The Serbian Part of the ELTeC Collection through the Magnifying Glass of Metadata

Cvetana Krstev

cvetana@matf.bg.ac.rs

University of Belgrade
Faculty of Philology
Belgrade, Serbia

ABSTRACT: In this paper we present the metadata assigned to the Serbian ELTeC sub-collection, analyse them and draw some conclusions about titling practices of Serbian narrative literature in the period 1840-1920, authors' gender and age, publication places, modes of publication, as well as about institutions and individuals that are due credits for the production of SrpELTeC.

KEYWORDS: metadata, TEI header, Serbian Literature, ELTeC corpus.

1 Introduction

The practice of assigning descriptive data to some piece of information has a long history, ever since librarians started to produce catalogues. According to Hodge (2001, 3) "Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data or information about information". These descriptive data was not always called "metadata", but, as pointed by (Duval et al. 2002): "The rapid changes in the means of information access occasioned by the emergence of the World Wide Web have spawned an upheaval in the means of describing and managing information resources. Metadata is a primary tool in this work, and an important link in the value chain of knowledge economies." Metadata assigned to information of any kind brings additional value to them: being well described, this information can be more easily retrieved by different users, for various purposes, over a long period of time.

The need to assign metadata to literary works in digital form was recognized many years before the WWW upheaval mentioned in (Duval et al. 2002). From the very beginning, and that means the first published guidelines in the early 1990s, the TEI prescribed, as mandatory, the use of headers

for all TEI conforming documents (Burnard 2013). The TEI header was seen as the primary source of information about a digital source, corresponding to a certain extent to a cover page of a printed book, and consequently to a library catalogue record. In the following section, we will explain which metadata can be found in headers of literary works of SrpELTeC, and which aggregated information can be deduced from them.

2   Metadata Header

Since all texts in the ELTeC corpus are TEI documents, each of them has its own TEI header[1] – The ELTeC Header.[2] The content of the header was agreed by the action's Working Group 1 – Scholarly Resources/ELTeC (WG1) with the aim of making it informative, but at the same time simple.

Each ELTeC header has three obligatory elements:

- `<fileDesc>`, which describes the digital, that is ELTeC edition (element `<titleStmt>`), its size (element `<extent>`), availability and licensing (element `<publicationStmt>`), and source(s) from which it was derived (element `<sourceDesc>`);
- `<profileDesc>`, which gives additional information about the ELTeC edition, about the language the text was written in (element `<langUsage>`), and particularly about its characteristics that bear upon the balance of the whole sub-collection (element `<textDesc>`);
- `<revisionDesc>`, which records all changes of the document after its first publication.

The element `<sourceDesc>` contains one or more bibliographic descriptions (element `<bibl>`). The bibliographic description for the first edition is obligatory (attribute value `firstEdition`). However, if the first edition was not used to prepare the ELTeC edition, than additional `<bibl>` elements are necessary: `<sourceEdition>` for the print source or `<digitalSource>` for the digital source. Of course, more than two `<bibl>` elements can occur if, for instance, the ELTeC edition was derived from some digital edition that came from a print edition that was not the first one.

Here, we give, as an example, the ELTeC header of the novel *Jarani* (Friends) (SRP19132) from SrpELTeC; some lines are shortened, and some elements deleted to make it more readable.

---

1. About the TEI header in TEI P5 guidelines TEI header.
2. About the ELTeC

```
<teiHeader>
<fileDesc>
  <titleStmt>
    <title>Јарани : приповетка : ELTeC издање</title>
    <author ref="viaf:136335">
       Ћоровић, Светозар (1875-1919)
    </author>
    <respStmt>
      <resp>Сканирање</resp>
      <name>
        Универзитетска библиотека "Светозар Марковић"
      </name>
    </respStmt>
    <respStmt>
      <resp>OCR и корекција текста</resp>
      <name>Тања Ћулафић</name>
      <name>Цветана Крстев</name>
    </respStmt>
    <respStmt>
      <resp>Кодирање за ELTeC</resp>
      <name>Цветана Крстев</name>
    </respStmt>
  </titleStmt>
  <extent>
    <measure unit="words">47623</measure>
    <measure unit="pages">199</measure>
  </extent>
  <publicationStmt>
    <publisher ref="https://distant-reading.net">
      COST Action "Distant Reading ..." (CA16204)
    </publisher>
    <distributor ref="https://zenodo.org/communities/eltec/">
    Zenodo.org
    </distributor>
    <date when="2021-04-09"/>
    <availability> <!-- about licence --></availability>
    <!-- references to various releases -->
  </publicationStmt>
  <sourceDesc>
```

```
    <bibl type="printSource">
      <author>Ћоровић, Светозар</author>
      <title>Дјела. Књ. 2</title>
      <publisher>Свјетлост</publisher>
      <pubPlace>Сарајево</pubPlace>
      <date>1953</date>
      <idno type="COBISS-SR.ID">54425863</idno>
    </bibl>
    <bibl type="firstEdition">
      <author>Ћоровић, Светозар</author>
      <title>Јарани : приповетка</title>
      <pubPlace>Нови Сад</pubPlace>
      <publisher>Матица српска</publisher>
      <date>1913</date>
      <idno type="COBISS-SR.ID">16904711</idno>
    </bibl>
  </sourceDesc>
</fileDesc>
<profileDesc>
  <langUsage>
    <language ident="sr-Cyrl" usage="100">српски</language>
  </langUsage>
  <textDesc>
    <authorGender xmlns="http:...ns" key="M"/>
    <size xmlns="http:...ns" key="short"/>
    <reprintCount xmlns="http:...ns" key="low"/>
    <timeSlot xmlns="...ns" key="T4"/>
  </textDesc>
</profileDesc>
<revisionDesc>
  <change when="2021-04-09">
    Converted by checkUp script for new release
  </change>
  <!-- other changes -->
</revisionDesc>
</teiHeader>
```

3 The Analysis of Data

Each novel in the SrpELTeC sub-collection has its own metadata header, as presented in the previous subsection. When preparing headers, care must be taken to use all its obligatory elements and fill in correct and consistent data. The use of elements and their sub-elements in the correct way is checked by the appropriate XML Schema, but the header producer has to take care about the content. For instance, the author's name should be given in the reverse order, and dates of her/his birth and death in parentheses.

As a result, we obtained consistent headers, one of which is presented in the previous subsection. Developers of most of other language sub-collections within ELTeC use similar headers, while some contain additional information. For instance, the headers of French sub-collection use the element `<textClass>` (within `<profileDesc>`) to indicate the narrative perspective and subgenre. The headers of Slovene sub-collection use the forth sub-element of a TEI header, `<encodingDesc >`, to describe briefly the process of transformation of existing digital edition to the ELTeC edition.

In the following subsections we will present some conclusions that can be derived about the SrpELTeC sub-collection from metadata assigned to each novel.

3.1 Titles

Titles of novels in the SrpELTeC sub-collection were taken from the first editions, whenever possible.[3] It was noted that the titles of subsequent editions sometimes deviate from first editions. For instance, the novel *Seljanka* by Janko Veselinović (SRP18932) was first published in instalments in 1884 in the journal *Otadžbina* as *Seljanka. Roman* (The peasant woman. Novel), its first book edition in 1893 (used for scanning) was entitled *Seljanka – pripovetka iz seoskog života* (The peasant woman – a story from rural life). The subsequent editions used both variants as well as simply – *Seljanka*. In rare occasions the first edition titles had to be adapted to modern script, as was the case with the novel *Đurađ Branković* by Jakov Ignjatović (SRP18590) that still did not use the reformed alphabet: *Ђурађъ Бранковићъ : (историчный романъ)*.

In paper (Patras et al. 2021) authors analysed titles in eleven ELTeC sub-collections, most of which were not complete at the time the article was

---

3. If the first edition was scanned, then the title was taken as stated on the front page; otherwise we relied on the library catalogue.

written. The Serbian sub-collection contained 50 novels and some observations were:[4]

− Titles of Serbian novels used only location entities that could be traced on the map, e.g. *Novac: roman iz beogradskog života* (SRP1906), which was not the case for other language sub-collections;
− Serbian novels had more "long" titles (6 words or more) than other sub-collections, approximately 25%;
− In Serbian novels, as well as Slovenian and Ukrainian, the existence of persons occurring in titles is more often than not impenetrable, meaning that their "roles" in novels have to be discovered by readers, e.g. *Milan Narandžić* (SRP18630) and *Radetića Mara: pripovetka iz seoskog života* (Radetića Mara: a story from rural life) (SRP18940).

While it would be interesting to extend this research on titles to the complete ELTeC collection, we will make here only some observations. The average length of titles in the 100-novels Serbian collection is 5.19 words, with only 8 one-word titles, e.g. *Sanjalo* (Dreamer) (SRP18880), and as many as 40 novels having titles of 6 or more words. The longest title has 18 words *Zločin jedne svekrve : kriminalna pripovetka : iz skore prošlosti / po autentičnim podatcima i sudskim aktima napisao K. D. Jezdić* (The crime of a mother in law : a crime story : from the recent past : written on the basis of authentic data and court acts by K. D. Jezdić) (SRP19062). This title also has as many as three subtitles. Only 40 novels from the whole sub-collection do not have subtitles, 52 have one subtitle, and 7 two.

The subtitles usually contain genre indicators – 53 out of 60 subtitles – and these are: *roman* (novel) 25, *pripovetka/pripovijetka* (short story) 21, *novela* (novella) 3, *slika/slike* (sketches) 3, *crta* (outline) 1. Although one would expect that all indicators except *roman* point to shorter narratives of a fragmented nature it is not always the case. As we already saw, the indicator of *Seljanka* (SRP18932) changed from one edition to another. Also, the indicator for *Pop Ćira i Pop Spira* (Father Ćira and father Spira) (SRP18941) in the first edition is short story (*pripovetka*), while today nobody would consider this work anything but a novel (*roman*).

3.2 Authors

The works represented in the SrpELTeC subcollection are written by 66 different authors: one author is represented with 4 novels (Jakov Ignjatović),

---

4. The whole data set and visualizations used in (Patras et al. 2021) can be found in (Patras et al. 2020).

12 with 3 novels (Svetozar Ćorović, Jelena Dimitrijević, Vladan Đorđević, Draga Gavrilović, Lazar Komarčić, Čedomilj Mijatović, Milan Đ. Milićević, Svetolik Ranković, Stevan Sremac, Borisav Stanković, Milutin Uskoković, Janko Veselinović), 7 with 2 novels (Andra Gavrilović, Dragutin J. Ilić, Đura Jakšić, Tadija P. Kostić, Branislav Nušić, Pera Todorović, Ivo Ćipiko), and 46 with one novel (see Figure 1. This distribution does not fit perfectly into the criteria for balanced representation of authors; the reasons for that are discussed in paper (Trtovac, Milnović, and Krstev 2021) of this issue.[5]

In the SrpELTeC sub-collection there are no works with common authorship of several authors, while the author of one work is unknown – *Beogradske tajne* (Belgrade's secrets) (SRP18923). The author's name is not given in the printed edition, neither was his/her name identified later. Only 8 novels are written by (4 different) "authoresses": Draga Gavrilović, Jelena Dimitrijević, Milica Janković and Isidora Sekulić. When compared with other ELTeC collections, which are complete comprising 100 works, one can observe that in all of them women are better represented: while English collection has a perfect balance (51 female authors), and German, French and Polish have a high representation of female authors (33, 34, 42, respectively), in other collections female authors are not so well represented, but nevertheless, better than in SrpELTeC: in Czech 12, Hungarian 21, Portuguese 17, Romanian 16, and Slovenian 11.

Twelve authors (not counting one unknown author), out of 66 different authors, are not represented in VIAF.[6] The authors' birth and death years were taken form the library catalogues, and where they were missing, we tried to retrieve them from some older encyclopedias (Петровић 1937; Бихаљи-Мерин 1959a, 1959b); they were, however, rarely found. Two authors with unknown birth and death dates are recorded in VIAF (Stevan Mamuzić and Dimitrije Tasić), while this data is known for three authors with no record in this database: Panta Popović (1843-1918), Živojin Jovičić (1837-1908), Boško St. Petrović (1869-1913). For one author recorded in VIAF, Dušan Rogić, the year of birth is known (1855) while the year of death remains unknown.

The author who was born the earliest, in 1817, was Jovan Subotić, while the author who was born the latest was Mladen St. Đuričić, in 1889. The

---

5. The final edition of SrpELTeC differs slightly from the one presented here; data about novels and authors in SrpELTeC at the time this article was written as well as in the final edition can be found in (Krstev and Stanković 2021) in this issue.

6. The Virtual International Authority File
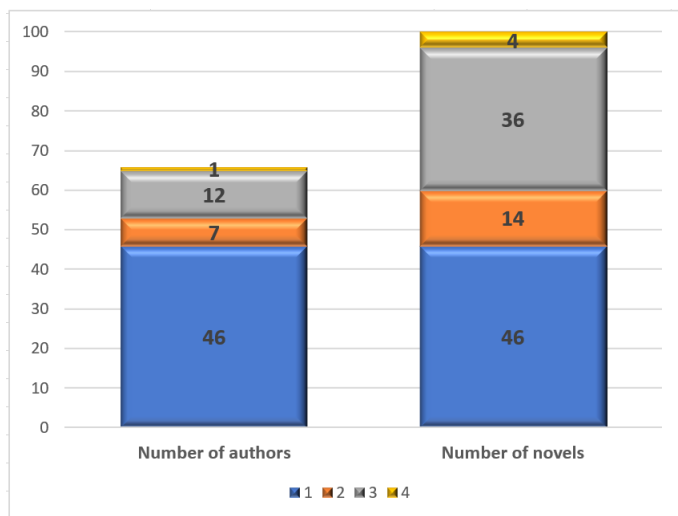
Figure 1. (a) Number of novels that are represented in SrpELTeC corpus by one to four novels; (b) number of novels by these authors in SrpELTeC.

author who died the earliest, in 1859, was Bogoboj Atanacković, while the author who died the latest is again Mladen St. Đuričić in 1987.[7] Milutin Usković, the author of three novels in SrpELTeC, lived the shortest – 32 years, while, as expected, Mladen St. Đuričić lived the longest – 99 years. The average life length of authors was 61.5 years.

The novel written at the youngest age of an author is *Kočina krajna* (Koča's frontier) by Vladan Đorđević (SRP18631), who was 20 years old at the time of the novel's publication. Other very young authors were Dragutin Ranković, who wrote the novel *Slavko* (SRP19023) at the age of 21, and Branko Mihajlović who wrote *Pred zoru* (Before dawn) at the age of 22. Stojan Novaković was the oldest when he published the novel *Kaluđer i hajduk : pripovetka o poslednjim danima Srbije u XV veku* (A monk and a

---

7. It is interesting to note that Mladen St. Đuričić wrote his novel, or rather novella *Kad šume talasi* (When the waves rustle) (SRP19141), in 1914, when he was 25 years old. He was a captain of river navigation – like the main character of his novella – and besides poems, stories, and essays, also wrote professional books, most notably *Istorija jugoslovenskog rečnog brodarstva* (History of Yugoslav river shipping) in 1965.

haiduk : a story about the last days of Serbia in 15th century) (SRP18730) – he was 72 years old. On the average, authors were 38.8 years old when their selected works were published for the first time.

3.3 The Year of the First Publication

As the year of the first publication of novels we have chosen the year in which the work was first published either as a book (83), or in instalments in literary journals (16). If the publication in instalments span over more than one year, we have taken as the year of publication the year when the last instalment was published. To obtain this information, we relied primarily on the OPAC library catalog, encyclopedias (Петровић 1937; Бихаљи-Мерин 1959a, 1959b), reference books (Деретић 1981, 1983; Милисавац 1972), and sometimes on the reference editions of a particular author's works. For the novel *Seljaci* (Peasants) by Đura Jakšić (SRP1974) we could not determine the mode of its first publication, but found the year of publication in the collected works of this famous author (Јакшић 1883).

The distribution of publications in instalments over time slots is as follows: one in T1 (1840-1859), i.e. 50.0% of all novels from that time slot, 7 in T2 (1860-1879), also 50.0%, 5 in T3 (1880-1899), i.e. 11.4%, and 2 in T4 (1900-1920), i.e. 5.0%. This data show that publication in instalments was losing popularity over time.

It is obvious that all years from time slots T1 and T2 do not have their representatives in SrpELTeC, since there are more years than published works for these slots. In time slot T3, in all years save three – 1883, 1885 and 1890 – one or several works were published that are represented in SrpELTeC. The situation in time slot T4 seems similar, as there are three years with no representative in the Serbian sub-collection: 1915, 1916, 1917, years that coincide with the World War I. That this was not the time for novel writing and publishing in Serbia is supported by the fact that these years of publishing do not occur neither among 11 novels of the extended sub-collection SrpELteC-ext nor among 39 novels waiting to be prepared for this sub-collection. In other 100-novel ELTeC sub-collections, one cannot observe something similar: Czech and Slovene have no gap between years 1914 and 1918, while for other sub-collections, one or two years are not represented. However, given the large literary production in these languages, this could be for some other reason than war.

3.4 Publishers and Places of Publication

As pointed in the previous subsection, 15 novels were published for the first time in the following journals:

– *Javor* (Јовановић Змај 1862–1893) (4);
– *Otadžbina* (Ђорђевић 1875–1892) (4);
– *Сербскій лѣтописъ* (Субботићь 1842–1855) (1);
– *Danica* (Рајковић 1860–1872) (1);
– *Matica* (Хаџић 1865–1870) (1);
– *Nova iskra* (Одавић 1899–1911) (1);
– *Srpski književni glasnik* (Поповић 1901–1941) (1);
– *Stražilovo* (Грчић 1885–894) (1);
– *Male novine* (Кимпановић 1888–1903) (1).[8]

The publisher that has published the most novels in book form – *Srpska književna zadruga*, with 11 novels, as well as publishers *Zadužbina Ilije M. Kolarca*, with 3 novels, and *Matica srpska* with 2, have all survived to the present day. The remaining publishers *Matica hrvatska*, *Književni jug*, *Srpski pregled*, *Braća Savić*, *Savić i komp.* are represented with only one novel in SrpELTeC.

In the time period covered by ELTeC the roles of publishers, printers and bookstores were not so clear-cut as today. Thus, besides 21 publishers mentioned before, as many as 42 novels, on their cover pages,[9] instead of a publisher, mention a printing house (*štamparija*), a total of 32 different, e.g. *Državna štamparija kraljevine Srbije* (State Printing House of the Kingdom of Serbia) or *Parna štamparija D. Dimitrijevića* (Steam printing house of D. Dimitrijević). In other cases bookstores are mentioned (5), e.g. *Knjižara Velimira Valožića* (Bookstore of Velimir Valožić), political parties (1) *Napredna stranka* (Progressive Party), or persons (11), where in one case the person mentioned is the author himself – *Bespuće* by Veljko M. Milićević (SRP19121). The publisher is unknown in two cases.

---

8. The only information about the publisher of *Beogradske tajne* (SRP18923) on the book cover and in the library catalog is the name *Zabavnik Malih novina* (The Entertainer of the Little Newspaper). It is not clear whether *Zabavnik* (The Entertainer) was related to the daily newspaper *Male novine* (Little Newspaper), although an exhaustive note in the catalog record for *Male novine* mentions that the subtitle for 1893 was "Dnevni list za svakoga sa zabavnim mesečnim dodatkom" (A daily newspaper for everyone, with a monthly entertainment supplement). Since the edition has the form of a book, we treated it as such.

9. In cases the first editions were not available we relied on the catalog data.

Most of the novels were published in Belgrade (59), Novi Sad (18) and Sremski Karlovci (4). One novel was published in each of the following places, which are within today's borders of Serbia: Veliki Bečkerek (today Zrenjanin), Vršac, Kikinda, Kragujevac, Mitrovica (most probably Sremska Mitrovica), and Niš. Some novels were published outside today's borders of Serbia: Zagreb (3), Sarajevo (2), Mostar (2), and one novel in Beč (Vienna), Budim (the western part of today's Budapest), and Dubrovnik. For three novels the place of publication is unknown (see Figure 2).
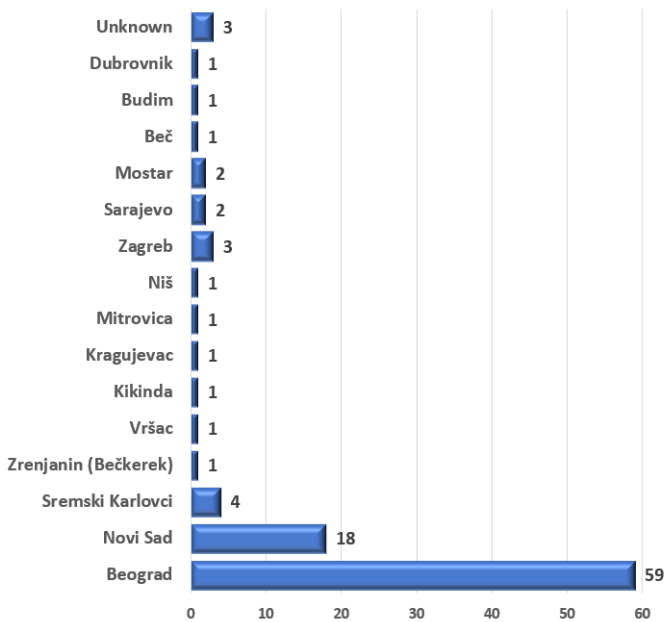
Figure 2. Publications per places of publication

3.5 Language and Alphabets

All novels are written in the Serbian language, although short passages in other languages appear almost in all of them, and are annotated with the

`<foreign>` tag. In only one novel, *Nove* (New women) (SRP19120), the passages in French and English appear more often and were roughly estimated to 1% of the whole text. This information is contained in the header.

All sources except two, whether printed or digital, used Cyrillic script. Two novels, *Pre sreće* (Prior to happiness) (SRP19180) and *U ćelijama* (In the cells) (SRP19192), both published in Zagreb, used Latin script.[10] The script of the sources was preserved in the electronic edition.

The Cyrillic alphabet in printed sources is the Serbian Cyrillic, as developed at the beginning of the 19th century by Vuk Stefanović Karadžić. Only the initial publication of the novel *Kaluđer : istina i poezija* (The monk: truth and poetry) (SRP1873), from the year 1881, used the Cyrillic letter Yat, which Vuk removed from the alphabet. Since this novel was never re-published, we used this initial publication for SrpELTeC and kept the letter Yat.[11] Several other novels used occasionally the soft sign (Ь), which was also removed from the alphabet or merged with other letters; we kept it as well. The use of these letters in the text is not recorded in the header.

3.6 Sources and Scanning

As explained in (Trtovac, Milnović, and Krstev 2021), we wanted to use the first editions for the SrpELTeC sub-collections whenever possible. This was, however, impossible for 15 novels, first published in journals, because not all journals (or issues) were available, or if they were already digitized, they were not suitable for further processing, as the OCR on scanned journals in general produces very poor results. In certain cases the first editions of books could not be used either, because they were not available or the OCR also produced very poor results. This was the case with *Opštinsko dete* (SRP19021), for which the first edition was retrieved and scanned, the OCR was done, but the result was so bad that we had to repeat the whole process with a later edition. As this was a very time consuming process, in four cases we decided to use already digitized novels, despite the fact that the solution was not optimal, as explained in (Trtovac, Milnović, and Krstev 2021).

---

10. One more novella used the Latin script for its first edition published in Belgrade in 1920: *Put Alije Gjerzeleza* (The Journey of Alija Gjerzelez) by Ivo Andrić, the only Serbian and Yugoslav Nobel Prize Winner for Literature. Unfortunately, it is shorter than 10K words and was included in SrpELTeC-ext.

11. We are particularly grateful to Aleksandra Davidović who read and corrected this novel, which was a very demanding task.

As a result, SrpELTeC comprises 56 novels scanned from the first editions, 40 novels scanned from a later edition and 4 novels that were already digitized. When we used later editions for scanning, we tried to use the edition that was time-wise the closest to the first edition. SrpELTeC comprises 14 later editions of novels published less than 10 years after the first edition, 5 later editions published less than 20 years after the first, and 2 later editions published less than 50 years after the first edition. Eleven later editions scanned for SrpELTeC were published between 50 and 100 years after the first edition, and 6 later editions more than 100 years after the first. For the novel *Nazareni* (Nazareans) (SRP18966) we used the second edition, which appeared the same year as the first, while the edition used for the novel *Đurađ Branković* (SRP18590) was published 129 years after the first. On the average, the print editions used for scanning that were not the first editions were published 34.8 years after the first.

The novels were procured from several libraries, but by far the most from the University Library "Svetozar Marković" – 68. The other libraries were: the National Library of Serbia (8), the Library of "Matica srpska" (3), the Library of the Serbian Academy of Sciences and Art (2), and the private library of Duško Vitas and Cvetana Krstev (12). Four already digitized novels were downloaded from the digital library *Antologija srpske književnosti* (The Anthology of Serbian literature), maintained by the Teacher Education Faculty at the University of Belgrade.[12] Two novels were retyped: *Đul-Marikina prikažnja* (Đul-Marika's narration) (SRP19012)[13] and *Došljaci* (Newcomers) (SRP19100).[14]

All novels were corrected and annotated by a number of volunteers, in line with the recommendations of the action's WG1 (as explained in (Trtovac, Milnović, and Krstev 2021)).[15] All volunteers and their contribution to the preparation of SrELTeC are listed in Table 1.[16] The contribution of readers is

---

12. Antologija srpske književnosti

13. The first edition of the novel could not be procured, so it was retyped from its scanned version available at the National Library of Kruševac Literary opus of Jelena Dimitrijević

14. This novel was previously retyped by students of Library and Information Sciences at the University of Belgrade, Faculty of Philology as part of their practical assignments. It was further checked, corrected and prepared for SrpELTeC.

15. The volunteers were mainly members or supporters of the Society for Language Resources and Technologies JeRTeH, as well as MA and PhD students at the University of Belgrade, for whom this was a part of their practical assignment.

16. It should be noted that some volunteers read additional novels, which were transferred to SrpELTeC-ext: Cvetana Krstev +4 (words 53,218; pages 237), Ivan

quantified here by the number of novels read, and the total number of words and printed pages. However, one important parameter is missing, and that is the quality of the text after OCR and automatic correction (see (Trtovac, Milnović, and Krstev 2021)). Namely, some shorter texts can be much harder to work on than longer texts, but we did not quantify this parameter.

Reader	bks	words	pages	Reader	bks	words	pages
Cvetana Krstev	32	1,669,442	7,572	Milena Mihajlović[*]	1	71,824	400
Duško Vitas	13	607,156	2,708	M. Ikonić Nešić	1	61,971	271
Ivan Obradović	9	592,497	2,534	Nenad Zekavica[*]	1	60,464	296
Ranka Stanković	7	255,718	1,192	A. Marković	1	59,480	204
B. Šandrih	4	189,575	805	Đorđe Stakić	1	51,288	276
Olivera Kitanović	4	130,493	559	Anica Milanović[*]	1	44,760	125
Biljana Lazić	3	120,890	549	Andrea Adamović[*]	1	37,860	181
A. Tomašević	3	111,913	593	Milica Antić[*]	1	29,543	125
A. Davidović	2	113,385	517	Vanja Radulović[*]	1	20,414	92
J. Dimitrijević	2	92,327	361	Tamara Radak	1	14,247	52
Tanja Ćulafić	2	68,481	296	Stefan Stepanović	1	12,023	36
J. Andonovski	2	60,262	213	A. Jovanović	1	10,095	80
Miloš Utvić	1	81,827	249	**total**	96	4,791,090	20,288

Table 1. List of readers that contributed to the preparation of the SrpELTeC. Master and PhD students are marked with an *, the others are JeRTeH volunteers.

4   Conclusions

In this paper we presented the metadata assigned to each novel of the ELTeC corpus, with the emphasis on the SrpELTeC sub-collection. From these data we could draw some conclusions about titling practices of Serbian narrative literature in the period 1840-1920, authors' gender and age, publication places, modes of publication, as well as about institutions and individuals that are due credits for the production of SrpELTeC.

We also believe that future users of ELTeC corpus, and SrpELTeC as its part, will benefit from the assigned metadata. We also hope that these users

---

Obradović +4 (words 81,227; 487), Duško Vitas +1 (words 20,286; pages 100), Stefan Stepanović +1 (words 7,105; pages 41), Sergej Adamov +1 (words 22,904; pages 122).

will enrich metadata with information related to their expertise. One can note that information about dialect and pronunciation used in novels, origin of authors, as well as novels' genre, are still missing, and should be added by experts.

References

Burnard, Lou. 2013. "The evolution of the Text Encoding Initiative: from research project to research infrastructure." *Journal of the Text Encoding Initiative,* no. 5.

Duval, Erik, Wayne Hodgins, Stuart Sutton, and Stuart L Weibel. 2002. "Metadata principles and practicalities." *D-lib Magazine* 8 (4): 1–10.

Hodge, Gail M. 2001. *Metadata made simpler.*

Krstev, Cvetana, and Ranka Stanković. 2021. "Novels and Authors of the Serbian ELTeC Collection." *Infotheca - Journal for Digital Humanities* 21 (2): 172–186. ISSN: 2217-9461.

Patras, Roxana, Carolin Odebrecht, Ioana Galleron, Rosario Arias, Berenike J Herrmann, Cvetana Krstev, Katja Mihurko Poniž, and Dmytro Yesypenko. 2020. *Dataset for ELTeC titles.* Zenodo. https://doi.org/10.5281/zenodo.4268669.

Patras, Roxana, Carolin Odebrecht, Ioana Galleron, Rosario Arias, Berenike J Herrmann, Cvetana Krstev, Katja Mihurko Poniž, and Dmytro Yesypenko. 2021. "Thresholds to the "Great Unread": Titling Practices in Eleven ELTeC Collections." *Interférences littéraires/Literaire interferenties* 25:163–187.

Trtovac, Aleksandra, Vasilije Milnović, and Cvetana Krstev. 2021. "The Serbian Part of the ELTeC Collection – from the Empty List to the 100 Novels Collection." *Infotheca - Journal for Digital Humanities* 21 (2): 7–25. ISSN: 2217-9461. https://doi.org/10.18485/infotheca.2021.21.2.1.

Бихаљи-Мерин, Ото и др., ed. 1959a. *Мала енциклопедија Просвета: општа енциклопедија. 1, А - Љутомер.* Тираж 30000. Просвета.

Бихаљи-Мерин, Ото и др., ed. 1959b. *Мала енциклопедија Просвета: општа енциклопедија. 2, М - Шчукин.* Просвета.

Грчић, Јован, ed. 1885–894. *Стражилово : лист за забаву, поуку и уметност.* Дружина око "Стражилова". http://ubsm.bg.ac.rs/cirilica/zbirka/novina/strazilovo-1885-1894.

Деретић, Јован. 1981. *Српски роман: 1800-1950.* Нолит.

Деретић, Јован. 1983. *Историја српске књижевности.* Нолит.

Ђорђевић, Владан, ed. 1875–1892. *Отаџбина : књижевност, наука, друштвени живот.* Владан Ђорђевић. http://ubsm.bg.ac.rs/cirilica/zbirka/novina/otadzbina-1875-1892.

Јакшић, Ђура. 1883. *Дела Ђуре Јакшића. Књ. 5, Приповетке.* У краљевско-српској држ. штампарији.

Јовановић Змај, Јован, ed. 1862–1893. *Јавор: лист за забаву и науку.* Јован Јовановић Змај. http://www.digitalna.nb.rs/sf/NBS/casopisi_pretrazivi_po_datumu/P_247%20http://digital.bms.rs/pub.php?s=RPSr-III-15.

Кимпановић, Ђорђе, ed. 1888–1903. *Мале новине : дневни лист за свакога.* Пера Тодоровић. http://ubsm.bg.ac.rs/cirilica/zbirka/novina/male-novine-1888-1903.

Милисавац, Живан, ed. 1972. *Приповедачи.* Vol. 66. Матица српска; Српска књижевна задруга.

Одавић, Риста Ј., ed. 1899–1911. *Нова Искра : илустровани лист.* Р. Ј. Одавић. http://ubsm.bg.ac.rs/cirilica/zbirka/novina/iskra-1899-1911.

Петровић, Петар М. и др., ed. 1937. *Свезнање: општи енциклопедиски лексикон у једној књизи.* Народно дело.

Поповић, Богдан, ed. 1901–1941. *Српски књижевни гласник.* Светислав С. Симић. http://www.digitalna.nb.rs/wb/NBS/casopisi_pretrazivi_po_datumu/Srpski_knjizevni_glasnik.

Рајковић, Ђорђе, ed. 1860–1872. *Даница : лист за забаву и књижевност.* Ђорђе Поповић. http://www.digitalna.nb.rs/sf/NBS/casopisi_pretrazivi_po_datumu/P_0151%20http://digital.bms.rs/pub.php?s=RPSr-II-15%20http://digital.nb.rs/scc/browse.php?collection=no-danica.

Субботићь, Іоаннъ, ed. 1842–1855. *Сербскій лѣтописъ.* Писмены Крал. Свеучилишта Пештанскогъ. http://digital.bms.rs/pub.php?s=RPSr-II-4.

Хаџић, Антоније, ed. 1865–1870. *Матица : лист за књижевност и забаву.* Матица србска. http://www.digitalna.nb.rs/sf/NBS/casopisi_pretrazivi_po_datumu/matica.