BULGARIAN ACADEMY OF SCIENCES
INSTITUTE OF LITERATURE

SCRIPTA & *e*-SCRIPTA

This issue of the journal is published with the financial support of the Alexander von Humboldt Foundation ('DigiPalSlav' Research Group Linkage Program).

Scripta & *e*-Scripta

The Journal of Interdisciplinary Mediaeval Studies

Volume 21

Sofia
2021

SCRIPTA
& e- SCRIPTA

21 / 2021

Editorial Board

Editorial Address
Institute for Literature
Bulgarian Academy of Sciences
52, Shipchenski prohod Str.
1113 Sofia, Bulgaria
E-mail: escripta.ilit@gmail.com

Executive editors of this issue: **Ralph Cleminson, Margaret Dimitrova, Jürgen Fuchsbauer, Anissava Miltenova, Petya Yaneva, Elissaveta Moussakova and Dilyana Radoslavova**

Proofreading: **Maria Cioata**
Guest editors: **Victor Baranov, Alexandr Moldovan, Achim Rabus**

ISSN 1312-238X
ISSN 2603-3364 (e-publication)

"Boyan Penev" Publishing Center
Institute of Literature, BAS
Grafic design and pre-press: ***Daniela Vasileva***

Scripta & e-Scripta

Volume 21
Institute of Literature, Bulgarian Academy of Sciences
Sofia 2021

Contents

e-Scripta

Scripta

Debuts

Personalia

Scripta & e-Scripta

Volume 21
Институт за литература, Българска академия на науките
София 2021

Съдържание

e-Scripta

Scripta

Дебюти

Personalia

SCRIPTA
& e- SCRIPTA

e-S C R I P T A

21 / 2021

Towards Fundamental Principles for Creating the Electronic Corpus of Serbian Medieval Charters and Letters

Vladimir R. Polomac

Abstract: The paper defines the elementary principles for creating an electronic corpus of Serbian medieval charters and letters. The commitment to the principle of maximum representativeness of the corpus of medieval charters and letters, determined entirely by the preserved written legacy (based on manuscripts, microfilms or photographs), excludes the indispensability of applying the principle of balance, while simultaneously satisfying the principle of reliability, since charters and letters known solely by the edition are not included in the corpus. The selection of texts is done according to the diplomatic criterion by excluding the transcripts and copies of documents already available in the original, as well as later transcripts, chronologically and linguistically distant from the assumed original. This approach to the selection of texts is justified by the size of the corpus, as well as by the exceptional cultural and historical significance of medieval charters and letters. The definition of the metadata about corpus texts is determined by their general diplomatic properties, as well as the corpus search needs for diatopic, diachronic and genre variations. Conversion of texts into electronic form strives for fidelity to the original, encompassing the preservation of abbreviations, superscript letters and original punctuation, as well as the absence of accent marks and contemporary rules of capitalization.

Keywords: Historical Corpus Linguistics, Old Serbian language, Serbian Church Slavonic, Serbian Medieval Charters and Letters, 12th–16th century.

1. Introduction

Although the beginnings of corpus linguistics in Serbia can be traced back to the end of the sixth decade of the 20th century and the compilation of the pre-electronic

historical corpus of the Serbian language at the Institute for Experimental Phonetics and Speech Pathology in Belgrade managed by Đ. Kostić, no significant attention was paid to this issue in Serbian linguistics in the following decades and even until today.[1] The project by Đ. Kostić, conceived fairly ambitiously at the time, was suspended as early as 1962 due to the cessation of funding, only to be renewed in the mid-1990s due to the incentive of his son A. Kostić (2003: 260–263). Collaborators on the renewed project digitalized the existing project resources, modernized the system of grammatical annotation of texts and published eight frequency dictionaries of Serbian Church Slavonic sources (Kostić 2003: 263–264) in a paperback version accompanied by texts in electronic form, searchable by various criteria using a special software application.[2] Besides the fact that this corpus is not publicly available on the Internet, its fundamental methodological drawback is the selection of texts which does not adequately represent the Serbian language of the 12th–18th centuries. The texts written in the Serbian Church Slavonic language dominate the corpus belonging to the 12th–18th centuries, while the representations of texts written in the Serbian language are based only on the edition of charters and letters of Lj. Stojanović (1929 and 1934).

Having the aforementioned shortcomings of Kostić's corpus in mind, as well as the previous experience in developing historical electronic corpora of Slavic languages,[3] the main goal of our study is (a) to define the historical corpus of the Serbian language and the position of medieval charters and letters in it especially with regard to the principles of representativeness, reliability and balance of historical corpora,[4] as well as (b) to define the fundamental principles for the preparation and collection of texts of medieval charters and letters for the corpus.[5]

[1] A brief history of corpus linguistics in Serbia is given by Utvić 2013: 51–55. On the need and principles of forming the corpus of the Old Serbian language, cf. Pavlović 2009.

[2] The software application was presented in detail in Kostić and Vitić 2015.

[3] Without claiming bibliographic exhaustiveness, here we refer only to Baranov 2019; Hansack et al. 2016; Kapetanović 2007; Kučera 1999; Kučera 2002: 249–250; Meyer 2012; Totomanova 2017, then on the historical subcorpora of the Russian NCRL: http://www.ruscorpora.ru/search-old_rus.html – and the Czech national corpus: http://ucnk.korpus.cz/diakorp.php, then to PROIEL – a parallel corpus of translations of the New Testament in Old Indo-European languages: http://foni.uio.no:3000/users/sign_in, in which the Zograf Gospel, the Gospel of Mary and the *Codex Suprasliensis* are presented, as well as on TOROT – https://nestor.uit.no – a continuation of the PROIEL corpus, in which Old Russian texts are attached to Old Church Slavonic manuscripts.

[4] See more details on these principles in Biber 1993; Nelson 2010: 56–60; Pavlović 2009: 136–137; Utvić 2013: 16–17, 20–21.

[5] The basic principles of orthographic normalization of the text, as well as the principles of lemmatization and annotation of medieval charters and letters will be considered in separate papers. The issue of corpus visualization, as well as the choice of search technology, deserves special consideration.

Having in mind that Serbian medieval charters and letters represent the most important and extensive subcorpus of the historical electronic corpus of the Serbian language, defining the aforementioned elementary principles for the compilation of this subcorpus would be useful for building the historical electronic corpora of the Serbian language as a whole. The ensuing remarks represent the result of the lengthy and extensive investigation of charters and letters of the 14th and 15th centuries from the perspective of historical dialectology of the Serbian language and the history of the Serbian literary language (cf. only Polomac 2016; 2017), as well as the engagement in the preparation of texts of medieval charters and letters for the Dictionary of Serbian language of the 12th–18th century, a long-term project realized within the Matica Srpska in Novi Sad and led by Jasmina Grković-Major.

2. Serbian Diachronic Corpus and Medieval Charters and Letters

2.1. Corpus Definition and Text Selection

The historical corpus of the Serbian language would consist of all the documents written in one of the folk idioms of the Štokavian dialect, regardless of the document script or the religion of the scribe. The lower chronological limit of this corpus is determined by the oldest preserved documents from the end of the 12th century, while the upper limit is the end of the 18th century, all the way to the documents included in the Dictionary of the Serbo-Croatian Language of the Serbian Academy of Sciences and Arts. Within the historical corpus of the Serbian language defined thus, the subcorpus of medieval charters and letters, with its antiquity, scope and dominant use of the Serbian language (in relation to Serbian Church Slavonic), stands out as the most important source for the history of the Serbian language in the Middle Ages. As such, it should consequently represent the outset for the compilation of a referential historical electronic corpus of the Serbian language, too.

When creating an electronic corpus of Serbian medieval charters and letters, it seems necessary to first consider the question of whether all the preserved texts should be included in the corpus or their selection is to adhere to the principles of representativeness and balance. Since the corpus of medieval charters and letters is primarily approached with the general goal of creating a specialized electronic resource that should enable (a) the development of reference grammars and lexicographic handbooks of the Serbian language (historical grammar, historical dictionary,

specialized historical dictionaries of personal and geographical names),[6] as well as (b) the comparison of the Old Serbian language with other Slavic languages (primarily Old Church Slavonic, Old Russian and Old Czech) in order to reconstruct the Proto-Slavic language structures maximum representativeness is sought by including all charters and letters preserved based on manuscripts, microfilms or photographs.[7] Excluding charters and letters known solely based on the edition would satisfy the principle of reliability of the historical electronic corpus. The ideal of maximum representativeness of the corpus of medieval charters and letters determined by the preservation of written heritage in its entirety excludes the application of the principle of balance.

The selection of texts could be done only according to the diplomatic criterion by excluding transcripts and copies of charters and letters from the corpus that are already available in the original or later transcripts, chronologically and linguistically distant from the presumed original. Such an approach to the selection of documents is justified by the size of the corpus, as well as the exceptional cultural and historical significance of medieval charters and letters, which accounts for the undoubted need to present every available text in its entirety in electronic form. The decision to include all the texts that meet the genre and diplomatic criteria in the electronic corpus – alongside texts written in their entirety or predominantly in Serbian – assumes the inclusion of texts written in Serbian Church Slavonic, as well. Potential preference of the linguistic over the genre criterion when choosing texts would be difficult to implement in practice since Serbian and Serbian Church Slavonic are used complementary in accordance with the principle of homogeneous diglossia in different types of texts (e.g. in charters addressed to monasteries),[8] where switching from one to another idiom often occurs within the same sentence using hybrid forms.[9] The principle of selecting texts according to the genre criterion does not exclude the possibility of separating Serbian and Serbian Church Slavonic material through different corpus search options. Due to the fact that the use of two linguistic idioms

[6] S. Pavlović (2009: 135) points to the considerable falling behind of Serbian linguistics in this domain, stating that despite the great achievements of structural linguistics in the 20th century, we remained deprived of corpus-based historical phonology, historical morphology, historical derivatology, historical syntax, and historical vocabulary, which positions us at the back of the line of European national philology.

[7] At this point, it is very difficult to precisely determine the size of the corpus. We assume that the corpus will contain at least 800,000 tokens.

[8] On the use of Serbian Church Slavonic and Old Serbian in texts of different genres of Serbian medieval literacy in accordance with the principle of homogeneous diglossia, for more detail see Grković-Mejdžor 2007.

[9] More about this in Ivić 2016: 109; and Polomac 2016: 495.

is invariably conditioned by the genre and diplomatic structure of the charter, by narrowing the search according to these criteria one can indirectly track the variation of a certain linguistic feature in the corpus according to the linguistic criterion. Thus – if the research question so requires – the search can be conducted on a certain genre exclusively, or even more specifically on smaller textual units within the genre (protocol, text and/or eschatocol (ending)). This approach requires providing texts with appropriate metadata in advance, as well as labelling textual units within charters and letters.

Along with the overview of the medieval linguistic situation among Serbs determined by the use of homogeneous diglossia, when selecting texts for the corpus, it is necessary to refer to the contemporary linguistic situation conditioned by the emergence of new states and standard languages after the breakup of the former Yugoslavia. The new linguistic situation had no implications for the selection of texts for our corpus. The selection of texts is determined by the script criterion (Cyrillic texts) and language (texts written in Serbian (Štokavian dialect) or the Serbian Church Slavonic language), employing the traditional understanding of this corpus as Serbian in European and Serbian Slavic studies of the 19th and 20th centuries,[10] as well as the mentioning of the Serbian name and language in the very charters and letters themselves.[11] This procedure, of course, does not deny that parts of this corpus may represent the written heritage of other cultures (Croatian, Bosnian and Montenegrin) whose standard languages originated from the Štokavian dialect after the break-up of the former Yugoslavia.

2.2. Text Registry Formation

After defining the criteria for selecting charters and letters to be incorporated in the corpus, it seems a prerequisite to start developing the registries of texts. The principal theoretical and methodological issue pertains to the definition of criteria based on which charters and letters would be classified and marked in the registry. The registries of charters and letters made within the project *Dictionary of the Serbian language of the 12th–18th century*[12] of Matica Srpska represent the starting point and foundation of the research. In these registries, charters and letters are marked with abbreviations indicating the archive in which they are located, followed by the century

[10] Cf. only Miklosich 1858; Stojanović 1929; Stojanović 1934.

[11] For charters and letters of medieval Bosnia, cf. Isailović 2018: 261–282.

[12] The authors of the registries are S. Pavlović (charters and letters until the 15th century) and V. Savić (Serbian charters from the archives of Mount Athos). We used our own registry for charters and letters of the 15th–16th centuries.

in which they were created and the ordinal number in the registry indicating their chronology of origin. Thus, the abbreviation Д XII 1 denotes the oldest document from the 12th century: The signature of the Grand Župan Stefan Nemanja and Prince Miroslav on a Latin document (September 27th, 1186) located in the State Archives in Dubrovnik. In addition to the abbreviation of the document, the registries in question also contain the name of the document (with genre explication provided, possibly with both the sender and the addressee), more precise time of creation (if known), diplomatic status (original, transcript, copy or translation) and information on the most famous or available edition. In relation to the mentioned model, the registry of texts for the electronic corpus could be significantly simplified by omitting data on the archive (not linguistically relevant) and the data on the century of creation (this will be stated in the metadata table for each document), i.e. by forming a document abbreviation merely by ordinal number in the registry (e.g. №1, №2, etc.) which would indicate the chronology of creation.

The registry of charters and letters for the corpus would cover the period whose lower chronological limit includes the oldest preserved texts from the end of the 12th century, while the upper limit would be determined by the discontinuation of Serbian medieval office activities (late 15th century) and the use of Serbian diplomatic correspondence of Turkish, Hungarian, Vlach and Moldavian rulers in the 16th century.

3. Towards Fundamental Principles for Preparing and Editing the Texts for the Corpus

3.1. Metadata Definition

During the definition of the corpus and the principles of making the registry of Serbian medieval charters and letters, the basic metadata about the texts were already underscored. It is now necessary to consider and expand them in more detail in accordance with the standards for coding medieval charters developed within the CEI project (see https://www.cei.lmu.de/) as well as in line with the latest achievements of Serbian diplomatics.[13]

[13] Cf. only the following monographs: Porčić 2012; Isailović 2014; Isailović 2015; and Vujošević 2015.

The types of metadata will be presented in the following table illustrated on the oldest preserved charter in the corpus:

Metadata	*Value*
Abbreviation	№2
Title	Charter of Ban Kulin to Dubrovnik
Author	Ban Kulin
Recipient	Dubrovnik
Diplomatic status	Original
General type	Charter
Specific purpose type	Contract
Date of creation	29th August 1189
Century of creation	12th
Place of origin	/
Region	Bosnia
Office	Offices of the first Bans
Scribe	Radoje Dijak
Edition	Mošin–Ćirković–Sindik 2011: 49–52
Recording	Đorđić 1991: 65
Text compiler	Vladimir Polomac
Date of text compilation	20th March 2020

Certain types of metadata require a more detailed commentary taking into consideration the nature of the corpus and search requirements relevant to the research on diatopic, diachronic, and genre variations in the corpus. Bearing in mind that the language of Serbian medieval charters and letters predominantly depends on who the charters and letters are addressed to, when defining the recipient of charters and letters addressed to Dubrovnik, we merely preserved the general determinant Dubrovnik (without specifying particular individuals), which could enable the search for linguistic features by this sole criterion. When it comes to the diplomatic status of texts, we have already mentioned that the originals, along with transcripts and copies (belonging to the time of the creation of the original) for documents whose original we do not know, as well as translations, are taken into account. In determining the typological (genre and sub-genre) affiliation of the texts, we were guided by the diplomatic criteria elaborated in Porčić 2012 and Isailović 2014. Considering the general form, the texts from the territory of medieval Serbia are defined as *charters* (documents which grant or confirm a permanent right or which oblige the author (originator) to respect the right regulated by a contract) or *letters* (documents relating to the current administration and correspondence, with a basic purpose of conveying

information), and considering forms for specific purposes, as *grants*, *contracts*, *receipts*, *notifications*, *powers of attorney*, *passes*, *orders* and *verdicts.*[14] The same criteria determine the texts from the territory of medieval Bosnia: *charters*, *letters*, *notes* and *wills* belong to general forms, while *grants*, *contracts*, *receipts*, *notifications*, *powers of attorney*, *verdicts* and *testaments* belong to the texts for specific purposes.[15] Since the date of creation is not always stated in charters and letters, the category of century of creation (XII, XIII, XIV, XV, XVI) was introduced, which allows us to search for texts according to this criterion and follow a certain linguistic phenomenon in a wider yet still limited time frame. As opposed to the time of creation, the place of creation of the document is very rarely stated in the text itself. Therefore, it was necessary to attach the data on geographical origin (region) to each document in order to be able to monitor diatopic variations in the corpus. The largest group of documents can be geographically linked to medieval Dubrovnik, then to the medieval Serbian states (Zeta, Hum, Raška, Despotate), as well as medieval Bosnia, and the smallest to the territory of medieval Dalmatia. Having in mind that the corpus also includes documents created in the 15th and 16th centuries outside the native Serbian territory, a smaller number of documents can be geographically linked to Albania, Bulgaria, Wallachia, Moldavia, Hungary and Turkey. Within the sets of documents related to geographical origin, a narrower differentiation can be established according to the offices where the documents originated.[16] Exceptions are charters and letters created in the Serbian office in Dubrovnik, for which a narrower differentiation can be made according to the scribes.[17] For each document the information on the most famous or most accessible edition is given, as well as the information on the photograph. Besides the long tradition of publishing and studying medieval charters and letters, it is especially important to underscore the availability of the largest number of photographs, both in the collections of archival institutions and scientific projects, and, more recently, in virtual archives and the increasing number of diplomatic studies.[18]

[14] More details in Porčić 2012: 238–351, 390–402.

[15] More details in Isailović 2014: 64–69, 595–608.

[16] The examples are the charters and letters of medieval Bosnia in which the offices of the first bans (Ban Kulin and Ban Matej Ninoslav), bishops of the Bosnian Church, the ruling family Kotromanić and aristocratic families Hrvatinić, Sanković, Nikolić, Pavlović, Kosača, Vlatković, Dinjičić, Borovinić, Stančić, Ozrisaljić can be singled out (cf. Isailović 2014: 588–595).

[17] It is known e.g. that about 300 letters can be ascribed to Rusko Hristiforović, the Serbian chancellor in Dubrovnik (1395–1423).

[18] The portal of the European virtual archive https://www.monasterium.net also contains collections of Serbian medieval charters and letters at http://monasterium.net:8181/mom/BISANU/collection and http://www.bisanu.rs/http://monasterium.net:8181/mom/SerbianRoyalDocumentsDubrovnik/collec-

3.2. Principles of Converting the Texts into Electronic Form

Since we have already emphasized that the corpus of Serbian medieval charters and letters will be formed in accordance with the principle of maximum representativeness, the determination to present entire texts in it, and not just selected paragraphs, seems natural. When converting the texts into electronic form, we strived to make the text as faithful to the original as possible, which means that the charters and letters were prepared in original graphics using the *BukyVede* font that supports the Unicode standard, respecting the following crucial principles: a) abbreviations and superscripts are adopted according to the original, b) the original punctuation is retained, where the dot is always brought to the middle of the line and separated by white spaces from the previous word, c) contemporary principles on the use of capital letters are not transferred, e) the new line is marked with a vertical line followed by a line number in the exponent, e) accent marks are omitted from the text. The edition can be illustrated by the example of the abovementioned charter (contract) of the Bosnian Ban Kulin to Dubrovnik from 1189:

† Ѕ име wца и сн̃а : и ст̃ога дх̃а ѣ бань : бо|[2]сьньски кѵ̀линь : присезаю ᵀб̊ѣ кнеже |[3] крьваш̀ : и вьсѣмь граѣамь дѵ̀бровьч|[4]амь : правы : приѣтель : быти вамь |[5] ѿь селѣ : и до вѣка : и правь гои дрьжати |[6] сь вами : и правѵ̀ : вѣрѵ̀ : до кола сьмь живь : вь|[7]си дѵ̀бровьчане : кире ходе : по моемѵ̀ владани|[8]ю : трьгѵ̀юке : годѣ си кто : хоке : крѣвати : го|[9]дѣ си кто мине : правовь вѣровь и правымь : ср̃ь|[10]дьцемь : дрьжати е : безь вьсакое зледи : раз|[11]вѣ цю ми кто : да б̊оевь воловь пок̂онь : и да имь |[12] не бѵ̀де : ѿь моихь : чьстьниковь : силе : и до колѣ : |[13] Ѕ мне бѵ̀дѵ̀ : дати имь : сьвѣть : и помокь какоре : и се|[14]бѣ коликоре моге : безь вьсега : зьлога примы|[15]сла : тако ми б̃же помагаи : и сие ст̃о еваньгелие : |[16] ѣ радое : диѣкь бань : писахь сию : книгѵ̀ : повеловь |[17] бановь : ѿь рожьст̊а : хр̃а : тисѵ̀ка : и сьто : и wсм|[18]ьдесеть : и деветь : лѣть : мц̃еца : авьгѵ̀ста : |[19] Ѕ дьвадесети : и деветы : дн̃ь : ѵ̀ствение : гла|[20]ве : иwвана : крьститела ⁜

tion. Among recent diplomatic studies, it seems indispensible to mention the papers published in the journals *Stari srpski arhiv* (engl. *Old Serbian Archive*) and *Građa o prošlosti Bosne* (orig. *Matériauh pour l' histoire de Bosnie*), the monograph by Porčić 2017, as well as the current project *Serbian Digital Collection of Charters and Letters*, which is being developed at the Balkan Institute of Serbian Academy of Sciences and Arts (SANU).

4. Prospective and Future Tasks

Further work on the conversion of texts of charters and letters in electronic form could be significantly accelerated by the use of technology for automatic text recognition in accordance with the principles set out in the paper by A. Rabus (2019). In parallel with this work, the definition of the principles of orthographic normalization, lemmatization and annotation of texts lie in perspective, as well as the publication of the first version of the corpus.[19]

REFERENCES

Baranov 2019: Баранов, А. Виктор. "Создание и использование исторических корпусов славянских письменных памятников." *Scripta & e-Scripta* 19 (2019), 33–57.

Biber 1993: Biber, Douglas. "Representativeness in Corpus Design." *Literary and Linguistic Computing* 8/4 (1993), 243–257.

ČNK: Český národní korpus – https://www.korpus.cz/ <accessed 30.05.2021>.

Grković-Mejdžor 2007: Грковић-Мејџор, Јасмина. "Диглосија у старосрпској писмености." У *Списи из историјске лингвистике*. Нови Сад–Сремски Карловци: Издавачка књижарница Зорана Стојановића, 2007, 443–460.

Hansack et al. 2016: Hansack, Ernst, Björn Hansen, Veronika Wald, Marijana Horvat, Sanja Perić Gavrančić. "Regensburški dijakronijski korpus hrvatskoga jezika – CroDi." *Rasprave* 42/1 (2016), 1–19.

Isailović 2014: Исаиловић, Невен. *Владарске канцеларије у средњовековној Босни*. Докторска дисертација. Филозофски факултет, Београд, 2014.

Isailović 2015: Исаиловић, Невен. "Дипломатичке особености владарских и великашких исправа уочи и након пада средњовековне босанске државе." У *Пад босанског краљевства 1463. године*. Ур. Невен Исаиловић. Београд – Сарајево – Бања Лука: Историјски институт у Београду – Филозофски факултет у Сарајеву – Филозофски факултет у Бањој Луци, 2015, 29–86 (Историјски институт Београд, Зборник радова 29).

Isailović 2018: Исаиловић, Невен. "Помени српског имена у средњовековним српским исправама." У *Српско писано насљеђе и историја средњовековне Босне и Хума*. Бања Лука – Источно Сарајево: Филолошки факултет у Бањој Луци – Филозофски факултет у Бањој Луци – Филозофски факултет на Палама, 2018, 261–282.

[19] The first version of the corpus should contain all preserved charters and letters from the 12th and 13th centuries, as well as a representative selection of charters and letters from the 14th-16th centuries. The publication and presentation of the first version of the corpus is planned for the second half of 2023 at the International Congress of Slavists in Paris.

Ivić 2016: Ивић, Павле. "Средњовековне српске повеље као документи о језику и култури." У *Целокупна дела Павла Ивића. 10. 3. Расправе, студије, чланци. О историји језика*. Ред. Милорад Ладовановић. Нови Сад – Сремски Карловци: Издавачка књижарница Зорана Стојановића, 2016, 99–111.

Kapetanović 2007: Kapetanović, Amir. "Digitalizacija korpusa starohrvatskih tekstova i kritika teksta." In *The Future of Information Sciences: INFuture2007 – Digital Information and Heritage*. Zagreb: Filozofski fakultet, 2007, 173−182.

Kostić 2003: Костић, Александар. "Електронски корпус српског језика Ђорђа Костића." *Зборник Матице српске за славистику* 64 (2003), 260–264.

Kostić and Vitić 2015: Костић, Александар, Зорица Витић. "Софтверска апликација за претрагу и анализу српских средњовековних текстова." У *Дигиталне библиотеке и дигитални архиви*. Београд: Филолошки факултет, 2015, 303–313.

Kučera 1999: Kučera, Karel. "The general principles of the diachronic part of the Czech National Corpus." In *Text, Speech and Dialogue. Proceedings of the 2nd International Conference on Text, Speech and Dialogue – TSD'99*. Berlin: Springer, 1999, 62–65.

Kučera 2002: Kučera, Karel. "The Czech National Corpus: Principles, Design and Results." *Literary and Linguistic Computing* 17/2, 2002, 245–257.

Meyer 2012: Meyer, Roland. "The construction and application of diachronic Slavonic corpora in linguistic research – RRuDi (Russian) and PolDi (Polish)." In *Diachrone Aspekte slavischer Sprachen. Für Ernst Hansack zum 65. Geburtstag*. München – Berlin – Washington D.C.: Verlag Otto Sagner, 2012, 223–242 (Slavolinguistica 16).

Miklosich 1858: Miklosich, Fr. *Monumenta serbica spectantia historiam Serbiae, Bosnae, Ragusii*. Viennae: apud Guilelmum Braumüller, 1858.

NCRL: Национальный корпус русского языка – https://ruscorpora.ru/new/index.html <accessed 30.05.2021>.

Nelson 2010: Nelson, Mike. "Building a written corpus: What are the basics?" In *The Routledge Handbook of Corpus Linguistics*. Ed. Anne O'Keeffe, Michael McCarthy. London–New York: Routledge, 2010, 53–65.

Pavlović 2009: Павловић, Слободан. "Електронско претраживање старосрпског језичког корпуса у светлу стандардизације старословенске ћирилице." У *Стандардизација старословенског ћириличког писма и његова регистрација у Уникоду*. Ур. Гордана Јовановић, Јасмина Грковић-Мејџор, Зоран Костић, Виктор Савић. Београд: Институт за српски језик САНУ, 2009, 135–146.

Polomac 2016: Поломац, Владимир. *Језик повеља и писама Српске деспотовине*. Крагујевац: ФИЛУМ, 2016.

Polomac 2017: Поломац, Владимир. "О правопису и језику повеља и писама Вука Бранковића." *Зборник Матице српске за филологију и лингвистику* 60/2 (2017), 59–71.

Porčić 2012: Порчић, Небојша. *Дипломатички обрасци средњовековних владарских докумената: српски пример*. Докторска дисертација. Филозофски факултет, Београд, 2012.

Porčić 2017: Порчић, Небојша. *Документи српских средњовековних владара у дубровачким збиркама: доба Немањића*. Београд: Балканолошки институт САНУ, 2017.

PROIEL: The PROIEL corpus – http://foni.uio.no:3000/users/sign_in <accessed 30.05.2021>.

Rabus 2019: Rabus, Achim. "Recognizing handwritten text in Slavic manuscripts: A neural-network approach using Transkribus." *Scripta & e-Scripta* 19 (2019), 9–32.

Stojanović 1929: Стојановић, Љубомир. *Старе српске повеље и писма*. 1. Београд: Српска краљевска академија, 1929.

Stojanović 1934: Стојановић, Љубомир. *Старе српске повеље и писма*. 2. Београд: Српска краљевска академија, 1934.

TOROT: The Tromsø Old Russian and OCS Treebank – https://nestor.uit.no <accessed 30.05.2021>.

Totomanova 2017: Тотоманова, Анна Мария. "Диахронный корпус болгарского языка: состояние и перспективы." *Filologija* 68 (2017), 223–242.

Utvić 2013: Utvić, Miloš. *Izgradnja referentnog korpusa savremenog srpskog jezika.* Doktorska disertacija. Filološki fakultet, Beograd, 2013.

Vitić and Kostić 2015: Витић, Зорица, Александар Костић. "Специфични захтеви анотације српскословенског у оквиру Електронског корпуса српског језика." У *Дигиталне библиотеке и дигитални архиви*. Београд: Филолошки факултет, 2015, 315–323.

Vujošević 2015: Вујошевић, Жарко. *Српска владарска канцеларија у средњем веку. Студија из упоредне дипломатике*. Докторска дисертација. Филозофски факултет, Београд, 2015.

[Baranov, A. Viktor. "Sozdanie i ispol'zovanie istoričeskih korpusov slavjanskih pis'mennyh pamjatnikov." *Scripta & e-Scripta* 19 (2019), 33–57.

Grković-Mejdžor, Jasmina. "Diglosija u starosrpskoj pismenosti." U *Spisi iz istorijske lingvistike*. Novi Sad–Sremski Karlovci: Izdavačka knjižarnica Zorana Stojanovića, 2007, 443–460.

Isailović, Neven. *Vladarske kancelarije u srednjovekovnoj Bosni*. Doktorska disertacija. Filozofski fakultet, Beograd, 2014.

Isailović, Neven. "Diplomatičke osobenosti vladarskih i velikaških isprava uoči i nakon pada srednjovekovne bosanske države." U *Pad bosanskog kraljevstva 1463. godine*. Ur. Neven Isailović. Beograd – Sarajevo – Banja Luka: Istorijski institut u Beogradu – Filozofski fakultet u Sarajevu – Filozofski fakultet u Banjoj Luci, 2015, 29–86 (Istorijski institut Beograd, Zbornik radova 29).

Isailović, Neven. "Pomeni srpskog imena u srednjovekovnim srpskim ispravama." U *Srpsko pisano nasljeđe i istorija srednjovekovne Bosne i Huma*. Banja Luka – Istočno Sarajevo: Filološki fakultet u Banjoj Luci – Filozofski fakultet u Banjoj Luci – Filozofski fakultet na Palama, 2018, 261–282.

Ivić, Pavle. "Srednjovekovne srpske povelje kao dokumenti o jeziku i kulturi." U *Celokupna dela Pavla Ivića*. 10. 3. *Rasprave, studije, članci. O istoriji jezika*. Red. Milorad Ladovanović. Novi Sad – Sremski Karlovci: Izdavačka knjižarnica Zorana Stojanovića, 2016, 99–111.

Kostić, Aleksandar. "Elektronski korpus srpskog jezika Đorđa Kostića." *Zbornik Matice srpske za slavistiku* 64 (2003), 260–264.

Kostić, Aleksandar, Zorica Vitić. "Softverska aplikacija za pretragu i analizu srpskih srednjovekovnih tekstova." U *Digitalne biblioteke i digitalni arhivi*. Beograd: Filološki fakultet, 2015, 303–313.

Nacional'nyj korpus russkogo jazyka – https://ruscorpora.ru/new/index.html.

Pavlović 2009: Pavlović, Slobodan. "Elektronsko pretraživanje starosrpskog jezičkog korpusa u svetlu standardizacije staroslovenske ćirilice." U *Standardizacija staroslovenskog ćiriličkog pisma i njegova registracija u Unikodu*. Ur. Gordana Jovanović, Jasmina Grković-Mejdžor, Zoran Kostić, Viktor Savić. Beograd: Institut za srpski jezik SANU, 2009, 135–146.

Polomac, Vladimir. *Jezik povelja i pisama Srpske despotovine*. Kragujevac: FILUM, 2016.

Polomac, Vladimir. "O pravopisu i jeziku povelja i pisama Vuka Brankovića." *Zbornik Matice srpske za filologiju i lingvistiku* 60/2 (2017), 59–71.

Porčić, Nebojša. *Diplomatički obrasci srednjovekovnih vladarskih dokumenata: srpski primer.* Doktorska disertacija. Filozofski fakultet, Beograd, 2012.

Porčić, Nebojša. *Dokumenti srpskih srednjovekovnih vladara u dubrovačkim zbirkama: doba Nemanjića*. Beograd: Balkanološki institut SANU, 2017.

Stojanović, Ljubomir. *Stare srpske povelje i pisma*. 1. Beograd: Srpska kraljevska akademija, 1929.

Stojanović, Ljubomir. *Stare srpske povelje i pisma*. 2. Beograd: Srpska kraljevska akademija, 1934.

Totomanova, Anna Marija. "Diahronnyj korpus bolgarskogo jazyka: sostojanie i perspektivy." *Filologija* 68 (2017), 223–242.

Vitić, Zorica, Aleksandar Kostić. "Specifični zahtevi anotacije srpskoslovenskog u okviru Elektronskog korpusa srpskog jezika." U *Digitalne biblioteke i digitalni arhivi*. Beograd: Filološki fakultet, 2015, 315–323.

Vujošević, Žarko. *Srpska vladarska kancelarija u srednjem veku. Studija iz uporedne diplomatike.* Doktorska disertacija. Filozofski fakultet, Beograd, 2015.]

About the author

Vladimir Polomac is an associate professor at the Department of the Serbian Language at the Faculty of Philology and Arts, University of Kragujevac (Serbia). For the monograph entitled *Језик повеља и писама Српске деспотовине* [The Language of Charters and Letters of Serbian Despotate] he received the "Pavle and Milka Ivić" award by the Serbian Slavic Association for the best book in the field of linguistic Slavic studies in Serbia in 2016. His current scientific interests include historical (corpus) linguistics (historical dialectology and onomastics of the Serbian language), philological and linguistic research of Serbian medieval literacy especially. He has been a member of the Onomastics Committee of the Serbian Academy of Sciences and Arts since 2015.