LINGUISTICA URALICA LVIII 2022 4 https://dx.doi.org/10.3176/lu.2022.4.01

JAAKKO RAUNAMAA, *ANTTI KANNER* (Helsinki)

CLUSTERING NAMES OF MEDIEVAL NOVGOROD: GEOGRAPHICAL VARIATION OF PERSONAL NAMES ATTESTED IN THE CENSUS BOOK OF VODSKAJA PJATINA

Abstract. This article attempts to investigate ancient personal names through digital methods. The article focuses on names attested in the census book of Vodskaja pjatina, one of the five administrative areas of late 15th century Novgorod. The research data is compiled digitally. Regional differences in naming conventions are studied through two hierarchical clustering procedures: one based Jaccard index with the average linkage and other based on Euclidean distance metric with the ward linkage. Overall, 35,726 names are collected, whereas the number of individual name variants is 2748. Almost all of the most common names are of Christian origin. Names containing Finnic features are a minority, approx. 2% of all. Distance measures and clustering results turned out to be useful to study ancient naming conventions particularly at the fine-grained level. At larger scale, the outcomes of both procedures are mostly in line with previous treatises of the study area's past: by the end of the 15th century, the southern region had already become Slavicized whereas the northern parts still had a significant Finnic population.

Keywords: onomastics, personal names, Finnic history, Russian history, digital humanities, clustering.

1. Introduction

In this paper, digital methods are used innovatively to compile and cluster personal names of medieval Northwest Russia. Computational clustering has rarely been used to investigate the differences in naming conventions. The rare studies where it has been used focus on contemporary names only.[1] According to our knowledge, this is the first time when this kind of methods are applied to study medieval naming conventions. Names analysed in the present work are collected from the census book of Vodskaja pjatina, which was one of the five administrative areas in late 15th century Novgorod (see the Map 1 below).[2] The material, which dates to 1499—1500, contains close to 36000 personal names

[1] Examples of studies based on modern name materials are Sousa, Ginzo Villamayor 2021 and Cheshire, Mateos, Longley 2011.

[2] Водская пятина 'Votic Fifth', named after Votes, a Finnic tribe, who were inhabiting parts of the region. See sections 1.1. and 6 for more details.

and covers large areas in present-day Finland and Russia.[3] The aim of clustering is to find regional differences in the use of names. Special emphasis is put on the Finnic features presented in the data. Observations made based on computational methods are compared with the results of other disciplines such as those of archaeology, history, historical linguistics and onomastics.

Since Northwest Russia has been remote and loosely populated before the modern era, there is only a limited amount of historically relevant sources, such as archaeological finds or written documents. For a long time already, researchers, from various fields of sciences, interested in history have used linguistics and onomastics in order to create a more comprehensive picture of Northwest Russia's past (e.g. Sjögren 1969 [1861]; Седов 1982; Рябинин 1997). However, as onomastics is not the main focus in any of these studies, the use of names as a source material is small scale and limited.

Lately, researchers such as Irma Mullonen (Муллонен 2008), Pauli Rahkonen (2013), Denis Kuzmin (2014) have studied place names and revealed that the cultural and linguistic history of Northwest Russia has been more diverse and richer than previously assumed. All these studies are done carefully and received well by the academic community. Despite this, they all have the very same shortcoming, which is that they rely mainly on contemporary name data. Even if names can have roots many centuries back, it is still more difficult to make conclusions about the past based on data that is several hundred years newer than the time studied. The same problem applies to studies that have addressed the past of Finnic and Finno-Ugric languages and peoples on the basis of contemporary genetic data (e.g. Tambets, Yunusbayev, Hudjashov, Ilumäe, Rootsi, Honkola, Vesakoski, Atkinson, Skoglund, Kushniarevich, Litvinov, Reidla, Metspalu, Saag, Rantanen, Karmin, Parik, Zhadanov, Gubina, Damba, Bermisheva, Reisberg, Dibirova, Evseeva, Nelis, Klovins, Metspalu, Esko, Balanovsky, Balanovska, Khusnutdinova, Osipova, Voevoda, Villems, Kivisild, Metspalu 2018) or lexical variation (e.g. Honkola, Santaharju, Syrjänen, Pajusalu 2019). Ancient written material has shortcomings as well and they will be explained more extensively in sections 2 and 3. It is, nevertheless, clear that the late medieval personal name data gives a possibility to have a contemporary view into the ancient anthroponymic system and accordingly into the past of Northwest Russia.

There are also studies, which focus on the names attested in the census books of Vodskaja pjatina. For instance, Aleksej Novožilov (Новожилов 2004) Boris Čibisov (Чибисов 2019) have investigated the spread of Finnic ethnicity within the fifths of Vodskaja and Šelonskaja on the basis of names linked to Finnic tribes. The results of these works indicate the areas where Baltic-Finns were located at the end of 15th century. Although both publications are valid researches, the fact is that they focus on a very little portion of the name data available within the material. For example, the study of Novožilov is based on 724 Finnic personal names that he has collected from the census book of Vodskaja pjatina. This number covers only approximately 2% of all the names presented in the material. As this study will indicate, the 98% left unanalysed in the study of Novožilov also contain significant information of the ethnic and linguistic past in Vodskaja pjatina.

[3] Original documents are dated to the year 7008 after the creation of the world according to the old Byzantine calendar. This is equivalent of the years 1499 and 1500 in the Gregorian calendar.

The present study is comprised of seven sections. The first one gives a short overview of the study area's history. The second section introduces the research material and discusses issues related to it. The third section focuses on research methods. We explain how the research data is compiled and modified for the purposes of this study. Furthermore, we present how the data is clustered. The fourth section gives an overview of the personal names used in the census book. Special emphasis is on Finnic names and features. The fifth section presents the clustering results in the form dendrograms and maps. Next, the clustering results are evaluated and interpreted, and, in the last section, the conclusions of this study are drawn together and discussed.

1.1. Historical background

It has been assumed that during the later stages of the Iron Age (500—1000 AD) Northwest Russia together with the area of Vodskaja pjatina was mostly populated by Finno-Ugric tribes (Рябинин 1997 : 236). It has been long debated when Slavic people arrived in Northwest Russia. According to Sedov (Седов 1995), the spread of Slavic settlement is connected to two kinds of burial styles called "long barrows" (= kurgans) and "sopkas". The first of them appeared in Northwest Russia at the middle of the first millennium AD and was found mostly on both sides of Lake Peipus and in the vicinity of Lake Ilmen (Седов 1995 : 211—213). Sopkas, burial mounds of slightly different shape than the long barrows, started to be used in 8th century and spread around Lake Ilmen and River Volkhov (Седов 1995 : 234—246). However, Sedov does not categorically connect these burial styles to the spread of Slavs, as he suggests that the local Baltic and Finnic tribes were probably adopting these customs as well.

During the past decades many researchers have been critical in linking specific burials styles (long barrows and sopkas) to the immigration of Slavs. Much of the discussion is presented by Anders Tvauri (2007). He claims that immigration of Slavs could not have been significant, as northern Russians have a closer genetic connection with the Baltic and Finnic populations than with the other Slavic groups (27). Regardless of all the criticism, it is clear that before the end of first millennium AD many areas in Northwest Russia had tight contacts with Slavic culture.

Similarly, there is clear evidence in archaeological data that Northwest Russia and especially market places and fortresses, like Rjurikovo Gorodishche and Staraja Ladoga, were influenced by Scandinavians (Androshchuk 2008). Scandinavians had strong impact on the formation of the first Russian states of Kievan Rus and Novgorod as well. For example, the first rulers of Novgorod and Kievan Rus had Scandinavian roots as their names indicate: *Rurik, Oleg* (< *Helgi*) and *Igor* (< *Ingvar*).

Whereas the Scandinavian influence was in decline during the first centuries of the second millennium, the Slavic cultural sphere continued to spread northwards in the form of agricultural colonization (see e.g. Рябинин 1997 : 3—7). As a result, regions south of the River Neva were multi-ethnic and multilinguistic in the end of the Middle Ages (see e.g. Рябинин 1997 : 3—81; Конькова 2008). Areas on the Karelian Isthmus and on the north-western coast of Lake Ladoga had, in turn, been less affected by Slavic influence (see e.g. Lang 2020 : 324). During the first centuries of the second millen-

nium, the Finnic populace of the region was probably divided into various tribes, but, in contemporary research, the situation is usually simplified by assuming that the main groups were the following ones: the Votes immediately east of the River Narva, followed by the Izhorians in the vicinity of River Neva, and the Karelians to the north (see e.g. Рябинин 1997 : 3—81; Frog, Saarikivi 2015; Lang 2020 : 317—326). A more difficult question is where exactly these groups were situated and how they evolved during the first centuries of the second millennium.

The city-state of Novgorod was established originally as a marketplace at the mouth of the Volkhov River on Lake Ilmen at the end of the first millennium AD (Ianin 2006 : 191—192). At first, Novgorod was a subject of Kievan Rus, but soon enough it gained political independence. From the 12th century onwards, local nobles called *boyars* together with the most powerful merchants increased their authority (Ianin 2006 : 195). The Orthodox Church was a significant political player in Vodskaja pjatina as well (Selart 2015 : 17). It controlled vast amounts of land in the countryside. In the county of Korela (in Finnish *Karjala*), the Church owned approximately half of the land in the end of 15th century (Korpela 2004 : 205).

Only very little is known about the government outside the city itself. That much is certain that the administrative system varied between different areas. It is known that, for example, the city of Staraja Ladoga was ruled by a *posadnik* appointed by Novgorod (Saksa, Uino, Hiekkanen 2003 : 458). In the countryside, however, leading officials of the city had only limited control, as nobility and the Church were the most significant powers. According to the census book (see section 2), most of the land was owned by these two. It has been suggested that the expansion of Novgorod to new areas was a result of land seizures made by the nobility and the Church (Витов 1962 : 51—54).

During the 15th century, Novgorod gradually lost its independence and was subordinated to the Grand Duchy of Moscow. Most of the land-property owned by the nobility and significant portions of the ecclesial possessions were confiscated in favour of Muscovites or the Grand Duke himself (Баранов 1999 : 5—6; Ianin 2006 : 206). If the land had previously been divided amongst a few large landowners, it was now distributed among the Grand Duke's many trustees. Changes in the ownership often meant changes in the peasantry as well. It is likely that the Slavinization of Northwest Russia had already started when Novgorod spread its influence on the area. Now, under the rule of the Grand Duke, immigrants started to move to the northwest from the heartlands of Moscow (Витов 1962 : 49—78; Баранов 1999 : 5—6).

As Moscow was strengthening its conquest of Novgorod, one form of subjugation was compiling census books (*писцовая книга,* 'scribal book'), in which the basis for the area's fiscal natures would be created (Баранов 1999 : 5—6).[4] This meant gathering information on taxation, land possessions, salaries, duties and such. The process of compiling the census book is better described in section 2. Furthermore, these administrative reforms were probably connected to the area of Novgorod being split into "fifths" (in Russian пятина, *pjatina*).[5]

[4] Translations are done by the authors unless noted otherwise.
[5] Since very little information has survived regarding the land division, nothing certain can be said about the origins.

1.2. The study area

The study area, namely Vodskaja pjatina (*Водская пятина*), was one the administrative fifths of medieval Novgorod. Map 1 displays the location of this region together with its outer borders. The visualization is mostly based on a map compiled by Sergij Tihomirov (Тихомиров 1905). Corrections were made when Tihomirov's map and the locations of settlements mentioned in the census book differed from each other.[6] The borders of other administrative areas presented in Map 1 are based on the map compiled by Konstantin Nevolin (Неволин 1853). Vodskaja pjatina's division against the other fifths, *Šelonskaja* in the west and *Obonežskaja* in east, is quite accurate because the borders were mainly based on natural formations, such as rivers and lakes, which have been easy to locate. The northern border is not very accurate and is drawn based on the

Map 1. **The study area (Vodskaja pjatina) together with other places and areas relevant to the study. Borders presented in the map are based on Неволин (1853), Тихомиров (1905) and Raunamaa (2020). All the maps are drawn by the authors. Base map is always Stamen Toner Background unless stated otherwise.**

[6] The locations of many settlements attested in the census book were determined for the purposes another study (Raunamaa 2020).

northernmost settlements mentioned in the census book of 1499—1500. Northern parts of the western border against the Diocese of Åbo (part of the Realm of Sweden) are also inaccurate due to the remoteness of the area. Only the southwestern part of the border in the Karelian Isthmus and Eastern Finland is indisputable as it is described in detail in the Treaty of Nöteborg (1323). A star on the map indicates the location of Novgorod.

Map 2, which is again based on Тихомиров 1905 and Raunamaa 2020, depicts locations and borders of those pogosts ('parishes') and towns that are covered in the study material. The names of numbered pogosts are found in Appendix 1. The orthography follows the 19th century editions. The blank area inside Vodskaja pjatina's southwestern part consists of the pogosts Greznevskoj and Orlinskoj, which were not included in the edited version of 1499—1500 census book.[7] There is another gap at the eastern border where a pogost Nikol'skoj s Gorodišča belonged to Obonežskaja pjatina although it was located on the western bank of the River Volkhov.

Map 2. **Pogosts and towns within Vodskaja pjatina. Numbers refer to pogosts whose names are explained in Appendix 1.**

Borders presented in Map 2 are only indicative. The biggest problem is that there are cases when a village, according to the census book, belongs to certain pogost but, in fact, it is located deep in a neighbouring one. This is the case, for example, of the pogosts Opol'skoj v Čudi (36)[8] and Jegor'jevskoj Radšinskoj (38)[9] and accordingly, the line between these two parishes marks a border zone rather than an exact border. Similarly, some villages belonging to the Voskresenskoj Gorodenskoj pogost (54) in the Karelian Isthmus are located inside the neighbouring pogost Kir'jažskoj (57).

2. Research material

This study is based on edited versions of the census book of Vodskaja pjatina. Editions were compiled in parts and the first two, which cover the northern-

[7] As seen in Appendix 1, the names of pogosts often include religious vocabulary, e.g. *Спасской* ('Saviour's') or *Воскресенской* ('Resurrection's'). To save space, these parts have often been excluded when referring to a pogost.
[8] The number inside the parentheses refers to respective pogost's number in Appendix 1.
[9] The pogost Jegor'jevskoj Radšinskoj was divided into two parts: the northeastern part belonged into Koporskoj uyezd ('county') and the southwestern one into Jamskoj uyezd.

most region, are called "Переписная окладная книга по новугороду вотской пятины" (1851; 1852) (POKV I—II).[10] The third part of the book series "Новгородские писцовые книги" (1868) (NPK III) covers the southern and western parts of the study area.[11]

As noted in the previous section, the Novgorod census books were compiled because the Grand Duchy of Moscow wanted to establish its authority over the subordinated area. Whereas political power in Novgorod was shared among nobles, merchants and the Church, the Grand Duchy of Moscow aimed to create a highly centralized and autocratic system where the Grand Duke had supreme powers over his subjects. This meant developing fiscal and bureaucratic infrastructures in order to have the control over the lands of the duchy. It is likely that a first census book was compiled quite soon after Moscow completed its conquest in 1478. This first inscription has not been preserved, but it is often referred to in the book of 1499—1500 (Баранов 1999 : 5—6; Korpela 2004).

It is uncertain how exactly the census book of Vodskaja was compiled, but it is known that the task was assigned to Dmitrij Vasil'jevič Kitajev and Nikita Semenov. The former was a boyar from one the most important noble families in Moscow, whereas the latter was a clerk often used in important diplomatic missions (Баранов 1999 : 10—11). It is unlikely that these two men with their high rank and political influence would have done all the documentation by themselves and therefore, they probably had assistants. Apparently, the officers given this task did not visit all the villages they documented. At least the authors of the Šelonskaja pjatina census book from 1584—85 wrote that local priests and trusted men had been used as informants (Ronimus 1906 : 6).

The form of documentation is similar through the whole census book. The inscription is divided into six parts according to local counties (Russian *уезд*, known in English as *uyezd*), which again consists of various numbers of parishes (*погост*, *pogost*). From one pogost to another, the census book describes the taxes, duties, possessions and salaries. Pogosts are divided into *volost*(s) (*волость*), which seem to be based on the possessions of Novgorodian nobles or ecclesiastical institutions, such as monasteries. Volosts consist of different kinds of smaller settlements, like *деревня* ('countryside villages'), *село* ('villages where there is a church or manor/homestead of a landowner') and *починок* ('new settlement') (Витов 1962 : 98). The settlements have a various number of homesteads or houses (*двор*), most commonly from one to three. There are, however, villages with dozens of homesteads, such as in village Pužzvina in Solomjanskoj (60) pogost, where the number of them is 75 (POKV II 180).

Each homestead is inhabited by one or more taxpayers and their families. The number of taxpayers in one homestead varies from area to area. In the southern parishes, it is common to have only one person named per homestead whereas in the north, one homestead can be inhabited by many taxpayers. For example, in the Pužzvina village mentioned above, the inscription read as follows: "(д) Кондратко Макавѣевъ, сынъ его Панкратко, братаничь его Огафонко Ивашковъ".[12]

[10] Translation: 'Census tax book of Novgorod's Votic fifth'.
[11] Translation: 'Scribe books of Novgorod'.
[12] Translation: '(house/homestead) Kondratko Makavejev", his son Pankratko, his nephew Ogafonko Ivaškov'.

The census book did not only keep record of the names of the peasants living in the countryside, but it also documented the residents of towns and ecclesiastical properties, as well as the servants in the manors/homesteads of landowners. It is also noteworthy that not all the people mentioned in the census book are taxpayers. Priests, artillery men, governors and communion bread bakers are among those who are named. These people did not pay any tax, but their salaries and duties are recorded, nevertheless. Women are mentioned only rarely. For example, communion bread bakers are always women and are found in almost all the parishes. Sometimes a widow is presented as the head of a homestead but even then, her deceased husband is usually named as well: e.g. "вдова Василиста Машковская жена" (POKV II 113).[13]

As the personal names mentioned in the census book are in the focus of this study, the following section will give an in-depth description of them. Typically, names of taxpayers are presented so, that the first thing mentioned is person's given name and then comes a patronym: e.g. "Ивашко Еремѣевъ" (POKV I 200). Sometimes a patronym is replaced by a name that can be considered something like a parallel given name or a clan name: "Ивашко Уской" (POKV I 218). Occasionally, a person can have three names and often the second one seems to be one of non-Christian origin: "Сменко Репуй Лентьевъ" (NPK III 543). Furthermore, a taxpayer could have important features that were worth mentioning, such as being the local tithe proctor[14] ("д е с я т ц к о й Офонасъ Спировъ", POKV I 199) or a headman of local peasants ("Андрейко Онтоновъ с т а р о с т а ", POKV II 144). Of particular interest for this study are the ethnonyms referring to different ethnicities. One can find names like "Игамасъ Ч ю д и н ъ" (NPK III 508) and "Царко И ж е р я н и н ъ" (NPK III 519).[15] Ethnonyms referring to Finnic tribes are discussed further in section 4.

2.1. Source-critical issues

In general, the most significant source-critical problems of the census book are connected to the obscurity of how the names are recorded. First of all, the orthography of names does not often correspond with the vernacular forms since the scribes were adapting the names to fit in with the traditions of local administration. The border region between the Diocese of Åbo and Novgorod (= the Grand Duchy of Moscow) is a good example of this. On the Swedish side, in Kivennapa parish (Swedish Kivinebb), one can find many peasants with personal names of Russian origin but most of them in different form compared to that how they are presented on the other side of the border, in Kujvošskoj (46) pogost: e.g. *Levoska* vs. *Левонко, Iffua*(*n*) vs. *Ивашко* and *Savo* vs. *Савка*.[16]

Scribes, who compiled the census book of Vodskaja, were clearly aware of the earlier records that were done in similar standards. A comparison with the edition made of the census book of Obonežskaja pjatina, originally compiled in 1496, implies that the form of the inscription, including the orthography of

[13] Translation: 'A widow Vasilista, wife of Maškov'.
[14] Tithe proctor = the one who collected ecclesiastical taxes.
[15] The exact meaning of the ethnonym *Chud* is still unsettled, but general consensus is that it has been connected to the Finnic speaking population living in Northwest Russia. See Grünthal 1997 for more details. The ethnonym *Iževjanin* refers to Finnic group that has been inhabiting Ingria, i.e. areas between the rivers Narva, Neva and Volkhov.
[16] Names have been obtained from a census document (Finnish *savuluettelo*) compiled in 1545 (Viipurin ja Porvoon läänien savuluettelo 1545—1545 (Signum VA 5006, pages 32—38)).

the names, is similar in both books (PKOP). In addition, despite the sheer size of Vodskaja pjatina, many forms are similar in different ends of the region, which again underlines how systematically the scribes were adapting the personal names to fit with their traditions.

On the other hand, there are also many examples of recorded personal names depicting local variances. This applies especially to the northern and western parts of Vodskaja pjatina that probably had a considerable Finnic speaking population. In those areas, one can find names that are probably Finnic forms of common Russian ones: e.g. *Миккуй* < *Мики*(*форик*), *Ивантуйка* < *Иван* and *Антуй* < *Андрей* (see section 4.2). Similarly, there is a small variation within the typical Russian names as well. For example, in some instances, the vowels *а* and *о* were used irregularly in the beginning of a name (e.g. *Андрейко* vs. *Ондрейко*) or in the end (*Ивашко* vs. *Ивашка*). It must be emphasized that this kind of variation might be caused by 19th century editors as well.

The original census books were found in the archives of Moscow in the 19th century. The Archaeographic Commission, which was a subgroup within the Academy of Sciences in St. Petersburg, started the editorial work. The first edition that was made (1836) covered the census book of Derevskaja pjatina from 1495—1496. Next, the commission began work with the census book of Vodskaja pjatina. First and second parts were published 1851 and 1852. The third part was printed in 1868.

The editions of 1851, 1852 and 1868 do not fully match the original documents. Original documents were written in language that is called a Russian chancery language. It was developed for the purposes of the Muscovite government and its need for bureaucratic documentation (Worth, Flier 2012 : 32). In order to make content of the census book comprehensible for the mid-19th century readers, editors could not copy original text as it was. Since some letters had changed their form or were not in use anymore, the editors transformed them into their mid-19th century equivalents. Editors have also supplemented words that had missing letters. Furthermore, one must remember that, as the original documents included over 300 years old handwritten text, it is likely that there are some letters and words that the editors have misunderstood. The process of editing is described more broadly by Nevolin (Неволин 1853 : 4, in Appendix 1).

As stated above, details of two pogosts were lost during the editing process. Furthermore, editors state a few times that some pages or details are missing. Likewise, the editors have misplaced some pages. Despite the errors and losses, most of the original census book has survived to this day.

3. Methods

Below we go through the most important methods used in the present study. In the first and second sections, we explain how the research material was refined and harmonized and processed into area-by-name frequency tabulations. In the third section, we describe how these tabulations are further processed into pairwise distance matrices, where each area is compared for similarity against each other area. To explore how different choices on a very basic level of modelling nomenclatures affect the clusters, we have compared results obtained from two different types of pairwise distance matrices; one where the distances are based on Jaccard index and the other where it is based on Euclidean distances.

3.1. Data compiling

First, editions of the census book were obtained as scanned PDF files (see example in Figure 1).[17] These were transformed as editable copies by using OCR (= Optical Character Recognition) software Abbyy. The program read the original mid-19th century Russian text adequately with its old Russian alphabet package. Nonetheless, there were many minor mistakes which required manual correction and cleaning. One of the biggest problems was the variation in the print quality. Some of the sections of the material had inferior quality, due to their age but also to variation in material aspects of the texts such as ink and paper quality. This caused the OCR to produce somewhat uneven results. The print quality was better in the 1868 edition (NPK III) but was not flawless there either. This particular edition, on the other hand, had two columns printed close to each other on the same page, which also caused problems for the OCR.

Деревня Погищы. (д) Олексѣйко Иевль, (д) Офремко да Куземко Ивашковы, (д) Ортюхно Ѳефиловъ, (д) Ивашко, да Гридка, да Июдка, да Ивашко жъ Карпиковы дѣти; (д) Ондрейко Овциформовъ, (д) Михалъ да Степанко Ивашковы, (д) Миколка Яшковъ, да сынъ его Терешко, (д) Родивонко Стехновъ, (д) Олексѣйко Вахтуй, (д) Тимошка Юхновъ, (д) Омельянко Демешкинъ Ускловъ, (д) Ивашко Кузьминъ, (д) Ондрейко Ивашковъ, (д) Гридка Киреша, да сынъ его Васько, (д) Гридка Яшковъ, (д) Яшко Устиновъ; двадцать и три луки.

Деревня Никола на рѣцѣ на Кирьемѣ. (д) Матѳѣйко Пахомовъ, (д) Степанко да Васько Исаковы, (д) Кирилко Новзѣевъ, (д) Левонко Степановъ, (д) Михалъ Исаковъ, (д) Яшко Кузьминъ, (д) Михѣйко Федотовъ; восмь луковъ.

Деревня подъ Городищомъ. (д) Яшко Ивашковъ, (д) Тимошка Овдѣевъ, (д) Еремка Микитинъ; три лука.

Figure 1. **An example of a page in the edition of the census book (POKV II 127).**

Excluding introductions, prefaces and such, these three editions consist of 1,111 pages, of which 480 pages are in NPK III and printed on two columns. Since personal names are in the focus of this study, we concentrated on correcting OCR-errors related to them. Some of the sections of the text were displaced by the editors and required some manual work to align the pogosts and their respective records.

After the initial corrections, a Python script was written to harvest the personal names. This was based on exploiting the systematic formalities in how most of the names were presented in the census book. The script looked for abbreviations *дв.* and *д.* and extracted all following capitalized words until section end markers ".", ";" or ":".[18] As an output, a name to pogost matrix was produced, which held the raw frequencies of each word in each pogost.[19]

[17] Pdf-files were retrieved from https://www.aroundspb.ru/perepisnaya-kniga.html.
[18] Abbreviations *дв.* and *д.* refer to word *двора* 'house, homestead'.
[19] The town Ladoga gorod (53) includes nearby town Voločok Svanskoj (in Finnish *Suvannon Taipale*) as well.

As the data had still some OCR-errors and expressions that were not names, a data wrangling program OpenRefine was used to do further corrections. For starters, all name forms shorter than four characters were removed as there were no personal names consisting of three or less letters. Furthermore, nouns that were not names were removed. This meant discarding expressions that described person's special feature or profession, such as being a widow ("вдова") or working as a deacon ("діакъ"). For some reason, editors followed inconsistent conventions in capitalizing these non-name nouns.

In addition, some orthographical and morphological harmonization was done on the data. The letter *ы* was cut from the end of bynames, where it denotes plurality: e.g. "Федко, да Сенка Михалевы" (POKV I 1). Similarity of so called soft and hard signs, *ь* and *ъ* caused some problems. The latter one was removed as it has no phonetic value in contemporary Russian and was not used in the original documents (Неволин 1853 : 4, in Appendix 1). The soft sign *ь* was also removed because it was absent in the original documents and it had been used inconsistently by the editors. The letter *ѣ* (yat) is rarely used in personal names but nevertheless, it was changed to *е* (as it is in contemporary Russian) since it was often confused with soft and hard signs (*ь* and *ъ*). Furthermore, the letter *ѳ* (fita) was often erroneously recognized as *о* or *е*. As it is only found in NPK III and only in the beginning of certain names, which all are also written with *Ф* (e.g. "Ѳедко" vs. "Федко"), it was replaced with *Ф*.

The second phase was more time consuming as most of the erroneous orthographies were corrected. Orthographies of all names, which had two or more occurrences after the above-outlined changes and corrections, were systematically evaluated. Names with only one occurrence were not scrutinized, as it would have been too time consuming. We do not describe all of the OCR-errors here that were found, but most of them were caused by the similarity of certain letters, which generated problems for the OCR-program (e.g. *и*/*ий* and *б*/*в*). In these cases, the correct orthography was sought in the census book editions and accordingly, OpenRefine was used to change erroneous forms to correct ones.

After the corrections were made, the number of name types (= name variants) was reduced from 4942 to 2748. The overall number of name tokens dropped as well: from 36,405 to 35,726. Of the name types, more than half (1484) have only one occurrence. The refined and harmonized name data is published as supplementary material in open-access repository Zenodo (Raunamaa, Kanner 2021).

3.2. The number of names in pogosts and the division of the name data into areas

The number of name tokens in each pogost varies immensely. Pogost Ižorskoj (51), for example, has almost 2000 name tokens whereas Kositskoj (10) has only 24. Such a significant difference in the number of names would obviously affect the distance measurements and clustering results. This mainly involves the metrics used for computing similarities between pogosts: Jaccard index especially, but also Euclidean to some degree, assigns more similarity to vectors or sets of roughly equal size (see the next section for more details about the clustering methods). This was also shown by a preliminary experiment: visualizing

the pairwise similarity matrices in two-dimensional charts using MDS (multi-dimensional scaling) corroborated that pogosts with low numbers of names were outliers (with both metrics) and, accordingly, shared little similarity with the rest. In order to have the name data divided into quantitatively more balanced groups, smaller pogosts were merged into larger groups.

The new division of areas was based on administrative factors, rather than linguistic or cultural. The merging was done by combining pogosts belonging to the same uyezds ('counties'), the administrative level above pogost, into groups of around 1000—2000 name tokens. Within the uyezds, adjacent pogosts were placed together. The only exceptions were Ilomanskoj (59) and Solomjanskoj (60) in the northern part of the study area. These two are joined although they are separated by the pogost of Serdovol'skoj (58). As a result, names were now distributed into 24 areas, depicted in Map 3. The map also displays the total name token frequencies across areas: dark areas have the most whereas lighter ones have less name tokens per area. Appendix 1 shows how the pogosts are distributed into areas.

Map 3. **Pogosts merged into 24 areas. Areas are numbered and explained in Appendix 1. The dark areas are the most densely filled with names per area whereas the lighter ones are less.**

3.3. Clustering methods

One of the main aims of this article is to see how medieval personal name usage can be studied through clustering algorithm. As far as we are aware, this has not been done before. Thus, the process has been much about comparing the results obtained using different parameters for our clustering workflow.

The data used for calculations is described in section 2. As we did not go through the orthography of those names systematically that had only one occurrence, we have excluded them from the clustering process. Thus, the total number of name tokens included is 34,242.

In previous studies, different distance metrics have been applied to different tasks when studying different linguistic phenomena. For example, many vector space models of word distributions (cf. e.g. Turney, Pantel 2010) rely on Euclidean or cosine distances, while Jaccard index is often used to compare similarity of word sets. Relevant to our study, Honkola, Santaharju, Syrjänen and Pajusalu (2019) used Jaccard index (= similarity coefficient) to evaluate the independence of the studied lexical features before distance based clustering. L1-based distance metrics (such as Manhattan distance) are rarer, as are metrics comparing probability distributions (such as Kullback-Leibler divergence).

We ended up having two different distance measuring metrices: Jaccard index and Euclidean distance. Both have their own peculiarities that must be taken into consideration while analysing the results. Jaccard index is a measurement of set similarity, where the two sets are compared by calculating the ratio between their intersection and union. We applied this measurement to compare the similarity of name types between individual areas. Since name distribution behaves quantitatively in many regards in a similar fashion to other word frequency distributions, it is not surprising to find a deep power curve in the name distributions. This means that in each area there is a small number of highly frequent names, relatively modest number of names from middle frequency band and a long tail of rare names. The Jaccard metric loses the information about this structure and treats each name type equally. This means it will relatively assign more weight to the low-frequency name as a group due to the simple fact that they far outnumber the more common names.

Euclidean distance, in turn, is based on frequencies (in this paper, we consistently apply the designation Euclidean distance to indicate Euclidean distance between L2-normalized unit vectors).[20] In Euclidean distance, the frequencies of each name type in each pogost are treated as vectors and the difference between two entries (areas) is measured as the difference between those vectors. Mathematically, this difference is equal to the distance between two points in a multidimensional coordinate system. Euclidean distance in normal vector space is computed as the square root of the sum of squared differences over the dimensions of the space. It is important to note, that due to this property, the Euclidean distance amplifies the weights of dimensions with big differences.

In conclusion, these two metrics weigh frequency bands of the name distributions in dramatically different ways. As Euclidean distance skews towards the high frequency band, it is suitable for inspecting the most common names (i.e., Russian names), while Jaccard gives relatively more weight to the frequency bands where the rarer names reside (i.e., Finnic names). The outcomes are presented in the form of pairwise distance matrices (Appendix 2a for Jaccard and Appendix 2b for Euclidean), where each area is compared for similarity against each other area.

[20] Instead of more commonly used cosine, we used Euclidean distance of L2-normalized vectors. They are not mathematically equivalent, but they do have rank-preserving relation. This means that the order of most similar to most dissimilar areas is not changed regardless of whether the calculation is done using cosine or Euclidean distance. The choice between the two might have a light effect on clustering scores, however, as the actual distances are not equal and the actual distances are used in clustering. Cosine slightly condenses the distances in near and far reaches of the scale compared to Euclidean. The Euclidean then results in somewhat more conservative scores in the extremes, which we saw as a good thing. In addition, the impact of each dimension on the overall distance is simpler to calculate in Euclidean than it is in cosine.

Clustering process was performed by Hierarchical Agglomerative Clustering (HAC) (Kaufmann, Rousseeuw 1990) algorithm from the Python machine learning library scikit-learn. It proceeds by taking the pairwise distance matrix (Appendix 2a or 2b) as an input and by coupling vectors to their closest neighbours. If the closest neighbour is a previously formed group, the exact procedure how the distance to that group is calculated is given as parameters to the algorithm. These parameters were selected to fit the slightly different nature of the two pairwise distance matrices: W a r d ' s for Euclidean matrix and a v e r a g e d for Jaccard. While not exactly mathematically equivalent, they both seek a kind of centre point for comparison when linking clusters together.

Hierarchical Agglomerative Clustering process has a known effect of sometimes introducing artificial cluster borders, especially when it is applied to cases where the studied phenomena show more gradual than discrete differences (Hyvönen, Leino, Salmenkivi 2007 : 283—285). Linguistic variation across geographical areas is usually this kind of gradual phenomenon. This feature of HAC is taken into account and discussed when we have felt it might have affected the outcome.

To avoid unnecessary repetition, expressions Jaccard procedure and Euclidean procedure are used from now on to refer to workflows, in which the distances are measured by Jaccard index or Euclidean distance. The linkage method (applied in hierarchical clustering) is either average or the Ward's method respectively.

4. Results: an overview of the personal names

In the following section, an overview of the personal names attested in the study material is given. First, we present the most common names, and then the names are shortly analysed in the light of semantics and morphology. Next, Finnic names and personal name suffixes attested in the data are introduced and last, a map with Baltic-Finnic ethnonyms is presented.

4.1. Most common personal names within the census book of Vodskaja pjatina

Table 1 depicts the 20 most common personal names within the study material. It is easy to notice that name forms *Ивашко* and *Ивашков*, which are derived from the name *Ivan*, are the most popular ones in each area. Similarly, it is notable that all the names are connected to Christianity (Петровский 1980; Суперанская 2010). The number of 20 most common personal names is 9398, which means that they cover about 26 percent of all names.

In the scope of this article, we cannot deal with the morphology or semantics of the names in depth, but some remarks are necessary. Starting with the latter, it can be stated that most of the names have been semantically obscure for their users, as most of them originate from Greek or Hebrew.

As seen in Appendix 1, most of the pogosts have saints mentioned in their names, e.g. Saint Nicholas in the name of *Nikol'skoj Sujdovskoj* (28) pogost. The saint name used in the name of the pogost was usually derived from the name of local church or monastery. It does not seem, however, that the name of a local saint would have affected people's way of giving names. For example, in Sujdovskoj pogost, the name *Nikola* and its variants (e.g. *Микулка*, *Никулин*, *Никутин*) were no more popular than in adjacent parishes.

Table 1

The 20 most popular personal names divided into areas (OA = Overall number of names)

Name	OA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Ивашко	1704	74	58	75	59	77	85	86	68	56	78	69	62	48	95	81	56	114	66	79	54	45	88	76	55
Ивашков	957	46	22	26	40	34	60	30	42	48	46	39	41	25	39	25	28	51	30	35	28	32	66	75	49
Федко	757	22	22	21	27	39	20	28	27	21	38	32	26	28	44	27	28	61	35	32	25	25	42	38	49
Васко	657	16	15	21	13	25	41	35	21	18	27	18	30	8	29	24	24	59	29	32	35	28	41	42	26
Гридка	532	12	13	28	11	20	29	29	15	14	18	19	9	13	20	29	12	30	13	22	15	22	60	55	24
Микитка	405	18	9	13	12	11	16	16	8	18	26	19	23	19	20	12	16	26	18	11	8	12	24	27	23
Васков	411	13	10	10	2	21	14	15	13	13	7	15	26	4	20	17	26	34	20	17	21	15	36	25	17
Федков	396	26	8	10	20	15	27	13	7	12	22	14	18	12	10	7	20	31	14	18	16	10	28	23	15
Андрейко	369	20	15	20	20	20	14	7	27	9	17	9	23	18	17	11	8	26	12	21	10	7	16	13	9
Куземка	360	11	12	10	12	10	16	14	16	20	12	15	14	11	27	9	19	27	16	19	14	11	18	19	8
Михал	326	14	10	12	15	11	13	15	23	6	6	19	19	8	10	13	18	15	19	19	17	10	10	13	11
Гридин	318	17	6	11	11	14	12	18	10	10	19	16	11	15	19	20	2	14	6	14	8	8	22	19	16
Палка	311	9	10	10	13	2	10	19	17	16	13	9	21	12	16	10	10	16	15	11	14	11	20	12	15
Степанко	311	10	10	14	12	12	14	17	9	8	17	7	13	7	9	13	5	24	9	12	15	11	24	25	14
Максимко	297	8	9	10	14	10	20	10	16	7	11	12	14	7	12	8	7	14	9	19	10	13	30	13	14
Якуш	294	9	8	17	19	16	8	5	23	11	14	14	9	4	6	7	5	5	14	13	5	7	36	17	22
Сенка	263	6	13	16	7	16	6	9	16	7	8	15	8	4	20	23	4	10	13	4	9	7	15	19	8
Кузмин	251	9	9	7	5	8	12	3	24	13	12	1	8	6	8	6	14	26	18	18	11	6	13	9	5
Тимошка	243	6	2	10	4	11	5	17	9	10	14	12	9	4	6	5	12	19	5	4	13	10	26	20	10
Сенкин	236	8	6	7	9	11	7	5	13	4	12	9	8	6	18	13	4	15	12	9	12	11	12	19	6

The habits of scribes played important roles and affected the morphology of personal names attested in the census book. As mentioned in section 2.1, scribes followed traditions according to which they adapted local variants of names into written forms. However, the adaptation was not always congruent. For example, in pogost Dudorovskoj (50), one can find eight occurrences of the name *Алексейко* and three occurrences of its variant *Олексейко*. These kinds of minor obscurities could be caused just because of medieval scribes or 19th century editors had made mistakes. On the other hand, there are numerous examples of intentional variation as well: like in the case of names *Васко* and *Васюк*, both of which derive from the name *Vasili*.

Almost all the names in Table 1 are diminutives. This emphasizes how the state officials named ordinary people in the late 15th century. Part of the recording tradition was that the social background of a person was underlined by the way their name was written (Суперанская 2010 : 14—15). A good example of this is the most popular name *Ивашко*, which is the diminutive form of *Иван* (*Ivan*). The name *Иван* is only seldomly used to name peasants. On the other hand, *Иван* is often mentioned within the census book but referring to landowners or to the Grand Duke Ivan III. Similarly, priests are not called by diminutives but always by their "proper names", e.g. *попъ Сава* (POKV II 33) and *попъ Кузьма* (POKV II 121).[21]

4.2. Finnic names and personal name suffixes in the names

The next section will deal with the Finnic names and features that are found in the material. Names searched were chosen on the basis of studies that have concerned the Finnic personal name elements found in census books regarding

[21] *Попъ* 'priest'.

the areas of Onega (e.g. Карлова 2014; Соболев 2017) and Votic fifth (e.g. Новожилов 2004; Чибисов 2019). Of all the possible Finnic names, we approved those that have an extensive distribution in former and modern Finnic areas, i.e. Estonia, Finland and Northwest Russia (cf. Stoebke 1964; Raunamaa 2020). Thus, the following 20 Finnic personal name elements were chosen: *Auvo, Hima, Hyvä, Iha, Ika, Ilma, Ilo, Kaipa, Kauka, Kyllä, Lempa, Mieli, Monta, Nousia, Päivä, Toivo, Uska, Valta, Vihta* and *Vilja*. Names were searched by reading through the census book page by page. Then, the different kinds of letter combinations, which could originate from Finnic personal name elements, were searched from the processed name data. Altogether we found 327 names that were considered Finnic. Table 2 presents these names and their overall numbers.

Table 2

The Finnic personal names attested in the material and their overall numbers

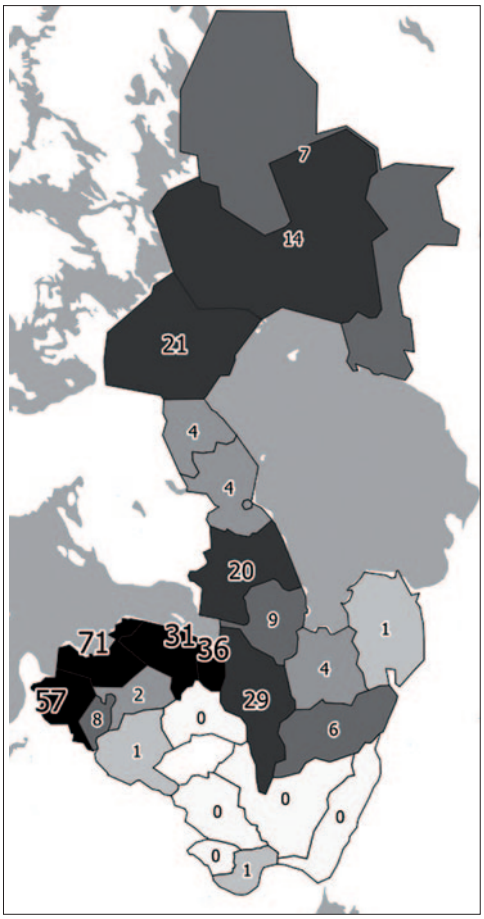
Name	OA									
Аввус	1	*Гавка*	1	*Игамов*	2	*Илойтеев*	1	*Лембит*	5	*Тойвал*	4
Аввусов	2	*Гимуй*	1	*Игамуев*	5	*Имонтеев*	4	*Лембитко*	2	*Тойвалов*	3
Аввушов	1	*Игавелев*	3	*Игамуй*	2	*Имоттеев*	1	*Лембитов*	17	*Тойвас*	3
Авусов	1	*Игавин*	1	*Игандов*	1	*Ичяпов*	3	*Лембіев*	7	*Тойвасов*	1
Валлитов	1	*Игайло*	2	*Игандуев*	10	*Кавгуев*	1	*Лембуев*	3	*Тойват*	4
Вигутов	1	*Игакин*	2	*Игандуй*	2	*Кавгуй*	1	*Мандин*	1	*Тойватов*	4
Виллуев	5	*Игала*	5	*Иганов*	1	*Каибин*	3	*Мандухин*	1	*Тойвол*	1
Виллуй	1	*Игалин*	8	*Игантов*	2	*Каибуев*	1	*Мелит*	1	*Тойвот*	3
Виляк	6	*Игалка*	3	*Игатуев*	1	*Какивалдов*	1	*Мелитов*	1	*Тойвуев*	6
Вилякен	1	*Игалкин*	2	*Игачей*	1	*Клюлетев*	1	*Новзеев*	2	*Тойвуй*	2
Вилякин	7	*Игалко*	23	*Игола*	10	*Клюлят*	1	*Новзей*	1	*Тойвутов*	4
Виляшов	1	*Игалков*	1	*Иголин*	7	*Ковко*	1	*Новзейко*	2	*Ускал*	5
Вихтеев	1	*Игалов*	5	*Иголка*	9	*Кюллетин*	1	*Новзин*	1	*Ускалин*	1
Вихтимко	1	*Игалтас*	1	*Икамел*	1	*Кюллята*	1	*Новзуев*	6	*Ускалко*	1
Вихтуев	5	*Игалтасов*	1	*Илменебуев*	1	*Кюллятев*	1	*Новзуй*	1	*Ускалов*	22
Вихтуй	6	*Игамас*	11	*Илмов*	1	*Лембей*	2	*Новзуйко*	1	*Уской*	1
Вихтуйко	1	*Игамасов*	2	*Илмуев*	3	*Лембейко*	2	*Пявзей*	1	*Чуллуев*	1
Гавгалцов	1	*Игамелев*	3	*Илой*	4	*Лембик*	1				

The distribution of the Finnic names is presented in Map 4. It shows the number of names in each area. The darker the shade, the more occurrences there are in the area. The map shows that the Finnic names concentrate in the northern and western parts of Vodskaja pjatina. The number of Finnic names is especially high in the two western areas consisting of the pogosts Kargal'skoj (22) and Toldožskoj v Čudi (37) and of the town Jama (35) (with its surroundings). This distribution is in line with the earlier studies regarding the spread of Finnic personal names (Новожилов 2004; Чибисов 2019). Nevertheless, Finnic names are only a minor part of all the names in the areas mentioned.

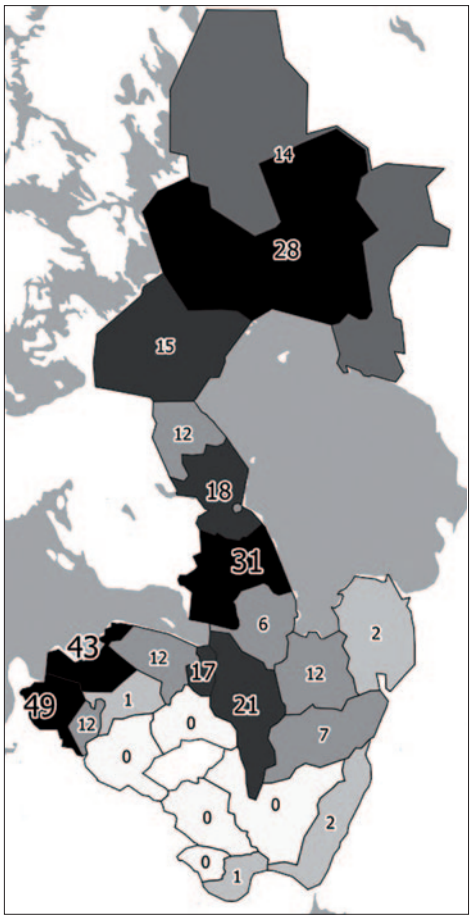
Finnic features can also be observed in the morphology of names. We detected that the distribution of names ending with *-уй* (*-uj*) (or in the case of patronym *-уев* (*-uev*)) is in accordance with the spread of Finnic personal names: Both concentrate in the northern part of Vodskaja pjatina and especially in its westernmost regions.[22] Clearly, the ending is comparable to the old Finnic

[22] Names containing the diphthong *-уй* (*-uj*) were reviewed manually, since some of them are probably not connected to the Finnic personal name suffix *-oi*, e.g. the name *Мануйлик*. Patronyms with the ending *-уев* were similarly examined. The rule of thumb was that if the name was attested as a given name as well with the ending *-уй*, it was then considered having Finnic ending: e.g. *Ишуев* > *Ишуй*.

personal name suffix *-oi*, which has been attested in many ancient personal names and place names (Муллонен 2008 : 185). Map 5 depicts the number of names with the Finnic personal suffixes *-уй* (*-uj*) and *-уев* (*-uev*) within the areas. The darker the shade, the more occurrences there are in the area. Altogether, there are 305 names containing this suffix, of which the most popular are: *Тимуев* (12 occurrences), *Игандуев* (10) and *Миккуй* (8). Of these, 77 names are also Finnic personal names, e.g. *Игандуев*, *Вихтуй*, *Новзуев* and *Тойвуй*.



Map 4. **The number of Finnic personal names within areas.**



Map 5. **The number of names with the Finnic personal name suffix *-uj* (or *-uev*) within areas.**

4.3. Finnic ethnonyms

The late 15th century distribution of Finnic people can be further investigated by placing the ethnonyms referring to Finnic tribes (*Ижерянин*, *Ингерев*, *Корелянин*, *Корелании*, *Чюдин* and *Вошко*) on the map (Map 6) according to the villages where they have been attested in the census book.[23]

[23] Some of the villages mentioned in the census book could not be placed accurately on the map. However, the census book often gives hints as to the whereabouts of villages. For example, it is possible to identify the neighbouring village or a natural landmark near and thus, locate the village relatively accurately on the map. If this has not been possible, a name is placed on the middle of the pogost.

Letter *C* refers to Chuds, *I* to Izhorians, *K* to Karelians and *V* to Votes. There are altogether 51 names in Map 6.

Names referring to Chuds (*Чюдин*) are problematic since their meaning is not certain. In the oldest Russian chronicles (e.g. Chronicle of Novgorod), expression *Chud* clearly refers to people who lived in the contemporary area of Estonia (Grünthal 1997 : 152). However, it has been suggested that the *Chuds* of the census book were actually Votes (e.g. Новожилов 2004; Чибисов 2019).[24] On the other hand, there is another ethnonym within the census book, *Вошко*, that probably refers to Votes as well (Новожилов 2004; Чибисов 2019). Even more confusing is that, according to the census book of Šelonskaja pjatina from 1498, there were persons named as *Вошко* and as *Чюдин* living in the

Map 6. **The distribution of Finnic ethnonyms attested in the study material. C = Chuds, K = Karelians, I = Izhorians and V = Votes (Voško).**

[24] Much of the discussion related to the connection between Chuds and Votes has been presented by Grünthal (1997 : 159).

town of Ivangorod (located near the western border of Vodskaja) (MIN 227—232). This indicates that these two expressions have had a different meaning.

It is also unclear if expressions such as *Ижерянин* or *Корелянин* always refer to a person's ethnic origin. As Appendix 1 shows, all these expressions are also used in the names of pogosts (e.g. *Toldožskoj v Čudi*, 37 and *Nikol'skoj Ižorskoj*, 51) and in the name of an uyezd (*Korel'skoj ujezd*). Accordingly, the above-mentioned ethnonyms could refer to a person's origins on a regional level rather than his ethnicity. However, contrary to this idea, *Chud*-names are also attested within the pogost Toldožskoj v Čudi. In addition, it is likely that many of the inhabitants of these areas have been Finnic, as it will be proposed in section 6.2. This means that a settler from the pogost Ižorskoj has likely been someone who can be considered both ethnically and linguistically an Izhorian.

Despite the issues, we can be fairly sure that the ethnonyms depicted in Map 6 refer to Finnic people. As the scarcity of names indicates, ethnonyms were used only in special occasions to underline a person's difference compared to his neighbours. This means that ethnonyms usually occur in places where their bearers were a minority (cf. Рябинин 1997 : 43—44). Thus, the ethnonyms located in the southern part of the study area are not indications of these areas being ethnically or linguistically Finnic, but they rather indicate that there has been an internal migration from the northern parts of Vodskaja pjatina to the southern regions. It is also interesting that none of southern persons with a Finnic ethnonym as a byname has a Finnic personal name (~ first name).

The main concentration of ethnonyms referring to Finnic people is in the western parts of Vodskaja. This is in line with the distribution of Finnic personal names and personal name features as depicted in Map 4 and 5. The distribution of different ethnonyms is further analysed in section 6.2.2 while discussing the connection between them and the clustering results.

5. Clustering results

This section focuses on the results of clustering. Since the results vary depending on the metrics used (Jaccard or Euclidean), this section is divided into two parts. First, we present the outcomes of Jaccard procedure and subsequently, results of Euclidean are introduced in a similar manner. Both outcomes are presented in the form of dendrograms and maps. Dendrogram visualizes how the algorithm proceeds by combining areas into clusters. Maps (Appendix 3a and 3b) depict the outcomes at given points. To limit the number of visualisations, odd counts of clusters and clusters larger than eight were excluded from the maps. Thus, the numbers of clusters (K) presented in the maps are two, four, six and eight.

5.1. Results based on the Jaccard procedure

Figure 2 shows how the pairwise distances obtained with Jaccard index are clustered hierarchically. It is noteworthy that the variation across pairwise distances obtained by Jaccard (Appendix 2a) is quite small: from 0.48 to 0.7. This means that the areas share considerable number of their names as there are no distances above 0.7 and, second, represent mostly quite unique blends of names (because no two areas have distance shorter than 0.48). Further, because of the small margins in the distances, the analysis is less robust to data errors such as OCR mistakes.

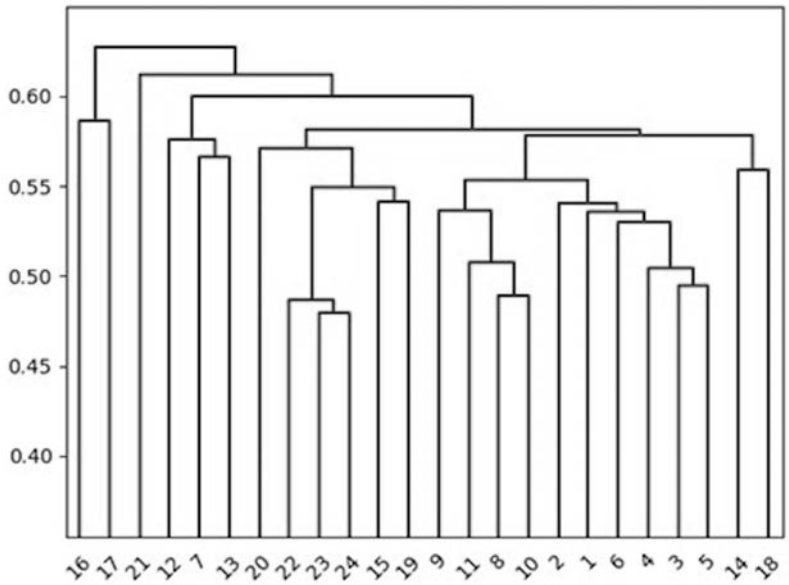


Figure 2. **Results of hierarchical clustering presented in the dendrogram. The outcome is based on the Jaccard metric and average linkage method. Numbers at the bottom of the figure refer to areas. The dendrogram is read bottom-up as it visualizes the process by which the algorithm proceeds by merging areas to larger clusters. The position of a merging of two branches in the y-axis shows the distance between these branches.**

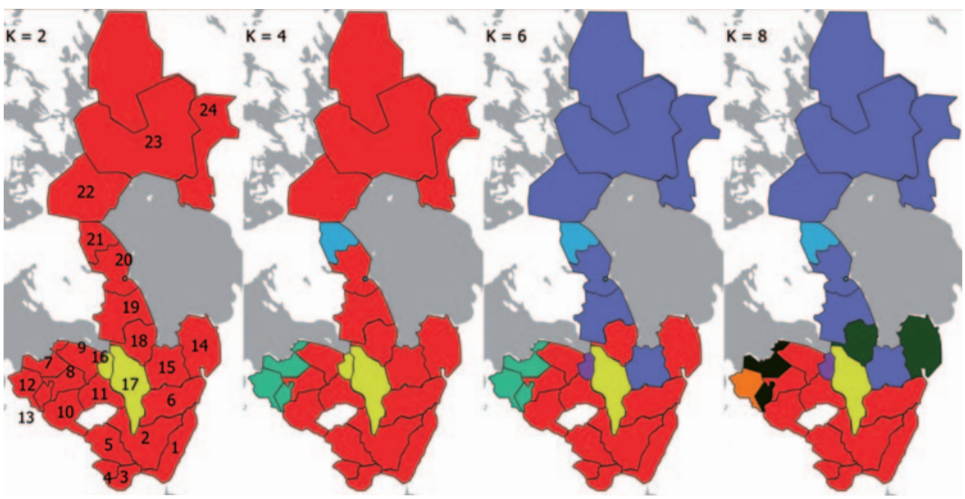


Figure 3. **Results of hierarchical clustering visualized in the maps. Clusters obtained by using the Jaccard metric and average distance. The number of clusters (K) from left to right: 2, 4, 6 and 8**

There are, however, some clusters that evidently share similar pools of name types. The group consisting of areas 22, 23 and 24 is one of them. Furthermore, areas 8, 10 and 11 in the western parts of Vodskaja pjatina together with areas 3, 4 and 5 form groups that are close-knit.

It is noteworthy that the last cluster group to merge with the others consists of two areas, 16 and 17, (the yellow cluster in Figure 3 when K = 2) that

are not very close to each other as their pairwise distance is 0.59. Basically, areas 16 and 17, while both outliers, are still closer to each other than to others.

The second to last cluster group to merge with the others is area 21 (blue in Figure 3 when K = 4 or larger) in the Karelian Isthmus, which consists of pogost Gorodenskoj (54) and towns Korel'skoj and Voločok Svanskoj (53). This outcome is interesting, since within the Euclidean results, area 21 forms a rather close-knit group with its neighbours (see Figure 4). However, area 21 is a special case because of the names used among the dwellers of the two towns mentioned above. A glance at their nomenclature shows that there are many names, both first names and bynames, which are not found elsewhere. In the town Korel'skoj, such names include for example *Искра, Дуравын*, *Берденев*, *Батков*, *Ластка* and *Обакумов*. Although, many of these names are not part of the clustering process as they occur only once, they indicate that nomenclature is different compared to neighbouring areas. A similar approach can be applied to the areas 14 and 18 (green in Figure 3 when K = 8) that form a cluster group that is far from others. The former of these contains the town of Ladoga whereas the latter has the town and fortress of Orešek. Both of these towns have many rare names and name forms.

In addition, it seems that the number of name types within each area affects the clustering results. The last two cluster groups to merge consist of areas 16 and 17 and of area 21. These three areas contain the least amount of name types (when names with one occurrence are excluded): first one has 280 name types and the second 333 and the third 332. The difference in set sizes (= number of name types within an area) contributes directly to the distance. The two sets can overlap maximally only to the extent that the smaller set covers the larger set when it resides entirely within the larger set.

In the last map in Figure 3, it is noteworthy that basically all the areas containing only few or no Finnic names and name forms at all (see Map 4 and 5) are clustered together (dark green). The only exception is area 9 on the coast of the Gulf of Finland. Most of the formed cluster groups are geographically cohesive, i.e. they are the result of combining areas adjacent to each other. Irregularities of this aspect are discussed at more length in Section 6.1.

5.2. Results based on the Euclidean procedure

Figure 4 visualizes a hierarchical clustering process based on Euclidean distance and Ward's linkage method. The range of distances is larger (from 0.36 to 0.7) (Appendix 2b) compared to those obtained with the Jaccard metric (Appendix 2a). This is to be expected, at least to some degree, because the high-frequency end of names is likely to be less noisy while at the same time likely to differentiate more between dissimilar areas.

As observed in the dendrogram obtained with the Jaccard procedure (Figure 2), the most close-knit cluster group is found in the northern region (yellow cluster in Figure 5, K = 2); especially the areas 22 and 23 have very similar naming conventions. In addition, similar to Jaccard, the areas 16 and 17 form a cluster group (dark green cluster) that is far stretched from others. This is to be expected as the nomenclature in these two pogosts has many unique features noticeable to the naked eye. For instance, the name *Матюк* is attested in these areas 11 and 14 times, respectively, whereas the average number for it is 2.7 in other areas.

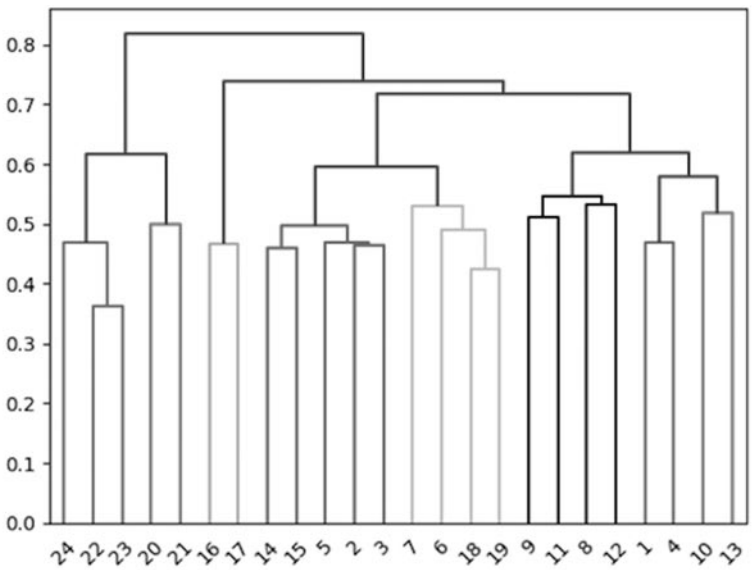


Figure 4. **Dendrogram structure: visualization of hierarchical clustering based on Euclidean metric and Ward's distance measuring. Numbers at the bottom of the figure refer to areas. The dendrogram is read bottom-up as it visualizes the process by which the algorithm proceeds by merging areas to larger clusters. The position of a merging of two branches in the y-axis shows the distance between these branches.**

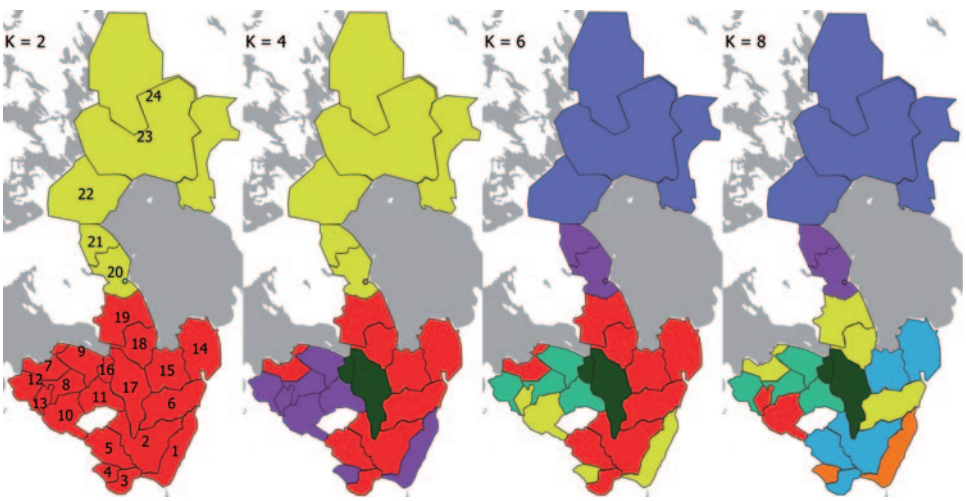


Figure 5. **Results of hierarchical clustering visualized in the maps. Clusters obtained by using Euclidean metric and Ward's distance. Cluster sizes from left to right: 2, 4, 6 and 8.**

There are another two close-knit cluster groups in the middle of the study area: one that consists of areas 18 and 19 and another one of areas 14 and 15. In the south, the two closest groups are formed from areas 2, 3 and 5 together with areas 1 and 4.

In general, the results depicted in Figure 4 and 5 indicate that cluster groups are geographically cohesive. This means that areas form clusters with their neighbours or with those that stand close by. There are, however, some excep-

tions, of which the most notable is the one consisting of areas 6, 7, 18 and 19, depicted as the yellow group on the fourth map of Figure 5. North from River Neva, areas 18 and 19 are the closest group and also adjacent to each other whereas area 6 is in the east and area 7 in the west. The case of area 7, which consists of the pogost Kargal'skoj (22) and the town Kopor'ja (21), is especially intriguing, as it is the location where the Finnic names and features occur the most whereas its counterparts within the cluster group have them significantly less (see Map 4 and 5).

6. Discussion

The following part discusses the results presented in the previous section. This is done in two parts. First (section 6.1), the success of the applied clustering procedures, Jaccard and Euclidean, is discussed from a broader perspective: how reliable they are mathematically, are the cluster geographically cohesive and how the results of two different procedures compare to each other. In the second section (6.2), we focus on the Finnic aspects of the clustering results at the more detailed level as it is one of the best ways to investigate if the outcomes are in line with other cultural-historical information.

In view of the above, it must be emphasized that there are not many historical resources that could be used to study Vodskaja pjatina's cultural-historical past. Basically, the best available source is the census book. This entails that most of the studies dealing with the history of Vodskaja pjatina are largely based on the same source as well. In addition, it should be remembered that analysing over 500 years old data is complicated. Thus, the depictions of Vodskaja pjatina's past differ among the researchers. There are, however, some generally accepted opinions that the comparative analysis can be based on.

6.1. Evaluating the success of the clustering methods

Based on the observations made in Section 5, the results obtained with Euclidean distance and Ward's linkage method are statistically more reliable and accurate than those of Jaccard (and average linkage). Unlike in Euclidean distance, where the pairwise distances between the areas are generally larger and the clusters seem to merge in intuitive intervals, Jaccard distances have generally smaller margins and the clusters are equally merged in tight margins of distance. This means that the results are not very robust to random variation: even few name types missing by chance or being erroneously allocated because of OCR have a greater probability of affecting the whole outcome.

The reliability of clustering results can be evaluated also based on the geographical cohesiveness of the produced clusters. Studies focusing on regional variation in the use of surnames (e.g. Shi, Li, Wang, Chen, Yuan, Stanley 2019; Sousa, Ginzo Villamayor 2021) indicate that the name-based clusters are usually formed of geographically adjacent areas. In the case of this study, both metrics produce some regionally fragmented clusters. Obviously, this does not mean that clustering would not have been successful, but rather it stresses that the results must be evaluated cautiously.

On the other hand, both clustering procedures produce fewer incohesive groups when the number of clusters increases. This means that the more closely related areas are also geographically related. Furthermore, all

the most close-knit cluster groups are formed of adjacent areas. Thus, it can be concluded that the results of clustering seem to be most valuable when looking for the closest clusters and areas from dendrograms and pairwise distance matrices. The clusters formed higher up in the hierarchy can in many cases be seen as artifacts of the hierarchical clustering algorithm.

Besides the mathematical and geographical aspects, the reliability of clustering results can be evaluated by comparing the outcomes of Jaccard and Euclidean procedures to each other. They are mathematically different, but their purpose is the same: to find differences and similarities in the name usage. However, they start from very different basic ideas about how the name usage should be modelled. To some extent, it can be expected that within two areas with similar name types (Jaccard), the most popular names would be alike as well (Euclidean).

A look at the five most close-knit clusters of Jaccard (Figure 2) and Euclidean (Figure 4) separately indicates that there are two groups common for both metrics. The first of these is formed from the northern areas 22, 23 and 24. Both results imply that these areas have shared similar names and naming conventions. Another close-knit group, common to both metrics, is found in the south, where areas 3 and 5 share similar nomenclature. Furthermore, the case of areas 16 and 17 is interesting as well. Although they are rather far stretched from each other according to the outcomes obtained with the Jaccard index, their common feature is being distanced from other areas and groups in the results of both metrics.

6.2. The clustering results and the Finnic past of the study area

Based on the results presented in section 5, it can be claimed that the naming conventions of those areas that are located within the supposed Finnic territory (i.e., areas with Finnic nomenclature, see Map 4 and 5) are not particularly similar and accordingly, there is no singular cluster group that would cover the whole Finnic region.

On the other hand, taking a closer look at the maps in Figure 3 (obtained through Jaccard), and especially at the fourth map where the number of clusters is eight, reveals that the southern areas form a group that contains approximately all the areas without significant amount of Finnic nomenclature. The only exception is area 9 in Ingria but it is not very closely connected to the other areas within the cluster group as seen in Figure 2. Based on this, it can be suggested that the pools of names were different outside the supposed Finnic region.

The above-presented suggestion that the name types of southern areas differed from those of central and northern parts of the study area could be at least partly explained on the basis of historical reasons. First of all, the southern areas had already been Slavinized before the end of first millennium (Рябинин 1997 : 3—7) whereas in the central and northern parts of Vodskaja pjatina, the level of Slavic influence varied by region. Central and northern areas (especially those south of the Karelian Isthmus) were culturally heterogeneous in other ways as well. For instance, archaeologists have suggested that, on the southern coast of the Gulf of Finland, medieval graves and/or artifacts in them have features that have been considered as Estonian, Izhorian, Karelian, Slavic and Votic (Kriiska, Tvauri 2007; Конькова 2008; Сорокин 2008).

However, it can be claimed that the described division into northern and southern naming conventions is not only based on the distribution of different ethnicities or linguistic groups but also on geographical differences. The landscape south of the Izhora uplands is characterized by many wetlands, which spread to the river Volkhov in the east. Geographical conditions are with all probability the main reason why the population density is so low in that region (see section 3.2). Wetlands formed natural borders between different cultural spheres and accordingly must have had an impact on the spread of naming conventions as well.

It is also good to notice the referring to those areas that contain Finnic personal name elements as Finnic region is questionable since the number of Finnic names and names with the suffix *-uj* is relatively small compared to other nomenclature. For instance, in Kargal'skoj pogost (22), the combined number of Finnic personal names and names with the Finnic personal name suffix *-uj* (+ patronyms *-uev*) is 119, whereas the overall number of names is 1529.[25] Thus, approximately eight percent of the names can be considered Finnic. It is evident that the Finnic names and name forms is so low that they have only limited impact on the outcomes of clustering and especially on those based on the Euclidean metric (see section 3.3). The results obtained with the Jaccard metric do not depict the spread of Finnic nomenclature with complete truthfulness either. Since the data used for clustering does not include names which have only one occurrence, many of the Finnic name variants have been excluded, as they are only mentioned once.

It is, nevertheless, likely that, especially within the northern and western regions, the Finnic population has been higher than what the personal name data suggests. This claim is partly based on the substantial number of Finnic place names in the census book (cf. Чибисов 2019). Furthermore, two historical events provide additional evidence for the claim. First, in the 1440s, the Teutonic Order transferred approximately 3000 inhabitants of Western Ingria to Courland. This Votic speaking group, which was called *Kreevins* by local Latvians, continued to exist near the town Bauska in Southern Latvia until the 19th century (Ränk 1960 : 10—16). Second, during the 17th century, many Orthodox Karelians emigrated from the Karelian Isthmus to the Tver region to escape continuous wars and Swedish occupation. Descendants of these people have preserved their Karelian identity to this day (Saloheimo 2010).

It is difficult to evaluate the impact the forced emigration of 1440s had on the cultural and ethnic situation of Western Ingria. Nevertheless, this event proves that, at least in the beginning of 15th century, the area had a substantial Finnic population. Furthermore, it is possible that the heterogeneity of the clustering results regarding the Western Ingria depicts the cultural and social turmoil that was caused by the settlement vacuum of 1440s. The case of Karelian emigration, however, is a clear evidence of the Karelian Isthmus being culturally and linguistically mainly Finnic at the end of 15th century when the census book was compiled.

In view of the above, it seems likely that most of the late 15th century Finnic inhabitants in Vodskaja pjatina did not have Finnic names or that these names were not used for administrative purposes. In other words, the majority of Finnic people had typical Russian Christian given names. This aspect is also addressed by Novožilov (Новожилов 2004).

[25] Some names are included twice since there are a couple of Finnic personal names containing the Finnic personal name suffix *-uj*.

An important question for the present study is whether the names (both Christian and Finnic names included) and their frequencies within the areas of supposed Finnic presence differ from the other parts of Vodskaja pjatina. As noted earlier, there is no singular cluster group that would depict the spread of Finnic nomenclature. On the other hand, at the more fine-grained level, there are some cluster groups that could be connected to specific Finnic tribes. These are discussed next.

6.2.1. Karelians, Karelian Isthmus and clusters

The correlation between a group of areas with similar naming conventions and a Finnic tribe is the most significant in the case of Korela uyezd ('county') (areas 20—24). Especially the Euclidean results indicate that these areas formed a close-knit group in light of personal names. This is in line with the historical assumptions as well. Karelian culture must have still been solid in that region, as it stayed relatively independent until the end of the 13th century when according to the Chronicle of Novgorod it was conquered (Forbes, Michell 1914; Lang 2020 : 224). In addition, descendants of those immigrants and refugees that migrated from the Karelian Isthmus to the Tver region in the 17th century have retained their ethnicity and language to this day (Saloheimo 2010).

The area of Karelian naming culture, the uyezd of Korela, is further split into two parts when the number of clusters increases (Euclidean). The southern part consists of the areas 20 and 21 where, according to the census book, the main tax unit is *obža* (= amount of seeds sown, *обжа*). In the three northern areas, 22, 23 and 24, the lifestyle was probably based more on hunting, fishing or slash-and-burn cultivation since the local taxation was not based on *obža* but on unit called *luk* (*лукъ*), which referred to an adult man capable of paying taxes (Ronimus 1906 : 64).[26]

It is intriguing that the region south of the uyezd of Korela, namely areas 18 and 19, do not have similar naming conventions with adjacent Finnic areas according to the clustering results (especially those of Euclidean metric). Areas 18 and 19 have their closest counterparts more in the south (see e.g. the third map, K = 6, in Figure 5). In addition, in area 18 at the River Neva, the number of Finnic names and features is small compared to its neighbours (see Map 4 and 5).

Traditionally, the southern Karelian Isthmus (i.e., areas 18 and 19) has been considered being ethnically Izhorian during the first centuries of the second millennium (e.g. Рябинин 1997 : 62—63). This assumption is reasonable, but there are, however, also reasons to believe that the region was ethnically more heterogeneous than previously thought and, as the clustering results suggest, had intricate connections with the southern regions.

First of all, the southern Karelian Isthmus is almost completely empty of such late Iron Age or medieval archaeological finds that could be connected to permanent settlement (Сорокин 2008). This does not mean necessarily that the region was uninhabited at the turn of the second millennium, but at least it indicates that it has differed culturally from the northern part of the Karelian Isthmus, where late Iron Age settlement was dense (Saksa, Uino, Hiekkanen 2003; Raunamaa 2020 : 127).

[26] Originally *лук* 'bow' probably referred to a man who is capable of shooting a bow.

Furthermore, it is obvious that the Neva River and its surroundings were militarily and commercially important to Novgorod and later the Grand Duchy of Moscow. Accordingly, the castle of Orešek together with the surrounding town at the head of the Neva River was an important place as well. It is likely that many of the local inhabitants had been assigned there. The census book separately names military experts such as gunmen (*пищальники*) and siege equipment engineers (or gatekeepers) (*воротники*). Furthermore, in the surrounding countryside in pogost Spasskoj Gorodenskoj (45), 20 homesteads were inhabited by special persons (nobles, knights etc.) who, in all probability, were transferred to the place for military reasons (see Martin 2007 : 239). As mentioned earlier in section 1.1, after finishing the annexation of Novgorod, the Grand Duke of Moscow replaced the previous landlords of Novgorod by his trustees from Moscow (Kepsu 2015 : 15). Thus, it is likely that these Muscovites were followed by people like military personnel and other trustees.

To conclude, in the light of above, there are reasons to believe that areas 18 and 19 were culturally and ethnically different from their neighbours. Neva River with its surroundings was commercially and militarily important and consequently, its ethnic, linguistic and anthroponymic composure was more heterogenous compared to areas of Korela uyezd (20—24).

6.2.2. Izhorians, Votes and clusters

The question about the connection between the clustering results and the two other Finnic tribes mentioned in the ancient Russian sources, Izhorians and Votes, is more complex. The first of these tribes, the Izhorians, is believed to have inhabited a region spreading from the southern Karelian Isthmus to some tens of kilometres to the south and southwest of the River Neva whereas the other tribe, the Votes, probably resided in the western parts of the study area (Frog, Saarikivi 2015). In addition, a group called *Chuds* could be added to the list as it is mentioned many times in written sources, of which the census book is a good example. Within the census book, the ethnonym *Chud* (*Чюдинъ*) clearly refers to people living in the western parts of Vodskaja pjatina against the border of Estonia and Livonia. It is possible that the *Chuds* of the census book were actually Votes (see Map 6 and section 4.3 for more details).

South from the Neva River lies the pogost Ižorskoj (51 = area 17), which, in all likelihood, has been originally the core area of the Izhorian ethnos (Lang 2020 : 225—226). It is also an area that is only grouped with its western neighbour, the pogost Dudorovskoj (49 = area 16) (by both clustering procedures). In addition, no ethnonyms occur in the pogost Ižorskoj that refer to Izhorians, and its neighbour has only one such (i.e., Izhorians were not considered as outsiders in these areas) (see section 4.3. for more details). In conclusion, the cluster formed from the pogosts Dudorovskoj and Ižorskoj might reflect Izhorian naming conventions.

The question of Votes is a difficult one. It is not easy to indicate any one area that would be in line with historical and linguistical assumptions on the subject, based on the name data and the clustering results. If the ethnonym *Chud* is connected to Votes as suggested in section 4.3., the westernmost part of the study area, namely area 12, can be considered as the core area of this tribe. Both clustering procedures produced results, according to which area 12

contains nomenclature that differs from adjacent areas (see Figure 3 and 5). Furthermore, area 7 (= pogost Kargal'skoj), the eastern neighbour of area 12, is the location where ethnonyms referring to Chuds and Izhorians are colliding (Map 6). Thus, it would be natural to assume that area 7 is the place where the sphere of Votes was limited in the east at the end of 15th century. On the other hand, one must remember that Western Ingria and the supposed territory of Votes experienced much turmoil during the 15th century, for example when the Teutonic Order plundered it and took around 3000 Votes as prisoners to Livonia (Ränk 1960 : 10—16). Accordingly, it is possible that so called Votic naming conventions did not exist anymore.

7. Conclusions

Our aim was to use computational methods for collecting and analysing medieval Russian personal name data. According to our knowledge, this kind of research has not been done before. Our main focus was on learning if computational clustering could be used to determine patterns of personal name usage in late 15th century Russian administrative area called Vodskaja pjatina. Furthermore, the purpose was to gather new information on the history of Northwest Russia and especially the area's Finnic past.

We discovered that modern digital methods, such as OCR-reading and Python based data collection, are useful for gathering and editing ancient personal name data. After the initial corrections and removals, our data contained 35,726 names.

The success of the data collection enabled the successful implementation of distance measures (through Jaccard index and Euclidean distance) and clustering (through Hierarchical Agglomerative Clustering) as well. Both Jaccard and Euclidean procedures produced mostly logical results and geographically cohesive cluster groups, but, in the case of the former, the pairwise distances between areas are so small that the clustering results must be viewed with caution. Furthermore, the results obtained with the Jaccard method were affected by the number of name variants in each area.

The reliability of the clustering outcomes was further analysed in light of other resources and studies related to the study area's cultural, ethnic and linguistic past. No critical ambiguities were found, but it was also noted that sources regarding the history of Northwest Russia are scarce and accordingly, it is impossible to know exactly what happened in the study area during the first centuries of the second millennium. The sheer size of Vodskaja pjatina was another problem. As researchers focused on Finnish and Finnic languages, our knowledge of the study area's southern parts was limited and because of that, we concentrated on the northern area in our analysis.

Overall, the clustering results implied that the study area can be divided into various subgroups according to personal name usage. Furthermore, in most cases, these results are in line with other sources and studies. We discovered differences between northern and southern naming patterns. This is largely due the fact that the southern areas formed large and united clusters with both clustering procedures. It can be claimed that the southern cluster(s) depicts an area where Slavs and Slavic/Russian names were in the majority, whereas in the northern part, many separate cluster groups imply the diversity in the naming conventions among the Finnic (and Slavic) inhabitants.

Nevertheless, we must emphasize that based on the clustering results no exact line can be drawn in the map to depict where the Finnic naming conventions ended and the Russian ones started. The fact is that more than 90 percent of the names are Russian. In other words, the clusters obtained, especially in the case of the Euclidean metric, are largely based on the differences in the usage of Russian names.

In many previous studies (such as Рябинин 1997; Новожилов 2004; Чибисов 2019), Finnic personal names and ethnonyms referring to Finnic tribes were used as one of the main arguments when defining the boundaries between Finnic and Russian cultures. It cannot be denied that Finnic names and name forms imply where people representing Finnic ethnicity have resided, but as this study has shown the situation has been more diverse than might be assumed. In this regard, one of the main results of our study is that, in most cases, differences in local naming conventions seem to be in line with cultural-historical knowledge, although the Finnic personal names or ethnonyms would not prove this. The best example of this is the close-knit Karelian cluster group within the Euclidean results that contains only a few dozen Finnic names or name forms.

In addition to the "Karelian" cluster group, the results imply that within the so-called Finnic region, there are subclusters that can be considered as groups of "Izhorian" and "Votic" naming conventions. First of them is located south from the River Neva and consists of the pogosts Dudorovskoj and Ižorskoj, whereas the one that can be connected to Votes lies in the west containing the town Jama with its surroundings and the pogosts Opol'skoj and Toldožskoj. However, the existence of "Izhorian" or "Votic" naming conventions is very questionable.

All things considered, we can conclude that those computational practises that we followed turned out to be mostly suitable for the purpose. Digital methods can be especially useful for studying the past of the northern regions where the historical sources are limited. Russian census books, for example, would offer many more possibilities for further research. Our study could be further refined using different clustering algorithms. This could mean, for example, using latent Dirichlet allocation (LDA) topic modelling. Furthermore, the reliability and informativeness of the clustering results would be higher if the data could be divided into smaller areal entities, in this case, into villages.

Acknowledgements. The publication costs of this article were covered by the Estonian Academy of Sciences.

Addresses

Jaakko Raunamaa
University of Helsinki
E-mail: jaakko.raunamaa@helsinki.fi

Antti Kanner
University of Helsinki
E-mail: antti.kanner@helsinki.fi

Abbreviations

MIN — А. М. А н д р і я ш е в ъ, Матеріалы по исторической географіи Новгородской земли. Шелонская пятина по писцовымъ книгамъ 1498—1576 гг. Списки селеній, Москва 1914; **POKV I, II —** Переписная окладная книга по Новугороду Вотьской пятины, Москва 1851, 1852; **PKOP —** Писцовые книги

Обонежской пятины: 1496 и 1563 гг, Ленинград 1930; **NPK III** — Новгородские писцовые книги, изданные Археографическою комиссиею. Т. 3. Переписная оброчная книга Вотской пятины, 1500 года. 1 половина, Санкт-Петербург 1868.

REFERENCES

Androshchuk, F. 2008, The Vikings in the East. — The Viking World, London, 514—542.

Cheshire, J., Mateos, P., Longley, P. 2011, Delineating Europe's Cultural Regions: Population Structure and Surname Clustering. — Human Biology 83, 573—598.

Forbes, N., Michell, R. 1914, The Chronicle of Novgorod 1016—1471, New York.

Frog, M., Saarikivi, J. 2015, De situ linguarum fennicarum aetatis ferreae. Pars I. — RMN Newsletter 9, 64—115.

Grünthal, R. 1997, Livvistä liiviin. Itämerensuomalaiset etnonyymit, Helsinki (Castrenianumin toimitteita 51).

Honkola, T., Santaharju, J., Syrjänen, K., Pajusalu, K. 2019, Clustering Lexical Variation of Finnic Languages Based on Atlas Linguarum Fennicarum. — LU LV, 161—184.

Hyvönen, S., Leino, A., Salmenkivi, M. 2007, Multivariate Analysis of Finnish Dialect Data — An Overview of Lexical Variation. — Literary and Linguistic Computing 22, 271—290.

Ianin, V. L. 2006, Medieval Novgorod. — The Cambridge History of Russia. Volume I. From Early Rus' to 1689, Cambridge, 188—210.

Kaufmann, L., Rousseeuw, P. 1990, Finding Groups in Data. An Introduction to Cluster Analysis, New York.

Kepsu, K. 2015, Moskovan ja Tukholman välissä: venäläiset pajarit Inkerinmaalla 1478—1722, Helsinki.

Korpela, J. 2004, Viipurin läänin historia II. Viipurin linnalläänin synty, Lappeenranta—Jyväskylä.

Kriiska, A., Tvauri, A. 2007, Viron esihistoria, Helsinki.

Kuzmin, D. 2014. Vienan Karjalan asutushistoria nimistön valossa. Doctoral dissertation, Helsingin yliopisto.

Lang, V. 2020, Homo Fennicus: itämerensuomalaisten etnohistoria, Helsinki.

Martin, J. 2007, Two Pomeshchiki From the Novgorod Lands: Their Fates and Fortunes During the Livonian War. — Russian History 34, 239—253.

Rahkonen, P. 2013. South-Eastern Contact Area of Finnic Languages in the Light of Onomastics. Doctoral dissertation, University of Helsinki.

Ränk, G. 1960, Vatjalaiset, Helsinki (SKST 267).

Raunamaa, J. 2020, The Distribution of Village Names Based on Pre-Christian Finnic Personal Names in the Northern Baltic Sea Area. — FUF 65, 98—152.

Raunamaa, J., Kanner, A. 2021, Refined Personal Name Data from the Census Book of Vodskaja Pjatina, Zenodo. http://doi.org/10.5281/zenodo.4436307.

Ronimus, J. V. 1906, Novgorodin vatjalaisen viidenneksen verokirja v. 1500 ja Karjalan silloinen asutus, Helsinki (Historiallinen arkisto XX. 1).

Saksa, A., Uino, P., Hiekkanen, M. 2003, Ristiretkiaika 1100—1300 jKr. — Viipurin läänin historia I. Karjalan synty, Lappeenranta, 383—474.

Saloheimo, V. A. 2010, Entisen esivallan alle, uusille elosijoille. Ortodoksikarjalaisten ja inkeroisten poismuutto 1500- ja 1600-luvuilla, Joensuu.

Selart, A. 2015, Livonia, Rus' and the Baltic Crusades in the Thirteenth Century, Leiden—Boston.

Shi, Y., Li, L., Wang, Y., Chen, J., Yuan, Y., Stanley, H. E. 2019, Regional Surname Affinity: A Spatial Network Approach. — American Journal of Physical Anthropology 168, 428—437.

Sjögren, A. J. 1969 [1861], Gesammelte Schriften I—II, Leipzig.

Sousa, X., Ginzo Villamayor, M. 2021, Surname Regions and Dialectal Variation in the Asturian Linguistic Space. — Journal of Linguistic Geography 8, 102—114.

Stoebke, D.-E. 1964, Die alten ostseefinnischen Personennamen im Rahmen eines urfinnischen Namensystems, Hamburg.
Tambets, K., Yunusbayev, B., Hudjashov, G., Ilumäe, A.-M., Rootsi, S., Honkola, T., Vesakoski, O., Atkinson, Q., Skoglund, P., Kushniarevich, A., Litvinov, S., Reidla, M., Metspalu, E., Saag, L., Rantanen, T., Karmin, M., Parik, J., Zhadanov, S. I., Gubina, M., Damba, L. D., Bermisheva, M., Reisberg, T., Dibirova, K., Evseeva, I., Nelis, M., Klovins, J., Metspalu, A., Esko, T., Balanovsky, O., Balanovska, E., Khusnutdinova, E. K., Osipova, L. P., Voevoda, M., Villems, R., Kivisild, T., Metspalu, M. 2018, Genes Reveal Traces of Common Recent Demographic History for Most of the Uralic-Speaking Populations. — Genome Biology 19. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1522-1.
Turney, P. D., Pantel, P. 2010, From Frequency to Meaning: Vector Space Models of Semantics. — Journal of Artificial Intelligence Research 37, 141—188.
Tvauri, A. 2007, Migrants or Natives? The Research History of Long Barrows in Russia and Estonia in the 5th—10th Centuries. — Topics on the Ethnic, Linguistic and Cultural Making of the Russian North, Helsinki (Slavica Helsingiensia 32), 247—285.
Worth, D., Flier, M. 2012, Language. — The Cambridge Companion to Modern Russian Culture, Cambridge (Cambridge Companions to Culture), 17—43.
Баранов К. В. 1999, Писцовые книги Новгородской земли, Москва.
Витов М. В. 1962, Историко-географические очерки Заонежья XVI—XVII вв. Из истории сельских поселений, Москва.
Карлова О. 2014, Карельская антропонимия нехристианского происхождения (на материале Писцовой книги Водской пятины 1500 года). — V Всероссийская конференция финно-угроведов «Финно-угорские языки и культуры в социокультурном ландшафте России». Материалы, Петрозаводск, 33—36.
Конькова О. И. 2008, Археологические находки на западе Ленинградской области и проблема происхождения ижоры. — Археологическое наследие Санкт-Петербурга. Выпуск 2. Древности Ижорской земли, Санкт-Петербург, 9—32.
Муллонен И. 2008, Топонимия Заонежья: словарь с историко-культурными комментариями, Петрозаводск.
Неволин К. А. 1853, О пятинах и погостах новгородских в XVI веке, с приложением карты, Санкт-Петербург (Из Записок Императорского русского географического общества. Кн. VIII).
Новожилов А. Г. 2004, Этническая ситуация на Северо-Западе Новгородской земли в XV—XVI вв. — Вестник Санкт-Петербургского университета. Серия 2: История, 79—92.
Петровский, Н. А. 1980, Словарь русских личных имен, Москва.
Рябинин Е. А. 1997, Финно-угорские племена в составе Древней Руси. К истории славяно-финских этнокультурных связей. Историко-археологические очерки, Санкт-Петербург.
Седов В. В. 1982, Восточные славяне в VI—XIII вв, Москва.
—— 1995, Славяне в раннем средневековье, Москва.
Соболев А. И. 2017, Антропонимы прибалтийско-финского происхождения в писцовых книгах Юго-Восточного Обонежья XV—XVI вв. — Вопросы ономастики 14, 7—34.
Сорокин П. Е. 2008, Археологическое изучение средневековых памятников в Приневье. Новые данные по археологии ижоры. — Археологическое наследие Санкт-Петербурга. Выпуск 2. Древности Ижорской земли, Санкт-Петербург, 88—127.
Суперанская А. В. 2010, Словарь народных форм русских имен, Москва.
Тихомиров С. 1905, Черты церковно-приходского и монастырского быта в Писцовой книге Водской пятины 1500 года (в связи с общими

условиями жизни). С приложением Алфавитного списка сел, селец, деревень, починков и пустошей, названных в Писцовой книге Водской пятины 1500 г. и Алфавитного списка помещиков, Санкт-Петербург.

Чибисов Б. И. 2019, Этнонимические прозвища жителей Водской пятины в конце XV века: этноисторический аспект. — Novogardia 2, 145—159.

Appendix 1. Pogosts, towns and areas

Number	Pogost/Town	Area			
1	Григорьевской Кречневской	1	**31**	Никольской Ястребинской	10
2	Никольской Пидебской	1	**32**	Григорьевской Льѣшской	8
3	Заверяжье	1	**33**	Богородицкой Врудской	10
4	Егорьевской Лузской	2	**34**	Егорьевской Вздылицкой	11
5	Дмитрѣевской Гдитцкой	2	**35**	Яма город	12
6	Климетцкой Тесовской	2	**36**	Воздвиженской Опольской в Чуди	13
7	Спасской на Оредежи	2	**37**	Никольской Толдожской в Чуди	12
8	Сабельской	3	**38**	Егорьевской Радшинской	13
9	Успѣнской Хрепельской	3	**39**	Ладога город	14
10	Косицкой	3	**40**	Пречистенской Городенской	14
11	Никольской Передольской	3	**41**	Ильинской на Волховѣ	14
12	Дмитрѣевской Городенской	4	**42**	Федоровской Песоцкой	14
13	Никольской Будковской	5	**43**	Егорьевской Теребужской	14
14	Ильинской Тигодской	6	**44**	Михайловской на Волховѣ	14
15	Солецкой на Волховѣп	6	**45**	Спасской Городенской	18
16	Андрѣевской Грузинской	1	**46**	Ивановской Куйвошской	19
17	Успѣнской Коломенской на Волховѣ	1	**47**	Воздвиженской Корбосельской	19
18	Антоновской на Волховѣ	1	**48**	Ильинской Келтушской	18
19	Петровской на Волховѣ	1	**49**	Егорьевской Лопьской	15
20	Иванской Переѣздьской на Волховѣ	1	**50**	Введенской Дудоровской	16
21	Копорья город	7	**51**	Никольской Ижорской	17
22	Каргальской	7	**52**	Никольской Ярвосольской	15
23	Егорьевской Радшинской	8	**53**	Корѣльской & Волочёк Сванской город	21
24	Ильинской Замозской в Бегуницах	8	**54**	Воскресенской Городенской	21
25	Покровской Дятелинской	9	**55**	Михайловской Сакульской	20
26	Дмитрѣевской Кипѣнской	9	**56**	Васильевской Ровдужской	20
27	Богородицкой Дягиленской	11	**57**	Богородицкой Кирьяжской	22
28	Никольской Суйдовской	11	**58**	Никольской Сердовольской	23
29	Покровской Озерѣтцкой	10	**59**	Ильинской Иломанской	24
30	Спасской Зарѣцской	11	**60**	Воскресенской Соломянской	24

Appendix 2a. Pairwise distances between the areas (Jaccard)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0	0.54	0.52	0.56	0.53	0.54	0.61	0.54	0.58	0.52	0.55	0.60	0.56	0.60	0.58	0.65	0.63	0.58	0.56	0.59	0.61	0.58	0.58	0.59
2	0.54	0	0.54	0.54	0.54	0.54	0.61	0.57	0.61	0.56	0.59	0.62	0.55	0.60	0.60	0.65	0.64	0.58	0.59	0.62	0.62	0.59	0.59	0.58
3	0.52	0.54	0	0.51	0.49	0.54	0.59	0.52	0.57	0.54	0.54	0.60	0.58	0.55	0.54	0.66	0.61	0.59	0.57	0.60	0.59	0.54	0.56	0.56
4	0.56	0.54	0.51	0	0.50	0.55	0.59	0.56	0.58	0.53	0.57	0.63	0.59	0.58	0.57	0.65	0.61	0.59	0.59	0.62	0.62	0.58	0.57	0.58
5	0.53	0.54	0.49	0.50	0	0.50	0.59	0.52	0.56	0.52	0.53	0.61	0.57	0.58	0.58	0.65	0.58	0.58	0.56	0.61	0.61	0.55	0.59	0.56
6	0.54	0.54	0.54	0.55	0.50	0	0.59	0.56	0.57	0.54	0.57	0.60	0.58	0.58	0.56	0.63	0.58	0.55	0.56	0.60	0.62	0.57	0.58	0.56
7	0.61	0.61	0.59	0.59	0.59	0.59	0	0.59	0.58	0.59	0.61	0.57	0.57	0.65	0.61	0.64	0.62	0.61	0.60	0.62	0.64	0.61	0.62	0.61
8	0.54	0.57	0.52	0.56	0.52	0.56	0.59	0	0.53	0.49	0.50	0.57	0.57	0.54	0.58	0.66	0.62	0.57	0.61	0.59	0.62	0.57	0.59	0.59
9	0.58	0.61	0.57	0.58	0.56	0.57	0.58	0.53	0	0.54	0.54	0.60	0.57	0.60	0.56	0.62	0.61	0.60	0.58	0.61	0.62	0.58	0.59	0.59
10	0.52	0.56	0.54	0.53	0.52	0.54	0.59	0.49	0.54	0	0.51	0.58	0.55	0.54	0.55	0.66	0.64	0.58	0.60	0.59	0.63	0.58	0.58	0.58
11	0.55	0.59	0.54	0.57	0.53	0.57	0.61	0.50	0.54	0.51	0	0.59	0.57	0.60	0.58	0.64	0.62	0.58	0.59	0.60	0.63	0.57	0.60	0.59
12	0.60	0.62	0.60	0.63	0.61	0.60	0.57	0.57	0.60	0.58	0.59	0	0.58	0.61	0.62	0.67	0.63	0.63	0.60	0.64	0.64	0.61	0.63	0.61
13	0.56	0.55	0.58	0.59	0.57	0.58	0.57	0.57	0.57	0.55	0.57	0.58	0	0.62	0.60	0.64	0.60	0.60	0.60	0.63	0.63	0.60	0.63	0.62
14	0.60	0.60	0.55	0.58	0.58	0.58	0.65	0.54	0.60	0.54	0.60	0.61	0.62	0	0.55	0.70	0.65	0.56	0.58	0.63	0.62	0.60	0.61	0.60
15	0.58	0.60	0.54	0.57	0.58	0.56	0.61	0.58	0.56	0.55	0.58	0.62	0.60	0.55	0	0.65	0.59	0.56	0.54	0.59	0.61	0.54	0.57	0.56
16	0.65	0.65	0.66	0.65	0.65	0.63	0.64	0.66	0.62	0.66	0.64	0.67	0.64	0.70	0.65	0	0.59	0.64	0.62	0.64	0.66	0.63	0.65	0.63
17	0.63	0.64	0.61	0.61	0.58	0.58	0.62	0.62	0.61	0.64	0.62	0.63	0.60	0.65	0.59	0.59	0	0.59	0.57	0.57	0.62	0.58	0.61	0.58
18	0.58	0.58	0.59	0.59	0.58	0.55	0.61	0.57	0.60	0.58	0.58	0.63	0.60	0.56	0.56	0.64	0.59	0	0.55	0.57	0.61	0.57	0.55	0.57
19	0.56	0.59	0.57	0.59	0.56	0.56	0.60	0.61	0.58	0.60	0.59	0.60	0.60	0.58	0.54	0.62	0.57	0.55	0	0.55	0.59	0.52	0.57	0.55
20	0.59	0.62	0.60	0.62	0.61	0.60	0.62	0.59	0.61	0.59	0.60	0.64	0.63	0.63	0.59	0.64	0.57	0.57	0.55	0	0.59	0.54	0.58	0.59
21	0.61	0.62	0.59	0.62	0.61	0.62	0.64	0.62	0.62	0.63	0.63	0.64	0.63	0.62	0.61	0.66	0.62	0.61	0.59	0.59	0	0.58	0.59	0.58
22	0.58	0.59	0.54	0.58	0.55	0.57	0.61	0.57	0.58	0.58	0.57	0.61	0.60	0.60	0.54	0.63	0.58	0.57	0.52	0.54	0.58	0	0.49	0.49
23	0.58	0.59	0.56	0.57	0.59	0.58	0.62	0.59	0.59	0.58	0.60	0.63	0.63	0.61	0.57	0.65	0.61	0.55	0.57	0.58	0.59	0.49	0	0.48
24	0.59	0.58	0.56	0.58	0.56	0.56	0.61	0.59	0.59	0.58	0.59	0.61	0.62	0.60	0.56	0.63	0.58	0.57	0.55	0.59	0.58	0.49	0.48	0

Appendix 2b. Pairwise distances between the areas (Euclidean)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0	0.50	0.52	0.47	0.48	0.47	0.56	0.54	0.55	0.53	0.52	0.52	0.53	0.50	0.54	0.63	0.55	0.50	0.48	0.59	0.58	0.57	0.58	0.59
2	0.50	0	0.46	0.50	0.47	0.52	0.55	0.54	0.58	0.56	0.51	0.56	0.59	0.47	0.50	0.62	0.54	0.47	0.45	0.56	0.57	0.57	0.58	0.60
3	0.52	0.46	0	0.51	0.47	0.54	0.53	0.56	0.59	0.54	0.50	0.58	0.58	0.49	0.47	0.67	0.58	0.52	0.50	0.59	0.56	0.52	0.56	0.59
4	0.47	0.50	0.51	0	0.51	0.51	0.58	0.55	0.55	0.53	0.51	0.55	0.56	0.53	0.58	0.65	0.59	0.51	0.50	0.63	0.58	0.55	0.57	0.54
5	0.48	0.47	0.47	0.51	0	0.50	0.54	0.54	0.59	0.56	0.48	0.56	0.60	0.46	0.48	0.63	0.54	0.48	0.49	0.56	0.54	0.51	0.55	0.54
6	0.47	0.52	0.54	0.51	0.50	0	0.50	0.56	0.51	0.56	0.51	0.55	0.61	0.54	0.55	0.60	0.51	0.51	0.44	0.56	0.52	0.53	0.51	0.57
7	0.56	0.55	0.53	0.58	0.54	0.50	0	0.60	0.55	0.57	0.51	0.53	0.59	0.53	0.49	0.61	0.52	0.52	0.50	0.58	0.52	0.54	0.57	0.60
8	0.54	0.54	0.56	0.55	0.54	0.56	0.60	0	0.52	0.54	0.54	0.53	0.59	0.54	0.58	0.65	0.58	0.54	0.51	0.60	0.60	0.61	0.60	0.58
9	0.55	0.58	0.59	0.55	0.59	0.51	0.55	0.52	0	0.55	0.51	0.54	0.59	0.56	0.60	0.57	0.52	0.56	0.56	0.59	0.57	0.57	0.54	0.56
10	0.53	0.56	0.54	0.53	0.56	0.56	0.57	0.54	0.55	0	0.53	0.59	0.52	0.56	0.59	0.70	0.60	0.54	0.55	0.60	0.62	0.59	0.57	0.53
11	0.52	0.51	0.50	0.51	0.48	0.51	0.51	0.54	0.51	0.53	0	0.54	0.58	0.51	0.51	0.60	0.55	0.53	0.51	0.56	0.54	0.53	0.54	0.50
12	0.52	0.56	0.58	0.55	0.56	0.55	0.53	0.53	0.54	0.59	0.54	0	0.56	0.51	0.58	0.61	0.52	0.52	0.52	0.56	0.53	0.57	0.57	0.56
13	0.53	0.59	0.58	0.56	0.60	0.61	0.59	0.59	0.59	0.52	0.58	0.56	0	0.56	0.60	0.68	0.59	0.58	0.58	0.68	0.63	0.65	0.64	0.60
14	0.50	0.47	0.49	0.53	0.46	0.54	0.53	0.54	0.56	0.56	0.51	0.51	0.56	0	0.46	0.60	0.51	0.46	0.49	0.56	0.53	0.56	0.55	0.56
15	0.54	0.50	0.47	0.58	0.48	0.55	0.49	0.58	0.60	0.59	0.51	0.58	0.60	0.46	0	0.63	0.53	0.50	0.49	0.57	0.56	0.55	0.56	0.59
16	0.63	0.62	0.67	0.65	0.63	0.60	0.61	0.65	0.57	0.70	0.60	0.61	0.68	0.60	0.63	0	0.47	0.56	0.55	0.59	0.60	0.64	0.67	0.67
17	0.55	0.54	0.58	0.59	0.54	0.51	0.52	0.58	0.52	0.60	0.55	0.52	0.59	0.51	0.53	0.47	0	0.49	0.45	0.53	0.51	0.54	0.54	0.55
18	0.50	0.47	0.52	0.51	0.48	0.51	0.52	0.54	0.56	0.54	0.53	0.52	0.58	0.46	0.50	0.56	0.49	0	0.42	0.53	0.52	0.54	0.56	0.54
19	0.48	0.45	0.50	0.50	0.49	0.44	0.50	0.51	0.56	0.55	0.51	0.52	0.58	0.49	0.49	0.55	0.45	0.42	0	0.51	0.51	0.52	0.55	0.53
20	0.59	0.56	0.59	0.63	0.56	0.56	0.58	0.60	0.59	0.60	0.56	0.56	0.68	0.56	0.57	0.59	0.53	0.53	0.51	0	0.50	0.57	0.55	0.59
21	0.58	0.57	0.56	0.58	0.54	0.52	0.52	0.60	0.57	0.62	0.54	0.53	0.63	0.53	0.56	0.60	0.51	0.52	0.51	0.50	0	0.46	0.46	0.53
22	0.57	0.57	0.52	0.55	0.51	0.53	0.54	0.61	0.57	0.59	0.53	0.57	0.65	0.56	0.55	0.64	0.54	0.54	0.52	0.57	0.46	0	0.36	0.46
23	0.58	0.58	0.56	0.57	0.55	0.51	0.57	0.60	0.54	0.57	0.54	0.57	0.64	0.55	0.56	0.67	0.54	0.56	0.55	0.55	0.46	0.36	0	0.43
24	0.59	0.60	0.59	0.54	0.54	0.57	0.60	0.58	0.56	0.53	0.50	0.56	0.60	0.56	0.59	0.67	0.55	0.54	0.53	0.59	0.53	0.46	0.43	0

Jaakko Raunamaa, Antti Kanner

ЯАККО РАУНАМАА, АНТТИ КАННЕР (Хельсинки)

КЛАСТЕРИЗАЦИЯ НАЗВАНИЙ СРЕДНЕВЕКОВОГО НОВГОРОДА: ГЕОГРАФИЧЕСКАЯ ВАРИАЦИЯ ЛИЧНЫХ ИМЕН В ПЕРЕПИСНОЙ КНИГЕ ВОДСКОЙ ПЯТИНЫ

Статья представляет одно из первых исследований, в котором используются цифровые методы для сбора и кластеризации личных имен с целью изучения тенденций именования. В данном исследовании анализируются имена, зафиксированные в переписной книге Водской пятины, которая в конце XV века была одной из пяти административных областей Новгородской земли. Материалы исследования собраны в цифровом виде. Для распределения имен по группам применялись два метода: метод средней связи (анг. average linkage) с использованием коэффициента Жаккара и метод Уорда (анг. Ward linkage) с использованием евклидовой метрики.

Общее количество собранных нами имен 35 726, среди них 2748 уникальны. Подавляющее большинство наиболее распространенных имен — христианского происхождения. Имена, в составе которых есть прибалтийско-финские элементы, составляют меньшинство — примерно 2% от всех собранных. Кластеризация оказалась продуктивным методом для изучения особенностей именования в средневековье. Результаты обеих метрик в основном соответствуют результатам предшествовавших исследований по истории региона: к концу XV века южный регион Водской пятины уже был славянизирован, в то время как в северных частях все еще проживало значительное прибалтийско-финское население.

JAAKKO RAUNAMAA, ANTTI KANNER (Helsingi)

KESKAEGSE NOVGORODI NIMEDE RÜHMITAMINE: VADJA VIIENDIKU RAHVALOENDUSRAAMATUS ESITATUD ISIKUNIMEDE GEOGRAAFILINE VARIEERUVUS

Artikkel on üks esimesi uurimusi, kus on kasutatud digitaalseid meetodeid isikunimede kogumiseks ja rühmitamiseks, et uurida nimeandmistrende. Analüüsitakse nimesid, mis on kantud XV sajandi lõpu Novgorodi Vadja viiendiku rahvaloendusraamatusse. Nimede rühmadesse jaotamiseks kasutati kahte meetodit: keskmise seose meetodit, kasutades Jaccardi koefitsienti, ja Wardi meetodit koos eukleidilise kauguse mõõtmisega.

Kogutud nimesid on 35 726, millest 2748 on ainulaadsed. Valdav enamus levinumaid nimesid on ristiusu päritolu. Nimed, mis sisaldavad läänemeresoome elemente, on suures vähemuses, neid on ainult ligikaudu 2 %. Klasteranalüüs osutus keskaegse nimetraditsiooni iseärasuste uurimisel viljakaks meetodiks. Mõlema meetodiga saadud tulemused on põhimõtteliselt kooskõlas piirkonna ajalugu käsitlevate varasemate uurimistööde andmetega: XV sajandi lõpuks oli Vadja viiendiku lõunapiirkond juba slaavistunud, samal ajal kui põhjapoolsetes osades oli veel olulisel määral läänemeresoome elanikkonda.