Daniel. “Multi-Word Expressions in the Spanish Bhagavad Gita, Extracted with Local Grammars Based on Semantic Classes”. In LREC ’2012 Workshop: LRE-Rel, Language Resources and Evaluation for Religious Texts, 88–93. 2012. Датум приступа: 19. октобар 2019, https://www.academia.edu/26035335/Linguistic_and_Semantic_Annotation_in_Religious_Memento_mori_Literature
- Taina, Jesse et al.. “Keywords in heavy metal lyrics: A Data-Driven Corpus Study into the Lyrics of Five Heavy Metal Subgenres”, 2014. Датум

- приступа: 19. октобар 2019, <https://helda.helsinki.fi/bitstream/handle/10138/136524/keywords.pdf?sequence=1>.
- Taylor, Charlotte. “What is corpus linguistics? What the data says”. *ICAME journal* Vol. 32 (2008): 179–200. Датум приступа: 19. октобар 2019, http://sro.sussex.ac.uk/id/eprint/53389/1/what_is_corpus_linguistics.pdf
- Whissell, Cynthia. “Traditional and Emotional Stylometric Analysis of the Songs of Beatles Paul McCartney and John Lennon”. *Computers and the Humanities* Vol. 30, no. 3 (1996): 257–265. Датум приступа: 19. октобар 2019, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.7171&rep=rep1&type=pdf>.
- Zhang, Shuo, Rafael Caro Repetto and Xavier Serra. “Understanding the expressive functions of jingju metrical patterns through lyrics text mining”. In *18th International Society for Music Information Retrieval Conference*, 397–403. 2017. Датум приступа: 19. октобар 2019, https://repositori.upf.edu/bitstream/handle/10230/32652/Zhang_ISMIR2017_unde.pdf?sequence=1&isAllowed=y
- Zörnig, Peter, Emmerich Kelih and Ladislav Fuks. “Classification of Serbian texts based on lexical characteristics and multivariate statistical analysis”. *Glottology* Vol. 7, no. 1(2016): 41–66. Датум приступа: 19. октобар 2019, http://homepage.univie.ac.at/emmerich.kelih/wp-content/uploads/2016_Zoernig_Kelih_Fuks_www.pdf.
- Арсенијевић, Александра, Милена Обрадовић and Михаило Шкорић. “Израда мултимедијалног документа „YU рок сцена””. *Инфотека – часопис за дигиталну хуманистику* Vol. 16, no. 1-2а(2016):113-129, датум приступа: 19. октобар 2019, https://infoteka.bg.ac.rs/ojs/index.php/Infoteka/article/view/2016.16.1_2.6_sr
- Божиловић, Никола. “Култура сећања и југословенски рокенрол”. *Култура* no. 152(2016): 257–280. Датум приступа: 19. октобар 2019, <https://scindeks-clanci.ceon.rs/data/pdf/0023-5164/2016/0023-51641652257B.pdf>
- Гајић, Златомир. “Рок поезија Бранимира Штулића и њени медијски одјаци”. PhD thesis, Универзитет у Новом Саду, Филозофски факултет, 2018. Датум приступа: 19. октобар 2019, <http://nardus.mpn.gov.rs/bitstream/handle/123456789/9926/Disertacija17602.pdf?sequence=1&isAllowed=y>
- Кешел, Владо and Данко Шипка. “Приступ изградњи стемера и лематизатора за језике с богатом флексијом и оскудним ресурсима заснован на обухватању суфикса”. *ИНФОтека – часопис за*

библиотекарство и информатику Vol. 9, no. 1-2(2008): 21–31. Датум приступа: 19. октобар 2019, http://infoteka.bg.ac.rs/pdf/Srp/2008/04%20Vlado-Danko_Stemeri.pdf

Раковић, Александар. “Бит мода, рокенрол и генерацијски сукоб у Југославији 1965-1967”. *Етноантрополошки проблеми* Vol. 6, no. 3(2011): 745–762. Датум приступа: 19. октобар 2019, <https://www.eap-iea.org/index.php/eap/article/view/604/594>

Раковић, Александар. “Рокенрол у Социјалистичкој Југославији: од забаве градске омладине до националне културе”. In *Сан о граду : зборник радова*, Андрићев институт, 427–439. 2018. Датум приступа: 19. октобар 2019, http://doi.fil.bg.ac.rs/pdf/eb_book/2018/ai_san_o_gradu/ai_san_o_gradu-2018-ch18.pdf