

A Frequency Dictionary of Proper Names in a Seventeenth-Century Original Text

Oksana Nika¹, Svitlana Hrytsyna¹

¹*Taras Shevchenko National University of Kyiv, 60 Volodymyrska St, Kyiv, 01033, Ukraine*

Abstract

In this paper, we have shown the features of our work at different stages of creating the Frequency Dictionary of Proper Names in a Seventeenth-Century Original Text – from preparing the ancient original text, compiling the Dictionary of Proper Names, classification of proper names into groups to the calculating the frequency of proper names in the text and presentation of the obtained results in a dictionary constructed in the descending frequency order. This is significant because: 1) the historical specificities of the textual material of the 17th century and its disclosure at the preparatory and subsequent stages of compiling this dictionary are taken into account; 2) groups of proper names are as detailed as possible, which makes it possible to establish both the ratio of proper names to the total number of words in the text and the frequency representation of each of the 24 groups of proper names; 3) the homonymy of proper names is removed by establishing hypero-hyponymic relations and contexts; 4) single- and multicomponent proper names are taken into account; 5) the problem of proper names variation is analyzed; 6) the figurative usage of proper names as part of stylistic tropes is also taken into account. The novelty of the scientific problem is in supplementing and adjusting certain stages in the algorithm for compiling a Frequency Dictionary of Proper Names in view of the historical specificities of the text (diacritics, variance, graphics, orthography, features of text generation), the necessity for correlation with texts of different genres, corpus related research, comparison of corpora.

Keywords

Proper names, frequency dictionary of proper names, usage frequency, seventeenth-century original text.

1. Introduction

Dictionaries of proper names are often compiled on the basis of historical materials (“Słownik staropolskich nazw osobowych” [1], I. Mytnik’s “Słownik historyczno-etymologiczny antroponimii ziemi chełmskiej 16 – 17 wieku” [2] and others), but less attention is paid to the creation of diachronic frequency dictionaries.

Frequency dictionaries are often compiled on the basis of an original text/texts and aim to characterize the onomastic space in a work/genre/style [3]; they take into account the special features introduced by interlingual interference [4]. Furthermore, “the statistics can be used in typography, stenography, psychology, psychiatry, language teaching, cryptography, software production, etc” [5, p. 1]. According to V. Perebyinis, there are a number of factors affecting the speech; “...language laws (laws of composition of language units used in speech), laws of language unit collocation in a speech chain, genre laws, the theme and purpose of utterance, the author’s taste, their psychophysical state at the moment of speech and others... speech composition will have certain features, which can be revealed by the statistics” [6, p. 8].

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland
EMAIL: nikaoksanaiv@gmail.com (O. Nika); sv.grytsyna@gmail.com (S. Hrytsyna)
ORCID: 0000-0001-6387-3835 (O. Nika); 0000-0002-8612-3372 (S. Hrytsyna)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

The potential of the stylometric approach is shown in the study of the ancient Greek text of the New Testament addressing the controversial issues of authorship of some of its parts [7].

Statistical methods are conventionally employed in textological and stylometric studies of written monuments to detect their authorship and time attribution.

The frequency of words used in the sermons was compared with various subcorpora of the historical corpus, which made it possible to identify meaningful words and intertextual connections. O. Nika and S. Hrytsyna described the peculiarities of compiling a frequency dictionary of common names used in the ancient text *Otpys* (Response) by Kliryk Ostrozkyi, a frequency dictionary of an original Cyrillic monument of the late 16th century, which was created for the first time [8, p. 277].

In its use of onomastic terminology this article relies on the following sources: “List of Key Onomastic Terms” [9], “Słowańska onomastyka. Encyklopedia” [10], dictionaries by D. Buchko and N. Tkachova [11] and others.

The centrality of the issue discussed in the article is linked to the necessity of theoretical justification of the diachronic specificities of compiling frequency dictionaries, detecting proper name text frequency, explaining the frequency discrepancies, as well as using the results (together with the text corpora) to describe thought and language structures depending on the episteme type.

2. Database and methodology

The theoretical basis of the study were works on mathematical linguistics [5, 6, 7, 12, 13, 14], onomastics [2, 3, 10, 11], historical linguistics [8].

This study aims to compile a frequency dictionary based on the seventeenth-century texts (sermons by Antonii Radyvylovskyi).

The object of the study is principles and methods of compiling a frequency dictionary of proper names in the early modern sermon of the 17th century. The subject of the study is diachronic specificities of compiling a frequency dictionary of proper names sorted in the decreasing frequency order.

The objectives of the paper are a) demonstrating the novelty of compiling a frequency dictionary of proper names in early modern texts; b) describing the stages of frequency dictionary compilation and their application to original ancient texts; c) specifying the preparatory stage of dictionary compilation and principles of tokenization of Cyrillic seventeenth-century texts; d) showing the specificities of stemming and lemmatization with regard to phonetic, orthographic and grammatical word and word-form variation; e) discussing the text frequency of proper name use, frequency discrepancies, division into groups and subgroups.

This frequency dictionary is historical: its compilation and results focus on describing the characteristics of the genre (a sermon), cultural model (baroque), period (Early Modern period, 17th century); monolingual (Ruthenian language with reference to ways of rendering proper names from other languages via Latin), a headword is a proper name (hyponym); its hyperonym is given if used in the texts.

The novelty of the scientific problem is in supplementing and adjusting certain stages in the algorithm for compiling a Frequency Dictionary of Proper Names in view of the historical specificities of the text (diacritics, variance, graphics, orthography, features of text generation), the necessity for correlation with texts of different genres, corpus related research, comparison of corpora.

The obtained results can be used for theoretical and applied works, as well as in the teaching of statistical, corpus linguistics, history of language, lexicography. Linguistic and statistical parameterization of the 17th century originals can be used to study other genres in synchrony and diachrony.

The main methods and techniques employed in the study are: *methods of quantitative and statistical analysis* combining *quantitative methods* (to determine the number of proper names (lexemes and word forms), frequency of their use in the text, their division into groups and subgroups, creating a Frequency Dictionary of Proper Names in a Seventeenth-Century Original Text, frequency discrepancies in proper name use);

distributional analysis (to identify the distribution of proper names in the text, collocation of single- and multicomponent proper names and their variation (phonetic, orthographic, structural), to differentiate the cases of homonymy and classification of proper names into groups and subgroups);

componential analysis (to establish the meaning of words and their belonging to a certain group or subgroup in the process of lemmatization and distinction of homonyms, to determine the primary and secondary nomination, to classify proper names into groups and subgroups);

descriptive method (for empirical research of language material, for description of quantitative characteristics of proper names in the studied original text of the seventeenth-century and their interpretation).

The frequency dictionary is based on the metagraphed seventeenth-century texts by Radyvylovskyyi, Index and Dictionary of Proper Names, containing headwords (with their graphic, orthographic and phonetic variants) and their meanings, both primary and secondary (contextual) with contexts given.

Text Representation. The genre of sermon was well-developed in Early Modern Europe. Oriented towards the audience of different social status, the sermon explicated precedent proper names. Antonii Radyvylovskyyi was one of the popular seventeenth-century authors, who combined in his writings European models and “Kyiv literary traditions”. The multilingual sources he referred to increased the enlightening and persuasive effect of his works, which represent the qualitative and quantitative ratio of proper names used in the Early Modern sermons.

The textual basis for the frequency dictionary of proper names is the publication *Radyvylovskyyi Antonii. Barokovi propovidi 17 stolittia* (Radyvylovskyyi Antonii. Baroque sermons of the 17th century) [15]. Radyvylovskyyi was unique in expanding the range of proper names used in his sermons, which was influenced by the broad variety of sources and Baroque character of presentation.

The seventeenth-century original texts (23 sermons covering 226 pages in total) have been reproduced through metagraphing (an ancient text publication with maximum saving of the original features). These texts come from the collections *Ohorodok Marii Bohorodytsi* (The Garden of Virgin Mary) (1676) and *Vinets Khrystov* (The Wreath of Christ) (1688), as well as their manuscript versions dating to 1671 and 1676–1683, which are kept in the Department of Early Printed Books and Rare Publications at the Manuscript Institute of the Vernadsky National Library of Ukraine [15, p. II]. As for their themes, most sermons are festive ones, while two of them are war-themed.

After creation The Dictionary of Proper Names, used in the text, the compilation process of the diachronic frequency dictionary had several stages: text preparation – tokenization, homonymy differentiation (the proper names identification in the text was made by Python script, based on The Dictionary of Proper Names, used in the text); frequency dictionary compilation procedure – identifying the ratio of proper names to the total number of words in the text, classifying proper names into 24 groups and subgroups, determining the frequency of proper names in the decreasing order within the groups. Frequency ranging demonstrates proper name coverage of the text, number of proper names and their variety in the text.

3. Text preparation

The work with the text started with its tokenization. Since proper names were of primary interest, all manipulations in the text, such as the removal of syntactic characters, the removal of the tilde (a superscript which was often drawn by the author over the names of persons to emphasize their holiness), etc., were made only with proper names. It should be noted that margin notes, so-called glosses, were not taken into account for the calculation purposes. They are usually abbreviated names of the sources cited, which are quite voluminous and expand the basic information, and therefore warrant a separate study. We did not take into account the title of the collection “Sermons from the Collections *Ohorodok Marii Bohorodytsi* (The Garden of Virgin Mary) (1676) and *Vinets Khrystov* (The Wreath of Christ) (1688) by Antonii Radyvylovskyyi” either, but the titles of all 23 sermons were taken into consideration.

Proper name homonymy is quite a frequent phenomenon, and it was addressed by adding a special character serving to distinguish homonyms. For instance, proper names **Нѡѡ**/Noie/Noah, **Нѡи**/Noi/Noah and word combinations **вторій Нѡѡ**/vtoryi Noie/the second Noah, **новий Нѡѡ**/novyi Noie/new Noah were put into different groups. In the first case, the proper name **Нѡѡ**/Noie/Noah, **Нѡи**/Noi/Noah belongs to the subgroup of *Biblical Anthroponyms*. In the second case, **вторій Нѡѡ**/vtoryi Noie/ the second Noah, **новий Нѡѡ**/novyi Noie/ new Noah is a metaphoric description of Theodosius of the Caves (**Фѡѡдосій Печерскій**)/Feodosii Pecherskyi), so it belongs to the subgroup *Hagionyms. Names of Ruthenian Saints*.

Homonymic proper names **Нилъ**/Nil/Nile were put into different groups, as their different meanings were clarified based on hyperonymic-hyponymic relationships – **рѣка Нилъ**/rika Nil/ river Nile and **Нилъ свѣтый**/Nil sviatyi/ Saint Nilus. Judging by its context, in the phrase **такъ мовитъ Нилъ свѣтый: Не можетъ(ъ) мѡлѣвы чн(с)тѡи до Бга преслати / tak movyt Nil sviatyi: Ne mozhет molytvy chystoi do Boha preslaty/ Thus says St. Nilus: he cannot send a pure prayer to God [15, p. 198] the proper name **Нилъ свѣтый**/Nil sviatyi/ Saint Nilus belongs to the subgroup *Hagionyms. Names of Saints and Martyrs. Names of Christian Theologians and Church Fathers*.**

To avoid errors in the calculation, multicomponent stemmas of proper names were placed at the beginning.

4. Frequency Dictionary compilation procedure

Lemmatization of proper names, used in the text, was made manually. Each lemma-proper name represents a set of stemmas, i.e. word forms that have the same lexical meaning and correspond to the same proper name. Due to different graphic representations of the same sound, stemmas belonging to the same lemma (**Мѡисей**/Moisei) could also look as follows: **Мѡисѣ**:: **Мѡисѣ**/Mois. Here the parallel spellings **ѡ/ѡ** are observed. Quite frequently, a variant spelling of the same proper name occurs in the same grammatical case, for example: **Гѣдѣѡна**/ Hedeona:: **Гѣдѣѡна**/Hedyona; **Дѡ(а)дѣ**:: **Дѡ(а)дѣ**:: **Дѡдѣ**:: **Дѡдѣ**/Davyd.

The theonym *Isus Khrystos/Jesus Christ* has the greatest number of phonetic and orthographic variants (85) in the text.

Sometimes the textual representation of a proper name was so variable that the set of stemmas consisted of word combinations: **Дѡа Мѡриѡ**/Diva Mariia:: **Црѣца Небѡнаѡ Мѡриѡ**/Tsarytsa Nebesnaia Mariia:: **Мѡтри Црѣца Не(с)наѡ**/Materi Tsaria Nebesnoho:: **Дѡѡ Прѣнайсѣтѣѡѡ**/Divo Prenaisviatishaia:: **Бѣѣ Дѡѡ**/Bohorodytse Divo:: **Бѣѣ и пр(с)нѡ Дѡѡ Мѡриѡ**/Bohorodyttsy i Prsnodivy Marii, or **Писмѡ Сѡтѡѡ**/Pismo Sviatoie:: **Писаніѣ Сѡтѡѡ**/Pysaniie Sviatoie.

The study proves that multicomponent proper names in the text may have up to 7 components, thus increasing the number of variants.

Orthographic variants include capital and small letters in spelling proper names in the original text. The ancient text also has a specific division into words. Two separate words **пр(с)нѡ Дѡѡ**/ *prisno Divy* corresponds to one word *Prsnodivy* according to modern orthographic rules.

On the basis of the Dictionary of Proper Names in Radyvylovskyi's Sermons, the proper names were divided into 24 groups, which are as follows: Biblical Anthroponyms; Theonyms. Christian Theonyms; Theonyms Denoting Gods of the Greco-Roman Pantheon; Theonyms. Names Denoting Slavic Deities; Hagionyms. Names of Biblical Prophets; Hagionyms. Names of Evangelists; Names of Angels and Archangels, Apostles; Names Denoting Biblical Demons; Hagionyms. Names of Saints and Martyrs. Names of Christian Theologians and Church Fathers; Hagionyms. Names of Ruthenian Saints; Anthroponyms. Names of Thinkers, Historians and Poets, including Classical Ones; Anthroponyms. Names of World History Figures; Anthroponyms. Names of Ruthenian Princes; Anthroponyms. Names of Parable Characters; Toponyms (geographical names). Astionyms (proper names of cities); Hydronyms (proper names of water bodies); Oronyms (proper names of any landforms); Horonyms (proper names of any territory, region, administrative and territorial unit); Insulonyms (proper names of islands); Ecclesionyms

(proper names of rite venues and places of worship of any religion; includes names of churches, chapels, crosses, and monasteries); Demonyms (names of inhabitants of a certain area, correlate with toponyms); Ethnonyms (names of ethnic groups); Biblionyms. Titles of Religious Texts; Names of the Seven Wonders of the World.

It is worth mentioning that we were interested in performing further break-up of proper names into groups and subgroups, as the aim was to single out each proper name in the text; here we relied on conclusions about the importance of frequency [13] for “understanding the text, its role in the statistical structure of the text” [3, p. 69], “revealing the functional significance of onyms in a given text” [4, p. 414].

5. Statistical data

1. Numerical data was obtained through a specially developed Python scripts. The total number of words in the text is 51,821 (N). 270 unique proper name lemmas (PN) were used for the analysis, which formed 811 (PN_f) unique word forms. The total number of proper names in the text (lemmas + their word forms) or their absolute frequency is 2,611 (PN_n). It means that a unique lemma or a unique word form could be presented several times in the text.

$$\begin{aligned} N &= 51,821 \\ PN &= 270 \\ PN_f &= 811 \\ PN_n &= 2,611 \end{aligned}$$

2. The percentage coverage of the text with proper names (PN_c), i.e. the ratio between the total number of proper names in the text (PN_n) and the total number of words in the text (N), is 0.0503.

$$PN_c = \frac{PN_n}{N} = \frac{2\,611}{51\,821} = 0.0503 \quad (1)$$

3. The diversity of the proper name vocabulary (B_{pn}), i.e. the ratio between the total number of proper name lemmas (PN) and the total number of proper names in the text (PN_n), is 0.1034.

$$B_{pn} = \frac{PN}{PN_n} = \frac{270}{2\,611} = 0.1034 \quad (2)$$

4. The onomastic diversity of the text (B_n), i.e. the ratio between the total number of proper name lemmas (PN) and the total number of words in the text (N), is 0.005.

$$B_n = \frac{PN}{N} = \frac{270}{51\,821} = 0.005 \quad (3)$$

5. The average repetition rate of proper names (A_{pn}), i.e. the ratio between the total number of proper names (PN_n) and the number of proper name lemmas (PN), is 9.6703.

$$A_{pn} = \frac{PN_n}{PN} = \frac{2\,611}{270} = 9.6703 \quad (4)$$

6. The calculations have also identified the least frequently used proper names. Therefore, we found 123 proper names with frequency 1, the so-called “hapax legomena”.

$$PN_1 = 123$$

7. The uniqueness index for proper names (E_{pn}), i.e. the ratio between the number of “hapax legomena” (PN_1) and the total number of proper names in the text (PN_n) is 0.047.

$$E_{pn} = \frac{PN_1}{PN_n} = \frac{123}{2\,611} = 0.047 \quad (5)$$

8. The vocabulary uniqueness index (Ed), i.e. the ratio between the number of proper names with frequency 1 (PN_1) and the number of proper name lemmas (PN), is 0.4555.

$$Ed = \frac{PN_1}{PN} = \frac{123}{270} = 0.4555 \quad (6)$$

9. The text proper name uniqueness index (Et), or the ratio between the number of the “hapax legomena” (PN_1) and the total number of words in the text (N), is 0.0023.

$$Et = \frac{PN_1}{N} = \frac{123}{51\,821} = 0.0023 \quad (7)$$

10. The vocabulary concentration index (Ecd), i.e. the ratio between the number of highest-frequency proper names in the text having frequency of at least 10 ($PN_{10} = 41$) and the number of proper name lemmas (PN), is 0.1518.

$$Ecd = \frac{PN_{10}}{PN} = \frac{41}{270} = 0.1518 \quad (8)$$

11. The specific proper name concentration index (Eco), i.e. the ratio between the number of highest-frequency proper names in the text having frequency of at least 10 (PN_{10}) and the total number of proper names in the text (PN_n), is 0.0157.

$$Eco = \frac{PN_{10}}{PN_n} = \frac{41}{2\,611} = 0.0157 \quad (9)$$

12. The text proper name concentration index (Ect), i.e. the ratio between the number of the highest-frequency proper names in the text having frequency of at least 10 (PN_{10}) and the total number of words in the text (N), is 0.0007.

$$Ect = \frac{PN_{10}}{N} = \frac{41}{51\,821} = 0.0007 \quad (10)$$

6. A Frequency Dictionary of Proper Names in a Seventeenth-Century Original Text

The computer calculations have also identified the most and the least used groups of proper names. The highest-frequency group is “Theonyms. Christian Theonyms” with 1265 word tokens. Then come “Biblical Anthroponyms” with 355 and “Hagionyms. Names of Saints and Martyrs. Names of Christian Theologians and Church Fathers” with 219. The next groups that are of a similar size are “Hagionyms. Names of Ruthenian Saints” with 134 word tokens, “Names of Angels and Archangels, Apostles” with 126 and “Hagionyms. Names of Biblical Prophets” with 117. The lowest-frequency groups are “Insulonyms” – 1, “Oronyms” – 4, “Theonyms. Names Denoting Slavic Deities” – 5, “Anthroponyms. Names of Parable Characters” – 6; and “Names of the Seven Wonders of the World” – 6.

For illustration purposes, we show the usage frequency of proper name groups in the following chart, with more precise data below it.

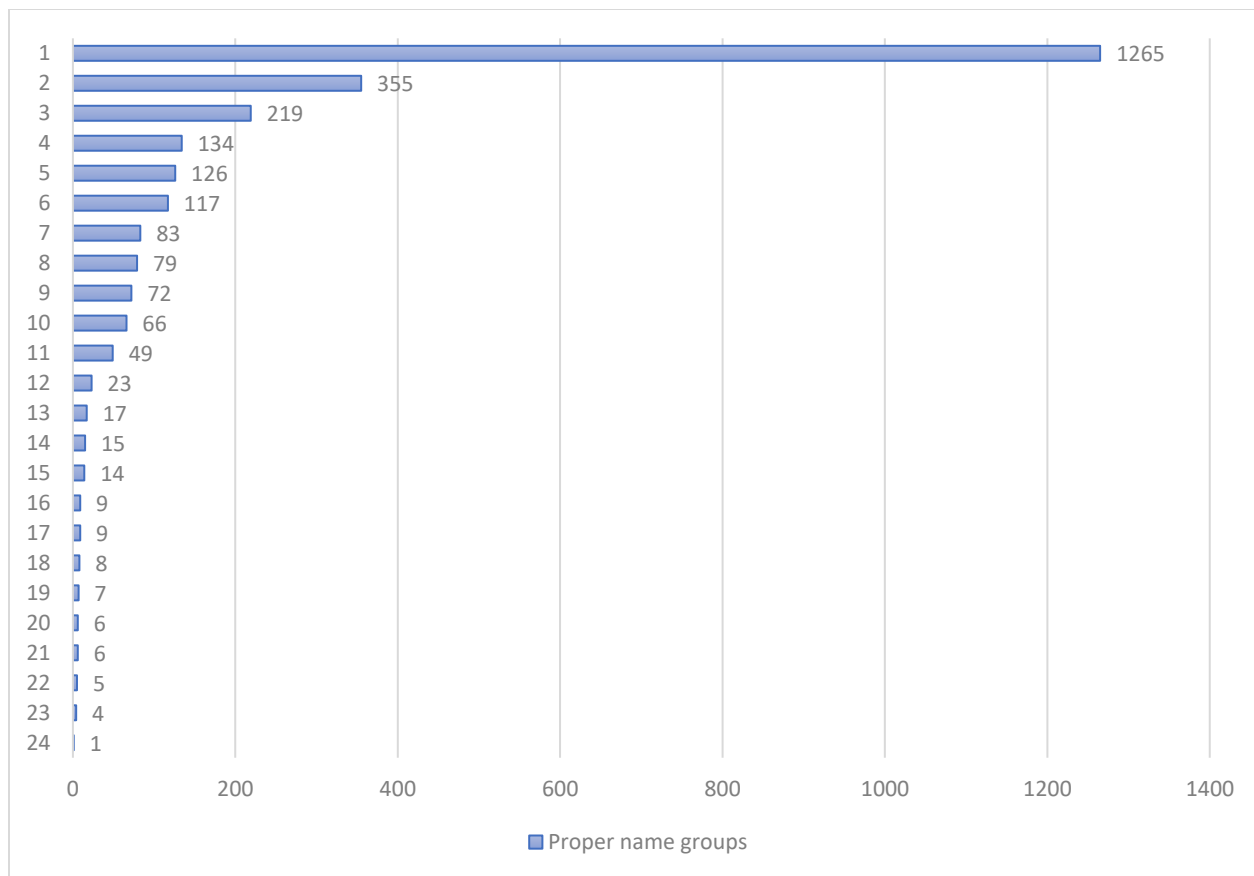


Figure 1: Usage frequency of proper name groups sorted in the decreasing frequency order:

1. Theonyms. Christian Theonyms – 1265; **2.** Biblical Anthroponyms – 355; **3.** Hagionyms. Names of Saints and Martyrs. Names of Christian Theologians and Church Fathers – 219; **4.** Hagionyms. Names of Ruthenian Saints – 134; **5.** Names of Angels and Archangels, Apostles – 126; **6.** Hagionyms. Names of Biblical Prophets – 117; **7.** Biblionyms. Titles of Religious Texts – 83; **8.** Anthroponyms. Names of Ruthenian Princes – 79; **9.** Toponyms. Astionyms – 72; **10.** Horonyms – 66; **11.** Anthroponyms. Names of World History Figures – 49; **12.** Demonyms – 23; **13.** Hagionyms. Names of Evangelists – 17; **14.** Ecclesionyms – 15; **15.** Anthroponyms. Names of Thinkers, Historians and Poets, including Classical Ones – 14; **16.** Hydronyms – 9; **17.** Theonyms Denoting Gods of the Greco-Roman Pantheon – 9; **18.** Ethnonyms – 8; **19.** Names Denoting Biblical Demons – 7; **20.** Anthroponyms. Names of Parable Characters – 6; **21.** Names of the Seven Wonders of the World – 6; **22.** Theonyms. Names Denoting Slavic Deities – 5; **23.** Oronyms – 4; **24.** Insulonyms – 1.

We calculate the proper name group coverage of the text (G_n) according to the formula (11)

$$G_n = \frac{G_o}{N} \times 100 \%, \quad (11)$$

where G_o is the proper name group, and N is the total number of words in the text.

We calculate the proper name group coverage of the proper name vocabulary (G_d) according to the formula (12)

$$G_d = \frac{G_o}{O_n} \times 100 \%, \quad (12)$$

where G_o is the proper name group, and O_n is the total number of proper names in the text. The calculation data are given in Table 1.

Table 1

The proper name group coverage of the text (G_n) and the proper name group coverage of the proper name vocabulary (G_d)

Group	G_n	G_d	Group	G_n	G_d
1	2.44 %	48.45 %	13	0.03 %	0.65 %
2	0.69 %	13.6 %	14	0.03 %	0.57 %
3	0.42 %	8.39 %	15	0.03 %	0.54 %
4	0.26 %	5.13 %	16	0.02 %	0.34 %
5	0.24 %	4.83 %	17	0.02 %	0.34 %
6	0.23 %	4.48 %	18	0.02 %	0.31 %
7	0.16 %	3.18 %	19	0.01 %	0.27 %
8	0.15 %	3.03 %	20	0.01 %	0.23 %
9	0.14 %	2.76 %	21	0.01 %	0.23 %
10	0.13 %	2.53 %	22	0.01 %	0.19 %
11	0.09 %	1.88 %	23	0.008 %	0.15 %
12	0.04 %	0.88 %	24	0.002 %	0.04 %

Table 2 shows the results of calculating the proper name frequency distribution on the basis of **A Frequency Dictionary of Proper Names in a Seventeenth-Century Original Text**, sorted in the decreasing frequency order.

Table 2

Calculation results of the proper name frequency distribution

Interval of frequencies	Number of Proper Names	Interval of frequencies	Number of Proper Names
500 - 999	1	20 - 29	7
400 - 499	1	10 - 19	21
300 - 399	0	9	4
200 - 299	0	8	3
100 - 199	0	7	5
90 - 99	0	6	4
80 - 89	0	5	11
70 - 79	3	4	7
60 - 69	1	3	22
50 - 59	2	2	53
40 - 49	2	1	123
30 - 39	3		

The table shows that 1 proper name with a usage frequency in the range of 500 – 999 and 1 proper name with a frequency in the range of 400 – 499 have been recorded in the text “Antonii Radyvylovskiy. Baroque Sermons of the 17th Century”. Interestingly, no lemmas fall in the range between 399 and 80, while the largest number of proper names have a usage frequency of 1, 2 and 3 (see Fig. 2, 3).

Therefore, examples of frequency-rated proper names are displayed for illustration purposes. The highest-frequency proper names are: **црѣ нб(с)ный**/Tsar Nebesnyi – 677, **црѣ нба и земли хс спситель**/Tsar Neba i Zemli Khrystos Spasytel – 428, **дх(м) сты(м)**/Dukhom Sviatym – 77, **мтки рж(с)тва сна бжегѡ**/Matky Rozhstva Syna Bozheho – 73, **павла**/Pavla – 72, **писмѣ стѡмѣ**/Pysmi Sviatom – 63, **антѡнїй печерскїй**/Antonyi Pecherskyi – 54, **илїа**/Iliia – 52, **ніколае**/Nikolaie – 49, **зарѡ**/Zaru – 46, **владимїрѣ**/Vladymir – 38.

Interestingly, almost one half of proper names are unique, with a usage frequency of 1: **їстинїанѣ**/Iustynyan, **днѣпромѣ**/Dniprom, **плутархѣ**/Pliutarkh, **їераполи**/Iierapoly, **алєкѡдрѣ царицѣ**/Alezandru tsarytsu, **геркулєсѡвѣ**/Herkulesov, **дїѡгенєсѣ фїлософѣ**/Diogenes filozof, **божница дїаннї богинїѣ ефескѡи**/Bozhnytsa Diany Bohyni Efeskoi, **мѡры вавилонскїѣ**/mury Vavilonskiie, **рѣка нилѣ**/rika Nil, etc.

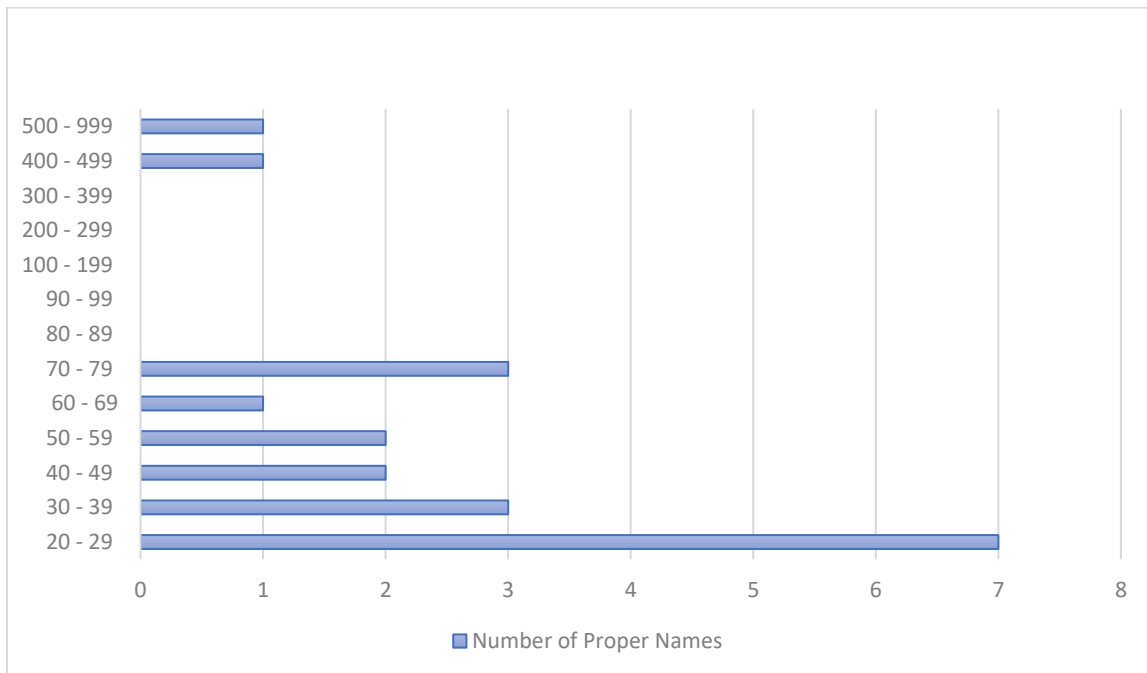


Figure 2: Graphic representation of the distribution of proper names having frequency between 999 and 20 in the decreasing frequency order, based on Table 2.

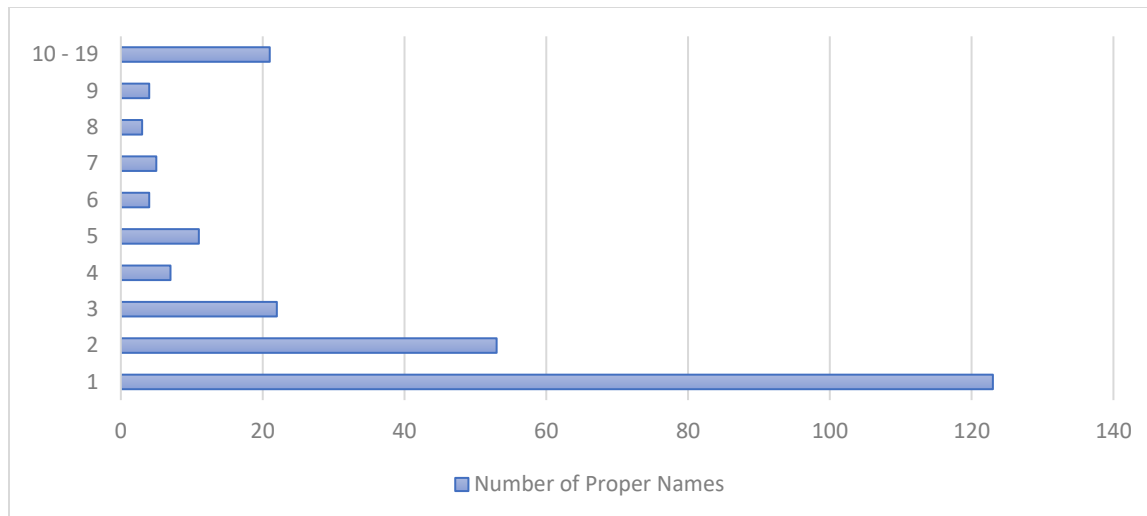


Figure 3: Graphic representation of the distribution of proper names having frequency between 19 and 1 in the decreasing frequency order, based on Table 2.

7. Conclusions

Working with an original seventeenth-century text, we have been able to reveal the specific historical features of the analyzed material at different stages of the study. Using a sequential algorithm, we compiled the Frequency Dictionary of Proper Names in a Seventeenth-Century Original Text, sorted in the decreasing frequency order (based on Radyvylovskyi’s sermons), described the volume of the text and its vocabulary, the variety of proper names in it, as well as the number of low frequency proper names (words with the frequency of 1 (hapax legomena)).

The onomastic space of historical records of “ruska mova” (the Ruthenian language) studied through the prism of quantitative measurement, allows us to draw conclusions about the worldview paradigms of the seventeenth-century intellectuals and identify proper names which were significant for the preacher and his audience. In particular, analysis of proper names attested in the publication *Radyvylovskyi Antonii. Barokovi propovidi 17 stolittia* (Radyvylovskyi Antonii. Baroque sermons of the 17th century) has shown that a Baroque preacher saw the Gospel as the main and undisputed authority among books. Thinking in the context of the Christian doctrine and being a zealous Christian, the author believed that the *Holy Trinity, composed of God the Father, Jesus Christ and the Holy Spirit*, was the supreme being, and ranked the *Virgin Mary* very high as well. This was also determined by the themes of the sermons, most of which were festive ones.

As a polyglot and erudite, Radyvylovskyi expanded the proper name range of sermons by using proper names of thinkers, historians and poets: **Авзоній/Avzonii**, **Аристотеле(с)/Arystoteles**, **Діогенесъ/Diogenes**, **Діоскоридесъ/Dioskorydes**, **Оригенъ/Orihen**, **Плѣнішъ/Pliniush**, **Плутархъ/Pliutarkh**, **Плато/Platon**, **Светоній/Svetonii**, (frequency 1), **Софоклесъ/Sofokles** (2) and world history figures: **Діоклитіанъ/Dioklytian** (9), **Юлішъ/ Yuliush** (3), **Непоціан/Nepotsian** – (1), etc.

The study is important for comparing the frequently used words or “hapax legomena” established in the Dictionary with the texts of different genres and historical periods in historical corpora to identify intertextual connections (for example, with the Bible), precedent names.

Further studies are warranted to compare the proper name frequency dictionary based on the seventeenth-century sermons with other texts of this genre in synchronic and diachronic perspectives, as well as with texts of other genres, which will add new data on the usage frequency of different proper name groups, development of their secondary meanings, ratio of one- and multicomponent proper names, etc.

8. Acknowledgements

We are grateful to Oleksandr Malin for his help with calculation of data, preparation of the text and indexes for calculating statistics.

9. References

- [1] Słownik staropolskich nazw osobowych [Dictionary of Old Polish Personal Names], pod red. W. Taszyckiego. Wrocław, 1965–1987. T. I–VII.
- [2] I. Mytnik, Słownik historyczno-etymologiczny antroponimii ziemi chełmskiej XVI–XVII wieku [Historical and etymological Dictionary of Anthroponymy of the Chełm Land in 16th–17th centuries]. Warszawa. 2017.
- [3] S. N. Buk, Onimnyi prostir romanu Ivana Franka “Perekhresni stezhky” [Proper-Name Space of “Perekhresni Stezhky (The Cross-Paths)”, a Novel by Ivan Franko] In: Onomastychni nauky, 2012. No. 4, pp. 68–76.
- [4] E. Vertheyko, Iz opyta sozdaniya Slovarya lichnykh imyen sobstvennykh belorussko-polskogo pogranichya [From the Experience of Creating the Dictionary of Personal Names of Belarusian-Polish Borderlands] In: Studia Wschodniosłowiańskie, T. 19, 2019, pp. 413-429. doi: 10.15290/sw.2019.19.29.
- [5] I.-I. Popescu, G. Altmann, P. Grzybek, B. D. Jayaram, R. Köhler, V. Krupa, J. Mačutek, R. Pustet, L. Uhlířová and M. N. Vidya, Word Frequency Studies. Berlin-New York, Mouton de Gruyter. 2009. doi: 10.1515/9783110218534.
- [6] V. I. Perebyinis, Statystychni metody dlia lnhvistiv: posibnyk. [Statistical methods for linguists: manual]. Vinnytsia, Nova Knyha, 2013.
- [7] Anthony Kenny, A Stylometric Study of the New Testament. Oxford University Press UK, 1986.
- [8] Oksana Nika, Svitlana Hrytsyna, Frequency Dictionary of 16th Century Cyrillic Written Monument. In: Jazykovedný časopis, Vol. 70, No. 2, 2019, pp. 276–288. doi: 10.2478/jazcas-2019-0058.
- [9] List of Key Onomastic Terms. International Council of Onomastic Sciences. URL: <https://adoc.pub/list-of-key-onomastic-terms.html>.
- [10] Słowiańska onomastyka. Encyklopedia [Slavic onomastics. Encyclopedia] / pod red. E. Rzetelskiej-Feleszko i A. Cieślíkovej. Warszawa; Kraków : Towarzystwo Naukowe Warszawskie, 2002–2003. T. 1–2.
- [11] Slovyk ukrainskoi onomastychnoi terminolohii. [Dictionary of Ukrainian onomastic terminology] / Uklad. Buchko D. G., Tkachova N. V. Kharkiv, Ranok – NT, 2012.
- [12] N. P. Darchuk, Kompyuterna lnhvistyka (avtomatyчне opratsyuvannya tekstu). [Computer linguistics (automatic processing of text)]. Kyiv: Kyiv university. 2008.
- [13] A. Kilgarriff, Putting frequencies in the dictionary. In: International Journal of Lexicography, Volume 10, Issue 2, 1997, pp. 135–155. doi: 10.1093/ijl/10.2.135_2
- [14] Reinhard Köhler, Gabriel Altmann, Rajmund G. Piotrowski (eds.), Quantitative Linguistik / Quantitative Linguistics. Ein internationales Handbuch / An International Handbook: DeGruyter. 2005.
- [15] Oksana Nika, Yuliia Oleshko. Radyvylovskiy A. Barokovi propovidi XVII st. [Baroque sermons of the 17th century]. Kyiv, Osvita Ukrainy. 2019.