

UNIVERSITY OF TARTU
FACULTY OF ARTS AND HUMANITIES
INSTITUTE OF ESTONIAN AND GENERAL LINGUISTICS
DEPARTMENT OF FINNO-UGRIC STUDIES

Seth Wilson

**The Paleo-European substrate in Saami: a critical harvest from the
literature**

Master Thesis

Supervisor Gerson Klumpp, PhD

Tartu, 2024

Affirmation of authorship/Autorsuse kinnitus

Kinnitan, et olen käesoleva lõputöö ise kirjutanud ning toonud korrekselt välja teiste autorite panuse. Töö on kirjutatud lähtudes Tartu Ülikooli eesti ja üldkeeleteaduse instituudi lõputöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

I confirm that I have written this thesis myself and have correctly cited the contributions of other authors. The work was written based on the thesis requirements of the Institute of Estonian and General Linguistics of the University of Tartu and is in line with good academic practices.

Table of contents

Table of contents	3
List of abbreviations	4
1. Introduction	5
1.1 An overview of the prehistory of Fennoscandia	5
1.2 Entrance of the Uralic Languages	6
1.3 On the origin of the Saami and their languages	7
1.4 Challenges of substrate studies	10
1.5 On linguistic substrata	11
2. Literature review	16
3. Phonological substrate	20
3.1 Consonant gradation	20
3.2 The vowel system of Saami	25
3.3 Preaspiration	29
3.4 Substrate features which can be identified via substrate vocabulary	31
4. Lexical substrate	34
4.1 Overview of lexical substrata	34
4.2 Reindeer vocabulary	34
4.3 Environmental vocabulary	35
4.4 Substrate toponymy	36
4.5 Kinship terminology and other substrate items of interest	40
5. Grammatical Substrate	42
6. Estimate of substrate languages in Fennoscandia	46
7. Conclusions and outlook	48
7.1 Conclusions	48
7.2 Outlook and further research	50
Bibliography	52
Kokkuvõtte	56
License to reproduce the thesis	58

List of abbreviations

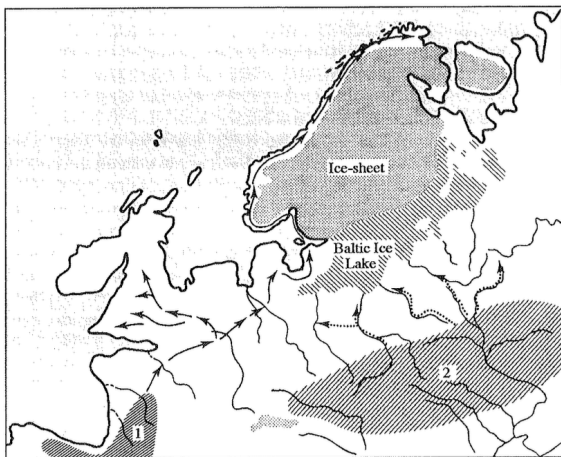
DEU	German
ENG	English
EST	Estonian
GAE	Irish (Gaeilge)
GEN	genitive
HUN	Hungarian
LIT	Lithuanian
NOM	nominative
PaLak	Paleo-Lakelandic
PART	partitive
PF	Proto-Finnic
PFS	Proto-Finno-Saami(c)
PFU	Proto-Finno-Ugric
PL	plural
PS	Proto-Saami
PU	Proto-Uralic
SaaLak	Lakeland Saami
SaaN	North Saami
SaaP	Pite Saami
SaaS	South Saami
SaaSk	Skolt Saami
SGr	strong grade
SOV	subject-object-verb
SVO	subject-verb-object
WGr	weak grade

1. Introduction

1.1 An overview of the Prehistory of Fennoscandia

The Saami are the indigenous people of northern Fennoscandia¹, the genetic descendants of the original settlers of the region, who are believed to have arrived in the region more than ten millennia ago. In the present day, Saami territory stretches from the Kola peninsula in northwest Russia across northern Norway, Finland, and Sweden, and south down the Scandinavian peninsula. To the outside world, the Saami are best known for their lifestyle in which reindeer play a major role, either hunted in the wild, or domesticated and used for transport as well as food. Although the Saami are the best-known reindeer-herding group, reindeer herding is not unique to the Saami, as it is practiced by other groups (Uralic and non-Uralic) in northern Russia, North America, and as far south as northeastern China.

Europe has been populated by humans since the Upper Paleolithic era, around 33000 BC. During the last glacial maximum (approx. 18000 BC) northern Europe, as far south as where Berlin is located today, was covered by an ice sheet, which rendered the entire region uninhabitable (Mailhammer 672). The human population was concentrated in southern Europe, and divided into two groups– western and eastern, which remained isolated from each other until the Mesolithic era (Mailhammer 672). Although we know nothing about the languages of these two groups, we can assume that they were only distantly related to each other, if at all. It is certain that the western complex, who are likely ancestors of the Saami, did not speak an Indo-European



Map 1: ice sheets, cultural complexes, migrations (Kallio 2003: 238)

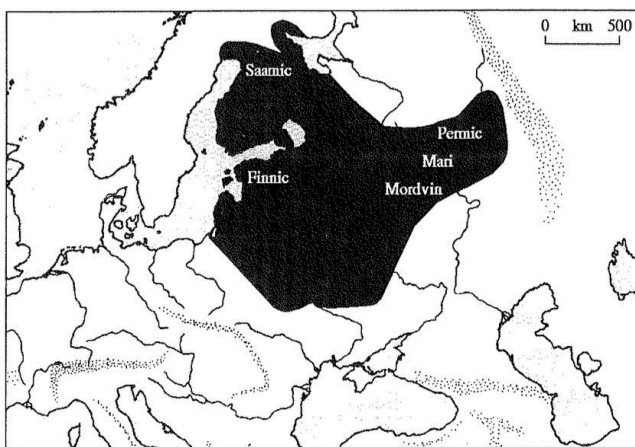
language (and of course could not have spoken a Uralic language

¹ A note on terminology: I will use “Fennoscandia” to refer to the region in northern Europe consisting of the Scandinavian Peninsula, Finland, Karelia, and the Kola Peninsula. “Scandinavia” does not include Finland. “Northern Europe” will refer to the northern half of Europe, including most of the area north of the Alps, including the British Isles.

either). It is believed that Basque² is the only remaining language of this prehistoric complex.

The ice sheet began retreating around 15000 BC, and with rapid increase in temperature from the 10th century BC, northward migration occurred, mainly from the southwest, but some from the southeast as well (Mailhammer 672-673). It was from the western cultural complex that the first settlers colonized the coast of Norway (10000-8800 BC) (Kallio 2003: 231), and from there moved further northward as climate conditions became more hospitable. It is these people who are believed to be the genetic ancestors of the Saami.

1.2 Entrance of Uralic Languages



Map 2: Distribution of the Pit-Comb Ware culture (Kallio 2003: 239)

The Uralic language family is dated to approximately the fifth or fourth millennium BC, and is believed to have arisen in the area of the Volga and Oka rivers in western Russia (Kallio 2003: 229; Saarikivi 2022: 57). Uralic arrived in the Baltic region between the late sixth and fourth millennia BC (Kallio 2003: 229). Archaeologically, the Uralic languages are associated with the Pit-Comb Ware cultural complex,

which stretched from the western Ural mountains in the east to the Baltic Sea in the west, and as far north as southern Lapland (Map 2). Of the Uralic languages, the most relevant here are the Saami and Finnic languages. Proto-Saami is believed to have originated in southeastern Finland and Karelia, in the area of Lake Ladoga (Saarikivi 2022: 33). Proto-Finnic is believed to have developed a bit to the south of Proto-Saami, in the approximate area of Saint Petersburg, or even further south, where Pskov and Tver oblasts are now located in Russia

² There have been attempts to link Basque to the languages of the Caucasus, in particular Georgian, due to shared systems of ergativity and a vigesimal number system. However these theories are not likely to be true, nor are they relevant here. The language of “Old European Hydronymy” (see section 1.5) is thought to be related to Basque (Mailhammer 676).

(Saarikivi 2022: 36). As human settlement in Europe predates the Uralic language family by several millennia, the ice-age inhabitants of Europe could not have spoken Uralic languages³.

1.3 On the origin of the Saami and their languages

The origins of the Saami people and their languages has long been a question in Uralic linguistics. On one hand, the Saami languages are undisputedly members of the Uralic family, whose closest relatives, both genealogically and geographically, are Finnic. On the other hand, the Saami people are anthropologically different from their closest linguistic relatives, the Baltic Finns, as can be seen by both their genetics and culture— the Pit-Comb Ware culture did not reach all the way to the north of Fennoscandia (Map 2), although there was doubtlessly contact between Uralic- and Paleo-European-speaking people further to the south. There is debate as to how Saami is placed on the Uralic family tree— whether it is an offshoot of Proto-Finno-Saami, or an independent branch, its similarities with Finnic being due to intensive long-term contact. This is investigated by Sylvi Lindström in “An analysis of the arguments for and against a common Proto-Finno-Saami language” (2021).

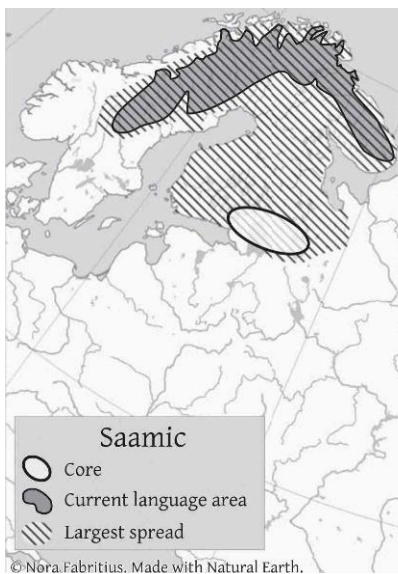
In short, this is what is known about the Saami people and languages:

- The Saami languages form a group within the Uralic language family
- The Saami language group has had most intensive contact with Finnic
- The Saami people show anthropological and genetic differences from other Uralic-speaking people
- The Saami languages contain a large number of lexical items which cannot be traced to a Uralic or Indo-European etymology
- Fennoscandia has been inhabited for nearly ten millennia, but the Uralic languages are believed to have emerged only six or seven millennia ago and thousands of kilometers away, in the Ural mountains. As Saami is a proven member of the Uralic family, some sort of shift to Uralic must have occurred, which resulted in Saami.

There are various hypotheses developed to explain the “contradictory” origins of Saamic (Korhonen 1981: 23): the language shift hypothesis, the loan contact hypothesis, and the Early Proto-Finnic hypothesis. The language shift hypothesis, developed by Swedish

³There is the question of “Para-Uralic”, which refers to the hypothesized extinct sister languages to Proto-Uralic. Peter Schrijver (2014) also suggests that Uralic and Indo-European are probably ultimately related deep in time. Of course this claim would be difficult to verify, and is not the point of this thesis.

linguist K. B. Wiklund in the late 19th century, assumes that the Saami were once a “linguistically and racially unknown population” (“*kielellisesti ja rodullisesti tuntematon väestö*”; Korhonen 23), who abandoned their native “Proto-Lappish”^{4 5} tongue in favor of a form of Proto-Finnic. According to the loan contact hypothesis (*lainakontaktihypoteesi*), the Saami and Finnic languages, though related, are not *recently* connected by a common ancestor, but have their similarities due to heavy contact. Finally the Early Proto-Finnic hypothesis (*varhaiskantasuomihypoteesi*) suggests that Saami is descended from a Finno-Saami proto-language.



Map 3: distribution of Saami languages (from Saarikivi 2022: 33)

No matter how exactly the Saami languages arose, it is known that there were people living in Fennoscandia millennia before Finno-Ugric languages were present in the region. As they were modern humans, they must have communicated via some language, but what could this language be? In short, we do not know. Owing to the sparse distribution of the population of northern Fennoscandia— both prehistorically and presently— it is likely that the Paleo-Fennoscandian population was not linguistically homogeneous (Kallio 2015: 77; Aikio 2012: 93-94). Saami likely replaced multiple languages, which are now lost deep in time.

Although the Saami languages form a branch of the Uralic family, the Saami people are genetic outliers— not only within the Uralic-speaking peoples, but within Europe (Kallio 2003: 231). It is believed that the Saami are the descendants of the earliest people to settle in northern Fennoscandia, who migrated into the region more than ten millennia ago from western Europe. These people of course spoke non-Uralic languages, but as their descendants shifted to Uralic languages in recent millennia, traces of their original languages, referred to as substrata, remained in the Saami languages spoken in the region today.

In this study, I will attempt to answer the following questions based primarily on the works of linguists Ante Aikio, Janne Saarikivi, and Peter Schrijver:

⁴ Another note on terminology: the term “Lapp” and its derivatives are nowadays considered pejorative. I will thus avoid using the term except for in direct quotations. The term “Lapland”, however, seems to still be in common use (cf. “Sápmi”).

⁵ The term “Proto-Lappish” is additionally problematic because it refers to a hypothesized unknown language from which the Saami shifted to Uralic, and not to a true Saami proto-language.

- *What influence have extinct Paleo-European⁶ languages had on Saami?*
- *Based on existing research, what can we determine about the Paleo-European languages of Fennoscandia?*
- *Do Finno-Saamic and Germanic share a Paleo-European substrate?*
- *How do the differences of Western and Eastern prehistoric Europe come into play with regard to the substrate languages in Fennoscandia?*
- *Can one estimate the number of Paleo-European substrate⁷ languages present in this region?*

Ideally, it would be possible to reconstruct the paleo-languages of Fennoscandia using the resources available. The ideal (though at this point unattainable) goal of this study, and potential future research, would be to create, as it were, a “Paleo-Laplandic Phrasebook” for travelers who find themselves in Pre-Uralic Fennoscandia. Unfortunately, as the ancient languages of Fennoscandia are long extinct, and as there are no written records, all that we can know about them is from the traces which remain as substrata in languages spoken in their previous territory. As an analogy, one could compare this study to the restoration of a palimpsest, in that traces of what was there before remain, but one cannot completely make out what was written. The objective of the restorer is to piece together the remnants into something somewhat intelligible.

The goal of this paper is to use what resources are available to determine what can be reconstructed of the paleo-languages of ancient Fennoscandia. Much of the literature on Saami and the historical linguistics of northeastern Europe has been published in Finnish and Russian, and has been a challenge to access. Parts of relevant sources have been translated.

Two authors in particular (who have published in English) have provided insights into the Pre-Uralic and prehistoric linguistic situation of Fennoscandia and the Baltic Sea region—Ante Aikio (Luobbal Sámmol Sámmol Ante) who discusses the ethnic and linguistic history

⁶ Yet another note on terminology: I will use “Paleo-/Palaeo-European” as a hypernym for all unidentified languages of Europe, which were replaced by Indo-European and Uralic languages. “Paleo-Laplandic” will refer to such languages spoken in northern Fennoscandia, in the area where Saami languages are present today. “Paleo-Lakelandic” will refer to the unidentified languages in central and southern Finland and Karelia—of course as it stands, it is unclear whether these languages are separate from Paleo-Laplandic, but due to geography, different names will be used. See Map 4 (section 1.4) for clarity.

⁷ “Substrate” here will generally be used as shorthand for “Paleo-European substrate”. Other relevant substrata (Saami substrate in Finnic, etc.) will be labeled accordingly.

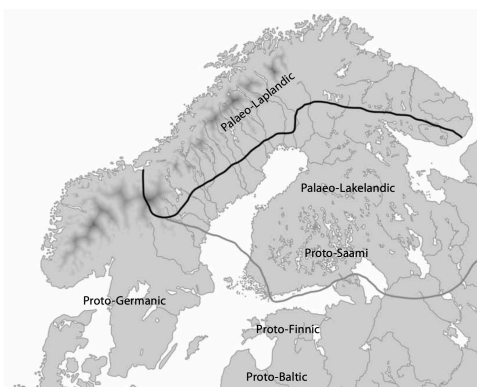
and prehistory of the Saami and their languages; and Peter Schrijver, who investigates the history of the Germanic languages, and their contacts in the Baltic Sea region.

The result of this thesis will not be a reconstruction of any paleo-language. Rather it is intended to provide a base from which to determine whether further research can be productively conducted, and if so, how it can be conducted.

Linguistic substrata are usually studied to reveal what effect they have had on the languages into which they have been absorbed. Substrata are often used as a sort of “excuse” for words and features whose origins are unclear. This study works in the opposite direction—to investigate what can be revealed about the substrate languages themselves, based on what is already known about the surface languages.

This study will follow roughly the methodology of Orest Tkachenko’s *Мерянский Язык* (1985). Tkachenko’s work is an attempt to reconstruct the extinct Merya language based on the substrate interference that it left on the toponyms and regional vocabulary of its former territory in western Russia, in the area of the Volga river. The primary difference between Tkachenko’s study and this one is that Merya was known to be a Finno-Ugric language, and thus certain hypothetical features could be assumed or ruled out. In the case of the Paleo-European languages in Fennoscandia, all that we can assume is that they are *not* Finno-Ugric or Indo-European, and thus we have nothing with which to compare them. Any features that we bring forward, even as educated guesses, remain speculations.

1.4 Challenges of Substrate Studies



Map 4: linguistic map of Fennoscandia before the expansion of Saami languages (from Aikio 2012: 64)

Substrate studies is a very uncertain field of research. Material from which to identify features of a substrate language is often quite limited, and only limited deductions can be made based on what “residue” remains in the replacing language. Non-lexical substrata (i.e., phonological, morphosyntactic) are especially hard to prove, since there is usually little to no evidence beyond the lexical level. As is discussed later, there can be some evidence for a phonological substrate, which may be deduced partially via the analysis of substrate lexical items. Much is still up to speculation, however, especially in cases where one is dealing with an extinct and unknown source language, as we are here. It is difficult even to

determine whether one is dealing with one or multiple substrate languages, as is the case with “Language X” or the Language of Geminates (see Chapter 2)– it is not known whether its features were part of a single language (unlikely), or multiple related or unrelated languages.

This map (Map 4), found in Ante Aikio’s “An essay on Saami ethnolinguistic prehistory” (2012), depicts the northern Baltic Sea area prior to the expansion of the Saami languages northward out of southern Finland. Lapland and Lakeland are marked with “Palaeo-Laplandic” and “Palaeo-Lakelandic” respectively, as Saami was not yet present in those areas. Those labels are misleading, as they imply that we know what was spoken there–concrete Paleo-Lapandic and Lakelandic languages.

1.5 On Linguistic Substrata

Although its precise definition remains undecided, a linguistic substrate or substratum can be roughly defined as a type of cross-linguistic influence caused by language contact (Saarikivi 2006a). In this instance, it refers to the “residue of an extinct language that is discernible in a present-day language” (Saarikivi 2006a: 12). This type of substrate comes about as a result of language shift (in this case Paleo-European to Saami) via incomplete shift to the target language. A major example is the Celtic substrate in English, which is responsible for many of the complexities of English syntax which set it apart from its other Germanic cousins (Schrijver 2014: 21). Substrate interference is described in Schrijver (2014) as the result of imperfect learning of a second language. The speakers of the substrate language learn the converging language, but will apply their native phonology and syntax–essentially speaking it with a substrate-language accent (Schrijver 2014: 17). Morphology and vocabulary are generally not carried over from one’s first language, unless one has learned the second language very imperfectly (Schrijver 2014: 17). Despite Schrijver’s statement of vocabulary not generally being carried over, substrata (such as the Paleo-European substrate in Saami) is most noticeable in the lexicon.

Substrate interference is in contrast with three other types of linguistic stratum–superstrate, adstrate, and perstrate interference (Saarikivi 2006a: 14). Superstrate interference occurs when a language takes influence, often in the form of vocabulary, from a culturally dominating language, such as the influence of French on English during the Middle Ages. Superstrate interference may even occur when speakers of the dominating language shift to the less prestigious one (Saarikivi 14). Adstrate interference occurs as the result of prolonged contact between languages, but without large-scale language shift. An example here would be

contact between Swedish and Finnish in Finland (Saarikivi 2006a:14) (e.g., loss of pitch accent). Perstrate interference, which is less relevant here, occurs due to contact of the vernacular with a literary language (Saarikivi 2006a : 14), such as the influence of Latin in many European languages, or of Classical Chinese on Japanese and Korean.

In “Lost Languages of northern Europe” (2001), Peter Schrijver outlines three proposed substrate languages that appear to have left their mark on the present-day languages of northern and western Europe. First is the language of “Old European Hydronymy”, which is considered to be the origin of many hydronyms in northern Europe, including the British Isles, Scandinavia, Germany, Poland, and the Baltic region. Due to its high frequency of resonants and *s* as opposed to stops, it is assumed to not be Indo-European, but rather possibly related to Basque (according to Theo Vennemann (1994)) (Schrijver 2001: 419). Second is the “Language of Bird Names”, which is also assumed to not be Indo-European. Proposed features include a stressed prefix *a-*, syncope of vowels, and certain phonemes not found in other languages of the region (ib.). This language is attested in Germanic, Celtic, Italic, and possibly Greek. Third and finally is the “Language of Geminates”, which already has been, and will continue to be discussed. It is notable for its variation in root consonants, and as a substratum appears to be heavily present in Germanic, Finnic, and Saami (Schrijver 2001: 420).

In “An essay on Saami ethnolinguistic prehistory,” Ante Aikio establishes the following criteria to determine whether there is likely substrate interference in a given language or language group:

Quantitative criterion: large number of unetymologized lexical items

Structural criterion: potential substrate items show non-native phonological or morphological structures

Irregular correspondence: potential substrate words show irregular sound correspondences between languages or dialects, indicating parallel borrowing

Semantic criterion: words of unknown origin clustered in specific semantic fields (natural environment, culture-specific topics)

Onomastic criterion: assumed substrate vocabulary paralleled by unknown-origin toponyms with similar non-native structures

The Saami languages contain a heavy stratum of lexical items which do not have an acceptable etymology, either from within the Uralic family, or from any known language family with which Saami was come into contact (i.e., Indo-European and its daughter

languages). Saami meets all five of Aikio's criteria for likely substrate presence, with a large number of unetymologized lexical items which show non-native phonotactic patterns, are semantically concentrated in fields concerning the natural environment, and have irregular correspondences between Saami languages.

Features in Saami that can be assumed to have come from unknown Paleo-European substrate languages are primarily phonological and lexical. Lexical items of assumed substrate origin include a large number (over 550, or more than a third of the reconstructed vocabulary) of common Proto-Saami roots without an acceptable etymology (Sammallahti 1998: 125; Aikio 2012: 83). With language- or dialect-specific words, the number increases further. Many of these items refer to area-specific items and concepts- names for animals and plants, words related to the local lifestyle, etc. Phonological features of assumed substrate origin are those found in unetymologized lexical items, and which also cannot be traced beyond Proto-Saami origins, due to phonotactic structures which are not permissible in Uralic-origin words.

Based on existing research, we can come up with a few theories about the features of the Paleo-European languages which were replaced by Saami. First is a list of lexical items which cannot be projected back to Uralic origins, or even to Proto-Saami. If we can determine at what stage in the development of Saami these items entered the lexicon, then we can likely determine the form of the item when it first entered the Saami lexicon, and assume that its form in the substrate language was similar. As for the phonological features of these substrate languages, we can assume that their phonemic inventories and phonotactics allowed for phonotactic structures which were not allowed in Saami, hence the unusual structure of many substrate words.

Peter Schrijver's book *Language Contact and the Origins of the Germanic Languages* (2014) argues for a common substrate between Germanic and Saami, which he refers to as the "Language of Geminates" or "Language X"⁸. Schrijver argues that Finno-Saamic, Saami, and Germanic share certain features which can be (certainly) traced neither to Proto-Uralic nor to Indo-European origins, including certain vowel rules, and complex consonant alternations (although the case for gradation in Uralic is rather complicated, and will be discussed later). Schrijver explains that the types of sound changes common to Germanic and Saami are likely due to substrate interference from the Language of Geminates, as Gothic

⁸ The terms "Language of Geminates" and "Language X" are used interchangeably. I will use whichever term is used in the work that I am discussing.

was not subject to some of the vowel changes (Schrijver 2014: 194). This limited distribution across language family boundaries supports the substrate theory. Any differences between the common features of Germanic and Saami (e.g., differing systems of consonant alternations) could be examples of Aikio's criterion of irregular correspondence, but applied to a phonological feature rather than to a lexical item.

We can assume that there are multiple substrate languages of non-Indo-European and non-Uralic origins present in Saami, mainly due to geography. As the prehistoric cultural complexes were isolated from one another, they certainly spoke different languages.

Saarikivi (2004) discusses the slim Paleo-European substrate in Finnic. There are various words in Finnic which are etymologically obscure, which also contain non-Uralic features, including final *-s* and lack of vowel harmony. Two examples are Finnic *ilves* 'lynx' and *jänis* 'hare' (Saarikivi 2004a: 208). Both are traceable to a Finno-Saamic root, but not further (ETY).

Other linguistic substrata in this region include the Finnic substrate in northern Russian dialects, the Saami substrate in Finnish, and the *Čud*⁹ (a pre-Slavic group in the Dvina basin) and Merya substrata in Russian varieties of the Dvina and Volga river areas (Saarikivi 2004a: 192; Tkachenko 1985). There have been borrowings between Finnic and Permic (Komi and Udmurt), as is outlined by Saarikivi (2006a: 33-38).

Finland, in particular the inland areas, has many place names of Saami origin. This is outlined in "The Study of Saami Substrate Toponyms in Finland" (Ante Aikio, 2007). Many of these toponyms are hard to detect or date due to phonological and etymological nativization. Most of these toponyms follow the typical Saami/Finnish structural type, consisting of two elements— a specific element followed by a generic (e.g., *Kukkas/järvi* 'long [specific] lake [generic]') (Aikio 168). Either element, or neither, may be borrowed from Finnish, but the majority have a borrowed (Saami) specific and a Finnish generic (Aikio 169). Some, however, originally consisted of more than two elements, which are then irregularly shortened when adopted into Finnish, rendering them etymologically opaque (Aikio 173). Many of the Saami substrate elements which Aikio presents are likely ultimately of Paleo-European origin, but an in-depth analysis is not possible here.

According to Saarikivi (2006b), there exists a Uralic (primarily Finnic and Komi (Permic)) substrate in the Russian dialects of the Arkhangelsk region of northwestern Russia

⁹ The term *Čud* (чудь; also *Chud*, *Chude*) refers to various Pre-Slavic medieval populations in present-day western Russia. See Saarikivi (2006b: 7) for more.

(Saarikivi 2006b: 5). Much of this substrate is in the form of toponyms, mainly hydronyms (Saarikivi 2006b: 12), such as the brooks *Нижний* ('lower') and *Верхний* ('upper') *Петручей*, the name of which likely comes from the Finnic **petäjä* 'pine' (Saarikivi 2006b:13). Many settlements have Finno-Ugric names as well, such as *Усигорка*, which may derive from the Finnic element **uusi* 'new' and the Russian element *горка* 'hill (diminutive)' (Saarikivi 2006b:13). Other common Uralic toponymic elements frequently occurring in this region are endings which presumably derive from the Proto-Finnic **neemi* 'cape' (realized as *-нень, -мень, -минь, -нема, -мена, -мина*) and **woja* 'brook' (*-ой, -ай, -оя, -ая*) (Saarikivi 2006b: 17). There are many more Uralic toponymic elements, which are explained by Saarikivi (Saarikivi 2006b: 29-44). What all this indicates is that the population of this region spoke Uralic languages (Finnic and Permic) before a shift to Russian. Saarikivi proposes Saami etymologies for various toponyms as far south as the Dvina river basin (Saarikivi 2006b: 45). The reflexes of supposedly Paleo-European substrate toponymic elements, such as SaaN *geađgi* and *njárga* are only found west of the Dvina basin (Saarikivi 2006b: 45) (see section 4.5).

2. Literature Review

The resources which form the “core” of this study are two articles by Ante Aikio (“An essay on substrate studies and the origin of Saami” (2004); “An essay on Saami ethnolinguistic prehistory” (2012)), one article by Janne Saariviki (“Is there Palaeo-European substratum interference in western branches of Uralic?” (2004)), and a book by Peter Schrijver (*Language Contact and the Origins of Germanic* (2014)).

Various other works serve as more “peripheral” resources, in terms of (pre-)historical background of the region, substrate studies, and information on Saami and other relevant languages. A full list of resources is provided at the end, of course, but the most significant will be introduced here.

Ante Aikio’s article “An essay on substrate studies and the origin of Saami” (2004) outlines the case for the existence of a heavy substrate of unknown origin in Saami. Aikio provides lists of lexical items which are likely of substrate origin, due to the criteria described in section 1.5. Aikio also lists various phonological features which occur frequently in substrate words, whose presence in Saami is likely due to contact with a substrate language (see section 3.4 for details).

Aikio’s second article, “An essay on Saami ethnolinguistic prehistory” (2012) is similar to the first one, but discusses the Paleo-European linguistic substrata in Saami in the context of general linguistic and cultural shifts in prehistoric Fennoscandia¹⁰. In addition to the Paleo-Laplandic substrate in Saami, Aikio also discusses the implications of archaeology in Fennoscandia, Saami contacts with Proto-Scandinavian, the position of Saami within the Uralic family, and the Saami substrate in northern Finnic. The conclusion relevant here is that the region was inhabited millennia before the Uralic language family even emerged (Kallio 2003; Mailhammer 2011), which means that at least one language shift must have occurred.

Peter Schrijver’s *Language Contact and the Origins of Germanic* (2014) discusses, as the title suggests, the role of language contact in the development of the Germanic languages. Germanic forms a branch of the Indo-European family whose core is located in northwestern Europe, and which is believed to have originated in the area around northern Germany, Denmark, and southern Sweden (Schrijver 2014: 158), by the western edge of the Baltic Sea.

¹⁰ Aikio brings up the problems of referring to an ethnic group by the language that they speak, and vice-versa. The prehistoric inhabitants of Lapland, although they were likely genetic and even cultural ancestors of today’s Saami, could not have been “Saami”, as they did not speak Saami languages. This is in contrast to K.B Wiklund’s theory that the Saami had always been “Saami”, but had previously spoken another tongue (see section 1.3).

Germanic is further divided into North Germanic (Scandinavian), West Germanic (High and Low German, Dutch, English, etc.), and East Germanic. East Germanic (e.g., Gothic) is extinct. As Germanic and Finno-Saami have been in contact in the Baltic Sea region, it can be assumed that both have had influence on each other. Schrijver specifically discusses the development of the vowel systems as well as “complex consonant alternations” in Germanic and Saami, and ascribes them to being of common origin. However, since the developments in the Proto-Saami vowel system predate the Germanic umlaut, the Proto-Saami vowel shift could not be directly due to contact with Germanic (Schrijver 2014: 192), and hence were possibly caused by contact with a now-extinct substrate language previously spoken in the Baltic Sea area, which Schrijver calls “Language X” or the “Language of Geminates.” Schrijver proposes that Germanic and Finno-Saami share a substrate (more heavily in Saami than Finnic), which is also responsible for various features found across northern European languages, from umlaut in Germanic and consonant gradation in Finno-Saamic to consonant mutations in Celtic. However, Schrijver provides a disclaimer that these similar features across languages in this region, along with a few words of unknown etymology, are the only remnants of this proposed substrate language. We do not know whether it was a single language (unlikely) or multiple languages which shared these features either by relation, contact, or coincidence.

Janne Saarikivi’s dissertation *Substrata Uralica: Studies on Finno-Ugrian Substrate in Northern Russian Dialects* (2006a) contains five articles on substrate studies in northeastern Europe. Most relevant here is “Is there Palaeo-European substratum interference in western branches of Uralic?” (2004a), in which Saarikivi argues for the existence of Paleo-European substrata not only in Saami, but also in Finnic, based on the large number of etymologically obscure words in both branches. Many of these words have been preserved in toponyms, the common structures of which may provide some clues about vocabulary and word structures in extinct languages. Also relevant here is “Über die saamischen Substratennamen des [sic] Nordrusslands und Finnlands” (2004b), which as its title suggests, provides an overview of Saami substrate toponymy in Russia and Finland, including structural features which will be discussed later. Beyond the study of Paleo-European languages, Saarikivi’s work is relevant from a methodological perspective, in that it deals with linguistic substrate in the same region, partially concerning the same language groups—Uralic and Balto-Finnic substrata in Russian, Saami substrate toponymy in Russia and Finland, and Paleo-European substrate in Uralic.

The following sources are more peripheral, but do merit their own discussion:

O. B. Tkachenko's *Мерянский Язык* ('Merya Language') (1985) is an attempt to reconstruct the extinct Merya (or Meryan) language which was spoken in the Volga region of western Russia until the 18th century. Tkachenko bases his reconstruction on Merya-origin lexical items in local Russian varieties— dialectal words, toponyms, and onomastics. Tkachenko manages to reconstruct most successfully phonological and lexical features, and less so morphosyntax, via the historical-comparative method. Merya was already known to be a Finno-Ugric language, although where in the Finno-Ugric family it falls exactly is not known. Since there are other closely-related language groups in the region (Finnic, Mari), one can already assume certain features and make educated guesses about what features (phonotactics, vocabulary, morphosyntax) were present in Merya, based on comparison with other Finno-Ugric languages. In contrast with this is the study of extinct Paleo-European languages— there are no living or well-attested languages with which to compare the substrate.

Michael Rießler's "On the Origin of Preaspiration in North Germanic" (2003) outlines the presence of preaspiration in various languages and languages groups of northern Europe, including Icelandic, Scottish Gaelic, and Saami. Rießler argues for a "non-Indo-European" origin of preaspiration, which he attributes to Proto-Saami. Unfortunately only the abstract of this article was available.

Eugene Helimski's article "Proto-Uralic gradation [sic]: continuation and traces" (1995) explains consonant gradation in Uralic languages (described in section 3.1), and evaluates the arguments concerning its origins. The origins of consonant gradation have long been debated in Uralic linguistics, as gradation occurs only in Samoyedic, Finnic, and Saami— the three most peripheral branches of the Uralic language family. Helimski argues in favor of a common origin for gradation in Uralic.

"Languages in the prehistoric Baltic Sea Region" by Petri Kallio (2003) describes the prehistoric patterns of migration and settlement in Europe, and their implications regarding the ancient linguistic situation in the Baltic region. Kallio links the major known linguistic groups in the Baltic Sea region (Uralic and Indo-European) to stone-age cultures— the Pit-Comb ware culture and the Corded Ware culture respectively. Like Schrijver (2014), Kallio also attributes the radical sound changes in Proto-Saami to "imperfect learning [of Uralic] by the non-Uralic aboriginal population", and mentions other types of sound changes— metaphony, preaspiration, preocclusion, denasalization, which occurred in Germanic around the same time (Iron Age) (Kallio 2003: 232).

Pekka Sammallahti's *The Saami Languages: an Introduction* (1998) is a general guide to the Saami languages (with a focus on North Saami) and their phonological, morphological, and syntactic features. It serves well as a source of general information, though is not the most user-friendly in parts (e.g., lack of index).

Mikko Korhonen's *Johdatus lapin kielen historiaan* ('Introduction to the history of the Lapp language'; 1981) describes the development of the Saami languages. Most relevant here is Korhonen's overview of the research history concerning the origin of the Saami, and the problems which came along with it (Korhonen 1981: 22-28); as well as his comparison of Proto-Uralic and Saami syntax (Korhonen 1981: 342-346).

Sylvi Lindström's master's thesis "An analysis of the arguments for and against a common Proto Finno-Saami language" (2021), as the title suggests, evaluates the question of whether Saami and Finnic can be traced to a recent common ancestor—Proto-Finno-Saami. Lindström concludes that a common Proto-Finno-Saami language is likely, but "it depends on the weight assigned to the different arguments" (Lindström 2021: 3). There is debate as to whether Saami and Finnic do in fact stem from a recent common ancestor or whether they are more distantly related, but that their sharing of features is due to long-term contact.

3. Phonological Substrate

In a language-shift situation, the replaced language often leaves traces of its sound system on the language to which its erstwhile speakers shift. As Schrijver (2014) describes it, a phonological substrate is essentially speaking the replacing language with a substrate accent. An areally or typologically unusual phonological feature could in turn indicate historical substrate interference. There are various aspects of Saami phonology which could be ascribed to substrate interference. The three major features discussed here are consonant gradation, the development of the Saami vowel system, and preaspiration.

3.1 Consonant Gradation

Consonant gradation refers to “a systematic morphophonological alternation of consonants on the border of two syllables, resulting in either weakening or strengthening of consonants” (Bakró-Nagy 2022: 859). In other words, it is a system of internal consonant mutations, in which the affected consonant alternates between “strong” and “weak” allomorphs, the underlying consonant undergoing either fortition or lenition. Gradation is heavily present in the Saami and Finnic branches of Uralic, as well as Nganasan, and to an extent Selkup¹¹ in Samoyedic. It is not present in other Uralic languages¹². There are two basic types of gradation— syllabic/radical and rhythmic/suffixal. In syllabic gradation, a consonant occurs in the strong grade before the vowel of an open syllable, and in the weak grade before the vowel of a closed syllable (Helimski 1995: 20), e.g., Finnish *leipä* (strong) : *leivän* (weak) ‘bread’. Rhythmic gradation is linked to the stress pattern of the word it affects. A strong-grade consonant appears after the vowel of an odd-numbered (stressed) syllable, and a weak-grade consonant after an even-numbered (unstressed) syllable, e.g., Finnish *púu.tă* vs. *ká.lě.Ōa* (Helimski).

Gradation originated as a form of word-internal sandhi, but over time has been grammaticalized, and in some languages (e.g., Estonian) has become merely a morphological

¹¹ Specifically Ket’ Selkup according to Helimski (1995). Although Wagner-Nagy and Szeverényi (2022: 662) write that gradation is only present in Nganasan, it is apparently present in Ket’ Selkup as well (Klumpp & Budzisch 2023: 903).

¹² In Komi (Zyrian), there is the regular alternation *l~v*, in which a stem-final *v* alternates with *l* when a vowel suffix is added (Kuznetsov 2022: 489). E.g. *вѡв* ‘horse’ > *вѡльсь* ‘his/her horse’. Although this may resemble gradation at first glance (and is incorrectly described as such on the English-language Wikipedia article on Komi), this alternation is not tied to the rhythmic or syllabic structure of the word, and thus is not consonant gradation in the sense that we are examining.

feature, rather than phonotactic: the grade alternations within a word are now productive and contrastive, with their original phonotactic conditions now disappeared and irrelevant. An example here is the Estonian word *leib* ‘(rye) bread’ (a cognate with the Finnish *leipä* above). It is inflected as follows: *leib* (NOM) : *leiva* (GEN) : *leiba* (PART). The genitive case was

Fig. 1: Rhythmic (top) and **Fig. 2:** Syllabic (bottom) gradation in Nganasan (Wagner-Nagy 2022: 758)

deep structure	H	T	K	S	Sʲ	NH	NT	NK	NS	NSʲ
strong grade	h	t	k	s	sʲ	ŋh	nt	ŋk	ns	ŋsʲ
weak grade	b	ð	g	ɟ		mb	nd	ŋg		ɲɟ

deep structure	H	T	K	S	Sʲ	NH	NT	NK	NS	NSʲ
strong grade	h	t	k	s	sʲ	ŋh	nt	ŋk	ns	ŋsʲ
weak grade	b	ð	g	ɟ		^c h	^c t	^c k	^c s	^c sʲ

marked with a final *-n* in older varieties of Estonian, which has disappeared in recent centuries (cf. Finnish). However, as the final syllable was closed with *-n*, it was subject to syllabic gradation, and the consonant underwent weakening gradation (*b* : *v*). Although the genitive suffix has long disappeared, the genitive case is still marked with (and only with) the weak grade of the internal consonant. In contrast, the partitive keeps the strong grade as the now-disused partitive suffix **-tA* formed an additional syllable rather than closing the final syllable (cf. **-n*). As

the consonant affected is after the stressed (odd-numbered) syllable, it is affected by rhythmic gradation.

Among Samoyedic languages, consonant gradation is present in Nganasan, and to some extent in Selkup. The gradation system of Nganasan rather resembles that of Finnic, in that it affects stops and sibilants, and involves two processes to form the surface form of a word, Rhythmic Gradation and Syllabic Gradation, which are akin to suffixal and radical gradation respectively in Finnic (Wagner-Nagy 2022: 758). Unlike in Finnic, however, the weak grades in Nganasan have different surface forms based on whether they are affected by Rhythmic or Syllabic gradation.

Nganasan additionally employs glottal stop alternation, truncation, nunnation, stem alternation, and vowel harmony, which makes its morphology considerably more complicated than that of Finnic languages (Helimski 23).

Selkup is believed to display traces of gradation, and only on stops (Bakró-Nagy 2022: 859). Selkup does display sandhi assimilation across morpheme boundaries, which is partially affected by morphological conditions (Kazakevič 2022: 784).

Gradation in Proto-Finnic closely resembles that of Nganasan, but lacks the other morphonemic features which mask the base form of a morpheme. Finnic gradation affects stops (*p*, *t*, *k* and their geminate forms), as well as possibly *s* (Laakso 2022a: 243). As in Nganasan, Finnic stem-internal consonants are subject to weakening when occurring in a closed non-initial syllable, and when occurring in a suffix in a non-stressed syllable.

Strong Grade	Weak Grade
<i>p</i>	<i>β</i>
<i>t</i>	<i>ð</i>
<i>k</i>	<i>ɣ</i>
<i>pp</i>	<i>p̃p̃¹³</i>
<i>tt</i>	<i>t̃t̃</i>
<i>kk</i>	<i>k̃k̃</i>
? <i>s</i> ¹⁴	<i>z</i>
? <i>c</i> ¹⁵	<i>s</i>

Fig 3: Proto-Finnic consonant gradation (based on Laakso (2022: 243))

This system is preserved in most of the daughter languages of Proto-Finnic, with the notable exceptions of Veps and Livonian, which lack gradation altogether (Laakso 2022a: 243).¹⁶ In modern Finnic languages, such as Estonian, while gradation remains noticeable on the surface, its motivating conditions have disappeared in many instances, leaving it as only a morphological phenomenon, with minimal phonotactic motivation.

Gradation is also present in most Saami languages, with the exception of South Saami. Although gradation in Saami is traditionally projected back to Proto-Finno-Saamic (due to the presence of gradation in Finnic as well), this is not certain, as Saami and Finnic gradation are rather different. The

two share only syllabic gradation (Koponen 2022: 106), and not rhythmic gradation. In addition, Saami grade alternations are based on length, while in Finnic they are based on lenition/fortition of stops (ib.). This is possible in Saami, as gradation has been extended to affect almost all consonants (Schrijver 2014: 194), and not only stops.

In addition, Finnic gradation is described as “weakening”, in that the default grade of the consonant in question is strong, but it “weakens” when affected by gradation (e.g., EST *auk* ‘hole’ : *augu* ‘hole.GEN’). In contrast, Saami gradation takes the form of “strengthening”, in that the weak grade is the default, and the strong is marked (Schrijver 171). Strong-grade

¹³ The diacritic ̃ indicates shortening, here a “short geminate” (Laakso 2022a: 243).

¹⁴ The *s*~*z* alternation is not certain (Laakso 2022a: 243).

¹⁵ Petri Kallio has also reconstructed the alternation *c*~*s* (Laakso 2022a: 243).

¹⁶ Livonian does have a system of stem alternations which resembles consonant gradation (and has even been referred to as “gradation,” but it is motivated differently from those of Nganasan, Saami, and Proto-Finnic (Laakso 2022b: 380).

consonants in Saami are also marked by (pre-)aspiration (e.g., SaaP WGr *d* : SGr *ht* (Wilbur n.d.)).

Consonant gradation is notably absent in South Saami, although it remains present in its closest geographic and genealogical neighbor, Ume Saami (Ylikoski 2022b: 113). South Saami is the southwesternmost of the Saami languages, and is spoken in central Scandinavia, approximately halfway down the Scandinavian peninsula. South Saami differs phonologically from other Saami languages, as it has a more conservative consonant system (Ylikoski 2022b: 116), lacks gradation, and has a vowel system which is also quite different from the Saami languages further to the north, in that it employs a “regressive metaphony that can be characterized as a kind of *umlaut*” (Ylikoski 2022b: 117). The oddities of South Saami will be discussed further in section 3.2.

The question of the origin of consonant gradation in Uralic languages remains unsettled. On one hand, it has been projected as far back as Proto-Uralic, owing to its presence in both Finno-Ugric and Samoyedic. However, the areal distribution of gradation calls this theory into question. As gradation is only present in the easternmost (Samoyedic) and westernmost (Finnic and Saami) members of the Uralic family, its absence in other more central branches make this theory less plausible. Although it may seem unlikely that such a feature was only preserved in the most peripheral groups of Uralic languages, one must consider the idea of a conservative periphery, in which members of a language family which are located farther from the center undergo less innovation, and thus remain conservative in their features (Austerlitz 1985). Eugene Helimski (1995) argues that consonant gradation is a Proto-Uralic phenomenon, due to the existence of the combination of syllabic and rhythmic gradation, which appears to be unique to Finnic and Samoyedic, and is simply too unlikely to have arisen separately in the two branches. On the other hand, Peter Schrijver (2014) compares gradation in Finnic and Saami to Verner’s law in Germanic (in which voiceless fricatives become voiced following an unstressed syllable), and claims that the consonant alternations in both language families came about due to substrate influence from the “Language of Geminates.” Thus we have two common hypotheses which are nearly mutually exclusive. First is the Proto-Uralic hypothesis, in which both Finno-Saami and Samoyedic inherited consonant gradation from Proto-Uralic. This would explain the similarities in the gradation systems in both branches, but is geographically problematic, as Finno-Saami and Samoyedic are located respectively on the western– and easternmost peripheries of the Uralic language area, and consonant gradation is absent in the languages in between. Second is the separate innovation hypothesis, according to which consonant gradation is not necessarily a

feature of Proto-Uralic, but rather arose separately in Finno-Saamic and Samoyedic. This hypothesis is geographically more plausible, due to the theory of Finno-Saamic gradation originating from contact with Germanic or a substrate language, but it does not account for the close similarities between Samoyedic and Finnic gradation.

Peter Schrijver (2014) claims that a substrate language of unknown classification (though neither Uralic nor Indo-European), which he refers to as “Language X” or the “Language of Geminates” is responsible for the rise of consonantal alternations in Saami, as similar features are apparent in other languages of northern Europe, including various Germanic languages (Scandinavian, Dutch, High German, English) and Insular Celtic languages (Irish, Gaelic, Welsh, Breton, etc.) (Schrijver 2014: 195).

One should note that while consonant gradation and Verner’s law apply to consonants in non-initial syllables, consonant mutations in Celtic are applied to the initial consonant of a word, and at least in modern Welsh and Irish, are highly morphosyntactically productive (e.g., GAE *a cat* ‘her cat’ vs. *a chat* ‘his cat’ vs. *a gcat* ‘their cat’). However, the mutated allomorphs almost never occur alone— they are usually preceded by a pronoun or an article, making the initial consonant of the affected word *not the initial consonant in the noun phrase*, hence a reflex of Verner’s law, or consonant gradation, is not out of the question here.

However, could both origin theories not be true? Gradation could well have been present in Proto-Uralic, and was inherited by Samoyedic, but not by Finno-Ugric, at least on the surface. Gradation does not appear in Finno-Ugric languages until the western edge of the language family, in Fennoscandia. One could hypothesize that gradation lay “dormant” in Finno-Ugric languages until Finnic and Saami came into contact with Germanic or the Language of Geminates, at which point gradation was “reactivated” due to contact with a language with a similar though not identical morphological phenomenon (i.e., Language X). The “revival” of such a feature would require there to be a common phonotactic feature or set of features across Uralic languages, which renders their consonant systems susceptible to the acquisition of such alternations. Helimski does not rule this hypothesis out— according to him, “viewing gradation as PUr [Proto-Uralic] is not incompatible even with the claim that consonant gradation was an allophonic phenomenon as late as in Proto-Finnic” (Helimski 1995: 29).

In accordance with how substrate influence from extinct languages usually functions, speakers of Language X likely shifted to Finno-Ugric, but continued to apply their consonant alternations to the language to which they shifted. Due to the phonotactic structure of Saami and Finnic, syllabic and rhythmic (in the case of Finnic) gradation were the most fitting way

to apply consonant alternations to this newly-acquired language. The differences between the gradation systems of Saami and Finnic could be attributed to differing intensity of contact between Finnic/Pre-Proto-Saami¹⁷ and the substrate language.

One could also hypothesize that consonant gradation as a phonological phenomenon is easily both gained and lost. Hence the presence of gradation in some Uralic languages, but the lack of it in many others (most of the Uralic family), including languages closely related to those in which gradation is present (Livonian, Veps, South Saami), implies that it does not take much for gradation as a feature to disappear.

The first hypothesis calls into question the traditional Uralic language tree, according to which Uralic split in a binary manner into Samoyedic and Finno-Ugric, and so forth, with Finno-Saamic finally splitting into Finnic and Saami most recently. This model would likely show more evidence of gradation in the various languages and proto-languages which split off in between. The other possible model is the comb model, according to which the branches all split individually from a common Proto-Uralic ancestor. The similarities in gradation between Finno-Saamic and Samoyedic support this model, however lexical evidence points toward the tree model. This question is addressed in Lindström (2021).

In conclusion, the origin of gradation in Finnic and Saamic remains uncertain. The two main theories of its origin are not entirely mutually exclusive: the phonotactic (rhythmic and syllabic) structure of Uralic languages could be of such a nature that it is easily susceptible to such allophonic alternations, which become grammaticalized over time. The rise of grade alternations in Finnic and Saami could have been triggered by contact with a substrate language with similar features.

3.2 The Vowel System of Saami

The Proto-Saami vowel system is noticeably more complicated than those of both

Fig. 4: vowel systems of West Uralic, Pre-Proto-Saami, and Proto-Saami (from Aikio 2012: 70)

West Uralic	Pre-Proto-Saami	Proto-Saami
i ü u	i u	ie i u uo
e o	e o	ea ë o oa
ä a	ä a	ā

Proto-Uralic and Proto-Finnic, its closest neighbor. While Proto-Uralic (Fig. 5) is reconstructed to have had an eight-vowel system based on frontness/backness, openness, and roundedness (Aikio 2022: 5), West

¹⁷ Aikio (2012) defines “Pre-Proto-Saami” as “the ancestral form of Saami languages that had already diverged from its Uralic sisters, including Finnic, but which had not yet undergone the Great Saami Vowel Shift”.

Fig. 5: Proto-Uralic vowel inventory (Aikio 2022: 5)

	FRONT		BACK	
	CARDINAL	NON-CARDINAL	NON-CARDINAL	CARDINAL
CLOSE	<i>i</i>	<i>ü</i>	<i>ɨ</i>	<i>u</i>
MID	<i>e</i>			<i>o</i>
OPEN	<i>ä</i>			<i>a</i>

Fig. 6: Proto-Saami vowels (from Koponen 2022: 104)

	long	short
stressed syllable (vowel centre)	<i>ie uo</i>	<i>i u</i>
	<i>ea oa</i>	<i>ɛ o</i>
	<i>ā</i>	
unstressed syllable (latus and vowel margin)	<i>ē ō</i>	<i>i u</i>
	<i>ā</i>	<i>ɛ</i>

Uralic had seven (Fig. 4), and Proto-Finnic nine¹⁸

based on similar criteria, Proto-Saami had a system based on length and stress. Some vowels were always long, some always short, some occurred only in stressed syllables, some in unstressed syllables (Koponen 2022: 104). Notably, diphthongs occur only as long vowels and in stressed positions, and only one vowel which is both long and stressed is not a diphthong (**ā*).

As the Proto-Uralic and Proto-Finnic vowel systems resemble each other closely, one can assume that it is the Proto-Saami system which underwent the most change after the divergence of Proto-Finno-Saamic into Proto-Finnic and Proto-Saami.

In addition, cognate lexemes between Finnic and Saami often have rather strange vowel correspondences (e.g., EST *kala* : SaaN *guolle* (< PSaa **kōlē*)). According to Aikio (2012), the reorganization of the Pre-Proto-Saami vowel system was the principal change which turned Pre-Proto-Saami into Proto-Saami. This most likely occurred when a small community who spoke a language with a completely different vowel system shifted to Pre-Proto-Saami. The community would have to have been quite small and tight-knit in order for the vowel shift to occur in as regular a manner as it did (Aikio 2012: 71).

Peter Schrijver (2014) reconstructs the developments in the vowel system of Saami, and presents them as follows:

1. **y > i*
**Vw > ɔ*
2. Vowel harmony disappeared, and was replaced with *umlaut*, in which the vowel quality of a non-initial syllable affects the quality of the vowel in the initial syllable (Schrijver 191). In the case of Proto-Saami, a second syllable close vowel would cause a mid vowel in the initial syllable to become close.

¹⁸ The status of the unrounded central vowel **ɛ* is not unanimously reconstructed as a phoneme in Proto-Finnic (Laakso 2022a: 241-242).

3. Low vowels **æ*, **a*, **ɔ* become long.
4. First-syllable **ā* and **ǣ* became **ō* and **ā* respectively.
5. Long *ō* and **ā* split like so:

<i>ā</i>	>	<i>ǣ</i>
	>	<i>ā</i> (before <i>i</i> in third syllable)
<i>ō</i>	>	<i>ō</i>
	>	<i>ū</i> (before <i>i</i> in third syllable)

The first three developments led the vowel system of Proto-Saami to closely resemble that of Proto-Germanic, and with the last changes, North and West Germanic (191-192). However, these changes cannot have been directly due to Germanic influence, as they occurred before the introduction of umlaut in Germanic (Schrijver 192). Schrijver claims that both the vowel shift which took Pre-Proto-Saami into Proto-Saami, and the development of umlaut in Germanic, are likely due to interference from a common substrate, Language X. According to both Schrijver and Aikio, the development of the Proto-Saami vowel system (and thus the development of Proto-Saami) probably occurred when “a population speaking a language with a completely different sound system switched to Finno-Saamic” (Schrijver 194). Based on these sound changes, one can make certain assumptions about the vowel system of Language X, besides simply the abstract idea that its vowel system differed from that of Finno-Saamic. Based on the first sound change, Language X did not have the vowel *y*, and so its erstwhile speakers mapped this phoneme to the closest one in their native vowel inventory, *i*. In addition, based on the change of **w* to **ɔ* following a vowel, it appears that a vowel+ (labial) semivowel sequence was not permitted. To judge by the position-based vowel alternations, present in both Proto-Saami and Germanic, one can assume that similar rules applied to Language X, in that vowel length was tied to quality, and quality to which vowels occurred later in the word. Interestingly, this tells us nothing about diphthongs, and the **Vw* > **ɔ* change implies that diphthongs might not even have been permitted. On the contrary, Sammallahti (1999) (in Aikio 2012) reconstructs this change to **Vw* > **oj*. It is interesting that the apparent influence of Language X on Saami ended before the development of the diphthongs so characteristic of the Saami vowel system. The diphthongization of some vowels (e.g., PS **uo* < FS **a*) is reconstructed to Proto-Saami (Sammallahti 1998: 186). How exactly this part of the “Great Saami Vowel Shift” was triggered is unclear, as it is not accounted for in Schrijver’s description of the substrate interference by Language X. However, this does not rule out the possibility of influence from the same language which triggered the beginnings of the vowel shift, but did not end up affecting Germanic in the same

manner it did Saami toward the end. The source of the extensive diphthongization in Saami could be an independent innovation, but could also be the result of further substrate interference, either from Language X or from another substrate language. At this point, one must be reminded that “Language X” or the “Language of Gemimates” is an extremely vague term– it is not clear whether it is one language (unlikely due to its wide spread across the Baltic region), a group of related languages, or various unrelated languages which share certain phonological features as a *sprachbund*. As the diphthongization appears from an exclusively Uralic perspective to be a continuation of the vowel shift triggered by contact with Language X, one could hypothesize that the diphthongization occurred as a result of later contact with another substrate language, which may or may not have fallen within the bounds of Language X features.

Table 1: *Vowel correspondences between Finno-Saami, Proto-Saami, and modern North Saami (from Sammallahti (1998: 189))¹⁹:*

Finno-Saami vowel	Finno-Saami	Proto-Saami	North Saami
*o	*ojva	*oajvē	oaivi
*o	*omti	*vuomDe	vuovda
*ō	*hōli	*hūola	njuolla
*ō	*kōsi	*kuose	guossa
*a	*ala	*vuolē	vuolli
*a	*kala	*kuolē	guolli
*a	*appi	*vuohpe	vuohppa
*ä	*päjvä	*peajvē	beaivi
*ē	*kēli	*kiele	giella
*ē	*sēji	*sieje	siedja
*e	*elä-	*ealē-	eallit
*e	*velji	*vielje	viellja

¹⁹ Note: due to keyboard constraints, the exact diacritics are not present on several letters, but the relevant ones are marked as in Sammallahti (1998).

Returning to South Saami, it is rather strange that South Saami lacks a system of consonant alternations so characteristic of its genealogical relatives, but instead has a system of metaphony, or umlaut, in which a first-syllable vowel is affected by the vowel in a syllable later in the word. This umlaut cannot be a reflex of Finno-Ugric vowel harmony, since vowel harmony was lost in Saami much earlier, during the development of Proto-Saami. This is productive in South Saami morphology (Ylikoski 2022b: 118). As South Saami is the southernmost of the Saami languages, situated around halfway down along the Scandinavian Peninsula, it has likely had the most contact with Scandinavian languages. Umlaut and vowel alternations are a common feature in Germanic languages and are extremely productive in nominal and verbal morphology (e.g., Swedish *finna* : *fann* ('to find' PRES : PRET)). The metaphony in South Saami could thus have arisen as a result of intensive contact with Scandinavian, as Swedish employs a system of complementary quantity, in which a long vowel in a syllable will be followed by a short consonant, and vice versa. On the contrary, as South Saami is geographically quite peripheral, it may also be affected by the “conservative periphery” (see section 3.1) in some way. The syntax of South Saami has been affected by its contact with Scandinavian, but it is still unique not only within western Uralic languages, but within the “Northern European Sprachbund”, as it uses subject-object-verb as its basic word order (Ylikoski 2022b: 125).

3.3 Preaspiration

Preaspiration is a notable areal feature in various languages of northern Europe. It is present in Skolt, Kildin, Lule, Pite, and North Saami (Bakró-Nagy et al. 2022),²⁰ as well as Faroese, Icelandic, and Scottish Gaelic (Riebler 2003). That it is present only in the two most peripheral Scandinavian languages indicates that it was likely present at least in western Scandinavian varieties (conservative periphery). Michael Riebler (2003) argues for a non-Indo-European (Saami) origin for preaspiration, as although it is traceable to Common Scandinavian, “no traces of PA [preaspiration] are found in other Germanic languages.” It has been traced to Proto-Saami²¹ (Koponen 2022: 105), however. Its presence in Scottish Gaelic has been traced to contact with Scandinavian as a result of Viking-era Scandinavian settlement in Scotland (hence its absence in Irish and other Celtic languages).

²⁰ Inari Saami has postaspiration of stops before vowels in non-initial syllables and in final positions (Valtonen, et al. 2022: 182), however this seems to be a reflex of preaspiration.

²¹ As preaspiration has been traced to Proto-Saami, its presence in modern Saami languages, or rather which Saami languages it is present in, is not of great relevance here.

Schrijver (2014), on the other hand, argues that the “preglottalized plosives” in English and Danish (e.g., the glottal stop in the Cockney [wɑʔv] “water”) are reflexes of preaspiration in Icelandic and Faroese (110). Not once does Schrijver mention Saami in this explanation.

Due to its limited distribution in Scandinavian and Saami, one could hypothesize that preaspiration is a feature acquired from a substrate language. At first glance, the presence of preaspiration in Scandinavian, Saami, and Celtic point to a possible origin from the Language of Geminates. However, the explanation of its presence in Scottish Gaelic being due to contact with Scandinavian makes this hypothesis less likely, though not impossible.

In Saami (Lule, North, Pite, Skolt), preaspiration occurs morphonemically, as part of the system of strong-weak grade alternations (e.g., SaaP SGr *áhpe* ‘big marsh’ : WGr *ábe* ‘big.marsh.GEN’) (Wilbur n.d.; Bakró-Nagy et al. 2022).

Could preaspiration indeed be another substrate feature from the Language of Geminates? This idea is supported on two bases. First is its areal distribution. Preaspiration is found in Saami and Scandinavian, both of which have had their consonant systems influenced by the Language of Geminates. Second is the grammatical function of preaspiration. In the Saami languages which have consonant gradation, the strong grade of a consonant is often marked by preaspiration (Bakró-Nagy 2022: 860) (e.g., SaaP SGr *hp* ~ WGr *b*). The connection of preaspiration with consonant gradation points to a possible common substrate connection— one which is not discussed by Schrijver. Due to the locations of the languages which have preaspiration (western and northern Fennoscandia), one can assume that its origins are in the west. Although consonant gradation is present in Finnic, preaspiration is not, which supports either the theory of inherited gradation, or that substrate gradation may have been acquired from a different language (within the “Language X” realm) from those in western Fennoscandia.

The laryngeal fricative **h* in Proto-Saami is believed to have originated from the preaspiration of geminate consonants. An example here is **vuohpe* ‘father-in-law’ (< Pre-Saami **appi*) (Koponen 2022:105). Its non-Saami cognates (e.g., Finnish *appi*) are not preaspirated (Kotimaisten kielten keskus), indicating that preaspiration entered the Saami phonological system at some point during the development of Proto-Saami from Pre-Saami. This change in consonant quality, combined with its use in nominal morphology (i.e., case inflection by gradation), and at what stage in the development of Proto-Saami it occurred, is again reminiscent of the consonant alternations which Schrijver (2014) ascribes to the Language of Geminates.

Aikio (2004) mentions the voiceless [M]²² as a frequent feature in proposed substrate words (e.g., SaaN *liehmu* ‘mild weather in winter’). However, he never links this feature to preaspiration in other Saami languages. In Pite Saami, for example, a preaspirated consonant may be realized as unvoiced, such as [M].

Whether or not preaspiration is a substrate feature is unclear, however the case for its substrate origin is strong. Its limited areal distribution (Saami and Scandinavian-contact areas), as well as its application to both etymologically-opaque words and known words whose cognates in ancestor languages (e.g., PS **vuohpe* < PreS **appi*) lack it, indicate that this feature is native to this region, but not native to the languages which are presently there.

3.4 Substrate Features which can be identified via substrate vocabulary

Ante Aikio (2004) provides a list of lexemes in North Saami which are of likely substrate origin. One of the identifying criteria for many of these supposed substrate words is their phonotactic structure— if a word contains features (phonemes or combinations) which are not permissible in Uralic or Proto-Saami, and if the word additionally lacks a satisfactory etymology, then it is likely of substrate origin.

The specific lexemes will be discussed in the section on lexical substrate.

Aikio divides the list into four categories based on phonotactic features (13-14):

1. *“Features which could in theory be projected back to Uralic reconstructions, but which do not appear in shared Uralic vocabulary”*

These (reconstructed Proto-Saami) lexemes often contain unusual, though permissible phonotactic features, including three-consonant clusters, two *r*’s or dental affricates in the same word, and consonant clusters unattested in the shared Uralic vocabulary, such as **ls*, **sn*, **pl*, **pc*, **čk*, **kv*, **ŋv* (additionally **js*, **sm*, **mn*, **lf* in toponyms (Aikio 2004: 19)). Additionally, the sequence *long vowel + single consonant + non-high vowel* implies a non-Uralic origin.

2. *“Features which cannot be projected back to Uralic reconstructions but which may derive from a phase of Pre-Saami anteceding the reorganization of the Proto-Saami vowel system”*

This is associated with the secondary sibilant *š* (e.g., *morša* ‘walrus’).

²² The voiceless [M] is represented by <*hm*> in North Saami orthography. It is not clear whether this is a reflex of preaspiration.

3. “Features which can only derive from the phase of Proto-Saami after the radical reorganization of the Uralic vowel system”

Lexemes in this category may contain vowel combinations which would have undergone sound change, had they already been present before the reorganization of the vowel system. These structures include *á-a*, *ea-a*, *oa-a*, *á-u*, *ie-i*, *ie-u* (last two, or first syllable vowels *i* and *u* with an internal consonant cluster), and bisyllabic stems ending in *-á*.

4. “Features which cannot even derive from common Proto-Saami”

This category includes the initial consonant cluster *sk-*, non-initial *f*, and unvoiced [M].

The lexemes containing these unusual features must have originated somewhere. The simplest and most likely explanation is that they were introduced from a substrate language whose phonotactic rules differed from those of Saami, hence the unusual structures. One can reverse-engineer, as it were, the features of these substrate lexemes to hypothesize what features were present in the substrate language(s). The most obvious features which were likely present in the substrate language are the consonants: initial *sk-* and non-initial *f*. The other consonant clusters (category 1) must have been present as well, although less certainly. As the features *sk-* and *-f* are of a completely non-native origin, one can assume that they have not undergone significant sound change from the form from which they were adopted into Saami. Additionally, all three of the examples which Aikio provides for initial *sk-* (*skuogga* ‘baleen’, *skuoggir* ‘ethmoid bone’, *skuolfi* ‘owl’) retain the *sk-* in all of their cognates between Saami languages (Kotimaisten kielten keskus). Words in Saami containing the initial *sk-* are of either Scandinavian or substrate origin. Both substrate and Scandinavian loans in western and central Saami varieties (including Inari) retain the *sk-*, while in eastern varieties (Skolt, Kildin, Kola) reflect it as simply *k-* (Aikio 2004: 25; Kotimaisten kielten keskus). Earlier Indo-European loanwords with initial *sk-* were also reflected in all Saami varieties with *k-*, which indicates that the presence of an initial *sk-* was a western Saami innovation, likely brought about via contact with Scandinavian. Besides the implications regarding the disintegration of Proto-Saami (see Aikio 2004: 26 for more), the correspondence of *sk-* and *k-* in substrate words muddies the water a bit in the attempted reconstruction of substrate words. Substrate items with initial *sk-* that entered Saami before significant contact with Scandinavian would be rendered (incorrectly) as beginning with *k-*. There could (hypothetically) have existed other initial consonant clusters (even more

complex than something like *sk-*), which were reflected in Saami as *sk-* or *k-* by a process of leveling.

The question of the unvoiced [M] is discussed above in section 3.3.

One more feature which does not occur in the common Proto-Uralic lexicon but is present in several proposed substrate words is the word-initial **r-* (Mayrhofer 2024). It is listed specifically as non-permissible by Aikio (2022: 9), however it is not listed as a substrate feature by Aikio in either his 2004 or 2012 articles, despite being present in multiple proposed substrate items (e.g., SaaN *rusta* ‘frost mist’ < PS **rusnë*; SaaN *riehkku*, *roahkka*, *rohka* (‘seal’) (Aikio 2004)).

Sound correspondences, like the initial *sk-* ~ *k-* correspondence above, between substrate words in different Saami languages can provide some clues about the languages from which they originated. For example, there is an apparently systematic correspondence between **s* in West Saami and **š* in East Saami (Aikio 2012: 85), e.g., SaaS *saasne* ‘half-dry, (<**sāsnë*) ~ SaaN *suostu* (<**suosnō*) ~ SaaSk *šošnn* (<**šošnë*) (all of which mean roughly ‘dead tree’). Via this, one can deduce that these words were borrowed into Saami from different substrate languages or varieties, which had different phonotactic rules concerning sibilants– the eastern language had *š* and the western language *s*. With this knowledge, we can assume that there were multiple substrate languages in the region which had an effect on Saami. Of course this assumption depends on the **s* ~ **š* correspondence *not* being due to a sound change which occurred in certain Saami languages after the substrate words entered Saami. Even then, the hypothetical sound change could have occurred due to substrate influence. Either way still, we do not know the relationships between the various substrate languages themselves. The languages implicated in the *s*~*š* variation could be closely-related dialects which are differentiated by some basic sound variations, or they could be languages from different families, who share vocabulary due to contact, or anything in between.

4. Lexical Substrate

4.1 Overview of lexical substrata

A lexical substrate is generally the simplest to identify, since lexical items are the aspect of a language most easily transferred via language contact (Matras 2009: 155), either through substrate interference or simple borrowing. Lexical substrate is also the best point at which to begin in substrate studies, as it is with lexical items that a hypothesized substrate can be identified (see the five criteria in Aikio 2004, 2012). It is from substrate vocabulary that one can deduce the features of a substrate language— vocabulary, phonology, and even the lifestyles of the people who once spoke that language. The grammar (i.e., morphosyntax) of the language, however, is more difficult to deduce based on lexical items alone, but this will be discussed later.

Aikio (2004) provides a list of lexical items in North Saami, which are of likely substrate origin. The items that Aikio provides are presumed to be of substrate origin due to them lacking an acceptable Uralic etymology, and fitting the five criteria (quantitative, structural, irregular correspondence, semantic, onomastic) indicating substrate interference, which are discussed earlier.

Aikio's first list (Aikio 2004: 10-12) consists of words which fall into semantic categories prone to substrate interference, in this case animals and animal body parts, topography, snow, ice, and weather. Aikio mentions that there is also a large amount of etymologically-opaque vocabulary related to reindeer, plant names, hunting and fishing, and butchering, but does not list it (for the sake of saving space) (Aikio 2004: 12). However, the mere existence of a heavy layer of substratum vocabulary in these fields can provide some clues about the lifestyles of the prehistoric people in northern Fennoscandia— food was obtained via hunting and fishing, and reindeer played an important role in the people's way of life. However, this is neither new knowledge nor particularly relevant here.

4.2 Reindeer Vocabulary

Aikio (2012) and Ole Henrik Magga (2006) provide lists of vocabulary in North Saami related to reindeer, most of which is of obscure origin (Aikio 86). Although these words are semantically very specialized (*luohpet* 'one-year old reindeer cow which has had a calf'), not all are exclusive to reindeer, but may be generally used for large mammals (Kotimaisten kielten keskus), of which reindeer were the most significant. Magga (2006)

notes that some terminology is reserved for animals, such as *gabba* ‘(all) white’ (cf. *vielgat* ‘white’ < ?PFU **wēlketa*). Interestingly, the word for ‘antler’, *čoarvi*, can be traced back to Proto-Uralic **śorwa* ‘horn’ (Kotimaisten kielten keskus), and is not a substrate word²³. It is difficult to draw conclusions based on substrate vocabulary due to the probability of semantic shift. *Gabba*, for all we know, could have once been the generic word for the color ‘white,’ but was supplanted by the Finno-Ugric word except for in the specific field of reindeer²⁴. *Biltu* is defined by Magga as ‘shy and wild (usually of female reindeer; sometimes of girls)’, which indicates a shift in usage between reindeer and humans (31). Another example of semantic shift is the South Saami *orra* ‘male reindeer’ and the related word *hurrä* ‘one year old male reindeer’, which are thought to both originate from the same root (Schrijver 2001: 421).

4.3 Environmental Vocabulary

Aikio (2004, 2012) also provides several Saami words pertaining to the natural environment, which are of etymologically-obscure origin. According to the *semantic criterion* (section 1.5; Aikio 2012: 83), this is a likely area to find substrate vocabulary. Many of these words pertain to features “that are typical of the natural environment of Lapland, but rare or nonexistent in more southern Finland and Karelia” (Aikio 2012: 85). Some concepts make sense, such as *ráš’sa* ‘high and barren mountain’ and *riehppi* ‘valley up in the mountains which is difficult to access’ (ib.), since these types of geographical features are common in the mountainous parts of northern and western Fennoscandia, but not in the flatter areas to the south where other Finno-Ugric languages are spoken. As the Pre- and Proto-Saami languages spread into the mountainous north from the flat south, earlier stages of these languages (which came into contact with paleo-languages of southern Finland (see Maps 3 & 4, sections 1.3, 1.4)) likely did not have an elaborate vocabulary for describing mountainous terrain, because they did not need it. Hence, words from northern paleo-languages were borrowed (or rather retained) once those words became necessary. In

²³*Čoarvi* ‘antler’ is an interesting word, in that not only does it have a Uralic etymology (as opposed to being of Paleo-European substrate origin), but it is a cognate with the Uralic word for ‘horn’, which is physiologically different from antlers. One would expect that languages with such elaborate vocabulary for reindeer would have a specific word for antler, unrelated to horn.

²⁴The development of specialized synonyms is not unique to Saami. In English, for example, many words pertaining to the equestrian world (horse anatomy, riding equipment, etc.) are not used for describing similar concepts unrelated to horses.

contrast is the element *Vuonamo-*, *Vuonos-*, *Vuonis-* ‘long and narrow bay; fjord’ (cf. SaaN *vuotna* <PS **vuonē*). As a bay of a lake is similar to a fjord on the sea, this sort of semantic expansion makes sense.

4.4 Substrate Toponymy

Toponyms are an area of lexica in which substrata are especially noticeable, and alongside a large number of appellative items of opaque origin, indicate a historical language shift (Saarikivi 2004a: 191).

As is discussed by both Aikio (2004) and Saarikivi (2004a), substrate lexical items are often preserved in toponyms, as a name which denotes a specific geographic feature is less likely to change than an appellative word. Substrate toponyms are noticeable in many areas where language shifts are known to have taken place, such as with Celtic place names in Britain and Ainu place names in Japan²⁵ (Saarikivi 2004a: 191). Saami has numerous toponyms of obscure origin, assumed to be substrate items, which alongside the large substrate appellative vocabulary, indicates a historical or prehistoric language shift (Saarikivi 2004a: 191).

Many Saami and other Finno-Ugric toponyms are compounds, consisting of at least two lexemes. Many of these compounds are composed of a borrowed (substrate) basis and native topographic noun as a generic (e.g., *Duikkášvuotna*, *Duikkáš* (unknown) + *vuotna* ‘fjord’) (Aikio 2004: 18). Toponyms which do not follow this pattern tend to be either completely monolexical (and of apparent substrate origin), or are marked by an identifier (e.g., *Stuorra Jeahkkáš* ‘Big Jeahkkáš’ (17). These tendencies may indicate that the compounds so typical of Saami toponymy, even those composed of multiple obscure elements (e.g., *Nussir*) may have been a Saami innovation. This is merely a speculation of course, and a more solid, though vague explanation is that the Paleo-Laplandic languages had a more flexible system of toponymy than Saami.

Saarikivi (2004b) also identifies three other patterns in Saami toponymy, which differentiate it from the standard specific+generic pattern. First are names containing three elements (as opposed to the typical two-element names), such as *Buoiddesguollejávri* ‘fat fish lake’ (Saarikivi 2004b: 182). Second are three-element names in which one element is shortened and resembles a suffix, such as *Fiellogahskáidi*, which is composed of *fiellu* ‘board’, *geađgi* ‘stone’, and *skáidi* ‘area between rivers’. However, the internal *geađgi* is

²⁵E.g. *Avon* (e.g., Stratford-upon-Avon) < Welsh *afon* ‘river’. Many places in Hokkaido and northern Honshu have names of Ainu origin. Common Ainu features include *-betsu* (‘river’), *-poro* (‘large’), *-shiri* (‘place’).

shortened to *gah* (Saarikivi 2004b: 183). The third type of toponyms are deverbal names, which contain a verbal element, such as *Gádjáriegádanjárvi* ‘Katja’s birth lake’ (Saarikivi 2004b: 183).

Aikio manages to isolate eight toponymic elements which are of obscure origin (Aikio 2004: 21-24):

1. *-ir ~ Vr* (<PS **-ērē ~ *Vr(ē)*) ‘mountain, uplands’. *Nussir, Jeahkir*, etc. This element appears in many toponyms denoting mountains and highlands, the names of which are almost always obscure.
2. *skiehč(č)-* (<PS **skieč(č)-*) ‘watershed’. *Skiehččajávri, Skiehččan*. All toponyms with this element are connected to bodies of water. The initial *sk-* indicates substrate origin.
3. *jeahk(k)-* (<PS **jeak(k)-*) ‘isolated (mountain)’. *Stuorra Jeahkir, Unna Jeahkkáš*. All referents are isolated.
4. *nuss-* (<PS **nus-*) ‘mountain top on the edge of a highland area, projection of a mountain’. *Nussir* (multiple occurrences, all in close proximity in Finnmark region (norgeskart.no)).
5. *čára-* (<PS **čārēk/n/m*) ‘uppermost (lake)’. *Čárajávrrit, Čárajávri*. All referents are lakes, uppermost in their water systems.
6. *sáll-* (<PS **sāl-*) ‘major island in the sea’. *Sállam, Sállir, Dávnesálla*. All referents are islands. Possibly related to Baltic (Proto-Baltic **salā*), which lacks an Indo-European etymology.
7. *Čiesti ~ Šesti* (<PS **čiestē ~ *šiestē*). “Both denote small fells with a steep cliff on the seashore” (24).
8. **iñč-* ‘outermost island’. *Ižžot* (<PS **iñčujě-k*), *Fávle-Iččát, Gádde-Iččát*.

The last two examples are less certain, as they occur in fewer place names.

Saarikivi (2004a) also comes up with a few toponymic features of apparent substrate origin. These include Finnish hydronyms which end in *-nkV* and *-nTe*, as well as those in Russia with the endings *-nga*, *-nda* (see also section 1.5). Although some of these toponyms have been etymologized, “not all... can be easily explained on the basis of attested languages and their predecessors” (Saarikivi 2004a: 209)²⁶. The Finnish name archive *Nimiarkisto* lists many place names containing the elements *-nkV* and *-nTe*, but none appear to refer to bodies

²⁶ Unfortunately Alpo Räisänen’s *Nimet mieltä kiehtovat: etnologista nimistöntutkimusta* (2003), cited here by Saarikivi, was not available for further examination.

of water, and many refer to microtoponyms, such as villages and individual houses (Nimiarkisto).

Various lexical items found throughout Finnic, as well as other Uralic languages, are also of opaque origin. These include *mäki* ‘hill’ (<PF **mäki*), *niemi* ‘cape, peninsula’ (<PF **neemi*), *suo* ‘swamp’ (<PF **soo*), *saari* ‘island’, *lampi* ‘pond’, *aho* ‘clearing, meadow’ (<PF **aho* < ?Pre-Finnic **ašo*), *oja* ‘ditch’, *järvi* ‘lake’, *korpi* ‘dense forest, wilderness’ (see Saarikivi (2004: 199-203) for details). The words which do not have an acceptable etymology (i.e., those listed above) may have their origins in extinct languages formerly spoken in the Finno-Ugric area. As several of the words have proposed parallels outside of Finno-Saamic (e.g., Finnic *järvi* : Mari *jer/jär*), whatever substrate language these words came from was spread further eastward than Finnic and Saami.

Järvi (<PF **järvi* <PFS **järvä*) ‘lake’ is especially interesting. Although a Baltic etymology is proposed (cf. Latvian *jūra* ‘sea’), it “appears problematic because in the Finno-Ugrian substratum toponymy of northern and central Russia the word-stem has been fixed as *jagr-* and *jahr-*” (Saarikivi 2004a: 202). **järvi* varies little throughout Finnic (Estonian *järv*, Veps *järv*, etc. (ETY)), however in Saami, the internal consonants undergo metathesis, hence North Saami *jávri* (< PS **jāvrē*). The metathesis is present throughout all Saami languages (Kotimaisten kielten keskus), indicating that the change (**rv* > *vr*) occurred in Proto-Saami. This may indicate that the cluster **rv* was not phonotactically permissible in the languages of the people who shifted to Proto-Saami. This is merely a speculation of course, as more examples are needed to form a sound claim. On the contrary, the cluster **lf*, which shares a place of articulation with **rv*, is identified by Aikio (2004) as a distinctly non-Uralic feature which is present in a few substrate toponyms. For example the element *Nalfa* is found in several toponyms in the Jukkasjärvi/Čohkkiras area in northern Sweden (Aikio 2004: 18,31; Sveriges Hembygdsförbund). Although the internal consonant cluster **rw* is attested in Proto-Uralic, **wr* is not (Aikio 2022: 9). This indicates that the cluster *vr*, such as in *jávri*, is not a Uralic feature. If, by chance, the cluster *vr* is a substrate feature mutually exclusive with **rv*, then it conflicts with the articulatorily-similar **lf* found in northern Sweden. By extrapolating, one can speculate that the substrate language from which **lf* originates is separate from that in the Proto-Saami area, via which Proto-Finno-Saamic **järvä* became Proto-Saami **jāvrē*. If this case of metathesis is not in fact an example of phonotactic adaptation, then the consonant alternation between the first and second syllable becomes reminiscent of the Language of Geminates.

Saarikivi (2006b: 45) touches on what appear to be Saami-origin toponyms in the Dvina river basin. Some of these toponyms contain elements which are presumed to be Paleo-European substrate lexemes— **kēδkē* ‘stone’ and **harkē* ‘cape’. It is surprising enough to see what appear to be Saami toponyms so far south (cf. Map 3, section 1.3), but even more interesting to see Paleo-European lexical elements not only this far south, but specifically *only to the west of the Dvina basin*. This may indicate that the substrate language from which Saami gained so many of its toponymic elements was present not only in Fennoscandia, but near the eastern shore of the Baltic Sea as well. If these elements entered the Saami lexicon before Saami (apparently) entered the Dvina basin, they should have a wider distribution in the area. Could this be the source of the Proto-Saami element **sāl-* as well? To find out more, one would have to investigate the Paleo-European substrate in Baltic as well. Saarikivi does not provide any concrete examples.

Aikio gives a short description of the extinct Lakeland Saami Language(s) (2012: 89). Lakeland (*Järvi-Suomi*) refers to a region in southern Finland notable for its high density of lakes. Saami was previously spoken in this region, until after the Middle Ages (Aikio 2012: 89). Lakeland Saami has since been replaced by Finnish, but has left a stratum of lexical items in the local vocabulary. Like with most substrate situations, the only remnants of these languages take the form of loanwords and toponyms in Finnish (Aikio 2012: 90). Lakeland Saami no doubt had substrate interference from “Paleo-Lakelandic”, but as those earlier contacts are buried under two substrate layers (PaLak < SaaLak < Finnish), there is little to be reconstructed. However, Aikio (2012: 91) does provide a list of toponymic elements which come from Lakeland Saami, but are ultimately of unknown origin. If there are noticeable differences in the etymologically-obscure vocabulary in Lakeland and Lapland Saami, then one could assume that there are different Paleo-European languages at play here. Of course Proto-Saami passed through Lakeland on its way to Lapland, meaning that the Lapland Saami languages likely also have a Lakelandic substrate. Most of the Paleo-Laplandic substrate vocabulary is probably related to specific Laplandic concepts (see section 1.5), and so there is likely little to compare between proposed Paleo-Lakelandic and Paleo-Laplandic. Aikio establishes that Lakeland Saami, like other Saami languages, used the suffix **-s* to mark attributive adjectives. This feature is typologically unusual, and its origin is unclear (90) (further discussed in Chapter 5).

4.5 Kinship terminology and other substrate items of interest

The system of kinship terminology in Saami is also worth taking a look at. Kinship terminology among Uralic languages is investigated in *Borrowability of kinship terms in Uralic languages* (Metsäranta, Milanova, Honkola 2023). It is found that the kinship system of Saami is quite different from those of the rest of the Uralic languages, as well as those of other European languages, to the extent that the kinship system in Saami in some aspects resembles more closely those of the indigenous languages of the Americas, Southeast Asia, and Australia (Metsäranta, et al. 2023: 146). Kinship terms are not considered especially prone to borrowing (Metsäranta, et al. 2023: 142), as most languages already have a native system for denoting relationships between people (one could say that kinship terminology is the opposite of the types of words which fit Ante Aikio’s “semantic criterion” for substrate influence). Specifically, Saami kinship terminology exhibits “alternate generation equivalence,” which refers to “kin-term pairs where the same lexeme or a derivation thereof is used to denote certain pairs of relatives”. The example given here is the North Saami *eahki*²⁷ ‘father’s elder brother’ and *eahkit*²⁸ ‘younger brother’s child’ (Metsäranta, et al. 2023: 146). A system of kinship terminology different from one’s closest linguistic relatives and a given language’s being such a typological outlier may indicate a shift from another language which had a very different system of kinship. Whether the kinship vocabulary in Saami is itself of Paleo-European origin is unclear, and requires more research. However, the “semantic gaps” for certain kinship roles had to be filled one way or another, either by retaining native (substrate) words or adapting borrowed words to new semantic spaces.

One substrate word of interest is the North Saami word *állat* (: *állap-*) ‘snow bunting’ (< PS **ällëppē*) (Aikio 2004: 10). Cognates (or parallel borrowings) exist in Lule Saami (*albba*²⁹) and “Norwegian Lapp”³⁰ (*alap*) as well (Kotimaiten kielten keskus). Due to the initial *a-* and the fact that it refers to a type of bird, there is a slim chance that it could have its

²⁷Reflexes of *eahki* are present throughout Saami languages (cf. SaaP *iehkje*), and is reconstructed to Proto-Uralic **ekä* (Kotimaisten kielten keskus), thus the lexeme itself is probably not of Paleo-European substrate origin.

²⁸The final *-t* in *eahkit* is morphologically productive. Here it appears to indicate one side of a reciprocal relationship. What it indicates more specifically, and what its origins are (as both a concept and a morpheme) require more research.

²⁹Although the syncope of vowels is one characteristic of the Language of Bird Names, the lack of internal vowel in Lule Saami *albba* (cf. PS **ällëppē*) is merely due to orthography, and is likely pronounced something like */a^lppa/* (Ylikoski 2022a: 135).

³⁰North Saami

origins in the Language of Bird Names (Schrijver 2001). Substrate vocabulary from the Language of Bird Names is present throughout multiple language groups in Europe, including Scandinavian (e.g., Old Icelandic *lævirki* ‘lark’ (< **laəwað-*, **a-lawð*) (Schrijver)), indicating that it was present on the Scandinavian peninsula. Although there is no evidence for direct contact between the Language of Bird Names and Uralic (Schrijver), it cannot yet be ruled out.

Aikio also mentions the toponymic element *sáll-*, which denotes a ‘major island in the sea’. *Sáll-* (<PS **sāl-*) resembles strikingly the Finno-Saamic word **salaw*, which is likely a loan from Baltic (LIT *salà* ‘island’ < Proto-Baltic **salā*), which in turn lacks an Indo-European etymology. If the resemblance is not coincidental, the word may be shared throughout Northern Fennoscandia and the eastern Baltic, either via borrowing or genetic relationship (24) (see Saarikivi 2004a: 208 for more). Although at this point there is not sufficient evidence to trace this item to a known substrate language, one can make an educated speculation. Peter Schrijver (2014: 177) describes the vowel system of Proto-Baltic as similar to that of Proto-Germanic, which in turn, is nearly identical to that of early Proto-Saami (Schrijver 191). As the convergence of Proto-Germanic and early Proto-Saami vowels are regarded as substrate interference from “Language X” (see section 3.2), contact between Language X and Baltic is plausible, and even likely, due to the location on the Baltic Sea. Hence, the Proto-Saami **sāl-* could be a trace of a prehistoric language (or language group) which spread from the eastern coasts of the Baltic Sea to the far north of Fennoscandia. However this is merely a guess.

Saami is also notable for having replaced high-frequency lexemes, common to most Uralic languages, with words of obscure origin, such as PS **kε:δkē* ‘stone’ (PU **kiwi*), PS **pēŋke* ‘wind’ (PU **tuyli*), PS **mōre* ‘tree’ (PU **puyi*) (Saarikivi 2004a: 189).

Another word of interest is that for ‘water’– *čáhci* in North Saami, which is reconstructed to Proto-Saami **čācē*, and further back to Proto-Finno-Ugric **šācā* (or **šāčā*) (Kotimaisten kielten keskus). It is also present in Proto-Ugric, Khanty, and Nenets, though is not attested in Proto-Uralic. As Khanty (Ob-Ugric) and Nenets are in close geographic proximity to one another along the Ob river, the root was likely transferred from one to the other via contact. By contrast, other Finno-Ugric languages use the root **wete*, which is likely ultimately an Indo-European loanword. Reflexes of **wete* are, however, present in three Samoyedic languages– Nenets (*jī*), Selkup (*ūt*), and Kamas (*bu*) (Kotimaisten kielten keskus). The Khanty reflex *šāδϕ* means ‘puddle’. This difference is not surprising, as semantic shift with regards to cognate words is quite common.

5. Grammatical Substrate

Peter Schrijver suggests that the Language of Geminates has almost certainly left a syntactic mark on the languages of northern Europe, but “nobody knows what to look for because nobody has ever tried” (Schrijver 2014: 196). As the Language of Geminates is spread across northern Europe, we may simply not be seeing the forest for the trees. To find out more, one would have to compare typological features of the languages of northern Europe (Germanic, Celtic, Finnic, Saami) with other languages outside of the geographical area. As syntax tends to leave a stronger mark on a second language (Schrijver 17) than even lexicon, there should logically be a *heavy* grammatical substrate left by the Language of Geminates in northern Europe.

As learners of a second language are unlikely to import morphology from their native language (Schrijver 17), evidence of a morphological substrate is likely less prominent than phonology or syntax. Unfortunately these types of substrate interference are much more difficult to identify than lexical substrate interference (see section 1.4).

Pekka Sammallahti (2012) provides a short list of morphological features in Saami languages which cannot be projected to other Uralic languages, and thus are potentially of substrate origin. These include the South Saami third-person present-tense suffixes *-s* and *-vies* (e.g., *leas* ~ *lievies* ‘they are’ (Sammallahti 98), which cannot be traced to a Uralic origin³¹ (cf. PS **-jā>Ø*, **-pēn*, **-jāk>-ēk* (Koponen 109)), as well as the Western Saami third-person dual element **-kaa/-Gaa-* (98). The dual number is marked both in nominal and verbal morphology in Proto-Uralic (Aikio 2022: 13, 15), and remains in Samoyed, Mansi, Khanty, as well as Saami (Aikio 2022: 16). As the dual is marked throughout Saami, why would the Western Saami languages need to adopt a non-native morpheme for a category which is marked with native morphology in related languages? As morphology is generally not transferred from a substrate language (Schrijver 2014: 17), this morpheme is particularly mysterious. If this morpheme (**-kaa/-Gaa-*) is indeed a Paleo-Laplandic feature, then what was its original purpose? Did it also serve as a dual marker in the substrate language, or did it originate from another morpheme whose meaning changed? In accordance with Schrijver, however, **-kaa/-Gaa* is more likely originally a Saami morpheme whose meaning was adapted to fit a semantic gap left by the receding substrate language (explained with **-s* below).

³¹The suffix *-vies* cannot be related to Finnic *-va/-vā-* for phonotactic reasons (Sammallahti 2012: 98 for more).

The attributive suffix **-s* in Saami, or rather its usage, may be due to substrate interference, as its origin is not agreed upon (Aikio 2012: 90; Sammallahti 1998: 72). This suffix is placed on certain adjectives when followed by a nominal which the adjective modifies (e.g., SaaN *duojár lea čeahppi* ‘the craftsman is skillful’ vs. *čeahpes duojár* ‘a skillful craftsman’ (Sammallahti 1998: 62). The attributive **-s* is not applied to every adjective (e.g., *nuorra nieida* ‘young girl’ (Sammallahti)). A description of Saami attributive forms is provided by Sammallahti (1998: 71). As such a feature is not present in other known languages of the region other than Saami, it may be an example of substrate influence. Korhonen (1981: 245-246) corroborates its lack of known origin. In contrast, however, Sammallahti traces the attributive **-s* to a Finno-Volgaic lative suffix (cf. EST inessive *-s*) (Sammallahti 1998: 71), but provides no source. If the attributive **-s* is indeed a Paleo-European substrate feature, it would have entered Proto-Saami no farther north than the Lakeland region of southern Finland, as **-s* is found in various Saami-substrate toponyms in Lakeland (e.g., *Kukasjärvi* (< PS **kukkē-s*) (Aikio 2012: 90)). The attributive **-s* in Saami is a morphological feature, which according to Schrijver (2014), would not likely have been transferred from a substrate language. However, if one looks past the suffix itself and rather at the presence of an attributive marker as a *syntactic* feature, then one can posit that whatever paleo-language left this substrate mark on Saami used a similar construction. One could please both parties— those on the side of Finno-Volgaic origin claim (i.e., Sammallahti), and those who claim that the suffix is of unknown origin (Aikio and Korhonen)— by hypothesizing that this is a case of substrate morphosyntax being applied to Uralic lexemes. This is in accordance with Schrijver’s theory of native (substrate) syntax being applied to the converging contact language due to imperfect learning. If this is the case, then we can assume that the speakers of this substrate language probably marked their attributive adjectives with a lative morpheme, but which was probably not *-s*. Could this feature be attributed to the Language of Gemimates? One must first remember that the “Language of Gemimates” does not refer necessarily to a single concrete language, but rather a set of shared features within a geographical area. Additionally, the Language of Gemimates has been studied as a *phonological* substrate, rather than a grammatical one, so its morphosyntactic effects on the languages of northern Europe are largely unknown.

Janne Saarikivi in “Substrata Uralica: Studies on Finno-Ugric Substrate in Northern Russian Dialects” (2006a) discusses the grammatical substrate left on Russian by Finno-Ugric languages in northwestern Russia. Here too, the substrate is most discernible in the vocabulary, and the grammatical substrate remains quite vague (42). A study on a

proposed Paleo-European grammatical substrate in Saami would yield even fewer, and less certain results, if any. Like Tkachenko's study of Merya, one already knows what types of features to look for between Finno-Ugric and Russian.

There are several ways in which Saami and Finno-Saamic differ from other Uralic languages, which could be attributed to contact with other languages. These have traditionally been ascribed to contact with Germanic, but such claims may be questionable.

First is basic word order. While many Uralic languages have retained subject-object-verb as their basic word order from Proto-Uralic, Finnic and Saami follow subject-verb-object. Although this change is usually attributed to contact with Germanic (Korhonen 342), the development of SVO order in Finnic and Saami predates that of Germanic (Laakso 2022a: 250), which is problematic for the common narrative. The presence of a contact-induced feature in Saami that is shared with Germanic, which cannot be caused by direct influence from Germanic due to chronology (i.e., present in (Finn)-Saami before Germanic) is reminiscent of the Great Saami Vowel Shift (see section 3.2). Could the SVO word order in Finnic and Saami be due to substrate influence as well? SVO as basic word order appears to be a mainly a feature of the Uralic languages spoken farther to the west (although many have rather flexible word order (Vilkuna 2022: 950)), but is not restricted to Finnic and Saami (Komi and Mordvin may variably use SVO, while South Saami uses SOV). Due to SVO word order being distributed considerably farther to the east than the other Language X features, it is rather unlikely to have entered Finnic and Saami via the Language X substrate. A Baltic origin is also unlikely, since Proto-Baltic used SOV order (Dini 2000: 109–100).³²

Korhonen (1981) and Lindström (2021) mention agreement as a notable feature in Saami and Finnic. Within Finno-Ugric languages, agreement is unique to Finnic and Saami (e.g., EST *suures linnas* vs. HUN *nagy(Ø) városban* (**nagyban*) 'in (a) big city'). Agreement is also present in Germanic (e.g., German) and Slavic. Although the presence of agreement in Finnic and Saami may be a contact feature, it is most likely due to contact with Germanic and/or Slavic, and not some Paleo-European language. Korhonen (1981: 344) provides an explanation of how agreement works in Saami.

Closely related to agreement is plural marking with numbers. In Proto-Saami, the quantified noun in a quantifier phrase with a small numeral is in the nominative plural (e.g., **kuoktē koatē-k* 'two huts') (Koponen 2022: 111), but sometimes in the partitive (ib.). The

³²Dini 2000: 109–100 according to https://en.wikipedia.org/wiki/Proto-Baltic_language#cite_note-3.

Eastern Saami languages use the nominative plural with large numbers (ib.). This feature differs from other Uralic branches, in which a noun phrase quantified with a number is placed in the nominative singular (e.g., HUN *két madár* ‘two birds’ (**madarak* ‘bird-PL’)). The differentiation between large and small quantities is reminiscent of Slavic³³, in which small plural amounts (less than five) are quantified with the genitive singular, while amounts five and over are in the genitive plural (e.g., Russian один рубль ‘one ruble’ : два рубля ‘two rubles’ : пять рублей ‘five rubles’). In Finnic, on the other hand, as well as North and Aanaar/Inari Saami, quantified nouns (more than one) are in the partitive singular. These systems appear to be an odd mix of Slavic, Finnic, and Germanic³⁴. In accordance with Schrijver’s theory of syntactic substrate (2014: 17; section 1.5), these systems of quantification could have been carried over from substrate languages as the speakers shifted to Saami. However, these systems of quantification are rather complicated, and since they resemble those of multiple language groups of the region, it is difficult to determine whether any are based on a Paleo-Laplandic system of quantification.

One minor piece of morphosyntax could *potentially* be gleaned from substrate vocabulary, particularly from toponyms. Due to the structure of Saami toponyms (discussed in section 4.4), one could deduce that like Finno-Ugric, compound nouns have head-final morphology (i.e., basis + generic). However, this remains pure speculation, as this structure occurs in Finno-Ugric and Germanic as well, and so could simply be a Saami innovation.

³³Latvian (Baltic) also uses the nominative plural with numbers. This is most unlike its Slavic cousins, which use the genitive case.

³⁴Nouns in Germanic are in the nominative plural when quantified with a number (e.g., ENG *two birds*; DEU *zwei Vögel*).

6. Estimate of Paleo-European substrate languages in Fennoscandia

Peter Schrijver (2014) speculates that there were “at least two” substrate languages present in the Finno-Saami and Germanic “contact” area, mainly due to the geographical size of the area (Schrijver 2014: 193). This is not very specific.

It seems that we are dealing with roughly two “axes” with regard to the spread of Paleo-European languages in the Baltic region and Fennoscandia. On the east-west axis is the Language of Geminates, or Language X, which spreads across northern Europe, affecting Celtic, Germanic, and Finnic, and Saami in a roughly east-west manner. On the north-south axis is the initial colonization of Fennoscandia, which occurred up the coast of Norway, from western Europe. It appears that it was on this axis that preaspiration found its way into Fennoscandia and later into Saami. Whether the languages on these two axes had any connection to each other remains a mystery, but the fact that both have some alternation of non-initial consonants (i.e., consonant gradation and preaspiration) points toward some sort of connection.

The original inhabitants of northern Fennoscandia had their origins in the western cultural complex, and although there was certainly contact between the people of the western and eastern complexes, they probably spoke different languages.

It is clear from phonological features (Aikio 2004) that substrate vocabulary did not enter the Saami lexicon all at once, but rather over different periods of the development of Saami. These lexical items likely came from different substrate languages or varieties, but the taxonomy of these languages— whether they were related to one another, or even varieties of one another, is not known.

Like in Schrijver’s claim (Schrijver 2014: 193), we can estimate that there were multiple Paleo-European languages present in northern Europe which left their mark on Saami. Here one must consider what is meant by “language”. Contrary to its name, “Language X” was almost certainly not a single language, but rather a language family or *sprachbund* which shared certain features. Irregular parallel borrowings from substrate languages (Aikio’s criteria of irregular correspondence (Aikio 2012: 83; see sections 1.5, 3.4)) may indicate influence from multiple substrate languages. A thorough study of lexical items of apparent substrate origin (e.g. those listed in Aikio (2004)), based on semantic and phonological features (Aikio’s semantic and structural criteria (section 1.5, Aikio (2012: 83))) may reveal some patterns which outline individual languages.

Aikio (2004) lists two non-native phonotactic structures which are of particular interest here: “features which cannot be projected back to Uralic reconstructions but may derive from a phase of Pre-Saami anteceding the reorganization of the Proto-Saami vowel system”, and “features which can only derive from the phase of Proto-Saami after the rather radical reorganization of the Uralic vowel system” (Aikio 2004: 13). As the reorganization of the vowel system is considered to have been caused by contact with Language X (Schrijver 2014), we can assume that there was lexical interference from the same source as the vowel shift itself.

Due to these phonological differences, it can be assumed that substrate lexical items entered Proto-Saami at different stages (before and after the vowel shift) and possibly even from different substrate languages— one with a vowel system closer to Pre-Proto-Saami, and one with a system closer to that of Proto-Saami (which was likely part of the Language X realm). So far, there are three factors by which we can hypothesize multiple substrate languages: vowels, consonants, and geography. The irregular *s* ~ *š* correspondence (section 3.4) may indicate parallel borrowing into Saami from two different (though probably related) substrate languages. As for geography, one can assume multiple substrate languages simply due to the vast area that we are dealing with (as explained earlier in this section). We cannot, however, link these distinguishing features to each other to form a list of features belonging to one substrate language, and a list of features belonging to another, at least at this point in research. A more thorough examination may yield a number more precise than “at least two.” As ancient Fennoscandia was even more sparsely populated than it is now, there was not likely a large number of languages spoken.

7. Conclusions and outlook

In this section, I present the results of this study in the form of answers to the research questions presented in section 1.3, and discuss potential research to further our understanding of the substrate situation in Fennoscandia.

7.1 Conclusions

•*What influence have extinct Paleo-European languages had on Saami?*

The Paleo-European languages of Fennoscandia had extensive substrate influence on Saami. We have discussed various features in Saami which may be of Paleo-European substrate origin; further digging would likely reveal more.

The Paleo-European substrate is most noticeable in the Saami lexicon. This is verifiable by the large number of non-Uralic and non-Indo-European vocabulary items, established by Aikio (2004, 2012). As is common in the case of lexical substrata, substrate words are concentrated in semantic fields which are specific or particularly relevant to their geographical area, such as words denoting geographical or environmental features. The phonological substrate is the second most noticeable. Substrate interference has left Saami with a vowel system very different from the rest of Uralic, and a system of complex consonant alternations which it partially shares with Finnic— these changes were essentially a result of speaking Finno-Saami with a substrate accent. In addition, some new phonemes and phoneme combinations became permissible via vocabulary adopted (or rather preserved) from the substrate languages.

•*Based on existing research, what can we determine about the Paleo-European languages of Fennoscandia?*

Some features of the extinct substrate languages can be deduced based on the comparison of the language(s) affected with related languages which have not been affected by the same substrata.

The two major phonological features in Saami which can likely be ascribed to substrate influence are consonant gradation and the development of the Proto-Saami vowel system. One can deduce that these features existed in the substrate languages (the Language of Gemimates in this case) which were replaced by (Finno-)Saami, and came about via an “imperfect” shift to Uralic— in other words, the people in the now-Finno-Saami area came to speak Uralic with a Language of Gemimates accent. One can assume that the substrate language had a “completely different sound system” (Schrijver 2014: 194) from

Proto-Finno-Saami, one which likely resembled that of Proto-Saami. This is corroborated by the similar developments in the Germanic vowel system as described by Schrijver (2014).

The development of complex consonant alternations may also be a substrate feature, as argued by Schrijver (2014). The case for this feature is weaker than that for the vowel shift (due to gradation in Samoyedic), but as it is spread across northern Europe in a similar manner, it could be ascribed to the same substrate. One could say with some level of confidence though, that the substrate language had *something going on involving consonant alternations*, the nature of which is not clear. Consonant gradation and umlaut often co-occur (e.g., SaaP *hāj̥bma* : *hiejma* ‘home’ (Wilbur n.d.)), which strengthens the case for their common origin.

Preaspiration is connected with consonant gradation in multiple Saami languages. As preaspiration is applied to geminate consonants, it is only logical to link it to substrate influence from the Language of Geminates.

The substrate most easily identifiable is the lexical substrate. It takes relatively little effort to reconstruct likely-substrate vocabulary to its earliest-possible form, and one can assume that it roughly resembled that form when it first entered the Saami lexicon. Of course there is the question of semantic shift, meaning that some words present in modern Saami may not have the same meaning or sense that they did in the substrate language. One can also examine the semantics of vocabulary to determine what concepts may be left over from a substrate language, but have been replaced lexically (e.g., kinship terminology, section 4.5).

A syntactic substrate likely has been left in Saami by Paleo-European languages, as syntax is an area in which non-native speakers will apply the rules of their native language. Saami, and only Saami among the language groups of northern Europe, has adopted a special marker for attributive adjectives. Due to the lack of such a feature in other languages of the region, one cannot link it to the Language of Geminates as a defining feature. However, some substrate language in the Proto-Saami area may have marked its adjectives in this way.

•*Do Finno-Saamic and Germanic share a Paleo-European substrate?*

The question of whether Finnic and Saami share a Paleo-European substrate with Germanic ultimately remains unproven, though Schrijver’s explanation of the similarities in vowel shifts in Proto-Germanic and Proto-Saami is convincing. The explanation of complex consonant alternations in Germanic and Celtic, though much more abstract, seems too similar to Uralic consonant gradation to be coincidental, especially in combination with the vowel shifts. The question of the origin of gradation in Finnic and Saami should not be treated in a binary manner: while it likely originated in Proto-Uralic, it could have been reinforced by

contact with substrate languages with similar such features. Even with less weight given to the gradation side of the argument for a common substrate, it remains quite strong when combined with the similarities in vowel shifts.

•*How do the differences of Western and Eastern prehistoric Europe come into play with regard to the substrate languages in Fennoscandia?*

There are almost certainly different substrate languages at play between the east and west of Fennoscandia, which have left separate traces on Saami. However at this moment no specific differences can be identified.

•*Can one estimate the number of Paleo-European substrate languages present in this region?*

There were almost certainly multiple non-Uralic, non-Indo-European languages spoken in prehistoric Fennoscandia. Some were likely assimilated and replaced by others which are also now extinct, long before Uralic and Scandinavian entered the region. The substrate paleo-languages are barely identifiable anyway, so to identify substrate layers *within* a substrate would be nearly impossible. We do not know the taxonomy of these languages, as in what families they were part of: rather we know what they were *not*— i.e., neither Indo-European nor Uralic.

However, we can make some estimates based on historical sound changes and correspondences. Based on the dating of substrate vocabulary to before and after the Great Saami Vowel Shift (Aikio 2004), as well as irregular sound correspondences (e.g. *s* ~ *š*), I think that it is safe to estimate that there were “more than three” substrate languages in Fennoscandia. A more thorough study could reveal more.

7.2 Outlook and further research

In order to understand the Paleo-European substrates in the languages of Fennoscandia to the full extent possible, further research must be done.

As there were surely multiple Paleo-European languages spoken in Fennoscandia prior to the arrival of Uralic in the region, an analysis of substrate vocabulary from a phonological perspective could be done to see if there is any correlation between phonological features and area distribution of substrate vocabulary.

To gain a more complete picture of the linguistic prehistory of northern Eurasia, one could perform a study comparing the substrate features of Saami with those (substrate or not) of other languages further to the east, in Siberia.

To better understand the grammatical substrate left by the Language of Geminates on the languages of northern Europe, a typological study of the languages with this substrate (Celtic, Germanic, Finnic, Saami) with other languages near and outside of the region could reveal some areal features which could be traced to a common substrate.

Another practical direction in which to proceed would be with a continued systematic analysis of substrate vocabulary. Most relevant may be the vocabulary pertaining to reindeer, as there are numerous semantically-similar words in this field, which could be analyzed in relation to each other. As many substrate words carry a very specific meaning, some are likely morphologically complex. A comparative study, similar to Aikio's analysis of substrate toponyms (2004: 18-25) could yield some information on nominal morphology of Paleo-Laplandic languages. An in-depth analysis of substrate vocabulary could also help us place where and when substrate words entered the Saami lexicon, and could give us a better understanding of how substrate languages affected the Saami lexicon, and where and when lexical items were borrowed.

As there are a apparently a few Saami toponyms as far south as the Dvina river basin (section 4.4), which lack further etymology, an investigation into Paleo-European substrate vocabulary in Baltic, and a comparison of it with that in Saami might reveal more about the prehistoric linguistic situation of the region. In addition, Baltic (Latvian) exhibits umlaut (the realization of *e* varies depending on the following vowel). In a short glance, Baltic seems to be on the periphery of the Language X substrate area. Of course further research would be required to support or refute this claim.

Bibliography

Note: all sources not in English were translated using ChatGPT

- Aikio, Ante 2004. "An Essay on Substrate Studies and the Origin of Saami." In: Hyvärinen, Irma & Petri Kallio & Jarmo Korhonen & Leena Kolehmainen (eds.) *Etymologie, Entlehnungen und Entwicklungen: Festschrift Für Jorma Koivulehto Zum 70. Geburtstag*. Mémoires de la Société Néophilologique de Helsinki 63. Helsinki: Société Néophilologique, 5-34.
- Aikio, Ante 2007. The study of Saami substrate toponyms in Finland. *Onomastica Uralica* 2007 (4), 159-197.
- Aikio, Ante (Luobbal Sámmol Sámmol Ánte) 2012. An essay on Saami ethnolinguistic prehistory. In: Grünthal, Riho & Petri Kallio (eds.). *A Linguistic Map of Prehistoric Northern Europe*. Suomalais-Ugrilaisen Seuran Toimituksia 266. Helsinki: Suomalais-Ugrilainen Seura, 63-117.
- Aikio, Ante 2022. Proto-Uralic. Chapter 1 in: Bakró-Nagy, M., Laakso, J., & Skribnik, E. (eds.). (2022). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press.
- Austerlitz, Robert 1985. The Permian Centre. *Linguistica Uralica*, 21:2, 99–109. <https://doi.org/10.3176/lu.1985.2.02>
- Bakró-Nagy, Marianne 2022. Consonant Gradation. Chapter 42 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 859-867.
- Bakró-Nagy, Marianne, Laakso, Johanna, & Skribnik, Elena (Eds.) 2022. *The Oxford Guide to the Uralic Languages*. Oxford University Press.
- Carpelan, Christian 1999, Jan. 8-10. *Late Palaeolithic and Mesolithic Settlement of the European North – Possible Linguistic Implications* [Paper presentation]. Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations. Helsinki, Finland. <https://www.sgr.fi/sust/SUST242.pdf>.
- Damm, Charlotte & Saarikivi, Janne (eds.) 2012. *Networks, Interaction and Emerging Identities in Fennoscandia and Beyond: Papers from the conference held in Tromsø, Norway, October 13–16 2009*. Suomalais-ugrilaisen Seuran Toimituksia 265. Helsinki: Suomalais-Ugrilainen Seura.
- Dini, Pietro U. 2000. Baltų kalbos. Lyginamoji istorija. Vilnius: Mokslo ir enciklopedijų leidybos institutas. pp. 109–110.
- [ETY] Eesti Etümoloogiasõnaraamat. (n.d.). <https://www.eki.ee/dict/ety/>
- Garner, Joshua 2023. "Lexical Evidence for a Common Non-Indo-European Substrate in the Celtic and Saami languages". Durham University.

- Helimski, Eugene 1995. Proto-Uralic gradation: continuation and traces. In: Leskinen, Heikki (ed.). *Congressus Octavus Internationalis Fenno-Ugristarum I: Orationes plenariae et conspectus quinquennales*. Jyväskylä: Gummerus, 17–51.
- Janhunen, Juha 2007. The primary laryngeal in Uralic and beyond. In: Ylikoski, Jussi & Ante Aikio (eds). *Sámit, sánit, sátnehámit. Riepmočála Pekka Sammallahti miessemánu 21. beaivve 2007*. Suomalais-Ugrilaisen Seuran Toimituksia 253. Helsinki: Suomalais-Ugrilainen Seura, 203–227.
- Kallio, Petri (2003). Languages in the Prehistoric Baltic Sea Region. Alfred Bammesberger & Theo Vennemann (Eds.), *Languages in Prehistoric Europe*, Pp. 227-244. Heidelberg.
- Kallio, Petri 2015. The Language Contact Situation in Prehistoric Northeastern Europe. In: Mailhammer, Robert & Theo Vennemann gen. Nierfeld & Birgit Anette Olsen (eds.). *The Linguistic Roots of Europe: Origin and Development of European Languages*. Copenhagen Studies in Indo-European 6. Copenhagen: Museum Tusulanum, 77–102.
- Kazakevič, Olga 2022. Selkup. Chapter 38 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 777-816.
- Klumpp, Gerson & Budzisch, Josefina 2023. Selkup. In: Abondolo, Daniel & Riitta-Liisa Valijärvi (eds.) *The Uralic Languages*. Second edition. London & New York: Routledge, 897–938. DOI: 10.4324/9781315625096-20.
- Koponen, Eino 2022. Saami. Chapter 7 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 103-112.
- Korhonen, Mikko 1981. *Johdatus lapin kielen historiaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kotimaisten kielten keskus. *Álgu-tietokannan toteutus*.
<https://kaino.kotus.fi/algu/index.php?t=etusivu>
- Kusmenko, Jurij, and Michael Riessler. “Traces of Sámi-Scandinavian contact in Scandinavian dialects.” *Studies in Slavic and General Linguistics*, vol. 28, 2000, pp. 209–24. *JSTOR*, <http://www.jstor.org/stable/40997166>. Accessed 11 Mar. 2024.
- Kuznetsov, Nikolay 2022. Komi. Chapter 26 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 487-506.
- Laakso, Johanna 2022a. Finnic. Chapter 14 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 240-253.
- Laakso, Johanna 2022b. Livonian. Chapter 22 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 380-391.

- Lindström, Sylvi 2021. *An analysis of the arguments for and against a common Proto Finno-Saami language*. Master's thesis, University of Vienna, Vienna.
- Magga, Ole. 2006. Diversity in Saami terminology for Reindeer, snow, and ice. *International Social Science Journal* 58. 25 - 34. 10.1111/j.1468-2451.2006.00594.x.
- Mailhammer, Robert 2011. The prehistory of European languages. In B. Kortmann & J. Auwera (eds.), *The Languages and Linguistics of Europe: A Comprehensive Guide*. Berlin & Boston: De Gruyter Mouton, 671-682. <https://doi.org/10.1515/9783110220261.671>
- Matras, Yaron 2009. *Language Contact*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Mayrhofer, Patrick 2024. The Voiceless Dental Fricative in Saami. Presentation at the REMODUS Winter School of Finno-Ugric Studies, Uppsala, February 5–10, 2024.
- Metsäranta, N., Milanova, V., & Honkola, T. (2023). Borrowability of kinship terms in Uralic languages. *Finnisch-Ugrische Forschungen*, 2023(68), 141–216. <https://doi.org/10.33339/fuf.120920>
- Nimiarkisto. (n.d.). <https://nimiarkisto.fi>
- Rießler, Michael 2003, November 7-8. *On the Origin of Preaspiration in North Germanic* [Abstract]. Fifteenth annual UCLA Indo-European Conference, Los Angeles, California, USA.
- Saarikivi, Janne 2004a. Is there Palaeo-European substratum interference in the western branches of Uralic? *Journal de la Société Finno-Ougrienne* 90, 187–214.
- Saarikivi, Janne 2004b. Über die saamischen Substratennamen des Nordrusslands und Finnlands. *Finnisch-Ugrische Forschungen* 58, 162–234.
- Saarikivi, Janne 2006a. *Substrata Uralica: Studies on Finno-Ugrian Substrate in Northern Russian Dialects*. Tartu: Tartu University Press, and Helsinki: University of Helsinki.
- Saarikivi, Janne 2006b. On the Uralic substrate toponymy of Arkhangelsk region: problems of research methodology and ethnohistorical interpretation. *Onomastica Uralica* 4. Debrecen 2006. [in print].
- Saarikivi, Janne & Lavento, Mika 2012. Linguistics and Archaeology: A Critical View of an Interdisciplinary Approach with Reference to the Prehistory of Northern Scandinavia. In: Damm, Charlotte & Janne Saarikivi, Janne (eds). *Networks, Interaction and Emerging Identities in Fennoscandia and Beyond: Papers from the conference held in Tromsø, Norway, October 13–16 2009*. Suomalais-ugrilaisen Seuran Toimituksia 265. Helsinki: Suomalais-Ugrilainen Seura, 177–217.
- Saarikivi, Janne 2022. The divergence of Proto-Uralic and its offspring. Chapter 2 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 28-58.

- Sammallahti, Pekka 2012. Bottlenecks and Contacts in the Linguistic Prehistory of the Saami. In: Damm, Charlotte & Janne Saarikivi, Janne (eds). *Networks, Interaction and Emerging Identities in Fennoscandia and Beyond: Papers from the conference held in Tromsø, Norway, October 13–16 2009*. Suomalais-ugrilaisen Seuran Toimituksia 265. Helsinki: Suomalais-Ugrilainen Seura, 93–104.
- Sammallahti, Pekka 1998. *The Saami Languages: an Introduction*. Kárášjohka: Davvi Girji OS.
- Schrijver, Peter 2001. Lost languages in northern Europe. In: Christian Carpelan and Asko Parpola and Petteri Koskikallio (eds.), *Early contacts between Uralic and Indo-European: linguistic and archaeological considerations: papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki, 8-10 January, 1999*. Helsinki: Suomalais-Ugrilainen Seura, 417-425.
- Schrijver, Peter 2014. *Language Contact and the Origins of the Germanic Languages*. Routledge.
- Sveriges hembygdsförbund. (n.d.). hembygd.se. <https://www.hembygd.se>
- Tkačenko, Orest B. 1985. *Merjanskij jazyk*. Kiev: Naukova Dumka.
- Valtonen, Taarna, Ylikoski, Jussi, & Aikio, Ante 2022. Inari (Aanar) Saami. Chapter 11 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 178-195.
- Vennemann, Theo 1994. Linguistic reconstruction in the context of European prehistory. *Transactions of the Philological Society* 92(2): 215-284.
- Vilkuna, Maria 2022. Word order. Chapter 49 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 950-960.
- Wagner-Nagy, Beáta 2022. Nganasan. Chapter 37 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 754-776.
- Wagner-Nagy, Beáta & Szeverényi, Sándor 2022. Samoyedic. Chapter 34 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 659-673.
- Wilbur, J. (n.d.). Pite Saami lexical database. <https://sjelex.keeleressursid.ee>
- Ylikoski, Jussi 2022a. Lule Saami. Chapter 9 in: Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 130-146.
- Ylikoski, Jussi 2022b. South Saami. Chapter 8 in: In Bakró-Nagy, Marianne, Johanna Laakso & Elena Skribnik (eds). *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press, 113-129.

Kokkuvõte

Fennoskandia on olnud asustatud üle kümne aastatuhande, samas kui uurali keeled on eksisteerinud vaid kuus-seitse ja Fennoskandias vaid kaks-kolm aastatuhandet. Piirkonna muinasaegsed elanikud rääkisid kindlasti mingit keelt – aga mis keel see võiks olla? Nende ammu väljasurnud keelte jäänused ehk substraat on jätnud saami keeltesse oma jälje, mis eristab neid nende soome-ugri sugulaskeeltest (1. peatükk).

Sellel teemal on kirjutanud märkimisväärsel hulgal kirjandust – keeleteadlased nagu Ante Aikio (Luobbal Sámmol Sámmol Ánte), Peter Schrijver, Janne Saarikivi, Pekka Sammallahti ja Petri Kallio (2. peatükk).

See uurimus püüab koondada seda, millist teavet saab Fennoskandia paleoeuroopa keelte kohta vastavast kirjandusest koguda. Substraadi võimalikud või tõenäolised tunnused on esitatud kolmes substraadi mõjuvaldkonnas – fonoloogilises, leksikaalses ja morfosüntaktilises. Neid omadusi käsitletakse seoses tõenäolisusega, et need pärinevad paleoeuroopa substraatkeeltest ja seoses sellega, mil määral saab neid kasutada substraatkeelte endi kohta teabe kogumiseks.

Kui substraadiuuringud keskenduvad tavaliselt sellele, milline oli substraatkeele mõju elavatele keeltele, siis see uuring läheneb teisest suunast: uurimistöös püütakse välja selgitada, mida saab öelda väljasurnud substraatkeelte endi kohta. Ideaalis võiks rekonstrueerida "paleo-fennoskandia" keele, kasutades saami keelde jäänud substraati, kuid kuna materjali on vähe, pole täielik rekonstrueerimine võimalik.

Selle lõputöö peamine eesmärk on vastata järgmistele küsimustele:

1. Millist mõju on väljasurnud paleoeuroopa keeled saami keelele avaldanud?

Paleo-euroopa mõju saami keelele on kõige märgatavam leksikoloogias: ligikaudu kolmandikul levinud protosaami sõnatüvedest ei ole rahuldavat päritolu selgitust. Kuna Fennoskandia paleo-euroopa keelte kõnelejad läksid üle soome-ugri keelele, jätsid nad oma algkeele jäljed ka saami keele morfosüntaksisse (5. peatükk) ja fonoloogiasse (3. peatükk), mis viis protosaami vokaalisüsteemi ümberkorraldamiseni, sellega kaasnesid mitmesugused kaashäälikuomadused, aga ka süntaktilised muustrid, nagu kokkulepe ja omistamine.

2. Mida on võimalik olemasolevate uuringute põhjal kindlaks teha Fennoskandia paleoeuroopa keelte kohta?

Saamikeelsete vastete põhjal saab tuletada mitmesuguseid paleo-euroopa keeltes esinevaid fonoloogilisi tunnuseid: algsaami omast täiesti erinev vokaalisüsteem, teatud vokaalikombinatsioonid, keerulised konsonantide vaheldused ja võib-olla ka eelspiratsioon (peatükk 3).

Leksikaalsel tasandil on suur hulk sõnu, millel ei ole sobilikku etümoloogiat. Neid kiputakse otsima semantiliselt väljadelt, mis on eriti aldis substraadimõjudele: kultuuri- ja piirkonnaspetsiifilised mõisted ja toponüümid. Paljud neist sõnadest sisaldavad ka võõrkeelseid fonotaktilisi tunnuseid (4. peatükk).

3. Kas soome-saami ja germaani keeled jagavad paleo-euroopa substraati?

Nagu Peter Schrijver (2014) väidab (3. peatükk), viitavad saami ja germaani vokaalisüsteemide sarnasused ning nende sarnased kaashäälikute vaheldused ühisele substraadile. Ehkki see küsimus jääb lõpuks lahendamata, leidub ühise substraadi jaoks kindel alus.

4. Kuidas tulevad mängu eelajaloolise Lääne- ja Ida-Euroopa erinevused Fennoskandia substraatkeelte osas?

Tõenäoliselt on erinevusi, kuid neist ükski ei olnud selles uurimisetapis tuvastatav.

5. Kas on võimalik hinnata selles piirkonnas esinevate paleoeuroopa substraatkeelte arvu?

Tulenevalt substraadi mõju geograafilistest ja piirkondlikest erinevustest oli kindlasti palju paleokeeli, mis asendati soome-ugri keelega (6. peatükk). Siin oleks kõige kindlam oletus, et neid on "rohkem kui kolm".

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Seth Wilson,

1. I grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis “The Paleo-European substrate in Saami: a critical harvest of the literature”, supervised by Gerson Klumpp, PhD.

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons’ intellectual property rights or rights arising from the personal data protection legislation.

Seth Wilson
27/05/2024