



DOI: 10.4312/mz.58.2.81-105
UDK 78 (497.11):070.447"1878/1941":004

Computational Research of Music Criticism between 1878 and 1941 in Serbian: What Can We Learn from the Digital Archive of the Historical Newspapers in the Svetozar Marković University Library?

Ivana Perković,^a Anđelka Zečević^b

^aUniversity of Arts, Belgrade

^bSerbian Academy of Sciences and Arts, Belgrade

ABSTRACT

The article deals with the critical writing on music in the digitized newspapers in the Svetozar Marković University Library in Belgrade. The multidisciplinary research aims to connect the quantitative and qualitative analysis of the musical critics in Serbian in the period between 1878 and 1941, and to explore how to combine both the computational and traditional musicological approaches.

Keywords: music criticism, natural language processing, text analysis, musicological canon

IZVLEČEK

Članek se ukvarja s kritiškimi zapisi o glasbi v digitaliziranih časopisih v univerzitetni knjižnici Svetozarja Markovića v Beogradu. Cilj multidisciplinarnе raziskave je povezati kvantitativno ter kvalitativno analizo glasbenih kritik v srbščini v obdobju med letoma 1878 in 1941 ter raziskati, kako združiti računalniške in tradicionalne muzikološke pristope.

Ključne besede: glasbena kritika, procesiranje naravnega jezika, analiza besedila, muzikološki kanon

Introduction

The process of digitization has affected many fields of musicological research, including the area of music criticism. However, computational research on the reception of musicians, musical pieces and performers, written in Serbian language, has not been done so far for a number of reasons, the most important of which include: (1) the limited availability of digital resources and (2) the requirements of the Natural Language Processing (NLP) systems regarding the language on music in historical sources. Furthermore, musicologists in Serbia often lack the computational and technical knowhow to conduct this type of research, since these fields of knowledge are not part of the regular academic curriculum.

The multidisciplinary research carried out for the purpose of this paper – to our knowledge, the first investigation of this kind related to Serbian resources – began with the definition of the repository, which consists primarily of writings related to music criticism in a digital format.

The repository of writing includes newspapers published across 63 years – between 1878 (the date of the independence of Serbia) and the start of the Second World War in Yugoslavia (1941). This time span is usually divided into two historical periods: the “long nineteenth century”¹ and the interwar period. A brief presentation of the historical and cultural background in these two periods follows, for a better understanding of the context.

The second half of the long nineteenth century was marked by frequent and complex political changes in the states inhabited by Serbs, including independence from Turkish domination (1878), the establishment of the Kingdom of Serbia (1882–1918), the change of the Obrenović and Karađorđević dynasties (1903), the Balkan Wars (1912–1913), and World War I. This is the time when important cultural and scientific institutions were founded, among which we find the Serbian Literary Guild (Srpska književna zadruga, 1892), and the *Serbian Literary Magazine* (*Srpski književni glasnik*, 1901–1914, 1920–1941). Significant impulses for the development of science came particularly from the Academy of Sciences and the Higher School (Velika Škola) that was transformed into the University of Belgrade (1905). Culture was also subjected to many changes, while national idealism and Romanticism were of special importance for cultural and artistic life in general.

The interwar period brought numerous transformations, within the Kingdom of Serbs, Croats and Slovenes (1918–1929) and later the Kingdom of Yugoslavia (1929–1941) resulting in aspirations towards unification, changes of various social structures, the economy, education and other fields. This was

1 The concept of the long nineteenth century in Western history, denoting the period between the French Revolution and the beginning of the First World War (1789–1914), relies on the historical-theoretical model of Eric Hobsbawm. For more information, cf. Phyllis Weliver and Katharine Ellis, eds., *Words and Notes in the Long Nineteenth Century* (Woodbridge, UK; Rochester, NY: Boydell & Brewer, 2013).

a time of high achievements of cultural history, and the key words here are certainly Europeanization and modernization. The renewal and establishment of institutions, whether cultural, scientific or educational, affected the overall level of prosperity, so that new tendencies in art, culture and science were synchronized with similar phenomena occurring in other European countries. The progressive and liberal journal *Nova Evropa* (*New Europe*) was founded in Zagreb in 1920, and a year later (1921) the *avant-garde* journal *Zenit* (*Zenith*, Zagreb and Belgrade) appeared, while in 1922 Cvijeta Zuzorić Society of Friends of Art was founded, promoting interest in the arts, to name just a few examples. The latest stylistic challenges were present in art (which does not mean that tradition was forgotten), and artists and scientists who had completed their education abroad successfully transferred their latest achievements to the Yugoslav environment. The increasingly developed bourgeoisie provided a good social framework for the establishment of numerous artistic associations, the development of amateurism and the provision of an educated audience.

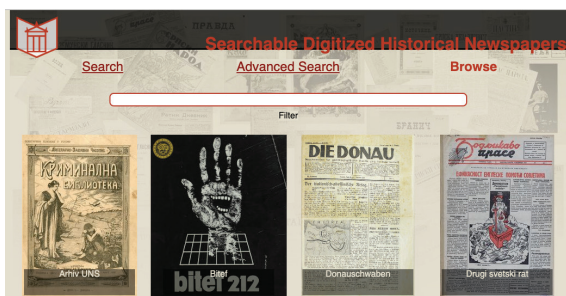
The main aim of the study is to explore how to combine the opportunities offered by computational quantitative analysis with traditional musicological qualitative research methods, and to indicate what this collaboration can bring to music research methods, particularly in terms of potential future research fields. Qualitative observations are combined with quantitative data in order to explore some of the selected topics relevant to the study of music criticism. For the purpose of this paper, three main topics were selected, as the most relevant in traditional musicological research of music criticism in Serbian: (1) the changes in the discourse in the selected historical period, (2) the language related to the emotional impact of certain performance or a new musical composition and (3) the enrichment of factual knowledge about music history. Possible further research does not exclude other themes and topics that might emerge from the huge amount of gathered material, rather, it is opened to – at the same time – broader and more refined insights, depending on the selected research discourse moving forward.

The Creation of the Corpus

Within the Europeana,² one of the largest projects intended to preserve the rich cultural heritage of Europe, the Svetozar Marković University Library contributed with digitized newspapers that uncover the social, political, and cultural life of the people of those states inhabited by Serbs. This effort included the digitalization of more than 500,000 pages in Serbian, dating from the beginning of the nineteenth century to the present through the lenses of dozens of periodicals. As such, the collection represents a valuable source of knowledge and stories embroidered around important historical events.

2 Europeana, <https://www.europeana.eu/en>.

All the digitized newspaper pages can be accessed via the official web application,³ where users can browse content and visually explore the pages using a number of functionalities.⁴ When it comes to the textual content, the web application allows users to search through the digitized collection by entering query terms and selecting the options for narrowing the specific newspaper source and/or its period. If a user enters more than one query term, by default newspaper pages containing all the terms are searched. In addition, the application supports an exact match search by enclosing the query terms in quotation marks (for instance, “koncertna muzika”), a search by a disjunction of terms by using the/or *OR* delimiter (for instance, “svirati”/“pevati”), as well as search for pages that do not contain the given query term by using a prefix - (for instance, “-muzika”). Although very helpful, these search mechanisms do not cover all the relevant retrieval scenarios, for example, search by all morphological word forms. The option “Search for similar words and expressions” allows an extended search based on the edit distance between given query terms and search terms, but offers no means of configuration or additional control.



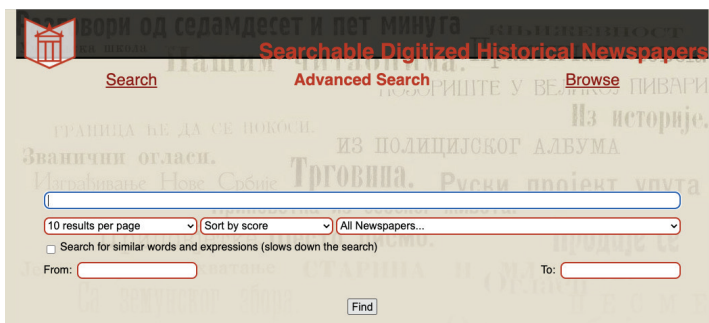
Browse panel



Preview panel (available in Serbian only)

3 Pretraživa digitalna biblioteka, <https://pretraziva.rs/search>.

4 As stated, these functionalities are partially based on the open source BnLViewer developed by National Library of Luxemburg available at “BnLViewer,” Sourceforge, <https://sourceforge.net/projects/bnlviewer/>.



Search panel

Figure 1: Application provided by the Svetozar Marković University Library.

Upon retrieval, the application generates a list of relevant newspaper pages followed by the number of results. Each newspaper page is described with the appropriate metadata and a link to the page image that can be visually explored by using functionalities such as page zooming, page rotation, or page scrolling. In order to collect and analyze the text content of the pages, we needed to develop an application programming interface (API) on top of the results list. As far as we know, no official API provided by the Svetozar Marković University Library supports this kind of content aggregation. Therefore, for the seed query terms, the selection of which will be described later, a pool of HTTP requests is sent to the web application: the HTTP requests are generated recursively following the links present in the result list in the spirit of the best crawling practices. The textual content of the newspaper pages is extracted from the HTTP response jointly with the metadata prepared for the interaction between the user and the application via a web browser.

The initial corpus encompassed the content of newspaper pages obtained using the query term “muzika” (music) with the default option for similar words and expressions search as Serbian follows an inflection paradigm with nouns characterized by cases, gender, and number. In total, the corpus contained a text content of 16,843 newspaper pages. As expected, some of the pages in the corpus referred to relevant morphological forms of the word “muzika” such as “muzike” or “muzikom.” However, there were a large number of pages containing word forms similar to the word “muzika” in respect to the edit distance, such as “fizika” (physics) or “jezika” (language) but with no overlapping semantics. Separate requests for each morphological form of the word “muzika” would generate a significantly larger number of results, in total 42,052 pages. As this corpus was intended for familiarization with writing on music, we decided to keep the smaller of the two but to perform separate searches for future retrievals.

By manual analysis of the pages in the corpus, we discovered plenty of resources related to musical education, theatre performances, ceremonies, and other similar activities not directly relevant to the study of music criticism. However, the analysis pointed us to the language of music critics – the past forms of the complete verbs such as “otpevao” ([he/she] sang), “odsvirao” ([he/she] played), “izveo” ([he/she] performed) as well as the expressions “izvedba je bila/izvođenje je bilo” (the performance was [...]), “nastupio je” (he performed) that were very frequent. This motivated us to use these verbs, their morphological forms, and both their Ekavian and Ijekavian pronunciation variants to create the new corpus. The total number of queries was 24 resulting in 3,442 pages. We were aware that this corpus is by no means complete but that it would provide us with a more accurate collection to test the usefulness of NLP methods in the study of musicology. For a number of collected pages, the metadata was missing, in most cases the exact publication date or the title of the periodical. We filtered the results to fit into the timeline range from 1 January 1878 to 6 April 1941. In total, 150 pages were eliminated.

As stated, one search result refers to a single page of a digitized newspaper that contains the given query terms. The query appearance is visually emphasized by a red rectangle around the query terms. From the web application, a user can click and single out only the page segment with the given query term, zoom in on it and read its textual content. Depending on the source and page layout, the segment is usually a text paragraph, a newspaper headline, a title, or an image caption. The page segmentation is not absolutely accurate, as we noticed numerous adjacent horizontal or vertical segments that merged. An example of such a scenario can be seen below followed by the extracted and intertwined text. The more complex the page layout and its visual presentation are, the more challenging the segmentation and NLP processing are.

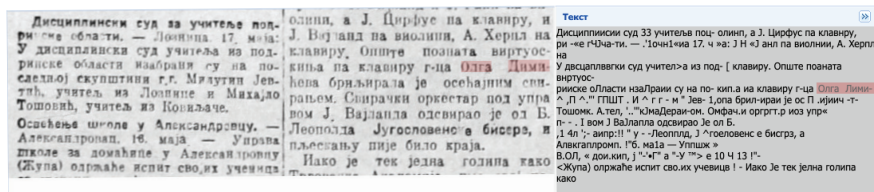


Figure 2: The page segmentation.⁵

5 *Vreme* (May 20, 1927), 4; http://istorijskenovine.unilib.rs/view/index.html#panel:pp|issue:UB_00043_19270520|article:div540|page:4|query:%D0%BE%D0%BB%D0%B3%D0%B0%20%D0%BB%D0%B8%D0%BC%D0%B8.

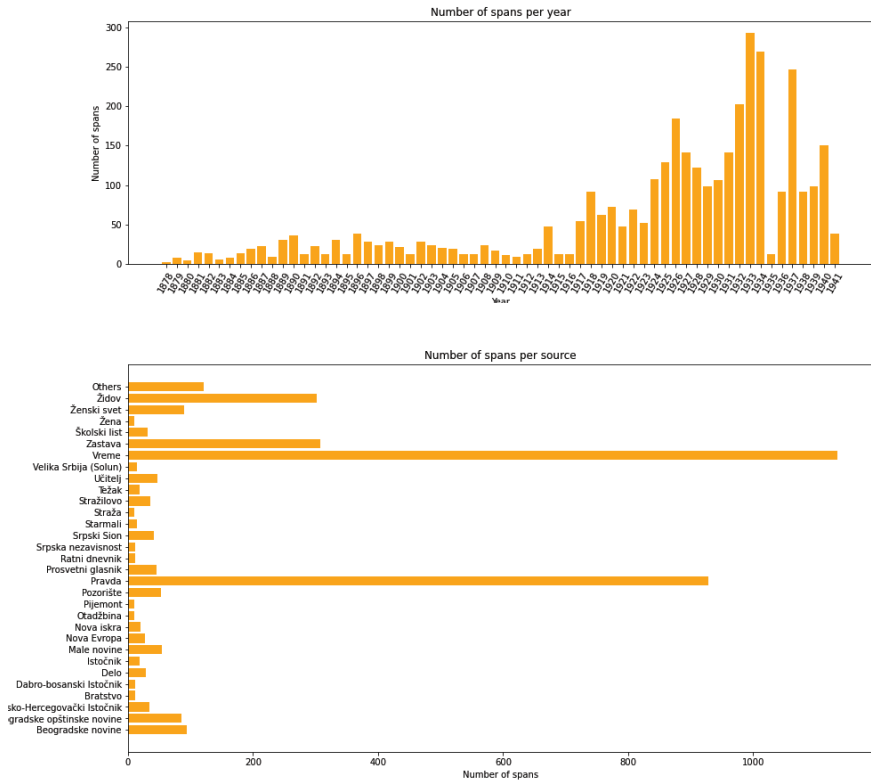


Figure 3: The distribution of spans with respect to the publication year and the source.

Each segment has its own unique identifier that becomes known only after a user clicks on it and selects it for detailed viewing. We observed this behaviour following the URLs in the application address bar. On some pages, segments with consecutive identifiers refer to two adjacent vertical paragraphs, while on other pages, this relates to two adjacent horizontal paragraphs. Therefore, it was not possible to correctly calculate the segment identifier and single out the exact text block it contains programmatically. In order to approximate the notion of the segment that corresponds to the query terms, we followed the left and right contexts represented by the number of characters. We will refer to these segment approximations as text spans throughout the rest of this paper.

We have experimented with longer text spans with a context length of 300 characters and shorter spans with a length of 150 characters. The longer spans turned out to be more informative for further analysis, but we have used the shorter spans too for the sentiment evaluation. We fixed the number of

characters instead of the number of tokens due to the errors caused by Optical Character Recognition (OCR) – many punctuation characters are misused and make the segmentation of text content in both sentences and tokens more difficult. The final repository contains 3,751 text spans in total. On the previous page, you can find the distribution of the spans with respect to both the publication year⁶ and the source.⁷

Text Preprocessing and Quality Analysis

The first text preprocessing steps were related to the replacement of HTML entities with appropriate characters (for example, “" was replaced with “) or their deletion (for example, the entities @ and ®). We also deleted those characters that do not belong to the extended ASCII coding scheme, which were primarily related to document layout (for example, various bullet points such as □) or Church Slavonic texts (for example, the letter IO).

As the digitization spans over an extended period, covering sources with diverse layouts, typography, and design, the number of OCR errors noticed in the created repository is not negligible. Similarly to other projects of the same kind,⁸ we have seen OCR errors related to:

- 1) Over-segmentation, that is the segmentation of proper words such as “klav ira” instead of “klavira” (piano);
- 2) Under-segmentation, that is the omission of delimiters such as, for example, “otpevaopesmu” instead of “otpevao pesmu” (sang the song);
- 3) Misrecognized characters, for example, “hi.mnu” instead of “himnu” (anthem);
- 4) A missing character such as “intelekualna elia” instead of “intelektualna elita” (intellectual elite);
- 5) Hallucination such as “otpevao jednu sv? I.mprovizašču” instead of “otpevao jednu svoju improvizaciju” (meaning [he/she] sang one of his/her improvisations).

Many OCR errors were connected to the misuse of punctuation marks. The relative percentage of the punctuation marks with respect to the total number of alphanumerical characters given below confirms this statement (the total percentage of punctuation marks is 4.64 %). Due to the importance of

6 For the year 1935, the number of digitized newspapers is lower than for the remaining years.

7 *Others* refers to those sources with fewer than ten spans per source. The total number of sources that satisfy this criterion is 44. The aggregation was done to provide a more accessible preview.

8 Elizabeth Soper, Stanley Fujimoto, and Yen-Yun Yu, “BART for Post-Correction of OCR Newspaper Text,” in *Proceedings of the Seventh Workshop on Noisy User-Generated Text (W-NUT 2021)*, (Association for Computational Linguistics, 2021), 284–290, <https://aclanthology.org/2021.wnut-1.31/>, DOI:10.18653/v1/2021.wnut-1.31.

punctuation for the NLP pipeline, particularly for the sentence splitting and tokenization stages, in the preprocessing stage, we kept only the elementary punctuation marks such as sentence delimiters (!?) and commas.

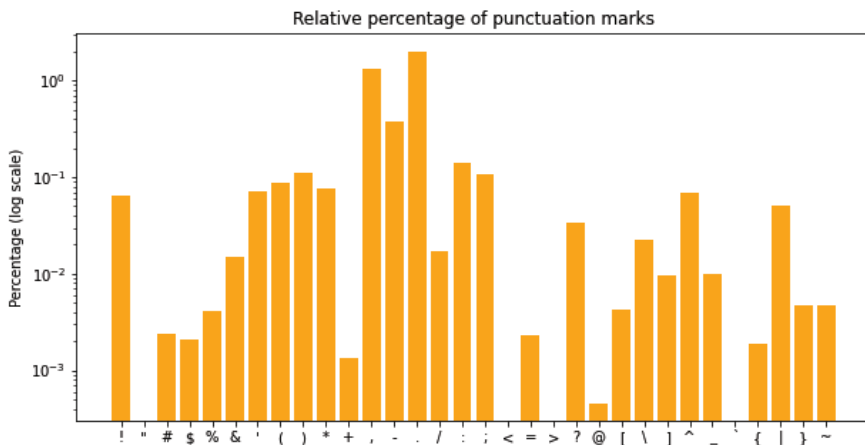


Figure 4: Relative percentage of punctuation marks.

In addition to varying publishing practices, the corpus of critical writing on music reflects diverse linguistic properties, as both Cyrillic and Latin alphabets are used for the transcription of Serbo-Croatian, as well as its distinct Ekavian and Ijekavian dialects. Therefore, the preprocessing stage required the transliteration of the collected spans into a common alphabet – for simplicity and alignment with other tools, we decided to use the Latin alphabet.⁹

To assess the quality of the text content, we analyzed the sets of short and long words. The experiment started with words of a length of less than four characters. This list was quite long, with a total of 85,165 words. However, the number of unique words was significantly lower, in total only 2,674. We manually checked the most frequent short words and handcrafted the rules for the correction of OCR errors wherever possible. We noticed many common prefixes and suffixes probably caused by over-segmentation due to newspaper layouts and column structures. In addition to short words, we analyzed the long words with a length of greater than 15 characters. This list contained 516 unique words. The majority of these words were related to hallucination, for example, *eieiemei6ieteiemeiem9i6teieiei9iei6k*, and helped us to identify the

9 For the transliteration we used the bi-directional transliterator for Python available at <https://github.com/barseghyanartur/transliterate>.

low-quality text spans and exclude them from further analysis. The rest of the words were primarily correct compound words.

Exploratory Analysis of the Corpus

In order to better understand the thematic structure of the extracted spans, we used k -means clustering.¹⁰ The k -means algorithm is traditionally used in exploratory data analysis to cluster the elements by a given similarity criterion into k distinct groups. We ran three experiments characterized by different span representations: n -grams of varying length, Tf-Idf, and Tf-Idf with lemmatized tokens. In all the experiments, we used the Euclidean distance as metrics and an elbow method to calculate the optimal number of clusters.

In n -gram-based experiments, we used n -grams of varying lengths, ranging from two to four. This approach is appropriate as it is not based on specific vocabulary and can adapt to various OCR errors. However, the results obtained in these experiments could not be interpreted easily. Still, the investigation was very informative as we singled out frequent n -grams that reflect OCR errors and that can be used for text quality improvement in the future.

Using Tf-Idf representations without prior lemmatization, we obtained the vocabulary with 9,330 tokens. All the tokens were lowercase and all punctuation marks were eliminated, as well as those tokens with fewer than three occurrences or more than 80 % occurrences. Analyzing the tokens with fewer than three occurrences, we noticed additional OCR errors and the morphological forms that refer to the same lemma. As the number of the tokens is quite large – close to 50,000 – once again we confirmed the importance of OCR improvement and the use of lemmatization. The tokens appearing in more than 80 % of spans were also manually analyzed and used for the extension of the final stopwords list. For example, indefinite pronouns such as “svaki,” “neki,” “ovako” and “ovo” were extracted from this list.

Following the exact setup, we used Tf-Idf representations with lemmatization.¹¹ The resulting vocabulary formed a total of 5,993 tokens. The lists of stopwords and out-of-vocabulary words were still quite large, with close to 35,000 words in total. We also partially manually examined the content of these lists and incorporated the conclusions into the extension of the final stopwords list and more fine-grained text cleaning practices.

For cluster visualization, we used word clouds.¹² Observing the visualization, we concluded the following:

10 We used Python library *scikit-learn* available at <https://scikit-learn.org/stable/>.

11 For lemmatization and part-of-speech tagging we used Classla Python package available at <https://pypi.org/project/classla/>.

12 We used Python library WordCloud available at https://amueller.github.io/word_cloud/index.html.

Multidisciplinary Considerations

Changes in the Discourse between 1878 and 1941

In this section, historical changes to the discourse in music criticism in Serbian between 1878 and 1941 are analyzed, based on the data gathered through the process described in the previous sections of the study. The purpose is to consider whether the canon that is shaped through traditional musicological research is reinforced or contradicted by the quantitative analysis of the selected digital corpora.

Regarding the canon, we rely on the work of Roksanda Pejović, the most prolific researcher of Serbian music criticism from the historiographic perspective,¹³ who identified certain fundamental changes occurred throughout the observed historical period, between 1878 and 1941. As a result, she asserts that there are notable differences between the approaches of the long nineteenth century, and the interwar period, respectively.¹⁴

Drawing on the wide selection of material at her disposal, and the methodology of the “close reading” of journals and newspapers, which is typical of musicological research,¹⁵ Pejović pointed out that “Serbian music criticism was in its infancy in the nineteenth century.”¹⁶ In her opinion, the critics of this period were mostly journalists with only modest musical knowledge, and their approach to music relied on theatrical criticism; otherwise, many of them simply repeated the words of authoritative commentators on music. In both cases, the emphasis was on the cultural event itself, while expert assessment was either missing or disregarded. Musical amateurs expressed national ideas and coloured their language with romantic nuances, while also taking care to enlighten and instruct their readers.

On the other hand, after the First World War, important changes were introduced through the professionalization of music critics, and Yugoslav musical

13 Roksanda Pejović published twelve books and many articles on musical performance and criticism in Serbian covering the period between the nineteenth and early twenty-first centuries. Some of them are: Roksanda Pejović, *Srpsko muzičko izvođaštvo romantičarskog doba* (Beograd: Univerzitet umetnosti, 1991); Roksanda Pejović, *Kritike, članci i posebne publikacije u srpskoj muzičkoj prošlosti (1825–1918)*, (Beograd: Fakultet muzičke umetnosti, 1994); Roksanda Pejović, *Opera i balet Narodnog pozorišta u Beogradu (1882–1941)*, (Beograd: [s. n.], 1996); Roksanda Pejović, *Muzička kritika i esejistika u Beogradu (1919–1941)*, (Beograd: Fakultet muzičke umetnosti, 1999); Roksanda Pejović, *Koncertni život u Beogradu* (Beograd: Fakultet muzičke umetnosti, 2004).

14 The borders between these two periods are permeable: certain elements of the new professional approach may be noted before 1914, and *vice versa* – certain relics are present in the interwar period.

15 We refer here to the concept of “distant reading” created by Franco Moretti, based on quantitative and computational methods, and used to study British novels of the eighteenth and nineteenth centuries. He introduced “distant reading” as an opposition to the practice of the “close reading,” which implies the careful analysis of the text, where attention is paid to critical interpretation and significant details. Cf. Franco Moretti, *Distant Reading* (London; New York: Verso, 2013).

16 Pejović, *Kritike, članci i posebne publikacije*, 2.

criticism came much closer to the typical style found in European journals and newspapers.¹⁷ The most important goal of the writers was, as Aleksandar Vasić noted in his study dedicated to Serbian music criticism in the first half of the twentieth century, related to *The Serbian Literary Magazine* (*Srpski književni glasnik*, 1901–1914, 1920–1941), to educate readers and to include the analytical perspective in their texts.¹⁸ The change in the profile of the critics is also important: instead of the journalists and mostly musical amateurs found during the previous period, in the interwar times music criticism was in the hands of professional musicians as well as educated intellectuals: literary critics, poets and writers, historians, and even lawyers.¹⁹

In the following table we offer a summary of some of the basic differences between the two periods in question.

Table 1: A comparative overview

	The Long Nineteenth Century	The Interwar Period
Critics	Amateurs, journalists specialized in theatre criticism	Professionals and educated intellectuals (writers, literary critics and historians)
The focus of critical writing	The cultural event	The musical piece and performance
Approach to the musical event	Descriptive	More analytical, with moderate and appropriate use of musical-technical attitudes
The typical language	Showing “romantic turns”	A literary style

Inevitably, a lot of questions arise regarding the computational exploration of the canon. Do digital corpora confirm changes in the type of description between the two periods, from mainly descriptive, generally nontechnical to a more music-specific and musically competent style of writing, or do they destabilize previous findings? What kind of aid can exploit the resources of digital newspaper archive offer to the qualitative research that has already generated certain conclusions? At this point in our research, bearing in mind all the methodological concerns and limitations imposed by the text quality, the main contribution which moves beyond the traditional approach is the ease with which digital corpora allow the searching of content and the increased breadth of the access available.

17 Pejović, *Muzička kritika*, 9.

18 Aleksandar Vasić, “Serbian Music Criticism in the First Half of the Twentieth Century: Its Canon, Its Method and Its Educational Role,” *Muzikologija*, no. 8 (2008): 202.

19 Pejović, *Muzička kritika*, 8.

To illustrate this point, we referred to the content analysis of the assembled digital corpus and conducted a search for the term *harmony*, including adjectives and adverbs related to this word. Critical comments related to this particular term were chosen because the competences related to music harmony are among those parameters that are demarcating two analyzed periods. As mentioned, music writers in the long nineteenth century had modest music-theoretical knowledge, and it is expected that their comments related to harmony are mostly of general nature. On the contrary, it is presumable that representatives of the interwar period demonstrated more sophisticated knowledge on the music harmony in their critical writings.

The term harmony appears seven times in the analyzed material: twice in the first, and five times in the second phase. In terms of the Serbian historical language, several variants of these terms were used: not just the adjective “harmonski,” but also “harmonijski” and even “harmoničeski” with all the propagations of the term. However, the word can have multiple meanings that are not differentiated by the NPL methods and tools: (1) it can pertain to choral singing, mostly related to the musical practice of the Orthodox Church,²⁰ (2) it can demonstrate performance qualities, in terms of a united and well-rehearsed ensemble (in the analyzed cases this meant choirs) and (3) a musical “vertical,” in accordance with the common understanding of the term. Examples of the first understanding are not present in the digital resources covering the selected time span, resulting in the exploration of the second and the third use of the word harmonija.

The following citation comes from the critical comment on the participation of the student choir in the divine service in the Church of the Ascension (Vaznesenjska crkva) in Belgrade in 1890. According to the anonymous critic, the music for the *Liturgy of St. John Chrysostom* composed by Kornelije Stanković was sung with great precision, in a church building full of faithful citizens.

*We are pleased to emphasize that only this liturgical music based on the folk tunes can arouse a feeling of piety among Serbs. Krestu tvojemu poklanjajemsja vladiko [At the Most Holy Cross of our Savior] is sung so beautifully in the traditional spirit and full, appropriate harmony (especially the bass was prominent), that one cannot stop listening to it.*²¹

Obviously, the romantically inclined critic was attracted by the coordination of the voices and musical skills of the group of performers; the syntagma

20 This is mostly related to terms “harmoničeski” and “harmonijski”. Cf. Ivana Perković Radak, *Od andeoskog pojanja do horske umetnosti: Srpska horska crkvena muzika u periodu romantizma (do 1914. godine)*, (Beograd: Fakultet muzičke umetnosti, 2008), 5–8.

21 “Sa zadovoljstvom naglašujemo, da samo ova narodna služba može u Srbinu pobuditi osećaj pobožnosti. – ‘Krestu tvojemu poklanjajemsja vladiko’ je tako divno otpevano u narodnom duhu i punoj, prikladnoj harmoniji (osobito se basovi pokazao), da se dovoljno naslušati ne mogosmo.” Anon., “Beogradske vesti,” *Male novine* (March 15, 1890): 3, http://istorijskenovine.unilib.rs/view/index.html#panel:pp|issue:UB_00031_18900315|page:3|query:otpevano.

appropriate (“prikladna”) *harmony*, particularly bearing in mind the comment on the bass part, is not related to the “vertical” musical component, but to the expressive and unified sound of the choral ensemble.

The second example is the critical review of the Academic Choral Society Obilić performance in February 1921. The music critic, Petar Krstić, offered a following comment:

Njest svjat [No One is as Holy] by Mokranjac and *Svjati Bože* [Holy God] by Hristić were performed, which sounded a bit thin in the “Obilić” choir, then *Romarska popevka* by Štolcer for a female choir and tenor, done homophonically, with interesting harmony and dynamics [...] Štolcer’s *Ej sastali se čapljinski Tatari* [Hey, the Tatars of Čapljina Met] (a folk song from Southern Serbia) was in unison, too, so that the composer did not offer anything of his own, except for the tones held in certain parts. Both the piano accompaniment and the harmonic arrangement are too primitive.²²

As previously mentioned, during the first phase the word *harmony* was used less often than later to denote the compatibility of the members of the musical ensemble (“appropriate harmony”) or general musical euphony. In the interwar period, the word “harmony” with its propagations was used more often, and its meaning mostly resonates with the contemporary implications of the word. As expected, and shown in the Krstić’s quotation, writers are commenting on harmony as it related to musical pieces much more often than in terms of performance. We presume that similar search of specialized musical terms, just like “harmonija,” would gain similar results.

In addition to the phenomenon of the word *harmony*, we explored the qualifiers, the distinct tokens, in the text spans for both periods and compared their frequencies. In our corpus, the long period is less covered by text spans with total of 290 spans. The interwar period is more elaborated, with 1,698 text spans. We used the part-of-speech tagger in order to single out the list of nouns and adjectives/adverbs related to the query term as appropriate. The numerous OCR errors propagated through the pipeline resulting, again, in approximate tags. Despite this fact, we managed to single out the qualifiers “besprekorno” (flawlessly), “solistički” (soloist), “filharmonija” (philharmonic), “tehnika” (technique), “homofono” (homophonic), “zvučno” (sonant) and “ton” (note of a certain pitch) as well as the voice types tenor, baritone, and bass that uniquely characterize the interwar period.

22 “Od crkvenih i pobožnih pesama izvođene su ‘Njest svjat’ od Mokranjca, ‘Svjati Bože’ od Hristića, koje je u horu ‘Obilića’ zvučalo nešto mršavo, zatim ‘Romarska popevka’ od Štolcera za ženski hor i tenor, rađena homofono, harmonski i dinamički zanimljiva [...] Štolcerova ‘Ej sastali se čapljinski Tatari’ (narodna pesma iz Južne Srbije) i suviše je unisono držana, tako, da kompozitor nije uneo ničeg svoga, do držanih tonova u pojedinim glasovima. I klavirska pratnja i harmonska obrada je suviše primitivna.” P. J. Kr., “Koncert Akademskog pevačkog društva ‘Obilić,’” *Pravda* (February 23, 1927): 5, http://istorijskenovine.unilib.rs/view/index.html#panel:pp|issue:UB_00042_19270223|page:5|query:otpevane.

Our findings confirm the previous findings that the nineteenth century approach was more informative and educational in nature, while music critics from the time of the first Yugoslavia paid more attention to the technical and stylistic layers of the musical pieces and performances. In other words, the corpus-wide analysis confirmed the musicological canon, or, as Estelle Joubert in her study dedicated to the “distant reading” practice of French music criticism, noticed, “it seemed to reinforce it.”²³ However, her suggestion that certain future machine learning projects may offer further opportunities for future quantitative research in music criticism is completely reasonable and equally applicable in terms of the findings of our research.

The Emotional Impact

Unfortunately, it was not possible to apply NLP methods to the sentiment analysis of music critical writings in Serbian, as a result of the many errors caused by the OCR tool, as already explained, and the domain shift of the existing models for sentiment analysis. It would be interesting to deal with the computational analysis of the emotions expressed in the text spans and the accompanying vocabulary. The exploration of emotional keywords and adjectives, used by Serbian critics in their writing on music, would be beneficial in a number of ways²⁴ but – regrettably – the current state of the created corpora does not provide sufficient opportunities for this type of research.

As is widely known, emotions are, as Juslin clearly states, regarded as the key aspect of the musician’s or listener’s “experience, even the main motive for listening to music.”²⁵ Even if the use of the corpus for research in this field is limited in the current phase of our research, it is intriguing to offer a glance at some of the music-induced emotions described and commented upon by musical critics who wrote in Serbian under the influence of the Romantic movement.

The collected data provided an interesting starting point for identifying evidence for one of the most fascinating themes in music research: the topic of music-induced tears/crying. As is well known, one of the most common strong experiences related to music is expressive behaviour and a physical reaction to music in the form of tears.²⁶ Tears, as Gabrielsson notes, are one of the most frequent physiological reactions, varying from a mild response (moist eyes) to

23 Estelle Joubert, “Distant Reading’ in French Music Criticism,” *Nineteenth-Century Music Review* 19, no. 2 (2021): 291–315, DOI:/10.1017/S1479409820000476.

24 Cf. Yao Liang and Hu Wang, “Sentiment Analysis of Music Criticism Based on Data Mining,” *Advances in Social Science, Education and Humanities Research* (Atlantis Press, 2018): 368–371, DOI:/10.2991/iceemr-18.2018.84.

25 Patrik N. Juslin, *Musical Emotions Explained: Unlocking the Secrets of Musical Affect* (New York: Oxford University Press, 2019), 8.

26 Alf Gabrielsson, “Emotions in Strong Experiences with Music,” in *Music and Emotion: Theory and Research*, Series in Affective Science, edited by Patrik N. Juslin and John A. Sloboda (New York: Oxford University Press, 2001), 549, DOI:/acprof:oso/9780199230143.001.0001.

intense crying. Moreover, it is interesting that tears can occur both in terms of showing positive and negative emotions: positive because of the quality of the music or the performance, and negative when “the music was associated with a sad event.”²⁷ Our selection was formed by several factors: the particular topic is underexplored (the most widely researched emotions are happiness, sadness, fear, valence, and arousal)²⁸ not only in the framework of national musicological research, but also from the historical perspective. It is also provocative from the contextual point of view, bearing in mind the complex historical, social and cultural situation in the selected time span.

Tears, as Sloboda points out, “may relate to emotions provoked by endings (whether of loss or the relief), and the precipitating musical structures may be those which encourage the listener to anticipate an impending ending or release of tension.”²⁹ However, nine examples found in our repository do not relate to the musical structure at all, but rather to extra musical reasons: religious belonging (three examples), the political importance of certain people (two examples), singing with a “gusle” accompaniment (two examples), national pride (one example) and institutional belonging (one example). However, the broadening of the research corpus might bring certain other results that would further confirm Sloboda’s thesis.

As an example, a critical comment on the concert given by Stevan Deskašev in Skoplje in 1896 illustrates well the idea of musical criticism as an agent of Romantic nationalism:

*Mr. Deskašev opened the concert with the Ottoman anthem, which was interrupted by a loud Padishah Chok Yasha! [Long live the Emperor!]. The crown of our artist’s concert was when he sang Serbian folk songs at the end. His patriotic act was very important for Serbian issues. Especially when he sang Gusle moje, the Serbs in the audience cried.*³⁰

Romantic lyrics written by the famous Serbian poet Branko Radičević in his poem dedicated to the national instrument – the “gusle” (the single stringed bowed instrument used for the accompaniment of Serbian epic poetry) on

27 Alf Gabrielsson and Siv Lindström Wik, “Strong Experiences Related to Music: A Descriptive System,” *Musicae Scientiae* 7, no. 2 (2003): 170, DOI:10.1177/102986490300700201.

28 Kazuma Mori and Makoto Iwanaga, “Two Types of Peak Emotional Responses to Music: The Psychophysiology of Chills and Tears,” *Scientific Reports* 7, no. 1 (2017): 460–463, DOI:10.1038/srep46063.

29 John A. Sloboda, “Music Structure and Emotional Response: Some Empirical Findings,” *Psychology of Music* 19, no. 2 (1991): 120, DOI:10.1177/0305735691192002.

30 “G. Deskašev je otvorio koncerat otomanskom himnom, koja je sa gromkim ‘Čok-jaša Padišah!’ usklikom prekidana. Kruna koncerta našeg umetnika beše, što je na svršetku otpevao srpske narodne pesme. Ovim njegovim patriotskim činom mnogo je koristio srpskoj stvari. Naročito kod stavke ‘Gusle moje’ Srbi su plakali.”

Anon., “Koncert St. Deskaševa u Skoplju,” *Pozorište* (October 3, 1896): 150, http://istorijskenovine.unilib.rs/view/index.html#panel:pp|issue:UB_00128_18960310|page:2|query:otpevao.

Ivan Zajc's music, caused a strong emotional response: the lyrics and music induced tears in the Serbian audience.

Former members of "Obilić," many of them already gray haired, mingled with current members, students, and sang the old national song *Hej, trubaču!* (*Hey, trumpet player!*). This song, the anthem of the choral society "Obilić," symbolically connected the old and new generations of "Obilić's people," caused tears in the eyes of many.³¹

In this case, the sense of belonging to the Obilić Academic Choral Society, founded in 1884, was reinforced by the performance of one of the distinguished conductors of the society, Josif Marinković, the composer of the patriotic song *Hej, trubaču!* The strong relationships among the members and the continuity of the institution, paired with the performance of Marinković's piece, resulted in music-elicited tears.

Enrichment of Factual Knowledge

The NLP analysis of digitized newspapers uncovered certain facts and trends that had not previously been known or observed in musicological studies. Moreover, gathering new evidence of the musical practice in Serbian society in the selected historical time span is one of the most important contributions of this type of research. It is precisely this which expands our knowledge, not just at the level of music-historical facts, but as a basis for either drawing certain new and unique conclusions for particular areas of research or else rethinking various canonical topics.

A few examples will illustrate our point: two of them based on a "close" and one on a "distant" reading of the digitized repository.

In the description of the Belgrade "slava"³² (patron feast) celebrated on 10 June 1926, an Orthodox procession ("litije") was organized after the solemn church service:

*At around eleven o'clock, a procession was formed to pass through the town. The Patriarch walked under the canopy, and the clergy in front of him. At the end of the procession was military music that played church hymns throughout.*³³

31 "Bivši članovi 'Obilića', mnogi među njima već sedi, pomešali su se sa sadašnjim članovima, studentima i studentkinjama, i otpevali staru nacionalnu pesmu 'Hej, trubaču!'. Ova pesma, himna pevačkog društva 'Obilić', koja je simbolično povezala stare i nove generacije 'Obilićevaca' izazvala je mnogima suze na oči."

Anon., "Slave beogradskih društava," *Beogradske opštinske novine* (1938): 9, http://istorijskenovine.unilib.rs/view/index.html#panel:pp|issue:UB_00005_19381101|page:65|query:otpevali.

32 *Krsna slava* is a particular Serbian religious celebration (a Serbian family's patron saint day). Many families have their *slava*, as a memory to the baptism day of their predecessors, but there are also *slavas* of institutions and settlements – cities or villages.

33 "Okolo jedanaest časova formirana je litija za prolazak kroz varoš. Pod nebom je išao Patrijarh, a ispred njega sveštenstvo. Na kraju povorke išla je vojna muzika koja je za celo vreme svirala crkvene arije."

The information that a military orchestra played church hymns, in the presence of the Patriarch and representatives of the senior Orthodox Church hierarchy attracts our attention. It is well known that Orthodox music is purely vocal, and that not a single musical instrument is used in church practice. Even if the town procession did not have such strict rules as the Orthodox rites, it still surprises us that an orchestra took part in it by playing ecclesiastical tunes. What kind of melodies were played by the military musicians? How did the Patriarch and priests respond to this? Was it the regular practice in the 1920s in Belgrade or not? All these topics are worth further exploration using NLP research methods based on broader digital resources or the “close reading” approach.

A different, but also interesting finding is the caricature of the Belgrade quartet drawn by Pjer Križanić in 1931 published in the Belgrade newspapers.³⁴ The members of the quartet were teaching staff at the Music School: the violinist Marija Mihailović, the viola player Jovan Zorko, the cello player Juro Tkalčić and the pianist Ćiril Ličar. They were one of the most prominent chamber ensembles in Serbian musical life during the eight years of their existence (1925–1933).³⁵ As far as we know, music historians have not previously identified this caricature, which is proof of their widespread popularity.

Another topic, where a “distant reading” approach would give valuable results, is the prominent presence of female performers, particularly choirs, in the musical life of the time. There are 58 references to female choirs in our dataset, mostly in terms of performers or, much less frequently, intended ensemble of the piece. School female choirs are notably present, whether in celebrations of a school *slava* where they took part in the church ritual, or on other occasions. Bearing in mind the common opinion that female performers and choirs were less visible, due to the historical social norms, the topic is worth further research. As an example, a short passage from the text devoted to a concert in honour of St. Sava, the first Serbian archbishop and the school patron, held at the Teacher Training School (Preparandija) in Sombor in 1889 is given. The second piece on the programme was Robert Tolinger’s

Anon. “Litija kroz Beograd,” *Vreme* (June 11, 1926), http://istorijskenovine.unilib.rs/view/index.html#panel:pp|issue:UB_00043_19260611|page:5|query:%D0%BB%D0%B8%D1%82%D0%B8%D1%98%D0%B0%20%D0%BA%D1%80%D0%BE%D0%B7%20%D0%B1%D0%B5%D0%BE%D0%B3%D1%80%D0%B0%D0%B4%20%D0%B7.

34 Kulundžić Zvonimir, “Karikaturista Pjer Križanić,” *Beogradske Opštinske Novine* 1 (1940): 89, http://istorijskenovine.unilib.rs/view/index.html#panel:pp|issue:UB_00005_19401001|article:pageDiv84|page:89|query:%D0%BF%D1%98%D0%B5%D1%80%20%D0%BA%D0%B2%D0%B0%D1%80%D1%82%D0%B5%D1%82%20%D1%82%D0%BA%D0%B0%D0%BB%D1%87%D0%B8%D1%9B.

35 Cf. Roksanda M. Pejović, “Kamerno muziciranje i njegovi zastupnici,” *Novi zvuk – internacionalni časopis za muziku*, no. 21 (2003): 95–103.



Figure 6: Krizanić's caricature of the Belgrade quartet.

choir *Ljubičice* (*Violets*) composed to lyrics by Jovan Grčić Milenko. “This song was sung by a female choir, and it can be said with full authority that they are not lagging behind the male choir at all. Tolinger’s art is vividly reflected in this piece as well.”³⁶

In addition to the female choirs, the corpus contains evidence of solo performances where female performers are referenced in various ways as “umetnice” (artists), “pijanistkinje” (pianists), “pevačice” (singers), or just with a title “gospođica” (Miss) or “gospođa” (Madam). The graph with the relevant span distribution is shown below.

The comparison with the male choir in terms of the quality of the performance is the guideline for future research of the musicological canon related to female ensembles in the long nineteenth century.

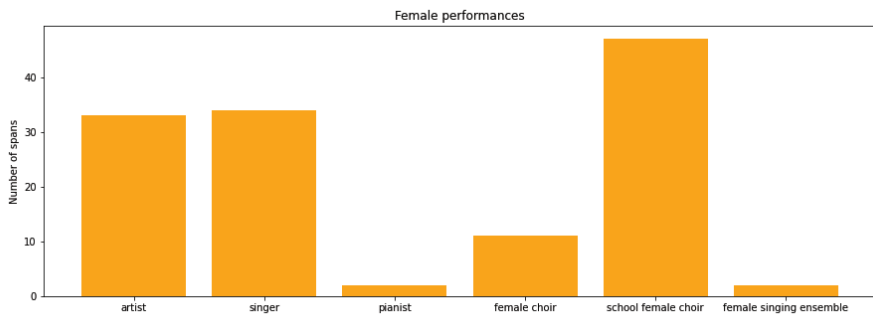


Figure 7: Statistics of female performances.

36 “Ovu je pesmu ženski lik otpevao i s punim pravom kazati se može da ni malo nije ustupio muškima. I u ovom komadu živo se ogleda umetništvo Tolingerovo.” IVAN., “Svetosavska Beseda u Somboru,” *Školski list* 15 (1889): 31, http://istorijskenovine.unilib.rs/view/index.html#panel:pa|issue:UB_00015_18890215|article:div115|query:otpevao.

Conclusion

The digital corpus selected for this study, after the procedure of the narrowing down of resources on critical writing in newspapers in Serbian, included 3,751 text spans in 3,442 pages available via the official web application of the Svetozar Marković University Library in Belgrade. The process of multidisciplinary collaboration aiming to connect the quantitative analysis of the digital resources and musicological research on the music criticism written between 1878 and 1941, showed that there are many challenges, but also great opportunities for both disciplinary areas. Even if the preparation of digital material was demanding and required a lot in terms of effort and time, our results encourage and support further study.

Some of the primary issues related to the creation of the corpus are the selection of adequate initial search queries, challenges in newspaper page segmentation, (too) many OCR errors, and consequently, out-of-vocabulary words that influenced the accuracy of the NLP tools used. Due to the corpus size, the intention was to perform as many as tasks possible automatically and in an unsupervised manner.

From the musicological point of view, narrowing the research terms was one of the challenges, as well as finding the right balance between broader concerns brought by distant reading methods and the intense examination of certain topics that came as a result of close reading. Furthermore, the abundance of available digital material was another concern, in terms of defining the most appropriate focus.

However, the main goal of the study showed that there are multiple and creative ways to combine computational and musicological approach to assess music criticism in Serbian, so as to identify new approaches to the topic.

The community interested in NLP and the Digital Humanities might find interesting the pipeline for the creation of sub-corpora on top of the reach and undiscovered resources provided by the Svetozar Marković University Library. Fortunately, there are initiatives³⁷ related to the correction of text after OCR, the restoration of diacritics, and switching between the different language variants for Serbian based on the cascades of finite state transducers and e-dictionaries, as well as more machine-learning guided approaches for other languages³⁸ which we would like to investigate and deepen in the future. In addition, sub-corpora related to music and music practices represent a vital starting point for the systematization and creation of the music-related language resources necessary for the analysis of both past and present music sources.

37 Cvetana Krstev and Ranka Stanković, "Old or New, We Repair, Adjust and Alter (Texts)," *Infoteka* 19, no. 2 (2019): 61–80, DOI:/10.18485/infoteka.2019.19.2.3.

38 Thi Tuyet Hai Nguyen et al., "Survey of Post-OCR Processing Approaches," *ACM Computing Surveys* 54, no. 6 (2021): 1–37, DOI:/10.1145/3453476.

Musicological analysis of the selected corpus directed towards discourse changes in the selected historical periods (before the First World War and in the interwar times) offered the possibility to reinforce the traditional findings that the nineteenth century approach was more informative and educational in nature, while in the interwar period more attention was paid to stylistic and technical aspects, as well as the layers of musical performances and pieces. Furthermore, even if the current state of the digital material does not provide sufficient opportunities for research in broader “strokes,” the collected data provided interesting prospect for one of the most fascinating themes in music research: the topic of music induced tears/crying. However, it was explained that, in the available critical writing in Serbian, strong emotions were stimulated by the feeling of belonging (whether national and religious), more than by the music itself. Finally, experiments with NLP methods demonstrated exciting possibilities for the gathering of new evidence regarding musical practice in Serbian society between 1878 and 1941. Although further work is necessary to develop/refine the computational tools designed for music-related texts – particularly in the field of NLP – and to create new open access repositories of digitized newspapers, this study has shown that it is not possible to deny the benefits of the digital endeavours in the field of musicology.

References

- [Anon.] “Beogradske vesti.” *Male novine* (March 15, 1890).
- [Anon.] “Koncert St. Deskaševa u Skoplju.” *Pozorište* (October 3, 1896).
- [Anon.] “Litija kroz Beograd.” *Vreme* (November, 1926).
- [Anon.] “Slave beogradskih društava.” *Beogradske opštinske novine* (November, 1938).
- Gabrielsson, Alf. “Emotions in Strong Experiences with Music.” In *Music and Emotion: Theory and Research*, 431–449. Series in Affective Science. New York: Oxford University Press, 2001.
- Gabrielsson, Alf, and Siv Lindström Wik. “Strong Experiences Related to Music: A Descriptive System.” *Musicae Scientiae* 7, no. 2 (2003): 157–217. DOI:10.1177/102986490300700201.
- IVAN. “Svetosavska beseda u Somboru.” *Školski list* 15 (1889).
- Joubert, Estelle. “‘Distant Reading’ in French Music Criticism.” *Nineteenth-Century Music Review* 19, no. 2 (2021): 291–315. DOI:10.1017/S1479409820000476.
- Juslin, Patrik N. *Musical Emotions Explained: Unlocking the Secrets of Musical Affect*. New York: Oxford University Press, 2019.
- Kr., P. J. “Koncert Akademskog Pevačkog Društva ‘Obilić.’” *Pravda* (February 23, 1927).
- Krstev, Cvetana, and Ranka Stanković. “Old or New, We Repair, Adjust and Alter (Texts).” *Infotheca* 19, no. 2 (2019): 61–80. DOI:10.18485/infotheca.2019.19.2.3.
- Kulundžić Zvonimir. “Karikaturista Pjer Križanić.” *Beogradske Opštinske Novine* 1 (1940).
- Liang, Yao, and Hu Wang. “Sentiment Analysis of Music Criticism Based on Data Mining.” *Advances in Social Science, Education and Humanities Research* (2018): 368–371. DOI:10.2991/iceemr-18.2018.84.

- Ljubešić, Nikola, and Kaja Dobrovoljc. "What does Neural Bring? Analysing Improvements in Morphosyntactic Annotation and Lemmatisation of Slovenian, Croatian and Serbian." In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*. Association for Computational Linguistics, 2019: 29–34. DOI: 10.18653/v1/W19-3704.
- Moretti, Franco. *Distant Reading*. London; New York: Verso, 2013.
- Mori, Kazuma, and Makoto Iwanaga. "Two Types of Peak Emotional Responses to Music: The Psychophysiology of Chills and Tears." *Scientific Reports* 7, no. 1 (2017): 460–463. DOI:10.1038/srep46063.
- Nguyen, Thi Tuyet Hai, Adam Jatowt, Mickael Coustaty, and Antoine Doucet. "Survey of Post-OCR Processing Approaches." *ACM Computing Surveys* 54, no. 6 (2021): 1–37. DOI:10.1145/3453476.
- Pejović, Roksanda. "Kamerno muziciranje i njegovi zastupnici." *Novi zvuk – internacionalni časopis za muziku*, no. 21 (2003): 95–109.
- Pejović, Roksanda. *Koncertni život u Beogradu*. Beograd: Fakultet muzičke umetnosti, 2004.
- Pejović, Roksanda. *Kritike, članci i posebne publikacije u srpskoj muzičkoj prošlosti (1825–1918)*. Beograd: Fakultet muzičke umetnosti, 1994.
- Pejović, Roksanda. *Muzička kritika i esejistika u Beogradu (1919–1941)*. Beograd: Fakultet muzičke umetnosti, 1999.
- Pejović, Roksanda. *Opera i balet Narodnog pozorišta u Beogradu (1882–1941)*. Beograd: [s. n.], 1996.
- Pejović, Roksanda. *Pisana reč o muzici u Srbiji: Knjige i članci (1945–2003)*. Beograd: Fakultet muzičke umetnosti; Signature, 2005.
- Pejović, Roksanda. *Srpsko muzičko izvođaštvo romantičarskog doba*. Beograd: Univerzitet umetnosti, 1991.
- Perković Radak, Ivana. *Od anđeoskog pojanja do horske umetnosti: Srpska horska crkvena muzika u periodu romantizma (do 1914. godine)*. Beograd: Fakultet muzičke umetnosti, 2008.
- Sloboda, John A. "Music Structure and Emotional Response: Some Empirical Findings." *Psychology of Music* 19, no. 2 (1991): 110–120. DOI:10.1177/0305735691192002.
- Soper, Elizabeth, Stanley Fujimoto, and Yen-Yen Yu. "BART for Post-Correction of OCR Newspaper Text." In *Proceedings of the 7th Workshop on Noisy User-generated Text (W-NUT 2021)*. Association for Computational Linguistics, 2021: 284–290. DOI:10.18653/v1/2021.wnut-1.31.
- Vasić, Aleksandar. "Serbian Music Criticism in the First Half of the Twentieth Century: Its Canon, Its Method and Its Educational Role." *Muzikologija*, no. 8 (2008): 185–202.
- Weliver, Phyllis, and Katharine Ellis, eds. *Words and Notes in the Long Nineteenth Century*. Woodbridge, UK; Rochester, NY: Boydell & Brewer, 2013.

POVZETEK

Računalniška analiza glasbene kritike med letoma 1878 ter 1941 v srbskem jeziku: Kaj se lahko naučimo iz digitalnega arhiva zgodovinskih časopisov v Univerzitetni knjižnici Svetozarja Markovića?

Proces digitalizacije je vplival na veliko področij muzikoloških raziskav, vključno s področjem glasbene kritike. Vendar se računalniška analiza o recepciji glasbenikov, glasbenih del ter izvajalcev le-teh v srbsščini do sedaj ni izvajala zaradi različnih vzrokov. Ključni med njimi so: 1) omejena količina digitaliziranih virov in 2) zahteve t. i. cevovoda procesiranja naravnih jezikov (*NLP pipeline*) v okviru jezika glasbe in zgodovinskih virov. Multidisciplinarna raziskava, ki je bila izvedena za potrebe tega članka – po našem vedenju prva te vrste, ki se ukvarja s srbskimi viri – se je začela z definicijo arhivskega depoja; v glavnem je šlo za kritično pisanje, ki je zdaj na voljo v digitalnem formatu. Sama definicija je bila iterativna, vodila so jo načini podatkovne analize in prakse nenadziranega strojnega učenja, morala pa se je ujemati tudi z glavnim ciljem študije: raziskati, kako združiti priložnosti, ki jih ponuja računalniška kvantitativna analiza, s tradicionalnimi kvalitativnimi muzikološkimi metodami raziskovanja ter nakazati, kaj lahko to sodelovanje doprinese k raziskovalnim metodam glasbe, posebej v smislu potencialnih področij raziskav.

Ustvarjeni depo predstavlja sub-korpus digitaliziranih časopisov v srbsščini, ki ga je pripravila Univerzitetna knjižnica Svetozarja Markovića v okviru projekta *Europeana Časopisi*. Izbor relevantnih virov, pred-procesiranje in analiza teksta so bile zaznamovane s številnimi metodološkimi izzivi; nekateri so bili povsem tehnični in v povezavi z zbiranjem podatkov ter pridobivanjem informacij ter drugi, bolj računalniški, povezani z ocenjevanjem kvalitete pisanja tekstov, izboljšavami optičnega prepoznavanja znakov (OCR) in zmožnostjo orodij za procesiranje naravnih jezikov, da sledijo glasbeni terminologiji in kronološkim spremembam v jeziku. Na osnovi tega depoja tekstov, ki so bili objavljeni v 63 letih – večinoma v časopisih – med letom 1878 (datum neodvisnosti Srbije) ter začetkom druge svetovne vojne v Jugoslaviji (1941), smo kvalitativne ugotovitve združili z rezultati eksplorativne analize podatkov z namenom, da bi preučili nekatere izbrane teme, ki so relevantne za raziskovanje glasbene kritike.

Za potrebe tega članka so bile izbrane tri glavne teme, ki so najbolj relevantne v tradicionalnih muzikoloških raziskavah o glasbeni kritiki v srbsščini: 1) spremembe v diskurzu v izbranem zgodovinskem obdobju, 2) jezik, ki se nanaša na emocionalni učinek določene izvedbe oz. nove glasbene kompozicije, in 3) obogatitev našega dejanskega znanja o glasbeni zgodovini. Možne smeri nadaljnjega raziskovalnega dela ne izključujejo drugih tem in področij, ki se lahko izluščijo iz ogromne zbirke zbranega materiala; nasprotno – v prihodnosti se odpirajo možnosti širšega ter hkrati bolj pretanjenega uvida, odvisno od izbire raziskovalnega diskurza.

ABOUT THE AUTHORS

IVANA PERKOVIĆ (ivanaperkovic@fmu.bg.ac.rs) is a musicologist and professor at the Department of Musicology of the Faculty of Music, University of Arts in Belgrade. She is author and co-author of five books (on Serbian religious music, history of Serbian music, Faculty of Music, music and interdisciplinarity), over 60 articles and chapters in peer-reviewed journals and monographs. Her research fields include Orthodox, particularly Serbian, music; music references in literature; interdisciplinary approach to music; topics of Self and Other in the context of European 18th century music; digital musicology, etc. She is member of the: IMS, SMS, ISOCM, editorial board of the Matica Srpska *Journal of Stage Arts and Music*, the Commission for Acquiring Scientific Title of the Ministry of Education, Science and Technological Development (MESTD) of the Republic of Serbia, Secretary General of the European Association of Conservatoires (AEC). She is experienced in creating/leading many national and international academic and research projects.

ANĐELKA ZEČEVIĆ (andjelkaz@mi.sanu.ac.rs) is a Ph.D. candidate in the domain of Natural Language Processing at the Faculty of Mathematics, University of Belgrade. She is currently employed as a researcher at the Mathematical Institute of the Serbian Academy of Sciences and Arts.

O AVTORICAH

IVANA PERKOVIĆ (ivanaperkovic@fmu.bg.ac.rs) je profesorica na Oddelku za muzikologiju na Fakulteti za glasbo Univerze za umetnost v Beogradu. Je avtorica ali soavtorica petih knjig (o srbski verski glasbi, zgodovini srbske glasbe, glasbeni fakulteti, glasbi in interdisciplinarnosti) ter več kakor 60 člankov in poglavij v recenziranih revijah in monografijah. Njena raziskovalna področja so pravoslavna glasba (zlasti srbska), glasbene reference v literaturi, interdisciplinarni pristop k glasbi, tema Sebe in Drugega v kontekstu evropske glasbe 18. stoletja, digitalna muzikologija idr. Je članica nacionalnih in mednarodnih združenj, kot so IMS, SMS in ISOCM, članica uredniškega odbora *Revije za uprizarjalno umetnost in glasbo* Matice Srpske, komisije za pridobivanje znanstvenih nazivov na Ministrstvu za izobraževanje, znanost in tehnološki razvoj (MESTD) Republike Srbije ter generalna sekretarka Evropskega združenja konservatorijev (AEC). Ima izkušnje s pisanjem predlogov ter vodenjem številnih nacionalnih in mednarodnih akademskih in raziskovalnih projektov.

ANĐELKA ZEČEVIĆ (andjelkaz@mi.sanu.ac.rs) je doktorska kandidatka na področju procesiranja naravnega jezika na Fakulteti za matematiko Univerze v Beogradu. Trenutno je zaposlena kot raziskovalka na Inštitutu za matematiko Srbske akademije znanosti in umetnosti.