



Open Library of Humanities



Part of the Ubiquity
Partner Network



Research

How to Cite: Netsunajev, Aleksei and Netsunajeva, Natalja. 2018. "On the Classification of the Slavic Menaia Manuscripts Dated from the 11th to 14th Centuries." *Digital Medievalist* 11(1): 1, pp. 1–17, DOI: <https://doi.org/10.16995/dm.75>

Published: 27 March 2018

Peer Review:

This article has been peer reviewed through the double-blind process of *Digital Medievalist*, which is a journal published by the Open Library of Humanities.

Copyright:

© 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Open Access:

Digital Medievalist is a peer-reviewed open access journal.

Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

RESEARCH

On the Classification of the Slavic Menaia Manuscripts Dated from the 11th to 14th Centuries

Aleksei Netsunajev¹ and Natalja Netsunajeva²

¹ Bank of Estonia and Tallinn University of Technology, EE

² Tallinn University, EE

Corresponding author: Aleksei Netsunajev (Aleksei.Netsunajev@fu-berlin.de)

In the present paper we analyze nine manuscripts from the 11th–14th century Menaia (Greek: μηνάιον), Old-Slavic hymnographic texts, using a vector space model. The analysis and classification of the manuscripts in previous studies have been rather subjective. In an attempt to be objective we use contemporary research methods. Vector analysis allows one to separate the Putyatina Menaion and the Menaion Q.п.1.25 from the set of analyzed texts, since both manuscripts share both textological and lexical similarities. Similar findings were reached in existing studies. Manuscript BAN 16.14.13 is shown to be quite similar to the set of analyzed texts. The results are new to the literature.

Keywords: Menaion; Old Slavic Manuscript; Classification; Vector Space Model

1. Introduction

The Menaion is a liturgical work containing texts that glorify saints and holy days. These texts are arranged according to the days in each month. The Menaion with its multiple manuscripts preserves a unity of existence as a text, while retaining differences on textual and language levels. Hence the importance of the classification of the collection of manuscripts. This collection is represented according to the Consolidated Catalogue (CC) of 68 manuscripts from the 11th–13th centuries (Consolidated Catalogue 1984). The largest number of manuscripts of that period, eleven to be exact, preserve the May Menaion.

The classification tradition of the hymnographic texts goes back to I. V. Yagich (1886). This tradition established the following lexical and textological parameters for the classification of the manuscripts: the composition of the manuscript (set of memorial texts), the structure of the service (the order of the hymns) and specific linguistic discrepancies. Among the contemporary researchers of the Menaion who adhere to the same approach, Maria Yovcheva should be mentioned. She describes the Service Menaion for April Sof.199 of the 12th–13th centuries from the collection of the National Library of Russia (NLR), taking into account the calendar features of the manuscript (set of memorial hymns), the structure of the service (order of the hymns), as well as the morphological, lexical and syntactic differences (Yovcheva 2014).

Textual typology and linguistic differences sort the May Menaia manuscripts into four types (Netšunajeva 2000):

1. Putyatina Menaion and its analogues (archaic type);
2. The manuscripts of the Studite type of the 11th–14th centuries – the main body of the CC;
3. The manuscripts of the early Jerusalem type described in the Preliminary List (PL) (1966), such as the manuscript of the May Menaion T.113 of the C14th;
4. The manuscripts of the Jerusalem type of the 14th–17th centuries, which are in most cases recorded in the lists of libraries. For example, manuscript KB Rålamb 4: 0 n: 0130 The Festival Menaion of the C17th from the collection of the National Library of Sweden (Stockholm).

The composition of the manuscript and the structure of services are the typical indicators for the typological characteristics of the manuscript; the set of memorial hymns and their order is associated with a particular statute. Taking into account the set of memorial hymns, the second-type manuscripts of the 11th–14th centuries correspond mainly to the Studite statute, while the manuscripts of the third and

fourth type correspond to the Jerusalem statute. In Menaia manuscripts three possible orders of hymns can be distinguished: 1) canon (Greek: κανων) – stichos (Greek: στιχηρὸν, pl.-ά) – sedalen (Greek: κάθισμα); 2) sedalen – stichos – canon; 3) stichos – canon interrupted after the third hymn by sedalen and after the 6th hymn by kontakion (Greek: κοντάκιον) and ikos (Greek: Ἰχος). The first two types of hymn-ordering are in accordance with the genre of the service; the third follows the order during the service. The first order of hymns is a typical feature of the archaic type Menaion. The collection of this type currently consists of four Old Russian manuscripts, including the Putyatin Menaion, five South-Slavic manuscripts and one Greek manuscript (Netšunajeva 2008).

The classification may become more detailed. For example, the May Menaion Sof. 203 of the 12th century can be distinguished from the Studite manuscripts. The manuscript shows that within a large array of similar Studite type manuscripts there exists a non-standard subtype. Manuscript Sof. 203 has textual features associated with a set of memorial hymns according to the Jerusalem statute, and specific discrepancies, such as the restoration of the Greek in the text: **ΩΤΕΕΜΒ** instead of **ΜΑΓΑΩΜΒ** (Netšunajeva 2011).

The researchers who follow the textual traditions established in the description of the manuscripts of the hymnography of the Old and New Testaments take into account the aforementioned differences in various ways (Alekseev 1999). However, these methods of classification do not fully exploit the information available in the manuscripts of the Menaion. In this article we propose to use methods of classification based on the properties of the data and, in this case, on the information concerning discrepancies contained in the manuscripts of the May Menaia. The sources for our analysis are the manuscripts of Menaion of the first and second type. The differences that we propose to speak in favor/against the similarity of texts are lexical and grammatical discrepancies. Phonetic differences at this stage of the analysis are excluded.

The article uses a new approach to the classification and analysis of the Menaion manuscripts, which we inherit from the literature on information

retrieval. We find Menaion manuscripts of type one and two that share many common elements with the help of a mathematical model. To formally analyze the manuscripts we set up a vector space model where the manuscripts are represented as vectors in a common vector space. Using this tool we can identify similar and different manuscripts, and classify manuscripts of the first and second types. Groups formed as a result of the analysis consist of Menaion manuscripts that have common lexical units and grammar.

In the considered manuscripts there are more lexical variations across fragments than there are phonetic or grammar variations. Lexical variations are important from the point of view of presence or absence of words in territorial and diachronic systems and in texts with different translation methods. Analysis of lexical variations allows one to make statements on the origin and date of manuscripts as well as topography. The grammar variations show the historical development of the language.

This paper shows that the methods of information retrieval can be successfully applied to the analysis of ancient texts. Obtained results allow us to identify similar and different lists of Menaion manuscripts and to relate the results to the existing literature. Specifically, we can (i) detect the manuscripts having textological similarities supported by specific lexical variations, (ii) show the manuscripts that deviate from the majority of the analyzed texts, (iii) show typical and atypical usage in the language and in the structure of the text.

The remainder of the article is organized as follows. Section 2 describes and explains the choice of Menaia and of the textual fragments. In Section 3 we construct the vector space model. Detailed analysis of the results obtained using the model is discussed in Section 4. The last section reports our conclusions.

2. The Manuscripts of the Menaion

For the vector model analysis we took nine manuscripts of the May Menaion of the 11th–14th centuries: seven are described in CC, two – in the PL. A brief description of the manuscripts and their storage is shown in **Table 1**.

Table 1: Menaia manuscripts dated from the 11th to 14th centuries.

Manuscript	№ in CC	№ in PL	Storage	Abbreviation used in this article
Service Menaion, May C 11 th . «Putyatina Menaion»	21	10	NLR, Соф. 202	PM
Festival Menaion, May, fragment. End C 12 th – beginning 13 th .	156	72	NLR, Q.п.1.25	Q.п.1.25
Service Menaion, May C 12 th .	90	107	NLR, Соф.203	Sof.203
Service Menaion, May, notated, C 12 th .	89	106	SHM, ¹ СИН.166	Sin.166
Service Menaion, May C 13 th .	282	436	NLR, Соф.204	Sof.204
Service Menaion with readings, May–June, C 13 th .	281	313	RASL, ² 4.5.10	BAN4.5.10
Festival Menaion. End C 13 th – beginning C 14 th .	454	434	NLR, Соф.382	Sof.382
Service Menaion, May. In PL C 13 th /14 th . In RASL beginning C 14 th .		435	RASL, 16.14.13	BAN16.14.13
Service Menaion, May C 14 th .		1062	RGADA, ³ ф.381 №112 Тип., №225	T.112

¹ State Historical Museum of Russia.

² Russian Academy of Sciences Library.

³ Russian State Archive of Old Acts.

To represent the manuscripts in vector form, we use two fragments of stich **СТН · ГЛА · Д · ПѢ · ДАСТЬ ЗНАМЕНИЕ**, which *follows* the canon of the service for May 21st in honor of holy Konstantin and Elena in PM and Q.п.1.25, but *precedes* the canon in the other manuscripts. As an example, we show the first fragment of text in three manuscripts: PM and Q.п.1.25, which refer to the archaic type, and manuscript BAN16.14.13, which refers to Studite type:

PM:

Ороужикъ крѣпко · црю ты дасть нашемоу · крѣсть свои чьстынѣи · имъ же црѣствова на земли правдыно · сипавъ вѣроу · и црю нбѣсъ съподоби са · млардикъмъ си · съ ними же ти млств но съмстреннє прославляемъ · ꙗе млстивѣи · спсе дшѣ нашихъ ·: (folio 88 recto)

Q.П.1.25:

Орлжикъ крѣпко црѣви дасть · нашемоу крѣсть твои чьстыни · имъ же црѣствова · на земли правды · ситъ въ вѣроу · црѣства нбѣ (...) съподоби · млардикъмъ си · съ нимъ же мларнокъ · ѡмотреннє тѣе слави · ꙗе въ селѣге ·: (folio 1 recto)

BAN16.14.13:

Ороужикъ крѣпко · црѣви нашемоу дасть крѣсть твои чьстыни · имъ же црѣствова на земли правды · сипавъ вѣроу · црѣства нбѣга съподоби са · млардикъмъ си · съ нимъ же ти чькѣмъ · ѡмотреннє слави · ꙗе млстивѣи · спсе дшѣ нашихъ ·: (folio 50 recto).

The aforementioned fragments are the only fragments to be present in the nine manuscripts of the analyzed period. They are thus unique in the sense that they offer a possibility for a full textual comparison of the manuscripts. It is worth noting that Q.п.1.25 is an outstanding manuscript. It consists of a single folio containing the end of the 9th song of the canon, two fragments of stich and the beginning of a sedalen. Therefore, the material for analysis is very limited due to the remaining fragment length of the manuscripts. For these reasons we believe that the analysis based on the selected fragments is representative and is using the parallel information available in all nine May Menaia in full.

3. Vector Space Model

For the analysis, we used a vector space model and represent the manuscripts as vectors in a common vector space (Salton, Wong and Yang 1975; Dubin 2004; Jockers 2014). Formally, we define:

$D = \{d_1, \dots, d_n\}$ – set (corpus) of documents

$T = \{t_1, \dots, t_m\}$ – terms contained in the documents, or a dictionary. In our case the term can be single words or phrases.

The next step was to determine the weight of a term in a document. Weight is the importance of the term within the specific manuscript. Vector representation of a document is a vector comprising the weights of each term in the document. If there is no term in the document, the weight is zero. For document i , for instance:

$W(i) = \{w_{1i}, w_{2i}, \dots, w_{mi}\}^T$, where w_{ji} – weight of j term in document i .

There are several standard ways of determining the weighting function (Leskovec et al. 2014). This paper uses the tf-idf (term frequency – inverse document frequency) weighting function. The weighting function is calculated as follows:

tf (term frequency) estimates the importance of the term t_i in document d_j
 $tf(t_i, d_j) = n_i$, n_i – number of entries for a term t_i in document d_j .

idf (inverse document frequency) – is the inverse of the frequency with which a term occurs in the corpus. It is calculated as follows:

$$idf(t_i, D) = \log_{10} \frac{|D|}{|d \in D: t_i \in d|}$$

where $|D|$ is the number of documents, and $|d \in D: t_i \in d|$ is the number of documents with the term t_i . Therefore, the tf-idf weight can be calculated as:

$$w_{ij} \equiv tfidf(t_i, d_j, D) = tf(t_i, d_j) idf(t_i, D).$$

Thus each document is now represented as a vector of weights in the space R^m . This allows us to calculate measures of distance between the vectors using certain metrics. There are a vast number of metrics to choose from (see Cha 2007 for comprehensive survey). In this paper we calculate the Euclidean distance between the manuscripts and cosine similarity. These are standard similarity measures used in vector space models and clustering (Cha 2007). From our point of view, the straight line measures are more suitable for text analysis than, for instance, block metrics. The former are more intuitive for our application as there should not be any obvious sources of non-linearity that prevent calculation of straight line distance between two vectors of sample manuscripts. The two measures – Euclidean distance and cosine similarity – address the same question but from a different angle. The first measure allows us to identify the manuscripts with the greatest amount of variety, while the second allows us to identify those containing similarities. The analysis of two distance measures may be seen as a complementary exercise. The distance is calculated as:

$$\text{dist}(d_i, d_j) = \sqrt{\sum_{k=1}^m (w_{ki} - w_{kj})^2}$$

The smaller the distance the closer the manuscripts are. The cosine similarity is calculated as:

$S(d_i, d_j) = \tilde{W}(i)^T \tilde{W}(j)$, where $\tilde{W}(i)$ is the length normalized vector $W(i)$ (we use L^2 norm to normalize the length). The larger the value of $S(d_i, d_j)$ the closer the manuscripts d_i and d_j are.

We implemented the vector space model as follows. We obtained the fragments of stich from the original manuscripts. This work was done partly in the libraries as only manuscripts PM, Sin.166, Sof.203, Sof.204 and Q.π.1.25 are available in electronic form. We assigned each word in the fragments an integer identification number (ID). The collection of the words and hence the ID numbers formed the dictionary. This allowed us to calculate vector of weights $W(i)$, $i = 1, \dots, 9$ for each manuscript. With these objects in hand the measures of distance were easily calculated. For example, the inner product of the length normalized vectors $W(i)$ s gives the cosine similarity. All calculations were performed in Matlab.

We are aware of other possible classification algorithms that may be applied for the analysis of texts, such as (hierarchical) clustering models. However, this paper focuses on the vector space model and leaves other setups for future research. Clustering may be problematic in this setup as the models may require larger text samples than studied here. As previously discussed, we were constrained by the length of readable text in the May Menaia.

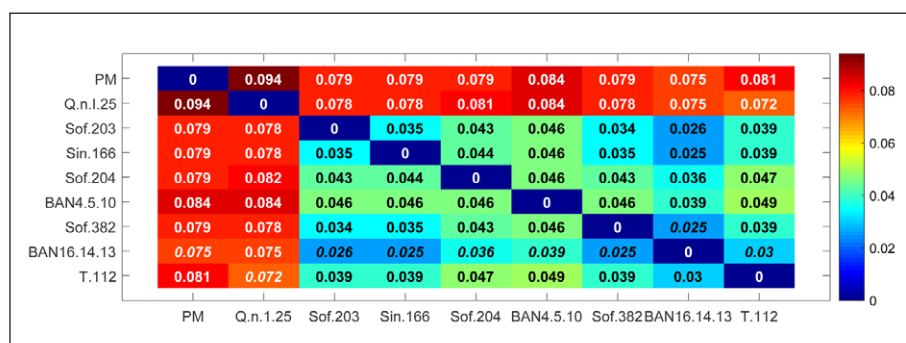
4. Analysis of Manuscripts

For the vector space model analysis we used two fragments of stich from nine manuscripts of the Menaion described in **Table 1**. We separately analyzed the similarity of the manuscripts in lexical units and in both lexical units and grammar. For this purpose we created two dictionaries; in the first cases, the dictionary contained 70 terms,

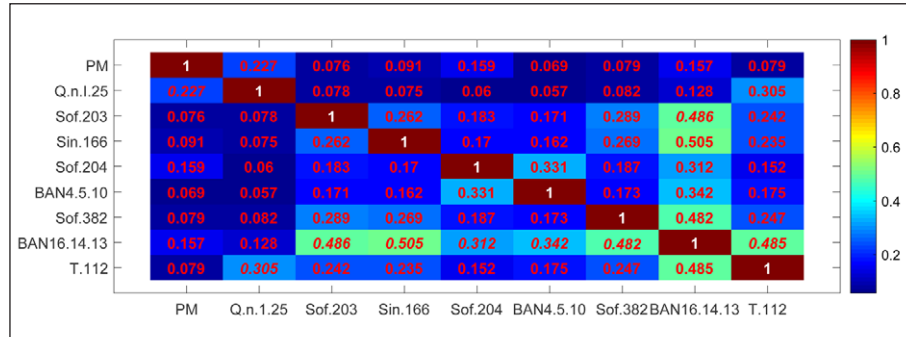
in the second 79 terms. In the first case we took into account lexical discrepancies, **СЛОУЗЪ - ОУГОДЬНИКОУ** and **МЛѢТВ'НО - МЛѢРНѢ - ЧВѢКОУБЬНОЮ**. In the second case, lexical and grammar discrepancies like **СИДѢВЪ - ВЪСИДѢВЪ - СИДА** were considered. As discussed in the previous section, we analyzed the two measures of distance – the Euclidean distance and cosine similarity. As a robustness check we calculated also the block measure – Manhattan distance – and obtained results similar to those of the Euclidean distance. The Manhattan distance was not chosen to be the primary measure of distance as it did not present straight line distance between vectors.

The separation of the analysis into lexical elements and lexical and grammatical elements proved important because it allowed to track the actual lexical variation and its relationship to certain manuscripts. Lexical and grammatical changes within a single linguistic unit showed the dynamics of some grammatical changes in diachrony and their connection with the lexical changes.

Consider the distance between the manuscripts when the lexical differences are taken into account. This distance is presented in **Heat map 1**. It is clear that PM and Q.π.1.25 have a greater distance from other manuscripts, while the shorter distance is observed for BAN16.14.13. The resulting measure of the cosine similarity for the manuscripts when lexical differences are considered is shown in **Heat map 2**.



Heat map 1: Euclidean distance between the manuscripts (lexical changes). The minimal value for each manuscript is highlighted in italics.



Heat map 2: Cosine similarity between the manuscripts (lexical changes). The highest value for each manuscript is highlighted in italics.

Next, we analyzed the lexical units of individual manuscripts to understand the results of **Heat maps 1** and **2**. Manuscripts PM and Q.π.1.25 differ from the others based on lexical unit differences: **МЛСТѢ НО** (PM) – **МЛСРЬНОСѢ** (Q.π.1.25) – **ЧВКЪЛКЪБЪНОСѢ** (Studite type manuscripts). Such discrepancies may be related to the Greek variation. Lexeme **ЧВКЪЛКЪБЪНОСѢ** is the Greek translation *φιλόανθρωπος* that allows a different translation in the Slavic text **МНЛОСТИВЪНЪН: ЗЪЛО СМДЬКО И МНЛОСТИВЪНО** (*φιλόανθρωπος*) (Sreznevskii 1989b). This option is reflected in PM as **МЛСТѢ НО**. The second term in the discrepancy **МЛСРЬНОСѢ** may be a communication of the Greek *φιλόανθρωπος*, as well as the Greek *εὐσπλαγχνος* **‘МНЛОСРЬДЪ’** (Dictionary of the Old Slavic Language 1999). Lexical **БЛГОВЕРЬНОУМОУ – БЛГОЧЬСТЕНУМОУ** are due to the single Greek lexeme *εὐσεβής* (Sreznevskii 1989a). The variation occurred on the Slavic level. Of the same origin is the discrepancy **СЛОУЗЪ** (PM, Q.π.1.25) – **ΟΥГОДЪНИКОУ** (manuscripts of Studite type). The results show that PM shares many similarities with Q.π.1.25 and Q.π.1.25 with T.112: **ТВѢ/ТВО** – (only in these two manuscripts). In the Greek text this pronoun is absent. As it appears only in the two manuscripts the translation might come from another Greek text.

The largest number of similarities is observed between BAN16.14.13 and all other manuscripts: **СВОИ** (PM) – **ТВОИ** (all other manuscripts). All manuscripts except PM conform to the Greek original: *τὸν τὸν σκαυρόν σου τὸν τίμιον*. The absence

of the lexeme in PM indicates the presence of another original text. Discrepancy **ПРОСЛАВ'ЛІКМЪ** (PM) – **СЛАВИМЪ** confirms this conjecture. The discrepancy comes from the Greek ἐνδοξάζειν – δοξάζειν **‘ПРОСЛАВИТИ-СЛАВИТИ’**. In the Greek text the lexeme δοξάζομεν appears and, as in the first case, is present in all manuscripts except PM.

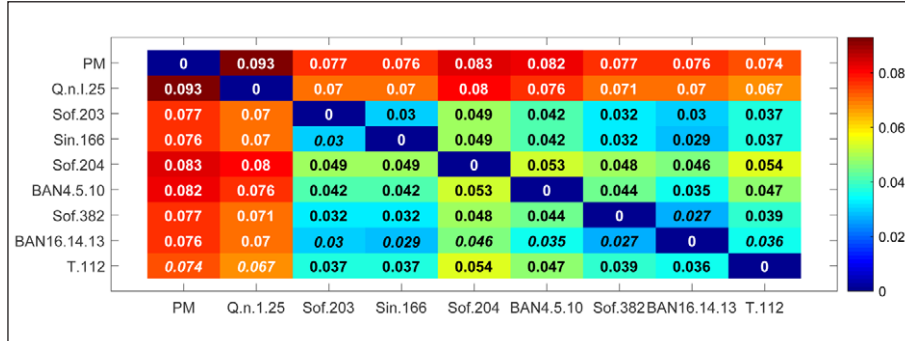
The manuscript BAN16.14.13 is very close to the manuscripts Sof.203, Sin.166, Sof.204, BAN4.5.10, Sof.382 and T.112. BAN16.14.13 and Sin.166 are the closest. Lexeme **КРЪПЬКО** is present in all manuscripts except Sof.382, in which **НЕПОДИМО** is a mutation of **НЕПОБЪДИМО**, and T.112, in which a grammar variation of the original lexeme **КРЪПЬКОК** appears. The overlap in the manuscripts of the second type is due to the use of the short form adjective **КРЪПЬКО**.

Therefore, it is possible to distinguish manuscripts PM and Q.π.1.25, which are separated from the others by lexical similarity. Cosine similarity indicates that these manuscripts are close. However, manuscript BAN16.14.13 has the largest number of similar elements to the rest of the group of the analyzed manuscripts. Previous studies based on a different methodology using textological and linguistic similarities concluded that PM and Q.π.1.25 stand out as a special archaic type. The similarity of manuscript BAN16.14.13 with the rest of the studied group is the new result.

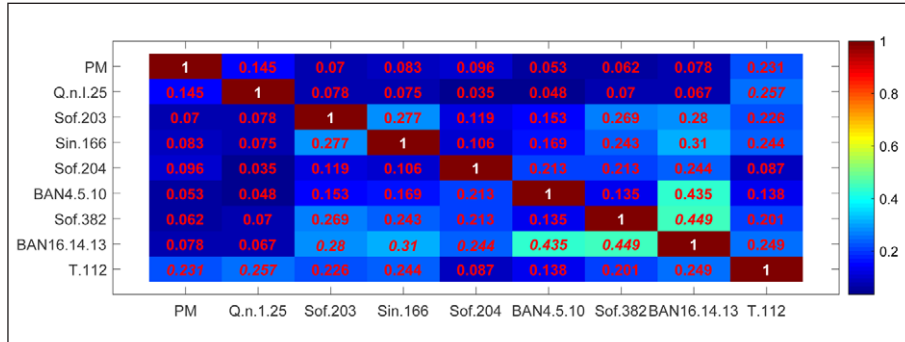
The results on the distance between the manuscripts based on the lexical and grammatical differences are shown in **Heat map 3**. The red columns/rows reveal that PM and Q.π.1.25 have the greater distance from the rest of the manuscripts. BAN16.14.13 is the closest to other manuscripts. These results are similar to those produced by the lexical variations analysis.

Heat map 4 shows the results of the comparison of the manuscripts when lexical and grammatical differences are taken into account. Furthermore, we analyzed the lexical units and grammar of individual manuscripts to reinforce the results of **Heat maps 3** and **4**. The following observations are worth noting.

1. The manuscripts Q.π.1.25 and PM have less similarity with the group of studied manuscripts: there is a difference in the choice



Heat map 3: Euclidean distance between the manuscripts (lexical and grammar changes). The minimal value for each manuscript is highlighted in italics.



Heat map 4: Cosine similarity between the manuscripts (lexical changes). The highest value for each manuscript is highlighted in italics.

- of case form: **ГЪ СТЫ** (PM) – **ГЪДВСТЪЖЦИХЪ ВЪСЪХЪ ГЪ**
(Q.п.1.25) – **ГЪДВСТВОУЩИМЪ СЪ НИМИ/НИМЪ** (other manuscripts).
- It is worth noting that PM has a greater lexical similarity to T.112: they share the aorist (Old Slavic past tense form **СПСЕ** (PM, T.112) and the case form **ДШЬ НАШИХЪ** (PM, T.112).
 - According to similarity the PM follows Q.п.1.25: they share the vocative form **ТСЕ**, while in the other manuscripts the form **ІАЪ – ИСЪ** is present.
 - Grammatical similarities between Q.п.1.25 and T.112 are revealed with **ТСЕ** while form **ІАЪ – ИСЪ** is used in all other manuscripts.

The light blue row/column shows that the highest level of similarity with the other manuscripts is observed in BAN16.14.13. This manuscript is the closest to Sof.203, Sin.166, Sof.204, BAN4.5.10 and Sof.382. For example, **СОЛОМОНЮ /СОЛОМОНЮ ПРМДРСТЬ** appears without discrepancies, compared to Q.п.1.25 **СОЛОМНЬ ПРЪМЪДРОСТЬ**.

Table 2 shows the most significant terms in the manuscripts. In the vector representation of the manuscripts the weight of these terms is maximized, so they are special for a manuscript and distinguish it from the rest of the studied group. Interestingly, the most important are the lexical discrepancies. Only for Sof.204 and BAN16.14.13 do the grammar discrepancies **ПРАВЪДЬНОК** – the full form of the adjective instead of short **ПРАВЪДЬНО**; and **СИДА**–participle in another form become very important. These manuscripts reflect the changes in the grammatical system of the language. In the discrepancies **СИДВЪ** (PM) – **СИГВЪ** (Q.п.1.25) – **ВЪСИДВЪ** (Sof.204) – **СИДА** (BAN4.5.10, BAN16.14.13) the Greek word λάμψας (λάμπω 'shine') is reflected in the archaic type manuscripts; in the Studite manuscript Sof.204 the prefix form of the participle is used. The semantics of prefix and non-prefix verbs coincide: **ВЪСИДАТИ = ВЪЗ+СИДАТИ**. The discrepancy reflects grammatical tendencies in the verbal system: the prefix, semantically not significant, becomes type-creating. Temporal variability is also reflected in the active

Table 2: The most important terms in the manuscripts.

Manuscript	Terms	
	Lexical Units	Lexical Units and grammar
PM	преслав лѡ кѡмъ	преслав лѡ кѡмъ
Q.п.1.25	въ сѣлѣгѣи	въ сѣлѣгѣи
Sof.203	дароуѣи	дароуѣи
Sin.166	богочѣстивоуѣмѡ	богочѣстивоуѣмѡ
Sof.204	въсидвѣ	правъдѣнок
BAN4.5.10	правдоу	правдоу
Sof.382	неподимѡ	неподимѡ
BAN16.14.13	чѣвколюбѣнок	сида
T.112	крѣпѣкоѣ	крѣпѣкоѣ

participle forms: the past tense is replaced by the present tense in the BAN4.5.10 and BAN16.14.13 manuscripts.

Thus, out of the entire group the PM and Q.π.1.25, manuscripts which are separated from the others by lexical and grammatical similarities can be identified. Lexically, these manuscripts are similar to each other. Given the lexis and the grammar, the most similar to the two manuscripts is T.112. One can then place the manuscripts Sof.203, Sin.166, Sof.204, BAN4.5.10, Sof.382, BAN16.14.13, T.112 into another group. Of special note is the BAN16.14.13 manuscript, which in our measurement is the closest to the majority of the studied manuscripts.

Previous studies using a different methodology concluded that the manuscripts PM and Q.π.1.25 form a special type based on the structure of the text and the linguistic discrepancies. We confirm these findings. The textological and lexical similarity of PM and Q.π.1.25 reveals their joint ancient history while the text reflects the state of the Menaia books in Russia prior to editing in the second half of the 11th century. These two Menaia represent a type of hymnographic manuscript that Verechaygin and Krysko (1999) name as pre-Studite in relation to the Ilyina Book (July Menaion 12th century), which we call archaic type (type 1).

Furthermore, a mathematical analysis of lexis and grammar confirms the similarity of manuscripts BAN16.14.13 and T.112 with the manuscripts of the first type (PM). Using a traditional lexical and textological method it is easy to show the PM to be the standard text for the first type. The standard text is historically formed and fulfills the requirements of the type (Kulik et al 2016). On the contrary, using traditional approach it is difficult to reveal the standard text for the second type of Menaion. However, the mathematical methods show that manuscript BAN16.14.13 can be seen as a standard text for the second type of Menaion. Manuscripts BAN16.14.13 and T.112 are the closest in our measurement to the main corpus of type II texts. Certain regularities in the functioning of language had been developed by the time the manuscripts were written. The analysis shows that one can predict the phrase in use in the manuscript depending on its association with the specific type.

5. Conclusion

In this article, nine manuscripts of the Menaion were analyzed using a formal vector space model. These manuscripts are PM, Q.π.1.25, Sof.203, Sin.166, Sof.204, BAN4.5.10, Sof.382, BAN16.14.13 and T.112. The analysis and classification of these texts in previous literature was based on the traditional lexical and textological parameters of the texts. In this article, the manuscripts are represented as vectors in a common vector space. The vector space model enabled us to explore similarities and differences between the manuscripts far more thoroughly than before. The results of the vector analysis allowed us to distinguish the PM and Q.π.1.25 texts as being apart from the set of analyzed manuscripts. The results are similar on the textual as well as lexical level. Moreover, manuscript BAN16.14.13 stands out as the most similar to the majority of the analyzed manuscripts.

In our setup, the manuscripts are close to each other or different from one another mostly because of lexical differences. These discrepancies reflect the evolution of the text from the archaic type to the Studite type and finally to the Jerusalem type. Grammatical difference usually show the dynamics of the language system. We observed that the grammatical differences are less important for the set of analyzed manuscript. This is explained by the fact that active changes in the grammar system of the Russian language took place later, in the 14th century.

Competing Interests

Financial support through German Research Council grant CRC-649 “Economic Risk” and Estonian Research Council grant MOBTP11 is gratefully acknowledged.

References

- Alekseev, A. A. 1999. *Текстология славянской Библии* [Textology of the Slavic Bible]. Saint-Petersburg: Böhlau Verlag.
- Cha, Sung-Hyuk. 2007. “Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions.” *International Journal of Mathematical Models and Methods in Applied Sciences* 4(1): 300–307.
- Consolidated Catalogue. 1984. *Сводный каталог славяно-русских рукописных книг, хранящихся в СССР XI–XIII вв.* [Consolidated Catalogue of the Slavic-

- Russian Manuscripts of the 11th–13th centuries that are stored in the USSR]. Moscow: Nauka.
- Dictionary of the Old Slavic Language. 1999. *Старославянский словарь (по рукописям 10–11 веков)* [Dictionary of the Old Slavic Language, based on C 10th–11th manuscripts]. Moscow: Izdatelstvo Russki Yazyk.
- Dubin, D. 2004. "The Most Influential Paper Gerard Salton Never Wrote." *Library Trends* 52(4): 748–764.
- Jockers, Matthew Lee. 2014. *Text Analysis with R for Students of Literature*. New York: Springer. DOI: <https://doi.org/10.1007/978-3-319-03164-4>
- Kulik, Alexander, Catherine Mary MacRobert, Svetlana Nikolova, Moshe Taube and Cynthia M. Vakareliyska. 2016. *The Bible in Slavic Tradition. Studia Judaeslavica* 9. Leiden: Brill. DOI: <https://doi.org/10.1163/9789004313675>
- Leskovec, Jure, Anand Rajaraman and Jeffrey David Ullman. 2014. *Mining of Massive Datasets*. 2nd edition. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139924801>
- Netšunajev, Natalja. 2000. *Минея как тип славяно-греческого средневекового текста* [The Menaion as a type of Slavic-Greek Medieval text]. Tallinn: TPÜ Press.
- Netšunajeva, Natalja. 2008. *История текста и списков XI–XIV вв* [History of the Text and Manuscripts of the C 11th–14th]. In *Славянское языкознание: покидая XX век. К XIV Международному съезду славистов*, edited by A. Dulichenko, 9–23. Tartu: Tartu University Press.
- Netšunajeva, Natalja. 2011. *Рукопись Минеи Соф. 203 XII в. из собрания РНБ как текст и как список* [Manuscript of the Menaion Sof.203 of the 12th century from the collection of the National Library of Russia as a text and manuscript]. In *Лингвокультурное пространство современной Европы через призму больших и малых языков. К 70-летию профессора А.Д. Дуличенко. Slavica Tartuensia IX*, I. Abisogonyan (ed.), 167–176. Tartu: Tartu University Press.

- Preliminary List. 1966. *Предварительный список славяно-русских рукописей XI-XIV вв., хранящихся в СССР* [Preliminary List of the Slavic-Russian Manuscripts of the 11th–14th centuries. that are stored in the USSR]. Moscow: Nauka.
- Salton, G. A. Wong and C.S. Yang. 1975. "A Vector Space Model for Automatic Indexing." *Communications of the ACM* 18(11): 613–620. DOI: <https://doi.org/10.1145/361219.361220>
- Sreznevskii, I. I. 1989a. *Словарь древнерусского языка* [Dictionary of the Old Russian Language Vol I]. Moscow: Kniga.
- Sreznevskii, I. I. 1989b. *Словарь древнерусского языка* [Dictionary of the Old Russian Language Vol II]. Moscow: Kniga.
- Verechagin, E. M. and V. B. Krysko. 1999. *Наблюдения над языком и текстом архаичного источника-Ильиной книги* [Observations on the language and text of the archaic source – Ilyina Book]. In *Voprosy Yazykoznania*, 3–27.
- Yagich, I. V. 1886. *Служебные минеи за сентябрь, октябрь и ноябрь в церковно-славянском переводе по русским рукописям 1095–1097 годов* [Service Menaia for September, October, and November in the Church-Slavonic Translation of the Russian Manuscripts for the years 1095–1097]. Saint-Petersburg.
- Yovcheva, Maria. 2014. *Старобългарският служебен миней* [Old Bulgarian Service Menaia]. Sofia: Bojan Penev Press.

How to cite this article: Netsunajev, Aleksei and Netsunajeva, Natalja. 2018. "On the Classification of the Slavic Menaia Manuscripts Dated from the 11th to 14th Centuries." *Digital Medievalist* 11(1): 1, pp. 1–17, DOI: <https://doi.org/10.16995/dm.75>

Published: 27 March 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Digital Medievalist is a peer-reviewed open access journal published by Open Library of Humanities.

OPEN ACCESS