Submitted on 2024, June 27th
Accepted on 2024, September 4th

"Studi Slavistici", XXI, 2024, 2: 37-60
DOI: 10.36253/Studi_Slavis-16301
ISSN 1824-7601 (*online*)
Articles

Vladimir Polomac
Achim Rabus

Serbian Early Printed Books from Venice. Quantitative Approach to Orthographic Variations*

1. *Introduction*

Recent advances in IT based on artificial intelligence and machine learning, applied to automatic text recognition (cfr. Rabus 2019), have opened new perspectives for philological and linguistic investigations of Serbian written and printed heritage. Traditional qualitative investigations based on smaller samples from rather voluminous individual medieval manuscripts and early printed books can now be complemented by quantitative investigations based on large, automatically generated textual corpora (Rabus, Petrov 2023: 25-26; Rabus *et al.* 2023). This paper represents an attempt to illustrate this methodological approach through the example of early Serbian books printed in Venice in the early-mid 16th century in Božidar Vuković's printing shop. Unlike the books printed in other early Serbian printing shops, which have been the subject of numerous studies (cfr. specifically Grković-Mejdžor 1994; Grbić 2008; Jerković 1968, 1970, 1972; Samardžić 2012, 2013, 2019), early Serbian books printed in Venice, despite the fact that they represent a large part (about two-thirds) of the early Serbian heritage which has been preserved (for a detailed list cfr. Lazić 2018: 178-182), have mostly been investigated from the perspective of their typographical (cfr. specifically Grbić 2020; Subotin-Golubović 2020), historical (cfr. specifically Lazić 2021, 2022) or theological aspects (cfr. Hrvaćanin 2017), and much less frequently from that of philological and linguistic ones (cfr. specifically Jerković 1967; Sindik 1986; Grković-Mejdžor 2012; Rabus, Petrov 2023: 32-36). This paper will cover the topic of variations in orthography in light of the well-known fact that orthographical differences between Serbian early printed books can be attributed to quantitative rather than qualitative factors, and that they are more manifested in inconsistent spellings or in the rate of occurrence of particular techniques, rather than in their adoption or rejection (cfr. Pešikan 1994: 164).

* This paper was written as a part of the international bilateral project called *Creating AI Models for Automatic Processing of Serbian Medieval Manuscripts* (2024-2025) funded by The Ministry of Science, Technological Development and Innovation of the Republic of Serbia and German Academic Exchange Service (DAAD).

Vladimir Polomac (University of Kragujevac) – v.polomac@filum.kg.ac.rs
Achim Rabus (University of Freiburg) – achim.rabus@slavistik.uni-freiburg.de
The authors declare that there is no conflict of interest

2. *Theoretical and Methodological Framework*

Our research involved five early books printed in Venice in Božidar Vuković's printing shop: *Liturgikon* (1520) (abbr. *Lit. 1520*), *Psalter with Appendices* (1520) (abbr. *Psal. 1520*), *Prayer Book (Miscellany for Travellers)* (1536) (abbr. *Misc. 1536)*, *Festal Menaion* (1538) (abbr. *Men.* 1538), and *Prayer Book (Euchologion)* (1538-1540) (abbr. *Euch. 1538-1540*). *Lit. 1520* and *Psal. 1520*, were printed by Hieromonk Pahomije of Montenegro in the first period of Božidar Vuković's activity as a printer[1]. The remaining books originate from the second period of his printing activity: *Misc. 1536* and *Men. 1538* were printed by Hieromonk Mojsije from Dečani, while *Euch. 1538-1540* was most probably printed by Hieromonks Teodosije and Đenadije[2]. We used copies of these books from the Matica Srpska Library in Novi Sad (for more on these books, see Grbić *et. al.* 1994: 25-30, 33-34, 41-60), which are also available as digital objects from the 15th-17th c. Serbian books collection in the Matica Srpska Digital Library[3]. Digital copies of the books were uploaded onto the *Transkribus* platform, without preprocessing. *Transkribus* is based on AI (artificial intelligence), machine learning and advanced neural networks, which have been established as a standard for automatic text recognition and the creation of searchable digital editions of medieval manuscripts and early printed books[4]. The process of automatic text recognition initially involved automatic layout recognition using a model previously trained on book material from various Serbian printing shops from the 15th and 16th centuries. Since the error rate manifested during the training of this model was 7.64%, it was necessary to manually check and correct the text regions and lines. We used a previously trained model called *Dionisio 2.0* for the automatic text recognition, which resulted in very good performance (cfr. Polomac 2022a, 2022b). For this paper, we trained the *Dionisio 3.0* model, which differs from the previous version of the model only in that it disregards accent marks (diacritics). This approach was justified under the assumption that the omission of accent marks from the model training process would yield even better results in automatic text recognition, and that it would substantially facilitate the further search for orthographic variations conducted in the automatically produced

[1] For more on *Lit. 1520* and *Psal.* 1520 see Pešikan 1994: 167-168, 193-195. Along with these two books, in the first period of Božidar Vuković's activity as a printer (1520-1521), the *Prayer Book (Miscellany for Travellers)* was printed in a concise and full version (Lazić 2021: 142). For a detailed account of Hieromonk Pahomije, the printer of these books, see Pešikan 1994: 162.

[2] For more on *Lit. 1536*, *Men. 1538* and *Euch.* 1538-1540 see Pešikan 1994: 122, 144-145, 147-148. Along with these books, during the second period of Božidar Vuković's activity as a printer (1536-1540), *Octoechos, Mode 5-8* (1537) was also published (Lazić 2021: 142). For more information on Hieromonks Mojsije, Teodosije and Đenadije, the printers of these books, see Lazić 2021: 147 and Pešikan 1994: 113, 198.

[3] Collection available at: <http://digital.bms.rs/ebiblioteka/publications/index/collection:4> (last access: 10.01.2025).

[4] For more details see <https://www.transkribus.org/> (last access: 10.01.2025).

text, which was conducted with the text-corpus analysis program *AntConc*[5]. During the training of the *Dionisio 3.0* model, we used the same set of data which had been used for the training of the previous version *Dionisio 2.0* (see Polomac 2022b: 159), but this time with prior automatic removal of accent marks. The characteristics and performance data of the *Dionisio 3.0* model are presented in TABLE 1:

TABLE 1. Performance of *Dionisio 3.0* model

Word count on the training set	Number of epochs[6]	CER on the training set	CER on the validation set
175 453	125	0.70%	1.81%

A comparison of this model with the previous version shows that the already very good performance has been improved further: the rate of incorrectly recognized characters in the *Dionisio 2.0* model validation set is 2.40% (including accent marks), while in the *Dionisio 3.0* model it is just 1.81% (without accent marks). The performance differential between these two models shows that the automatic recognition of accent marks represents a smaller problem for the *Transkribus* algorithm than expected.

Carrying out the automatic text recognition using the *Dionisio 3.0* model yielded a large corpus for study. Its characteristics are presented in TABLE 2. In order to gain a clearer picture regarding the quality of the corpus created in such a manner and used for the purposes of performing quantitative philological and linguistic investigations, we manually corrected small samples of all the printed books and compared them with the automatically recognized texts. Quantitative results are presented in TABLE 3.

TABLE 2. Corpus characteristics

Book	Pages	Tokens
Lit. 1520,	458	49,795
Psal. 1520	692	73,872
Misc. 1536	471	52,597
Men. 1538	861	243,058
Euch. 1538-1540	557	64,650
Total	**3,039**	**483,972**

[5] See <https://www.laurenceanthony.net/software/antconc/> (last access: 10.01.2025).

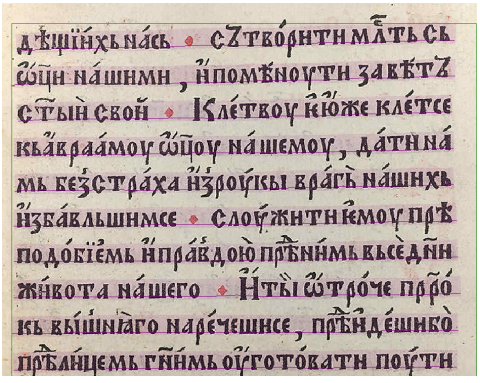
[6] In machine learning, the term epoch refers to one complete presentation of the data to a learning machine (cfr. Burlacu, Rabus 2021: 1).

TABLE 3. Quantitative indicators of the *Dionisio 3.0* model

Book	CER
Lit. 1520	0.96%
Psal. 1520	0.83%
Misc. 1536	1.05%
Men. 1538	1.44%
Euch. 1538-1540	0.69%

TABLE 3 shows that CER (character error rate) is even smaller than in the validation data set employed during the training of the *Dionisio 3.0* model. A noticeably lower rate of incorrectly recognized characters (below 1%) is observed in *Euch. 1538-1540*, *Lit. 1520* and *Psal. 1520*. The rate of incorrectly recognized characters is just a little higher in *Misc.* 1536, while in *Men.* 1538, it is somewhat higher than in other books (1.44%), but it can still be deemed exceptionally low. A true picture of the success of this model can be obtained only when the quantitative indicators are complemented by a qualitative analysis. As an illustration, we take a fragment of page 152r from *Psal. 1520* and the corresponding automatically recognized text using the *Dionisio 3.0*, as shown in TABLE 4:

TABLE 4. Qualitative indicators of success pertaining to *Dionisio 3.0* model

Psal. 1520, page 152r,	*Dionisio 3.0,*
	1–1 дѣщїихъ насъ · сътворити мл҃ть сь– 1–2 wц҃и нашими, и помѣноути завѣтъ 1–3 с҃тыи свои · клетвоу ѥюже клет се 1–4 кь авраамоу wц҃оу нашемоу, дати на– 1–5 мь безъ страха изъ рукы врагъ нашихъ 1–6 избавльшим се · слоужити ѥмоу пр҃ѣ– 1–7 подобїемь и правъдою пр҃ѣ нимь вьсе дн҃и 1–8 живота нашего · И ты wтроче прр҃о– 1–9 кь выш҃наго нарѣчеши се, пр҃ѣ идеши бо 1–10 пр҃ѣ лицемь г҃нимь оуготовати поути

The table above suggests that the sparse errors in the process of automatic text recognition are due to spaces between words: сь- wц҃и 1, безъ страха (5) instead of the expected сь wц҃и (1), безъ страха (5). Apart from this category, errors can also be found in superscript letters, the pajerak mark and the titlo mark, and highly infrequently in individual letters.

Judging by the quantitative and qualitative indicators of the success of the *Dionisio 3.0* model presented here, we believe that the resulting texts can be used for further quan-

titative investigations of orthographic variation without additional manual correction. In recent years, this methodological approach has been advocated by Besters-Dilger, Rabus 2021, Rabus, Petrov 2023 and Rabus *et al.* 2023; these studies have underscored the feasibility of deploying of automatically obtained data without additional manual correction for quantitative philological and linguistic analyses of medieval Slavonic manuscripts and early printed books. In all these papers, the authors prove the hypothesis that quantitative analysis of medieval Slavonic manuscripts and early printed books can also be successfully based on texts that have been recognized and morpho-syntactically annotated automatically, without additional manual correction (see specifically Besters-Dilger, Rabus 2021: 87-90; Rabus, Petrov 2023; Rabus *et al.* 2023: 113). Although results obtained in this manner may prove not worse than those obtained with a traditional approach, which includes a qualitative analysis of smaller portions of manually transcribed manuscripts or early printed books, Rabus, Petrov 2023 point to limitations of both of these approaches and emphasise the necessity of combining the two – more precisely, of applying an approach in which quantitative and qualitative methods are complementary.

Following automatic recognition of Serbian early printed books with the *Transkribus Dionisio 3.0* model, the next methodological step involved downloading the recognized texts as separate plain text files, and then removing the hyphens in words that break across lines, along with all other punctuation marks, in the text editor *Notepad++*[7]. The texts prepared in this way were first submitted to computer-based stylometric analysis with the *R stylo* package (see Eder *et al.* 2016), followed by a quantitative analysis of the most important individual orthographic variations, conducted with the program *AntConc*. The first approach was meant to provide an insight into orthographic macrovariations and general relations among the books from the corpus, while the second approach was supposed to reveal the most important microvariations (among individual orthographic features) observed in the books.

3. *Computer Stylometry and Orthographic Macrovariations*

The aforementioned computer-based stylometric analysis using the *stylo* package for the *R* statistical programming language entailed automatic extraction and quantification of different linguistic features from large textual corpora accompanied by a corresponding statistical analysis (see Eder *et al.* 2016: 108). In recent years, such an approach to texts has been applied in various areas of linguistics, most commonly in studies concerning various aspects of authorship (authorship attribution, author identification, individual style variations and the like) (Eder 2011, 2016: 107), but also in studies on diachronic variation in language (see specifically Górski *et al.* 2019) and studies concerning standard language stratification (see specifically Von Waldenfels, Eder 2016; Lahjouji-Sepällä *et al.* 2022). The main goal of the computer-based stylometric analysis conducted in this section was to validate

[7] See <https://notepad-plus-plus.org/> (last access: 10.01.2025).

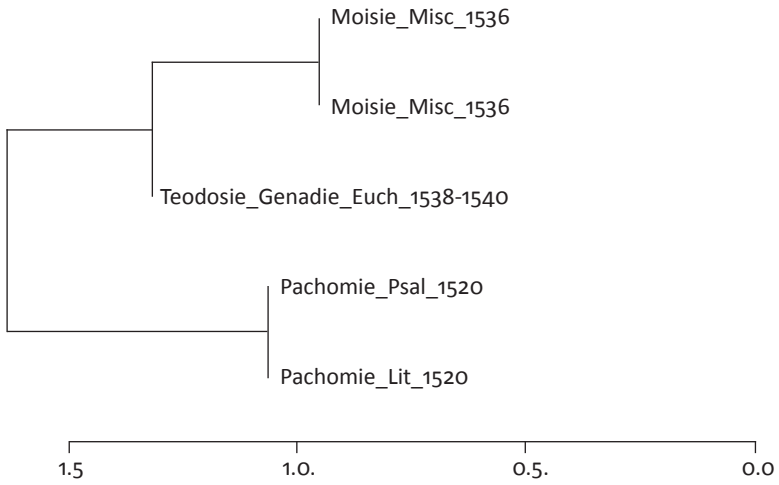
the hypothesis that orthographic variations in the books from the corpus are the result of their being typeset by different printers during two periods of Božidar Vuković's activity in Venice: in the first period, Hieromonk Pahomije printed *Lit. 1520* and *Psal. 1520*, while in the second period, Hieromonk Mojsije printed *Misc. 1536* and *Men. 1538*, and Hieromonks Teodosije and Đenadije printed *Euch. 1538-1540*.

The key question in testing this hypothesis was the choice of linguistic features to be automatically selected and quantified (statistically analysed) in the texts. The *R stylo* package enabled us to compute the most frequent *n*-grams on either the character or the token levels (Eder *et al.* 2016: 108). Which of these approaches one ought to choose, as well as the optimal *n*-gram length (two, three, four or more), mostly depends on the language of the text and the research question itself (Stamatatos 2009: 541-542). Since orthographic variations are the subject of our research, we chose the *n*-gram-based approach on the character level, as the choice of an *n*-gram on the token level would most probably be more suitable for, say, an analysis of the genre of the books mentioned above. When establishing the optimal length of an *n*-gram, we considered the fact that orthographic variations occur most frequently in the connexions between two characters (for instance, in the prepositions and prefixes **вѣ**(-), **сѣ**(-) or **вь**(-), **сь**(-); in combinations of graphemes **їа** or **иꙗ**; or in combinations of the graphemes **ꙗ** or **ѥ** at the beginning of a word or after a vowel), so we based our stylometric analysis on the automatic generation and quantification of the most frequent bigrams on the character level.

The second important question in the stylometric analysis concerned the choice of suitable metrics for the calculation of mutual text distance. The *R stylo* package allows one to choose among different metrics for calculating the distance of texts based on multivariate statistical analysis (Eder *et al.* 2016: 109). In our experiment, we used Cosine Delta Distance (or Würzburg Delta Distance) (see Eder *et al.* 2017), an enhanced version of Delta metrics, the most common and most frequent metrics in stylometry (see Burrows 2002; Büttner *et al.* 2017). The results of the automatic computer stylometry analysis are shown in GRAPH 1 as a dendrogram, in which the books from the corpus were classified into clusters in terms of their mutual distance.

By observing the graph, we can conclude that the automatic computer-based stylometric analysis using 100 most frequent character-based bigrams confirmed the initial hypothesis that orthographic variations in Serbian early printed books from Božidar Vuković's printing shop in Venice depend on the time in which they were printed and the identity of the individual printer himself. The graph shows that the books were first classified into two clusters corresponding to the two periods of Božidar Vuković's activity as a printer (*Lit. 1520* and *Psal. 1520* in the first, *Misc. 1536*, *Men. 1538* and *Euch. 1538-1540* in the second cluster), and that the second cluster was subsequently additionally divided according to different printers (*Misc. 1536* and *Men. 1538* were printed by Hieromonk Mojsije, and *Euch. 1538-1540* by Hieromonks Teodosije and Đenadije).

GRAPH 1. Serbian early printed books from Venice grouped into clusters



4. *Orthographic Microvariations*

The general pattern of variation presented in the previous section using computer stylistics will be supplemented here by consideration of the most important individual orthographic features observed in the books from the corpus. Previous investigations of 16th-century Serbian early printed books indicate the presence of a post-Resavian orthography, in which the features of the Resava and Raška schools overlap (cfr. Jerković 1970: 17; Grković-Mejdžor 1994: 214; Grbić 2008: 263; Samardžić 2013: 120). Our analysis took into account some of the most important orthographic categories reflecting either the Resava or the Raška tradition (following Jerković 1980): the a) spelling of the phonemic group [ja] in postvocalic position, b) spelling of the phonemic group [je] in initial and postvocalic positions, c) spelling of the vowel /i/ before /j/, d) spelling of the graphemes ы and ъ; and e) the etymological or conventional use of the grapheme s in certain roots. The spelling of the graphemes ѯ, ѱ, ѳ, and ѵ was not investigated, since it was expected that their use in non-Slavonic words was codified by explicit rules (cfr. Grković-Mejdžor 1994: 42-44). The omission of accent marks and punctuation from the analysis has already been explained in terms of the difficulties encountered during the automatic look-up and statistical processing of the text (see § 2 above). The quantitative analysis of the most important orthographic features in the books from Božidar Vuković's printing shop was performed

using *AntConc* by creating specific queries using the combinations of graphemes, or regular expressions and combinations of graphemes, and then by extracting the examples and processing them statistically.

4.1. *Group [ja] in Postvocalic Position*

Contrary to initial position, where the ligature ꙗ is expected to be used uniformly, in postvocalic position, we can also expect the grapheme a in accordance with Resava tradition. Since previous research on the marking the [ja] group in Serbian 15th-16th c. early printed books (cfr. Jerković 1968: 92-93, 1970: 9-10; Grković-Mejdžor 1994: 22-24; Grbić 2008: 226-228; Samardžić 2012: 161) indicates that orthographic variations depend on the preceding vowel, we can plausibly classify the material from early books form Božidar Vuković's printing shop in Venice using said criterion.

Performing a look-up using the regular expressions [ѧ,ꙗ]ѧ (ѧ after ѧ or ꙗ) (Resava tradition) and [ѧ,ꙗ]ꙗ (ꙗ after ѧ or ꙗ) (Raška tradition), and then by manually filtering selected examples[8], we created a statistical representation of the spellings of the [aja] group, presented in TABLE 5.

TABLE 5. Spelling of the [aja] group in the corpus

Book	[ѧ,ꙗ]ѧ	%	[ѧ,ꙗ]ꙗ	%
Lit. 1520	233	92.10%	20	7.90%
Psal. 1520	558	94.73%	31	5.26%
Misc. 1536	660	95.94%	29	4.06%
Men. 1538	4,015	96.47%	147	3.53%
Euch. 1538-1540	275	48.24%	295	51.75%

The table shows that the [aja] group is almost uniformly marked by the grapheme a in the vast majority of the investigated books, in line with the Resava tradition[9]. *Euch. 1538-1540* manifests a significant deviation from the situation mentioned above, as the Raška and Resava traditions are almost equally represented in it.

[8] The examples of writing the aa group were filtered manually because this group can also represent ā in certain categories: for instance, in the gen. sg. masc. definite form of adjectives, forms of the imperfect, personal names like *Avraam*, *Aaron*, *Isaak* and the like.

[9] Such orthography is also characteristic of the *Psalter with Appendices* from Crnojević's printing shop (Grković-Mejdžor 1994: 22), while the samples from the books of the Goražde and Mrkšina Crkva printing shops exhibit the Resava tradition uniformly (Jerković 1968: 92- 93, 1970: 9; Samardžić 2012: 161).

Looking up the graphemic combinations ѣа (Resava tradition) and ѣꙗ (Raška tradition)[10], followed by the manual filtering of selected examples[11], gave us a statistical overview of the spelling of the [eja] group in Slavonic words from the corpus, which is presented in TABLE 6.

TABLE 6. Spelling of the [eja] group using ѣа or ѣꙗ in Slavonic words from the corpus

Book	ѣа	%	ѣꙗ	%
Lit. 1520	29	37.17%	49	62.82%
Psal. 1520	1	2.43%	40	97.56%
Misc. 1536	1	2.38%	41	97.61%
Men. 1538	70	28.68%	174	71.31%
Euch. 1538-1540	3	7.31%	38	92.68%

The table shows that books from Božidar Vuković's printing shop can be divided into two groups according to this orthographic feature: the first group includes *Psal. 1520*, *Misc. 1536* and *Euch. 1538-1540*, with their almost fully uniform use of the ligature ꙗ[12], while the other group includes *Lit. 1520* and *Men. 1538* with a predominant use of ligature ѣа, but with a substantial number of examples with the grapheme а corresponding to the Resava tradition.

By performing the look-up of the graphemic combinations їа and ию, we obtained a statistical overview of the spelling of the [ija] group in the corpus, which is presented in TABLE 7.

TABLE 7. Spelling of the [ija] group in the corpus[13]

Book	-їа	%	-ию	%
Lit. 1520	812	96.55%	29	3.44%
Psal. 1520	1 352	98.83%	16	1.16%
Misc. 1536	897	98.89%	10	1.10%
Men. 1538	3 856	85.40%	659	14.60%
Euch. 1538-1540	1 058	94.29%	64	5.70%

[10] Spellings of the [eja] group as ѣа or ѣꙗ were omitted from statistical processing because the number of attested examples was very small.

[11] Non-Slavonic words have also been omitted from the analysis (predominantly personal names), in which [ja] was almost uniformly marked by the grapheme а.

[12] Such an orthography is also observed in the *Psalter with Appendices* from Crnojević's printing shop (Grković-Mejdžor 1994: 23).

[13] The analysis excluded rare examples in which the [ija] group was spelt as їꙗ or иа.

The table shows that the [ija] group is almost uniformly marked by ѩ in the investigated books, in accordance with the Resava tradition[14]. In most of the books, cases where the иа group adheres to the Raška tradition only occur as isolated individual examples, while only in *Men. 1538* do we find more examples of this type (around 15%).

By performing a look-up of graphemic combinations оа and оиа, we obtained a statistical overview of spellings of the [oja] group, which is presented in TABLE 8.

TABLE 8. Spelling of the [oja] group in the corpus

Book	оа	%	оиа	%
Lit. 1520	–	–	151	100%
Psal. 1520	2	0.45%	439	99.54%
Misc. 1536	1	0.46%	213	99.53%
Men. 1538	41	5.93%	650	94.06%
Euch. 1538-1540	2	0.83%	231	99.14%

The table indicates a uniform (*Lit. 1520*) or (in the remaining books) almost uniform spelling of оиа in accordance with the Raška tradition[15].

A look-up using the regular expressions [оу,ѳ]а (а after оу or ѳ) (Resava tradition) and [оу,ѳ]иа (иа after оу or ѳ) (Raška tradition) led to the finding that in the books from Božidar Vuković's printing shop the [uja] group was uniformly marked in the Raška manner[16].

4.2. *Group [je] in Initial Position*

This phonemic group can be written in initial position using the ligature ꙗ in accordance with the Raška orthographic tradition, but also using the graphemes є and ѥ in accordance with Resava tradition. Previous research into Serbian early printed books shows that the relation between Resava and Raška traditions is governed by distinct principles when it comes to Slavonic words, while non-Slavonic words mostly contain the graphemes

[14] Such an orthography is also observed in the *Psalter with Appendices* from Crnojević's printing shop (Grković-Mejdžor 1994: 22-23), books from the Goražde printing shop (Grbić 2008: 227; Samardžić 2012: 161), as well as the books from the printing shop of Mrkšina Crkva (Jerković 1968: 93, 1970: 10).

[15] Such an orthography is also a feature of the *Psalter with Appendices* from Crnojević's printing shop (Grković-Mejdžor 1994: 23), the *Liturgikon* and *Euchologion* from the Goražde printing shop (Samardžić 2012: 162), as well as books from the printing shop of Mrkšina Crkva (Jerković 1968: 93, 1970: 9-10). It is interesting that the Resava tradition prevails in the *Psalter with Appendices* from the Goražde printing shop (Grbić 2008: 228).

[16] The same orthography is also observed in the *Psalter with Appendices* from Crnojević's printing shop and the *Psalter with Appendices* from the Goražde printing shop (Grković-Mejdžor 1994: 23; Grbić 2008: 229).

ѥ and є in line with the orthography of the Greek source texts (cfr. Jerković 1970: 7-8; Grković-Mejdžor 1994: 28; Grbić 2008: 231).

Performing a look-up in the corpus using the regular expressions ^ꙗ (the ligature ꙗ at the beginning of a word) and ^[ѥ,є] (ѥ or є at the beginning of a word) yielded a statistical overview that is presented in TABLE 9.

TABLE 9. Word-initial spelling of [je] group in the corpus

Book	^ꙗ	%	^[ѥ,є]	%
Lit. 1520	1,250	58.74%	878	41.25%
Psal. 1520	1,531	58.25%	1,097	41.74%
Misc. 1536	738	46.03%	865	53.96%
Men. 1538	1,894	24.09%	5,968	75.90%
Euch. 1538-1540	1,809	82.60%	381	21.06%

The table indicates that the same orthography was employed in the first period of Božidar Vuković's activity as a printer, but also that there are considerable variations in the books printed in the second period of his shop's operation. The Raška tradition with the ꙗ ligature slightly prevails over the Resava tradition with the graphemes ѥ or є in *Lit. 1520* and *Psal. 1520*, in *Men. 1538* there is a considerable predominance of the Resava tradition, in *Euch. 1538-1540* Raška considerably prevails, while in *Misc. 1536* both traditions are represented in approximately even proportions, with Resava having a slight prevalence. A detailed comparison with early books from other printing shops was not possible due to the use of different research methodology and presentation of data. Our statistical overview encompasses both Slavonic and non-Slavonic words, in contrast to other studies in which this phenomenon in non-Slavonic words is investigated separately (cfr. for example Grković-Mejdžor 1994: 24-25, 27-28; Grbić 2008: 229, 231- 232). However, an analysis of the most frequent words containing these graphemes leads us to the conclusion that in the books of Božidar Vuković's printing shop in Venice, the *Psalter with Appendices* from Crnojević's printing shop and the *Psalter with Appendices* from the Goražde printing shop (Grković-Mejdžor 1994: 28, Grbić 2008: 231), the ligature ꙗ is used in non-Slavonic words only as an exception.

4.3. *Group [je] in Postvocalic Position*

We can expect to find the graphemes ѥ or є (in line with the Resava tradition) or the ligature ꙗ (according to the Raška tradition) marking the group [je] in postvocalic position. In a way similar to the marking of the group [ja] in postvocalic position (see § 3.2.), we organized the material by the vowel preceding the group [je]. Due to the absence or extremely small number of occurrences, we excluded examples with the grapheme є in this position from the quantitative analysis.

After performing a look-up using the regular expressions [ѧ,ꙗ]ѥ (ѥ after ѧ or ꙗ) (Resava tradition) and [ѧ,ꙗ]ꙗ (ꙗ after ѧ or ꙗ) (Raška tradition), we obtained a statistical overview of the spelling of the group [aje] in the corpus, as presented in TABLE 10.

TABLE 10. Word-initial spelling of the [aje] group in the corpus

Book	[ѧ,ꙗ]ѥ	%	[ѧ,ꙗ]ꙗ	%
Lit. 1520	131	25.43%	383	74.56%
Psal. 1520	224	27.93%	578	72.06%
Misc. 1536	174	30.85%	390	69.14%
Men. 1538	2,594	60.39%	1,701	39.60%
Euch. 1538-1540	293	28.64%	730	71.35%

The table shows that in the overwhelming majority of the books, we find the same orthography for the [aje] group, where the Raška tradition with the ligature ꙗ predominates significantly (69-75%)[17]. *Men. 1538* deviates considerably from the situation mentioned above, in that the Resava tradition with the grapheme ѥ predominates in it.

Performing a look-up using the regular expressions [ѥ,ѣ]ѥ (ѥ after ѥ or ѣ) (Resava tradition) and [ѥ,ѣ]ꙗ (ꙗ after ѥ or ѣ) (Raška tradition), we obtained a statistical overview of the spelling of the group [eje] in the corpus, as presented in TABLE 11.

TABLE 11. Spelling of the [eje] group in the corpus

Book	[ѥ,ѣ]ѥ	%	[ѥ,ѣ]ꙗ	%
Lit. 1520	42	16.47%	213	83.52%
Psal. 1520	54	16.07%	282	83.92%
Misc. 1536	35	17.94%	195	84.78%
Men. 1538	682	47.00%	769	52.99%
Euch. 1538-1540	28	11.15%	223	88.84%

The table indicates that the proportions are similar to those of the group [aje] within the corpus under study. In the vast majority of the books (*Lit. 1520, Psal. 1520*, *Misc. 1536,* and *Euch. 1538-1540*), there is a considerable predominance of the Raška tradition, where

[17] The predominance of the Raška tradition in the spelling of this group has also been observed in the *Psalter with Appendices* from Crnojević's printing shop (Grković-Mejdžor 1994: 23), as well as in the books from the Goražde printing shop (Grbić 2008: 230; Samardžić 2012: 164).

the ligature is written as ꙗ[18], while only in *Men. 1538* do the two spellings occur in approximately the same ratio[19].

By performing a look-up of the two-letter sequences їє (Resava tradition) and иє (Raška tradition), as well as їꙗ and иꙗ, we obtained a statistical overview of the spellings of the group [ije], as presented in TABLE 12[20].

TABLE 12. Spelling of the [ije] group in the corpus

Book	їє	%	иє	%	їє/иє	%
Lit. 1520	1,015	81.26%	223	17.85%	2/9	0.88%
Psal. 1520	1,523	84.05%	252	13.90%	9/28	2.04%
Misc. 1536	1,147	79.65%	205	14.23%	66/22	6.11%
Men. 1538	6,705	79.65%	1392	16.53%	61/260	3.81%
Euch. 1538-1540	1,394	74.34%	370	19.73%	67/44	5.92%

The table shows that in all the books, the phonemic group [ije] is most frequently written using їе in accordance with the Resava tradition, yet considerably less frequently using the graphemes ие in the Raška tradition, albeit with certain isolated hybrid cases. This situation in the books from Božidar Vuković's printing shop generally corresponds to the situation in the books from other 15th-16th c. early Serbian printing shops investigated to date (Jerković 1970: 8, 1972: 93-94; Grković-Mejdžor 1994: 25-26; Grbić 2008: 230-231).

Looking up the regular expressions [ѡ,ѡ]є (є after ѡ or ѡ) (Resava tradition) and [ѡ,ѡ]ꙗ (ꙗ after ѡ or ѡ) (Raška tradition), we obtained a statistical overview of the spellings of the group [oje] in the corpus, as presented in TABLE 13.

TABLE 13. Spelling of the [oje] group in the corpus

Book	[ѡ,ѡ]є	%	[ѡ,ѡ]ꙗ	%
Lit. 1520	255	31.40%	557	68.59%
Psal. 1520	702	39.06%	1,094	60.91%
Misc. 1536	415	32.24%	872	67.74%
Men. 1538	2,823	78.83%	758	21.23%
Euch. 1538-1540	593	36.46%	1,032	63.46%

[18] Such an orthography is also found in the *Psalter with Appendices* from Crnojević's printing shop (Grković-Mejdžor 1994: 25), as well as in the *Liturgikon* from the Goražde printing shop (Samardžić 2012: 164).

[19] This orthography is also a feature of the *Psalter with Appendices* from the Goražde printing shop (Grbić 2008: 230).

[20] Isolated cases of sequences with the grapheme ҿ have been omitted from the statistical analysis here.

The table shows that in the majority of the books (*Lit. 1520, Psal. 1520*, *Misc. 1536,* and *Euch. 1538-1540*), the Raška tradition prevails with the ligature ꙗ[21], while in *Men. 1538*, just as it is the case with writing the groups [aje] and [eje], the Resava style significantly predominates over that of Raška.

Performing a look-up using the regular expressions [оу,ѫ]є (є after оу or ѫ) and [оу,ѫ]ꙗ (ꙗ after оу or ѫ), we obtained a statistical overview of the spellings of the group [uje] in the corpus, as presented in TABLE 14.

TABLE 14. Spelling of the [uje] group in the corpus

Book	[оу,ѫ]є	%	[оу,ѫ]ꙗ	%
Lit. 1520	20	17.39%	95	82.60%
Psal. 1520	23	17.69%	107	82.30%
Misc. 1536	25	21.92%	89	78.07%
Men. 1538	237	38.78%	374	61.21%
Euch. 1538-1540	32	19.75%	130	80.24%

The table shows that in the majority of the books (*Lit. 1520, Psal. 1520*, *Misc. 1536,* and *Euch. 1538-1540*), the Raška tradition considerably prevails over Resava when it comes to the spelling of the phonemic group [uje], while in *Men. 1538,* the Raška tradition predominates to a somewhat lesser degree[22].

4.4. *Spelling of the [iju] Group in Non-Initial Position*

To the previously described orthography of phonemic groups [ija] and [ije] (cfr. §§ 3.1. and 3.3.), we may add the orthography of the phonemic group [iju] when it is not in initial position. By performing a look-up of the two-letter sequences їю (Resava tradition) and ию (Raška tradition), we obtained a statistical overview of the spelling of this group in the corpus, which is presented in TABLE 15.

The table shows that in the books from Božidar Vuković's printing shop, the Resava tradition also prevails in spellings of the [iju] group: to a large extent in *Lit. 1520, Psal. 1520*, *Misc. 1536,* and *Euch. 1538-1540*, and to a somewhat smaller extent in *Men. 1538*[23].

---

[21] This is also a feature of the *Psalter with Appendices* from Crnojević's printing shop (Grković-Mejdžor 1994: 26).

[22] Said situation in books from Božidar Vuković's printing shop generally corresponds to the situation in the books investigated so far from other early 15th-16th century Serbian printing shops (Grković-Mejdžor 1994: 26-27; Grbić 2008: 231; Samardžić 2012: 166).

[23] The almost fully uniform spelling of їю in this role has also been observed in the *Psalter with Appendices* from Crnojević's printing shop (Grković-Mejdžor 1994: 34) and *Psalter with Appendices* from the Goražde printing shop (Grbić 2008: 236-237).

TABLE 15. Spelling of the [ije] group in the corpus

Book	їю	%	ию	%
Lit. 1520	262	88.51%	34	11.48%
Psal. 1520	349	90.64%	36	9.35%
Misc. 1536	332	84.91%	59	15.08%
Men. 1538	1,771	70.95%	725	29.04%
Euch. 1538-1540	436	84.66%	79	15.33%

4.5. *Grapheme* ы

Former research into the orthographies of Serbian 15th-16th century early printed books indicated that the use of the grapheme ы was not codified, namely that ы was written in its etymological position, but also that there was mixed usage of the graphemes ы and и (Jerković 1970: 13-14, 1972: 99-100; Grković-Mejdžor 1994: 37; Grbić 2008: 244-245; Samardžić 2013: 118-119). A quantitative method for researching the use of the graphemes ы and и is not quite suitable, because orthographic variations are not conditioned by the graphemic environment, so they cannot be efficiently looked up with a query in a corpus analysis program such as *AntConc*. Quantitative research cannot provide reliable data about variations in the stems and forms of individual words, since the corpus was not lemmatized or annotated. As we possess for now only the automatically recognized text from the books, we can only provide a statistical overview of the frequency of the use of the grapheme ы in this same corpus (TABLE 16):

TABLE 16. Frequency of the use of the grapheme ы in the corpus

Book	Number of hits	Number of characters	Frequency
Lit. 1520	3,558	291,603	1.22%
Psal. 1520	4,751	443,694	1.07%
Misc. 1536	3,734	324,658	1.15%
Men. 1538	20,970	1,579,925	1.32%
Euch. 1538-1540	4,695	395,309	1.18%

The table shows that in all the books form Božidar Vuković's printing shop, the grapheme ы is used with almost the same frequency.

4.6. *Grapheme* ъ

Prior research into the orthography of 15th and 16th century Serbian early printed books indicates that the situation is variable in terms of the use of grapheme ъ. The observation

of P. Đorđić (1991: 190) that there was an orthography with one front yer in the Crnojević, Belgrade and Skadar printing shops[24], as well in the printing shop of Jerolim Zagurović in Venice, was amended by Pešikan's (1994: 164) finding that the grapheme ѣ cannot be found either in short excerpts of books from the printing shop of Mrkšina Crkva, nor in the *Psalter with Appendices* (1557) from the Mileševa printing shop[25], nor in the *Gospel* from Rujno and Jakov Krajkov's *Miscellany*. For books from other printing shops, Pešikan (1994: 164) states that there was "partial use of the back yer, most commonly in originally semivocalic monosyllabic morphemes", hence claiming that the impact of these letters was regarded more as "free graphical and calligraphic variation than orthographic differentiation". Research into the orthography of books from the Goražde printing shop points to the presence of both graphemes only in the *Psalter with Appendices*, while the other books are characterized by a single-yer orthography (Grbić 2008: 245- 246; Samardžić 2013: 115).

A look-up of the graphemes ь and ѣ with *AntConc* enabled us to create a quantitative statistical overview of the frequencies of these graphemes in the books from the corpus, as presented in TABLE 17.

TABLE 17. Number of hits and frequencies of the graphemes ь and ѣ in the corpus

Book	ь	ѣ	Characters	Frequency of ь	Frequency of ѣ
Lit. 1520	12,140	1,607	291,603	4.16%	0.55%
Psal. 1520	17,264	2,502	443,694	3.89%	0.56%
Misc. 1536	13,343	1,398	324,658	4.10%	0.43%
Men. 1538	63,126	3,747	1,579,925	3.99%	0.23%
Euch. 1538-1540	18,603	831	395,309	4.70%	0.21%

The table shows that the grapheme ѣ can be found in all the books from Božidar Vuković's printing shop, but that its distribution is not even. The highest frequency was observed in *Psal. 1520* and *Lit. 1520*, followed by *Misc. 1536*, while in *Men. 1538* and *Euch. 1538-1540* the frequency is at most half of that in the two previously mentioned books.

The occurrence of the grapheme ѣ in specific categories has not always lent itself well to automatic quantitative analysis. By analysing the list of the most frequent words containing ѣ and ь in *AntConc*, we concluded that the two graphemes are most commonly used in the prepositions вѣ/вь, съ/сь, and кѣ/кь. A statistical overview of the use of these graphemes in these prepositions in the books from the corpus is shown in TABLES 18 and 19.

[24] A rare instance of the grapheme ъ in the *Belgrade Gospel* by the printer Mardarije is attributed by Jerković (1972: 94-95) to the manuscript basis from which the books were copied.

[25] According to research by V. Jerković (1972: 94-95), the orthography of the printer Mardarije from Mrkšina Crkva was of a single-yer type, which is confirmed by the *Gospel*, while the presence of the graphemes ь and ѣ in the *Triod*, and in the *Gospel*, can be associated with the basis from which the book was copied.

TABLE 18. Spelling of the prepositions вѣ and вь in the corpus

Book	вѣ	%	вь	%
Lit. 1520	411	31,88%	878	68,11%
Psal. 1520	735	34,09%	1421	65,90%
Misc. 1536	255	23,67%	822	76,33%
Men. 1538	744	15,38%	4092	84,61%
Euch. 1538-1540	198	12,88%	1339	87,11%

TABLE 19. Spelling of the prepositions сѣ and сь, кѣ and кь in the corpus

Book	сѣ	%	сь	%	кѣ	%	кь	%
Lit. 1520	84	26.00%	239	73.99%	81	25.15%	241	74.84%
Psal. 1520	129	39.69%	196	60.30%	36	23.68%	116	76.31%
Misc. 1536	74	24.02%	234	75.97%	39	16.18%	202	83.81%
Men. 1538	154	13.12%	1,019	86.87%	125	10.77%	1,044	89.53%
Euch. 1538-1540	43	12.87%	291	87.12%	38	13.14%	251	86.85%

Unlike in the *Psalter with Appendices* from the Goražde printing shop – in which the graphemes ѣ and ь are equally present in the prepositions вѣ/вь and сѣ/сь, while the grapheme ь is more frequent in the preposition кѣ/кь (Grbić 2008: 247-248) – spellings with the grapheme ь in all the investigated books from Božidar Vuković's printing shop prevail in these prepositions. Orthographic variations in individual books and this category generally fit with the overall relations presented in TABLE 17. The highest frequency of occurrences of the grapheme ѣ was registered in *Lit. 1520* and *Psal. 1520*, a slightly lesser frequency in *Misc. 1536*, and the least of all in *Men. 1538* and *Euch. 1538-1540.* Except in *Euch. 1538-1540*, where the use of the grapheme ѣ is balanced in all three prepositions, in the other books, the grapheme ѣ is more often found in the prepositions вѣ and сѣ.

The use of ѣ in other categories typical of the Resava orthography (prefixes вѣ-, вѣз-, and сѣ-/сь-, pronominal and adverbial root вѣс-) does not readily lend itself to look-up and automatic extraction of examples. By performing a look-up by using the regular expressions ^вѣ\S and ^вь\S (вѣ at the beginning of a word and any other character other than a blank space, or вь at the beginning of a word and any character other than a blank space), we can automatically obtain examples of the spellings вѣ and вь when they are not a part of a preposition, but not data about the relations in the categories of the prefixes вѣ-/вь-, вѣз-/вьз-, вѣс-/вьс- or the pronominal and adverbial root вѣс-/вьс-. By potentially narrowing down the look-up with the regular expressions ^вѣс\S or ^вьс\S (вѣс at the beginning of a word and any other character than a blank space, or вьс at the beginning of a word and any other character than a blank space), we would obtain data where there would still be overlap between examples with the prefixes вѣс-/вьс- and ones with the pronominal and adver-

bial root вѫс-/вьс-. Narrowing down the look-up using the regular expressions ^вѫз\S or ^вьз\S (вѫз at the beginning of a word and any other character other than a blank space) would, however, not only allow us to obtain examples with the prefixes вѫз-/вьз-, but would also yield examples such as вѫзвати, in which we find the prefix вѫ- before a verb beginning with з. By using the regular expressions ^сѫ\S or ^сь\S (сѫ at the beginning of a word and any other character than a blank space, or сь at the beginning of a word and any character other than a blank space), we obtained a result with the smallest number of 'false' hits. The largest number of resulting examples did correspond to the prefixes сѫ-/сь-, whereas a smaller number of examples could also belong to other categories (for instance, the noun сьнь, pronoun сьи, adverb сьда and the like). As with all the categories mentioned above, it is necessary to conduct a qualitative analysis after the automatic extraction and to filter the resulting data manually.

4.7. *Grapheme* ѕ

Prior research into the orthographies of early Serbian printed books indicate that the grapheme ѕ was most commonly used in the words ѕѣло, ѕвѣзда, ѕвѣрь, as well as derived terms, and less often in corresponding grammatical cases with the Proto-Slavic second palatalization (cfr. Grković-Mejdžor 1994: 41-42). Performing a look-up of the graphemic combinations ѕѣл-/зѣл-, ѕвѣзд-/звѣзд-, and ѕвѣр-/звѣр-, we obtained a statistical overview of the spelling of the grapheme ѕ in these words from the corpus, as represented in TABLE 20.

TABLE 20. Writing the words ѕѣло, ѕвѣзда and ѕвѣрь in the corpus

Book	ѕѣл-/зѣл-	ѕвѣзд-/звѣзд-	ѕвѣр-/звѣр-
Lit. 1520	9/–	8/1	/
Psal. 1520	39/1	19/1	1/15
Misc. 1536	21/–	12/1	5/6
Men. 1538	33/–	158/7	55/12
Euch. 1538-1540	–/11	–/6	–/2

The table shows that in the vast majority of the books from Božidar Vuković's printing shop (*Lit. 1520, Psal. 1520*, *Misc. 1536*, and *Men. 1538*), the grapheme ѕ can be found with a higher or a lower frequency when writing the words ѕѣло, ѕвѣзда, and ѕвѣрь (and words derived from them), while only *Euch. 1538-1540* manifests a consistent spelling of these words with the grapheme з. In the first group of books, the grapheme s was registered uniformly or almost uniformly in the spellings of the words ѕѣло and ѕвѣзда, while the spelling of ѕвѣрь points to an unstable practice: we see an almost uniform spelling of звѣрь in *Psal. 1520*, an equal distribution of the spellings of ѕвѣрь and звѣрь in *Misc. 1536*, and the predominance of the spelling ѕвѣрь in *Men.* 1538. The grapheme ѕ is also found with a

somewhat higher frequency only in *Misc. 1536* and *Men. 1538* in ѕмїѧ and derived words; in *Misc. 1536*, the use of ѕ and з in this word is balanced, while in *Men. 1538*, there is a considerable predominance of ѕ. Outside the words mentioned above, the grapheme ѕ in the books form Božidar Vuković's printing shop can be found only in rare or isolated instances.

5. *Concluding Remarks*

By analyzing the orthographic variations observed in Serbian early printed books from Božidar Vuković's printing shop in Venice, this paper has demonstrated that philological investigations of the early Serbian written and printed heritage can be based not only on a qualitative analysis of small-scale textual samples, but also on a quantitative analysis of entire automatically transcribed texts. Thanks to an HTR model for the automatic text recognition of Serbian early printed books previously trained on the *Transkribus* platform, we were able to conduct a quantitative investigation of orthographic variations in Serbian early printed books from Božidar Vuković's printing shop in Venice using a vast corpus of automatically recognized texts that were not corrected manually. The omission of manual correction did not have a significant impact on the results of the quantitative analysis, since CER for automatic recognition was extremely low (1-2%) and any errors were mostly related to spacing.

Orthographic variations in the textual corpus produced in this manner were analysed at the macrolevel using computer stylometry (*R stylo* package), as well as at the microlevel by means of look-up, automatic extraction, and statistical processing of examples of the most important orthographic features of post-Resava orthography, using the program *AntConc*. Automatic computer stylometric analysis of the 100 most frequent bigrams on the character level using the *R stylo* package confirmed the initial hypothesis that orthographic macrovariations in Serbian early books from Božidar Vuković's printing shop in Venice generally depend on the period in which they were printed and the individual printer.

The quantitative analysis of the most important individual features of the post-Resava orthography conducted using *AntConc* confirmed the previous hypothesis only in the instance of *Lit. 1520* and *Psal. 1520*, the books printed by Hieromonk Pahomije in the first period of Božidar Vuković's activity as a printer. The statistical relations in the most important individual orthographic features in these books usually match, with the sole exception that the combinations of graphemes ѣа and ѣѩ are used to write out [eja] in Slavonic words. A further partial exception was noted with regard to the frequency of the preposition съ/сь. In both books, the Resava orthography prevailed or was used almost uniformly to spell the phonemic groups [aja], [ija], [ije], and [iju]. The reverse – a considerable predominance of Raška spellings or almost uniform use of them – was registered for the phonemic groups [oja], [uja], [aje], [eje], [oje] and [uje]. A balance between the two or a moderate prevalence of the Raška manner was detected in both books with respect to the spelling of the [je] group in initial position. Both books are characterized by a similar frequency of use of the graphemes ы and ъ, as well as of ѕ in the words ѕѣло and ѕвѣзда.

Variations in the most prominent features of the post-Resava orthography in *Misc. 1536*, *Men. 1538*, and *Euch. 1538-1540* – the books printed in the second period of Božidar Vuković's activity in Venice – were considerably more noticeable. Apart from a similar frequency of the grapheme ы, the only shared features noted among these books were the spelling of [ije] almost uniformly in the Resava manner, as well as the phonemic groups [oje] and [uje] being spelt uniformly or almost uniformly in accordance with the Raška orthographic norm. While *Misc. 1536* and *Euch. 1538-1540* are characterized by the almost uniform use of Resava orthography when marking the phonemic groups [ija] and [iju], and of Raška orthography when it comes to the groups [eja] and [uje], in *Men. 1538* these features are prominent to a somewhat lesser extent. *Misc. 1536* and *Euch. 1538-1540* can also be placed within a single group with respect to the prevailing spelling of the phonemic groups [aje], [eje], and [oje] in the Raška manner, in contrast to *Men. 1538*, in which the Resava orthography predominates for these categories. A different situation is noted in the spelling of the phonemic group [aja]: in *Misc. 1536* and *Men. 1538*, we observed an almost uniform use of the Resava orthography, while in *Euch. 1538-1540*, both orthographic manners were employed with almost equal frequency. *Misc. 1536* and *Men. 1538* can be placed into a single group with regard to the uniform or almost uniform use of the grapheme s in the words ѕѣло and ѕвѣзда, and partially in ѕвѣрь, while in *Euch. 1538-1540* the grapheme ѕ was almost never used. Because they both employ the grapheme ъ less than half as often as *Misc. 1536*, we can classify *Men. 1538* and *Euch. 1538-1540* together within the same group. The level of variation in the books typeset during Božidar Vuković's second period in Venice is especially well illustrated by the different spellings of the phonemic group [je] in initial position: in *Men. 1538*, there is a considerable predominance of the Resava tradition, in *Euch. 1538-1540* the Raška version considerably prevails, whereas in *Misc. 1536* we can note an even distribution of the two spellings.

Our investigation has confirmed the hypothesis that the use of the most prominent features of the Resava and Raška orthographies was normalized in the Serbian books from Božidar Vuković's printing shop in Venice, and it has also for the first time presented more precise statistical data in relation to competing orthographic solutions in 16th-century Serbian printed books, which had previously not been possible with qualitative methods applied to smaller textual samples. The computer-assisted, quantitative approach using tools applied here for the first time to Serbian printed books (automatic text recognition with a *Transkribus* HTR model, automatic look-up and extraction of examples using *AntConc*, stylometric analysis with the *R stylo* package) can be applied to other orthographic and linguistic features of Serbian early books from Božidar Vuković's printing shop in Venice, as well as to other medieval Serbian manuscripts and early Serbian printed books. The limitations of the approach presented in this paper can be overcome by combining it with qualitative methods (as part of a so-called *mixed-methods approach*), and by developing specific AI models for automatic lemmatization and morphosyntactic annotation of Serbian Church Slavonic manuscripts and printed books, which could in turn lead to the creation of an electronic reference corpus of Serbian Church Slavonic.

Sources

Lit. 1520: *Liturgikon* – Novi Sad, Matica Srpska Library, catalogue no. PCp II 12.1.

Psal. 1520: *Psalter with Appendices* – Novi Sad, Matica Srpska Library, catalogue no. PCp II 1.1.

Misc. 1536: *Prayer Book (Miscellany for Travellers)* – Novi Sad, Matica Srpska Library, catalogue no. PCp I 3.1.

Men. 1538: *Festal Menaion* – Novi Sad, Matica Srpska Library, catalogue no. PCp IV 1.1.

Euch. 1538-1540: *Prayer Book (Euchologion)* – Novi Sad, Matica Srpska Library, catalogue no. PCp II 8.1

Literature

Bester-Dilger, Rabus 2021: J. Besters-Dilger, A. Rabus, *Neural Morphological Tagging for Slavic: Strengths and Weaknesses*, "Scripta & e-Scripta", XXI, 2021, pp. 79-92.

Burlacu, Rabus 2021: C. Burlacu, A. Rabus, *Digitising (Romanian) Cyrillic using Transkribus: New Perspectives*, "Diacronia", XIV, 2021, pp. 1-9.

Burrows 2002: J. Burrows, Delta – *a Measure of Stylistic Difference and a Guide to Likely Authorship*, "Literary and Linguistic Computing", XVII, 2002, 3, pp. 267-287.

Büttner *et al.* 2017: A. Büttner, F. M. Dimpel, S. Evert, F. Jannidis, S. Pielström, T. Proisl, I. Reger, C. Schöch, T. Vitt, *Delta in der stylometrischen Autorsschaftsattribution*, "Zeitschrift für Digitale Geisteswissenschaften", 2017, DOI: 10.17175/2017_006.

Đorđić 1991: P. Đorđić, *Istorija srpske ćirilice*, Beograd 1991.

Eder *et al.* 2017: M. Eder, J. Rybicki, M. Kestemont, *Stylometry with R: a Package for Computional Text Analysis*, "R Journal", VIII, 2016, 1, pp. 107-121.

Eder 2011: M. Eder, *Style-Markers in Authorship Attribution: A Cross-Language Study of the Authorial Fingerprint*, "Studies in Polish Linguistics", VI, 2011, pp. 99-114.

Górski *et al.* 2019: R. Górski, M. Król, M. Eder, *Zmiana w języku. Studia kwantytatywno-korpusowe*, Kraków 2019.

Grbić 2008: D. Grbić, *Ortografske odlike Goraždanskog Psaltira sa posledovanjem iz 1521. godine*, in: D. Barać (ur.), *Goraždanska štamparija 1519-1523*, Beograd-Istočno Sarajevo 2008, pp. 225-264.

Grbić 2020: D. Grbić, *Tipografske odlike i varijante izdanja* štamparije *Božidara Vukovića u Veneciji*, "Ricerche slavistiche" N.S., III (LXIII), 2020, pp. 75-91.

Grbić *et al.* 1994: D. Grbić, K. Minčić-Obradović, K. Škorić, *Ćirilicom štampane knjige 15-17. veka Biblioteke Matice srpske,* Novi Sad 1994.

Grković-Mejdžor 1994: J. Grković-Mejdžor, *Jezik "Psaltira" iz štamparije Crnojevića*, Podgorica 1994.

Grković-Mejdžor 2012: J. Grković-Mejdžor, *Funkcionalne odlike predgovora i pogovora iz štamparije Đurđa Crnojevića i Božidara Vukovića,* in: J. Stojanović (ur.), *Srpsko jezičko nasljeđe na prostoru današnje Crne Gore i srpski jezik danas*, Nikšić 2012, pp. 27-36

Hrvaćanin 2017: R. Hrvaćanin, *Istorijsko-bogoslovska analiza srpskih štampanih liturgijara XVI veka,* Beograd 2017.

Jerković 1967: V. Jerković, *Osobine glagolskih oblika u Služabniku Božidara Vukovića*, "Godišnjak Filozofskog fakulteta u Novom Sadu", X, 1967, pp. 155-160.

Jerković 1968: V. Jerković, *Cvetni triod iz 16. veka: pravopisne i fonetske odlike*, "Zbornik za filologiju i lingvistiku", XI, 1968, pp. 85-96.

Jerković 1970: V. Jerković, *Dva četvorojevanđelja štampara Mardarija iz Mrkšine crkve*, "Prilozi proučavanju jezika", VI, 1970, pp. 1-17.

Jerković 1972: V. Jerković, *Cvetni triod štampara Mardarija iz Mrkšine crkve*, "Godišnjak Filozofskog fakulteta u Novom Sadu", XV, 1972, 1, pp. 89-104.

Jerković 1980: V. Jerković, *Srednjovekovne ortografske škole kod Srba*, "Jugoslovenski seminar za strane slaviste", XXXI, 1980, pp. 19-28.

Lahjouji-Sepälä *et al*. 2022: M. Z. Lahjouji-Sepälä, A. Rabus, R. von Waldenfels, *Ukrainian Standard Variants in the 20th Century: Stylometry to the Rescue*, "Russian Linguistics", XLVI, 2022, pp. 217-232.

Lazić 2018: M. Lazić, *Venice and Editions of Early Serbian Printed Books*, "Thesaurismata", XLVIII, 2018, pp. 161-192.

Lazić 2021: M. Lazić, *Izdavačka delatnost Božidara Vukovića i Vićenca Vukovića u Veneciji (1519-1561)*, Beograd 2021.

Lazić 2022: M. Lazić, *Božidar Vuković: između istorije i imaginacije*, Beograd 2022.

Pešikan 1994: M. Pešikan, *Leksikon srpskoslovenskog* štamparstva*,* in: M. Pešikan, K. Mano-Zisi, M. Kovačević (ur.), *Pet vekova srpskog štamparstva 1494-1994: razdoblje srpskoslovenske štampe XV-XVII veka*, Beograd-Novi Sad 1994, pp. 71-218.

Polomac 2022a: V. Polomac, *Serbian Early Printed Books from Venice: Creating Models for Automatic Text Recognition using Transkribus*, "Scripta & e-Scripta", XXII, 2022, pp. 11-29.

Polomac 2022b: V. Polomac, *Serbian Early Printed Books: Towards Generic Model for Automatic Text Recognition using Transkribus*, in: D. Fišer, T. Erjavec (eds.). *Proceedings of the Conference on Language Technologies and Digital Humanities*, Ljubljana 2022, pp. 154-161.

Rabus 2019: A. Rabus, *Recognizing Handwritten Text in Slavic Manuscripts: a Neural-Network Approach using Transkribus*, "Scripta&e-Scripta", XIX, 2019, pp. 9-32.

Rabus, Petrov 2023: A. Rabus, I. Petrov, *Linguistic Analysis of Church Slavonic Documents: A Mixed-Method Approach*, "Scando-Slavica", LXIX, 2023, 1, pp. 25-38.

Rabus *et al.* 2023: A. Rabus, E. Arnold, A. Jouravel, P. Lendvai, M. Meindl, V. Polomac, E. Renje, *Developing a Pipeline for Automatic Linguistic Analysis of Historical Manuscripts and Early Printings: The Pre-Modern Slavic Case*, in: A. Baillot, T. Tasovac, W. Scholger, G. Vogeler (eds.), *Digital Humanities 2023: Book of Abstracts. Digital Humanities 2023. Collaboration as Opportunity (DH 2023)*, Graz 2023, pp. 112-113.

Samardžić 2012: B. Samardžić, *Ortografske odlike knjiga Goraždanske štamparije (obilježavanje spojeva* ja, je, ju, jo, lja, nja, lje, nje, lju*),* "Filolog", V, 2012, pp. 159-169.

Samardžić 2013: B. Samardžić, *Ortografske odlike knjiga Goraždanske štamparije (obilježavanje poluglasničkog znaka, vokala* u*,* o *i* jeri*, udvojenih slova, vokalnog* r *i vokalnog* l*, sekvence* št*, glasovne vrijednosti* ć *i* đ *i pitanje upotrebe slova)*, "Radovi Filozofskog fakulteta", XV, 2013, 1, pp. 113-123.

Samardžić 2019: B. Samardžić, *Starostavne goraždanske knjige: ortografija i jezika knjiga Goraždanske štamparije*, Beograd-Banja Luka 2019.

Sindik 1986: N. Sindik, *Književne odlike predgovora i pogovora u izdanjima Božidara Vukovića*, in: J. Milović, R. Vujošević, Č. Vuković (ur.), *Štamparska i književna djelatnost Božidara Vukovića Podgoričanina*, Titograd 1986, pp. 117-130.

Stamatatos 2009: E. Stamatatos, *A Survey of Modern Authorship Attribution Methods*, "Journal of the American Society for Information Science and Technology", LX, 2009, 3, pp. 538-556.

Subotin-Golubović 2020: T. Subotin-Golubović, *Sastav Prazničnog mineja Božidara Vukovića i njegovi rukopisni izvori*, "Ricerche slavistiche" N.S., III (LXIII), 2020, pp. 261-278.

Von Waldenfels, Eder 2016: R. von Waldenfels, M. Eder, *A Stylometric Approach to the Study of Differences between Standard Variants of Bosnian/Croatian/Serbian, or: is the Hobbit in Serbian more Hobbit or more Serbian?*, "Russian Linguistics", XL, 2016, pp. 11-31.

Abstract

Vladimir Polomac, Achim Rabus
Serbian Early Printed Books from Venice: A Quantitative Approach to Orthographic Variations

The paper analyses the most significant orthographic variations in early printed Serbian books from Božidar Vuković's printing shop in Venice. The research was conducted based on automatically obtained transcripts using a previously trained model for automatic text recognition in the *Transkribus* software platform. Orthographic variations were examined at the macro level using the *stylo* package in the statistical programming language *R*, and at the micro level by extracting the most important features of post-Resava orthography using the program *AntConc*. In addition to confirming the initial hypothesis that orthographic variations generally depend on the period in which the books were printed and individual printers, the paper demonstrates precise quantitative relationships of competing orthographic solutions in the corpus, which was not possible to achieve with the traditional qualitative method based on smaller text samples.

Keywords

Serbian Early Printed Books from Venice; Serbian Church Slavonic; Post-Resava Orthography; Automatic Text Recognition; Computer Stylistics.