

UNIVERSITY OF TARTU
FACULTY OF ARTS AND HUMANITIES
INSTITUTE OF ESTONIAN AND GENERAL LINGUISTICS

Riku Tapani Kuusijärvi

RELATIVE CLAUSE CONSTRUCTIONS IN SOIKKOLA INGRIAN

Master's Thesis

Supervisors:

associate professor Elena Markus, associate professor Fedor Rozhanskiy

TARTU 2026

Author's Affirmation

I confirm that I have written this thesis independently and have correctly cited all sources. The thesis complies with the requirements of the Institute of Estonian and General Linguistics and with good academic practice.

Riku Tapani Kuusijärvi

Abstract

This thesis investigates the syntax of relative clause constructions in Soikkola Ingrian. The constructions were manually extracted from several text collections spanning from the end of the 19th century to the early 21st century. The extracted sentences were annotated for a range of variables to find out which syntactic functions are relativizable, how the type of head, animacy and concreteness affect the choice of relative pronouns, whether the relative pronouns agree in number with their heads, and what the internal syntactic structure of the constructions is like.

By analyzing 248 relative clause constructions, the study shows that the relative pronoun strategy can be used to relativize any participant on the Noun Phrase Accessibility Hierarchy at least up to the possessor function. Typical Soikkola Ingrian relative clauses are externally headed, postnominal, headed by a full noun, have a clause-initial relative pronoun and immediately follow the head. Deviations from these patterns are rare and appear to be associated with information-structural factors. The relative pronouns *migä* and *kumba* show number agreement with the head, whereas *ken* does not in 46 % of contemporary examples. Inverse attraction is not attested in spontaneous speech samples, but it is found in elicited examples from several speakers.

Keywords: Ingrian, syntax, complex sentences, clause combining, relative clauses

Acknowledgements

First and foremost, I want to thank my supervisors, Elena Markus and Fedor Rozhanskiy, for their professional guidance in every phase of this work. Their insightful comments and suggestions have pushed me in the right direction and helped me develop the idea for the topic of the thesis and improve the work to its final form.

I am also grateful to my family and friends for their continuing support, encouragement and understanding throughout the process of writing this work.

The completion of this thesis was financially supported by scholarships from the Kindred Peoples' Programme and the Kadri, Nikolai and Gerda Rõuk Memorial Fund.

List of Glosses

1 / 2 / 3	person	LOC	locative
ABL	ablative	NEG	negation
ACC	accusative	NOM	nominative
ACT	active	NPST	non-past tense
ADE	adessive	O	objective conjugation
ALL	allative	PART	partitive
AN	action nominal	PASS	passive
ART	article	PERF	perfective
CNG	connegative	PL	plural
COM	comitative	POSS	possessive suffix
COMP	comparative	PRCL	particle
DAT	dative	PROL	prolative
DEF	definite conjugation	PRON	pronoun
DU	dual	PRS	present tense
ELA	elative	PST	past tense
FREQ	frequentative	PST2	second past tense
GEN	genitive	PTCP	participle
GER	gerund	Q	question marker
ILL	illative	REL	relativizer
IMP	imperative	RES	resultative
INE	inessive	S	subjective conjugation
INF	infinitive	SG	singular
INSTR	instrumental	SUP	supine
IPS	impersonal	SUPERL	superlative
LAT	lative		

Table of Contents

1. Introduction.....	8
1.1. Thesis structure and ethical considerations.....	8
1.2. Glossing principles.....	9
2. Ingrian.....	10
2.1. The Ingrians	10
2.2. Soikkola Ingrian	12
2.3. Previous research on Soikkola Ingrian.....	14
3. Theoretical background of the study.....	16
3.1. Clause combining typologically.....	16
3.2. Clause combining in Ingrian	19
3.3. Relative clause constructions	20
3.3.1. Definition	20
3.3.2. Common argument and relativization strategies.....	21
3.3.3. Headedness of the relative clause	22
3.3.4. Restrictiveness	24
3.3.5. Internal syntax.....	24
3.3.6. Noun Phrase Accessibility Hierarchy	25
3.4. Relative clause constructions in Finnic.....	27
3.4.1. Relative pronouns	27
3.4.2. Choice of relative pronoun.....	28
3.4.3. Number agreement.....	29
3.4.4. Inverse attraction.....	30
3.5. Relativization strategies in Ingrian.....	31
3.6. Research questions	36
4. Data and method	37

4.1. Description of data sources	37
4.1.1. Data from the 19 th century	37
4.1.2. Data from the 20 th century	38
4.1.3. Data from the 21 st century.....	39
4.1.4. Evaluation of the data sources	40
4.2. Method	43
4.2.1. Extracting the data	43
4.2.1. Annotating the data	45
5. Results.....	48
5.1. General overview	48
5.2. Syntactic function of the common argument	51
5.3. Choice of relativizer	53
5.3.1. Type of head	54
5.3.2. Animacy	58
5.3.3. Concreteness	59
5.4. Number agreement	60
5.5. Internal syntax	64
5.5.1. Position of relativizer	64
5.5.2. Position of the relative clause	66
6. Discussion	69
7. Conclusions.....	75
References.....	77
Relatiivlausekonstruktsioonid isuri keele Soikkola murdes. Kokkuvõte	85

1. Introduction

In existing descriptions of the Uralic languages, syntax generally receives only minimal attention, if any (Klumpp et al. 2018: 11). Although the importance of syntax had already been recognized in the 20th century, many grammars and studies of Uralic languages continued to follow the research tradition of the 19th century. Research on the syntax of the Uralic languages has properly begun only in recent decades (Ovsjannikova & Rozhanskiy 2024).

When it comes to the Finnic languages, only the syntax of Finnish and Estonian has been thoroughly studied, whereas the syntax of the smaller languages remains unexplored for the most part (Norvik 2016: 177). The situation is not better in Ingrian. Currently, there is no synchronic grammar of Ingrian, and the existing works often lack a separate chapter dedicated to syntax altogether (Rozhanskiy 2024: 187–188). Some syntax-related aspects of Ingrian are discussed in various sources, for example Grünthal (2021), Markus (2024), Markus and Rozhanskiy (2022a), Rozhanskiy and Markus (2025b).

The syntax of Ingrian complex sentences has not yet been investigated, and this thesis aims to fill part of the research gap by examining the syntax of relative clause constructions in the Soikkola dialect of Ingrian. More specifically, the focus is on which syntactic functions are relativizable, how the relative pronoun is chosen, whether the relative pronouns agree with their heads in number, and what the internal syntactic structure of the relative clause construction is like.

1.1. Thesis structure and ethical considerations

This thesis is structured as follows: Chapter 2 provides background information about Ingrians and their language and specifies what makes Soikkola Ingrian distinct. The theoretical background and the key concepts needed for the analysis of relative clause constructions are laid out in Chapter 3. Based on the terms and approaches described in the chapter, the research questions of the study are formulated in 3.6. The sources from which the relative clause constructions are taken as well as the method of extracting and annotating are described in Chapter 4. Chapter 5 presents the result in terms of qualitative notes and descriptive statistics about the annotated variables. Chapter 6 focuses on finding explanations for deviations from the common patterns observed in the results. Finally, concluding remarks as well as the main findings of the study are summarized in Chapter 7.

The language samples used in this study come from previously published sources or unpublished field data collected with the informed consent of the speakers. Personal information that is not publically available is not gathered or given about the speakers in this work. The research design and its results do not raise ethical concerns.

1.2. Glossing principles

All the language examples presented in this work are glossed according to the Leipzig Glossing Rules.¹ For previously published samples, transcriptions and glosses are kept the same as in the original sources. However, additional markings are added to each example containing a relative clause construction. The head of the relative clause is shown in bold, the relativizer is marked with REL, and the relative clause is enclosed in square brackets.

Examples that have not been published previously or that have not been glossed in the source materials are glossed so that inflectional morphemes are segmented, but derivational ones are not. The present tense and singular number of verbs as well as the nominative case of nouns are not explicitly marked by default, but they are glossed where appropriate. In addition, case marking on pronouns is not segmented.

¹ Available at <https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>

2. Ingrian

The aim of this chapter is to provide background information about Ingrians and their language as well as to briefly present the current state of research on Soikkola Ingrian. In section 2.1., the Ingrian people are described from a historical and sociological point of view. What makes the Ingrian language and especially the Soikkola dialect distinct is discussed in section 2.2. An overview of major works on Soikkola Ingrian is presented in section 2.3.

2.1. The Ingrians

The Ingrians are a Finnic people who have long inhabited the western parts of the historical region of Ingria.² Nowadays, the Ingrians live in the Kingisepp district of the Leningrad Oblast in Russia (Markus & Rozhanskiy 2022a: 308). According to the Soviet census of 1926, there were 26,137 Ingrians at the time, but the number reduced rapidly during the following decades, and in 1989, only 820 people had registered themselves as Ingrians (Raudalainen 2021: 38). According to the latest estimates, there are no more than ten fluent speakers of Ingrian left today (Rozhanskiy & Markus 2025a: 449–450). For a more detailed description on the decline of the number of Ingrians, see for example Ernits (2007).

In written sources, Ingrians are known under several different ethnonyms. They were mentioned in Russian chronicles for the first time as a separate people, *ižerjanje*, in 1228 (Raudalainen 2021: 36). Other names used in literature include, but are not limited to *Ižorians*, *Ischoren*, *Ingrier*, *inkerikot*, *inkeroiset* and *isurid*. The Ingrians have many terms for describing themselves, such as *ižorid*, *ingerläišt*, *šōmalaišt*, *šoikkolašt* and *mäkküläišt* (Rožanskij & Markus 2013: 270). As the last two endonyms show, sometimes Ingrians do not use a collective term to refer to their people but rather name their nationality in terms of their place of birth or residence. Ingrians have also been observed to call themselves *karjalaiset* ‘Karelians’, which has led researchers into believing that the Ingrians originate from Karelia. This conclusion has later been put under question (Grünthal 2021: 125–126).

Throughout this work, the term *Ingrian* will be used to refer to both the language and to members of the speech community. If not specified otherwise, all the remarks on the Ingrian language and people are to be understood as referring to the Soikkola dialect, a variety of Ingrian which is described in more detail in section 2.2.

² For descriptions of the history of Ingria, see for example Häikiö (2022) and Raudalainen (2021).

In addition to Ingrians, there are three other Finnic peoples living in Ingria: Votes, Estonians and Ingrian Finns (Kuznetsova et al. 2015: 128). Note that Ingrians should not be confused with Ingrian Finns, who are sometimes called Ingrians in literature. The Ingrian Finns are descendants of Finns who moved to the area when Ingria came under Swedish rule after the Treaty of Stolbovo in 1617.³ They consist of two groups, *äyrämöiset* and *savakot*, who came from the Karelian Isthmus and Savonia to replace the Orthodox population in Ingria who fled to inner Russia in fear of being persecuted by the Lutheran Swedes (Mantila 2025: 139). Traditionally, Ingrian Finnish is considered to be a member of the Southeast Finnish dialects (Itkonen 1983: 196), whereas Ingrian is categorized as a language of its own nowadays.

As Patrakka (2021: 96–97, 105–106) reports, the beginning of the 20th century was a time when the Ingrian culture and language were still well alive. The language was passed down to children and until the 1930s they usually did not speak Russian before going to school. Even the working language in the *kolkhozes* (collective farms) was Ingrian in the beginning. Things took a radical turn after the Second World War, when most Ingrians who had been deported to Finland during the war were sent back to the Soviet Union. Instead of going back to their homes, they were banished to other parts of Russia, and their homes were either destroyed or given away to Russians who moved into the area. This is, of course, only a very brief description of a much more complicated story.⁴ There were deportations already before the war and eventually, after Stalin’s death, the Ingrians were allowed to move back to their homes.

The Ingrians are traditionally divided into four different groups based on the dialect they speak. These are the Hevaha, Oredeži, Soikkola and Lower Luga dialects, of which only the last two are still alive today (Markus & Rozhanskiy 2022a: 308). The historical speaking areas of the different Ingrian dialects are showcased in Figure 1 below (from Rantanen et al. 2021, see also Rantanen et al. 2022).

³ There was a small Finnish population living in Ingria already in the 1580s. It is estimated that the Finns made up not more than five percent of the population of Ingria in the beginning of the 17th century. These people had moved there most likely due to the years of famine that were ravaging in Finland at the time. For more details, see Kokko (2007: 15–16).

⁴ Often, the Ingrians were grouped together with Ingrian Finns, who were the largest Finnic people in Ingria at the time. For a comprehensive description of the history of deportation of Ingrians and Ingrian Finns, see Flink (2010). For an overview of the situation of the Ingrians and their language in post-war years, see Rozhanskiy & Markus (2025a).

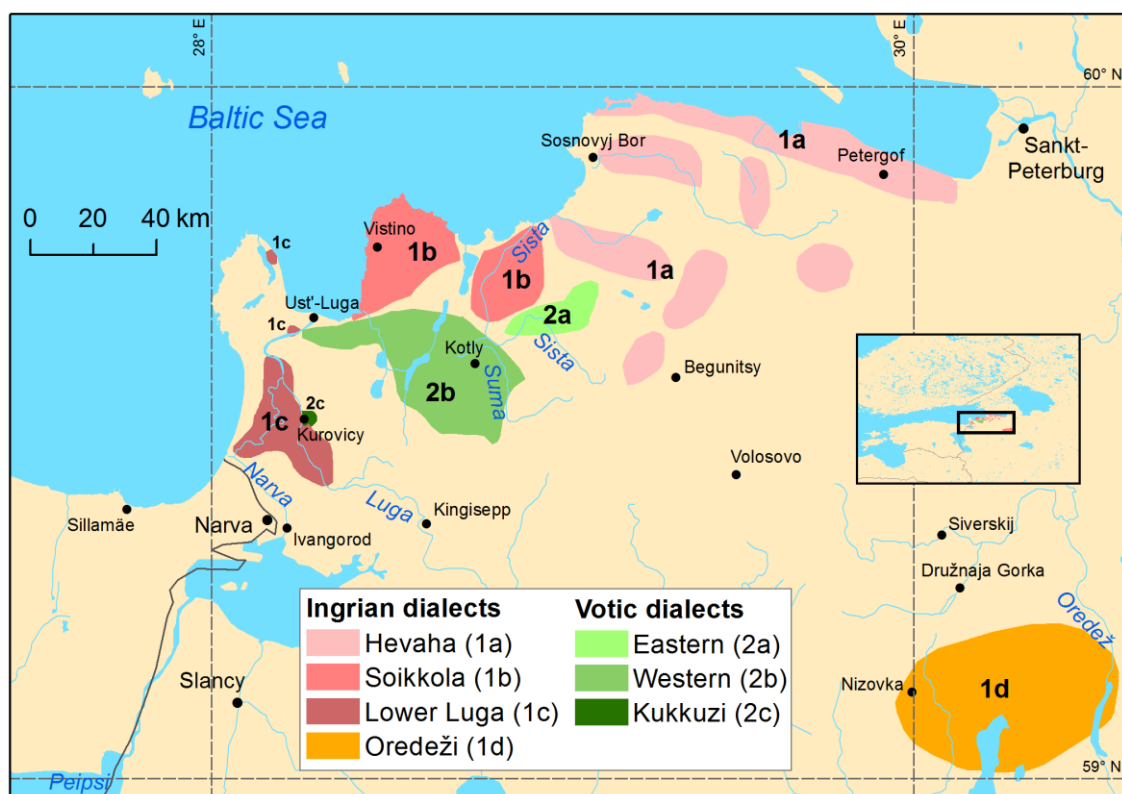


Figure 1. Traditional speaking areas of Ingrian and Votic dialects.

2.2. Soikkola Ingrian

Soikkola Ingrian is a dialect of Ingrian spoken mostly on the Soikkola (Rus. *Soikinsky*) peninsula and its vicinity along the river Sista (see area 1b in Figure 1). Note, however, that during linguistic fieldwork conducted in 2006, only one Ingrian speaker was found in the Sista river speaking area (Kuznetsova et al. 2015: 139). One significant factor that differentiates Soikkola Ingrian from Lower Luga Ingrian, the other dialect of Ingrian still alive, is the fact that Lower Luga Ingrian is likely a mixture of Ingrian and Votic languages. As Markus and Rozhanskiy (2012: 93) argue, Lower Luga Ingrian is a convergent variety which has developed from the speech of Ingrians who have adapted their way of speech to be understood by Votes and from the language of Votes who have started speaking the mix of Ingrian and Votic. The fact that the dialects of Ingrian are clearly distinct and that Ingrian lacks a common standard language are reasons why Ingrian is best described and studied through the lens of dialectal developments, as stated by Grünthal (2021: 122). This is why this work focuses only on the Soikkola dialect of Ingrian.

There are minor differences within the Soikkola Ingrian dialect spoken in different parts of the peninsula. Kuznetsova (2009: 19–20) divides Soikkola Ingrian into northern and southern subdialects, with the village of Viistinä (Rus. *Vistino*) functioning as a dividing line and the language there as a central transitional subdialect. She also points out that the northern subdialect has been subject to more Finnish influence due to long-lasting interactions related to fishing and the presence of Ingrian Finnish settlements nearby. The southern subdialect, on the other hand, is more innovative and has been influenced by Votic and Lower Luga Ingrian. However, as the reported differences are based mainly on lexicographic evidence, different subdialects of Soikkola Ingrian are not analyzed separately by default in this work.

A phonological feature that makes Soikkola Ingrian typologically interesting and distinct among the Finnic languages is worth mentioning, namely the ternary quantity contrast of consonants. Soikkola Ingrian differentiates between single consonants, short geminates and full geminates in intervocalic positions, as shown in Table 1 below (from Markus & Rozhanskiy 2022a: 311). Although this feature is attested also in related languages, such as Estonian and Livonian, its specialty in Soikkola Ingrian is that short and full geminates can produce minimal pairs, such as *tappà* ‘catch:3SG’ and *tappà* ‘kill:INF’ (Kuznetsova 2015: 100).

Table 1. Ternary length opposition of consonants in Soikkola Ingrian.

single consonant	short geminate	full geminate
<i>kana</i> ‘hen’	<i>kañnaa</i> ‘hen.PART/ILL’	<i>linnaa</i> ‘town.PART/ILL’
<i>maadaa</i> ‘sleep.IPS.PRS’	<i>püüñtää</i> ‘catch.INF’	<i>šaattaa</i> ‘accompany.INF’
<i>lagada</i> ‘leave.INF’	<i>lakkaamaa</i> ‘sweep.SPN’	<i>lakkaamaa</i> ‘leave-SPN’

Ingrian used to have a written standard in the 1930s, and it was mostly based on the Soikkola dialect at first. Later on, the aim was to develop it so that it would be understood by both Soikkola and Lower Luga Ingrians. Features from both dialects were adapted and a common orthography was introduced for both dialects (for more details, see Saar 2017: 18–19). As Grünthal (2021: 144) reports, the Ingrian Finnish researcher Väinö Junus compiled the written variety for Ingrian as per the *korenizatsiia* policy in the Soviet Union. The idea was to create a written standard for minor languages in order to publish materials promoting the communist ideology that the people could read in their own language. Moreover, as Raudalainen (2021: 39) points out, even though many Ingrians had positive experiences from the written language, it was not fully adopted by the community since it contained a lot of Finnish influence and did not fully reflect the spoken language of the time. In any case, the written variety of Ingrian was

short-lived and its use in schools and books had already been banned by 1938 (Markus & Rozhanskiy 2022a: 308).

2.3. Previous research on Soikkola Ingrian

Academic research on Soikkola Ingrian started in the 1880s with the publication of Volmari Porkka's doctoral dissertation *Über den Ingrischen Dialekt mit Berücksichtigung der übrigen finnisch-ingermanländischen Dialekte* ["On the Ingrian dialect with respect to other Ingrian Finnish dialects"] (1885). The next major research work that touches upon Ingrian came out 40 years later with the publication of Julius Mägiste's master's thesis *Rosona (Eesti Ingeri) murde pääjooned* ["The main features of Rosona (Estonian Ingria) dialect"] (1925). Another important work from the first half of the 20th century is Antti Sovijärvi's monograph *Foneettis-äännehistoriallinen tutkimus Soikkolan inkeröismurteesta* ["A phonetic and historical-phonological study of the Soikkola Ingrian dialect"] (1944), which focuses on the phonetics and phonology of the language with a historical perspective in mind. Dialectal features and historical-phonological aspects are in focus in many other major publications of the 20th century, such as Arvo Laanest's *Ižorskie dialektu: Lingvogeografičeskoe issledovanie* ["Ingrian dialects: A linguistic-geographical study"] (1966b) and *Istoričeskaja fonetika i morfologija ižorskogo jazyka* ["Historical phonetics and morphology of the Ingrian language"] (1978, 1986 in Estonian). Ingrian phonology is covered in more recent publications as well, such as Natalia Kuznetsova's doctoral dissertation *Fonologičeskie sistemy ižorskich dialektov* ["The phonological systems of Ingrian dialects"] (2009). The latest larger publication about Soikkola Ingrian is Eva Saar's doctoral dissertation *Isuri keele Soikkola murde sõnamuutmissüsteem* ["The inflectional system of Soikkola Ingrian"] (2017).

In the 20th and 21st centuries, the focus of researching Ingrian has strongly been in phonology and morphology.⁵ Compared to the other Finnic languages, Ingrian remains among the least studied ones to this day. For a long time, Finnish linguists in particular did not consider Ingrian a language of its own but rather thought of it as a dialect of Finnish. According to Grünthal (2021: 145), Ingrian is quite a typical Finnic language with a small number of distinctive features, at least in terms of phonology and morphology, and this can explain why Ingrian has not received as much attention as its relatives.

⁵ For further discussion on previous studies on these fields, see for example Grünthal (2023) and Saar (2017).

The syntax of Ingrian has not been thoroughly studied yet. As Rozhanskiy (2024: 187–188) observes, none of the major works on Ingrian contain a chapter dedicated to the syntax of the language. A brief overview of the word order, clause and phrase structures as well as clause combining can be found in the chapter describing Ingrian in *The Oxford Guide to the Uralic Languages* (Markus & Rozhanskiy 2022a: 324–328). Different aspects of syntax are also covered in several publications (e.g. Grünthal 2021; Markus 2024; Rozhanskiy & Markus 2025b). Other than some minor remarks in the aforementioned works, the syntax and nature of relative clause constructions have not been described for Ingrian. The morphology of relative pronouns has been studied (Porkka 1885; Junus 1936; Schwarz & Rozhanskiy 2024), but other aspects of the construction are still undiscovered.

3. Theoretical background of the study

In this chapter, the theoretical background and key concepts relevant for the thesis are presented. Section 3.1. and 3.2. give a brief overview of clause combining strategies, first in general and then specifically for Ingrian. The structures of different types of relative clause constructions are showcased in 3.3. In 3.4., aspects of relative clause constructions are presented from the point of view of Finnic languages, especially Finnish and Estonian. The different relativization strategies in Ingrian are presented in section 3.5. Finally, section 3.6. formulates the research questions of this thesis on the basis of the theoretical background and previous research on relativization strategies in Ingrian.

3.1. Clause combining typologically

Defining a **clause** in typological terms is difficult since the minimal requirements for a unit to be considered a clause vary among different languages (for further discussion, see for example Laury et al. 2017). Therefore, a somewhat vague working definition for a clause is usually adopted. As Longacre (2008: 372) puts it, clauses can be roughly understood as representations of individual predications on the surface level of language. A **sentence**, on the other hand, is considered a structural unit of the surface level that is minimally the size of a clause but smaller than a paragraph. If the clause stands on its own, it is considered a **simple sentence**. Moreover, a **complex sentence** is a sentence consisting of at least two clauses (Gast & Diessel 2012: 3). Examples of a simple and a complex sentence are given in (1) and (2), respectively.

- (1) Simple sentence, Standard Finnish

Mene-n auto-lla.

go-1SG car-ADE

‘I’m going by car.’

- (2) Complex sentence, Standard Finnish

Mene-n auto-lla ja sinä pyöräile-t.

go-1SG car-ADE and 2SG ride.bicycle-2SG

‘I’m going by car and you’ll ride a bike.’

The term **clause combining** refers to the grammatical strategies used to show syntactic or semantic relations between clauses (Skribnik 2022: 996). However, as Gast and Diessel (2012: 4) point out, the term clause combining leads to believe that it refers to two clauses combined

with each other. While this is often true, this definition disregards cases where a clause is combined with a noun, for example (such as in *the fact that he was angry*). Therefore, a better definition for clause combining would be the combination of a clause with some other constituent. In this work, clause combining refers to the ways in which clauses are connected to other constituents in a complex sentence.

The two main strategies of clause combining are **coordination** and **subordination**. Complex sentences that show coordination consist of at least two independent clauses and in subordination, one sentence is embedded in another. According to Gast and Diessel (2012: 5), clause combining strategies can be further divided into **syndetic** and **asyndetic** ones. If there is an explicit linking device, such as a complementizer or a conjunction connecting the clauses, the clause combining strategy can be called syndetic. In cases where the linking device is not present, the clause combining strategy is asyndetic. Sentence (2) above shows a case of syndetic coordination, since the coordinating conjunction *ja* is present and functioning as the linking device between the clauses. An example of asyndetic coordination is given in (3). In the sentence, the last two clauses with finite verbs are coordinated without a conjunction or any other coordinator between them.

(3) Eastern Khanty, Borise & Kiss 2023: 194⁶

<i>tʃi suːltə-m-em-e</i>	<i>ulək</i>	<i>nʉːr</i>	<i>mɐː-n</i>	<i>kətʃəy-nət</i>
so slip-PTCP.PST-1SG-LAT	harness	tether	1SG-LOC	knife-COM
<i>ɛβətəm-i</i>	<i>mɐː-nə</i>	<i>βɐːləy</i>	<i>qoβit uːtnəm</i>	<i>quːyl-əm.</i>
cut-PST.PASS.3SG	1SG-LOC	driving-pole	along up.to.bank	climb-PST.1SG

‘Upon me having slipped, the harness tether was cut with a knife by me, I climbed along the driving pole up to the bank.’

Subordinate clauses are traditionally divided into **complement clauses**, **adverbial clauses** and **relative clauses**, depending on their syntactic function in the complex sentence. According to Laury and Suzuki (2011: 2), complement clauses are arguments of the predicate in the main clause, adverbial clauses modify the verb or the whole clause, and relative clauses modify noun phrases. Gast and Diessel (2012: 6) go on further to distinguish between nominal and verbal complementation (compare the clauses *the fact that he was angry* and *he said that he was angry*). However, as will be shown in the next chapter, this taxonomy is not always straightforward. For example, the line between a relative clause and a complement clause is

⁶ In their paper, Borise and Kiss show that asyndetic coordination was the main combining strategy for clauses and phrases in the earliest Khanty sources. Syndetic coordination and conjunctions came to use only after Russian influence became more dominant in the 20th century.

sometimes fuzzy. In some cases, relative clauses appear in constructions that could be classified as complement clauses.⁷ In this work, they are treated as relative clauses in non-canonical relative clause constructions. Sentences (4) through (6) below present examples of a complement, relative and adverbial clause, respectively.

- (4) Mari, Bradley & Luutonen 2023: 564

valja erla tol-ef manən kol-ən-am
 Valya tomorrow come-3SG that hear-PST2-1SG
 ‘I heard that Valya is coming tomorrow’

- (5) South Saami, Söder 2023: 270

Vuartesjh dujtie niejtide [gïeh dusnie tjahkesjeminie]!
 look.IMP.2SG those.ACC girl.ACC.PL REL.PL over.there sit.GER
 ‘Look at the girls who are sitting over there!’

- (6) Erzya Mordvin, Hamari & Ajanki 2022: 428

kodak molï-an kola-v, sïesïte korta-n
 when go-PRS.1SG school-LAT then talk-PRS.1SG
tonavtitsïa-nïtï marto
 teacher.GEN.DEF.SG with
 ‘When I go to school, then I’ll talk with the teacher.’

The most important type of subordinate clause concerning this work is the relative clause. As mentioned above, Laury and Suzuki’s (2011: 2) functional definition focuses on the relative clause as a modifier of a noun phrase. An alternative definition for a relative clause is presented by Andrews (2007: 206) as follows: “A relative clause (RC) is a subordinate clause which delimits the reference of an NP by specifying the role of the referent of that NP in the situation described by the RC.” However, as will be shown later, sometimes relative clauses do not delimit the reference, but only provide more information about the referent.

In this work, a distinction is made between **full noun phrases**, which are formed around a proper noun or a common noun, and **pronominal phrases**, which are noun phrases that have only pronouns. Relative clauses can modify full noun phrases, pronominal phrases and demonstratives. In terms of identifying relative clauses, an operational definition is adopted here: any clause containing a **relativizer**, usually a relative pronoun, is considered a relative clause.⁸

⁷ For further discussion on the traditional taxonomy and alternatives on the analysis of subordinate clauses typologically, see for example Gast & Diessel (2012a).

⁸ Sometimes, relative pronouns and relativizers are considered separate categories (cf. Glossary of Linguistics Terms 2015), but in this work, relativizer is used as an umbrella term that covers relative pronouns, relative proadjectives and all linking devices used between the main clause and the relative clause.

3.2. Clause combining in Ingrian

In Ingrian, a typical clause consists minimally of a verbal predicate along with its complements. An example of a simple Ingrian sentence is provided in (7) below. Note that, depending on the context, the subject of the sentence may not be overtly expressed as in (7), but it can be omitted. A case of subject omission is presented in (8). It is also possible to omit the predicate in case it is the verb *olla* ‘to be’. However, this phenomenon is not observed frequently and it has possibly emerged due to contact-induced influence from Russian (Markus & Rozhanskiy 2022a: 325). An example where the verb *olla* is omitted is given in (9).

- (7) Markus & Rozhanskiy 2022a: 324

miä ša-i-n haugi-n
1SG get-PST-1SG pike-GEN
‘I caught a pike.’

- (8) Porkka 1885: 130

Män-i hōmukse-st kottii.
go-PST.3SG morning-ELA home.ILL
‘He went home in the morning.’

- (9) Markus & Rozhanskiy 2022a: 326

näi-l tee-löi-l veži
these-ADE road-PL-ADE water
‘There is water on these roads.’

In general, syndetic coordination is the main strategy for coordinating clauses in Ingrian. However, asyndetic coordination is also possible. An example of that is shown in (10) below.

- (10) Markus & Rozhanskiy 2022a: 328

höö itse möö-väd petterii-št tulloo-d maššiina-d
3PL oneself sell-3PL St.Petersburg-ELA come-3PL car-PL
ottaa-d kaig i möö-väd
take-3PL all and sell-3PL
‘They sell (it) themselves; cars come from St. Petersburg, take everything, and sell.’

Conjunctions, relativizers and pronominal adverbs are used for marking subordinated clauses. Examples of a complement clause, a relative clause and an adverbial clause in Ingrian are given in (11) through (13) below, respectively.

(11) Markus 2024: 169

a šiiž meḡhe-lle läkkää-n što aida ono kubehee-l
 and then husband-ALL talk.PRS-1SG that fence be.PRS.3SG side-ADE
 ‘And then I tell my husband that the fence is (lying) on its side.’

(12) Porkka 1885: 131

Sël ol-i se lindu, [migä varast-i omen-i-a].
 there be-PST.3SG that bird REL steal-PST.3SG apple-PL-PART
 ‘There was the bird who was stealing apples.’

(13) Markus & Rozhanskiy 2022a: 328

a konž miä ol-i-n pikkarain miä nä-i-n paljo šuži-loj-a
 and when 1SG be-PST-1SG small 1SG see-PST-1SG many wolf-PL-PART
 ‘When I was small, I saw a lot of wolves.’

3.3. Relative clause constructions

3.3.1. Definition

Dixon (2010: 314) defines a **canonical relative clause construction** consisting of three parts: a main clause, a relative clause and an argument that is shared between the two clauses. This argument, called the **common argument**, can be expressed on the surface level in the main clause, in the relative clause, in both clauses, or neither of them. Sentence (14) presents an example of a canonical relative clause construction where the common argument, *the boy*, is expressed in the main clause as a full noun phrase. In the relative clause, the common argument is represented by the relative pronoun *who*.

(14) English, Dixon 2010: 313

MAIN CLAUSE				
<u>A ARGUMENT</u>	<u>PREDICATE</u>	<u>O ARGUMENT = NP</u>		
I	know	<u>ARTICLE</u>	<u>HEAD</u>	<u>RELATIVE CLAUSE</u>
		the	boy	who ate the mango

Other authors often include only the relative clause and the phrase it modifies in the construction (see for example Whaley 2011; Comrie & Kuteva 2013a; Shagal 2023). The key difference compared to Dixon’s definition is that the main clause is absent from the construction. In this work, the broader definition including the main clause in the construction is adopted, although not all of the variables inspected concern the main clause directly.

3.3.2. Common argument and relativization strategies

Even though the term *common argument* might suggest that both the main clause and the relative clause have the same argument, such as a subject or an object, this is not necessarily the case. It is possible that the common argument functions, for example, as a subject in both the main clause and the relative clause, but more often than not, it fulfills different syntactic functions in each of them. The notion of a common argument refers to the fact that there is something shared by the main clause and the relative clause, and that this something functions as one of the arguments in the main clause and as one of the arguments in the relative clause. Therefore, in the underlying structure, it is rather the entity behind the arguments that is shared by the clauses, not any of the arguments themselves. This is illustrated by examples (15) and (16) below. In both of them, the common argument is a man who is moving to Tallinn, and it is functioning as the subject in the relative clause in both examples. The difference is that the common argument functions as the subject of the main clause in (15), and as the object in (16), indicated by the nominative and partitive case markings on the word *mees*.

(15) Standard Estonian

Seal kõnni-b mees, [kes koli-b homme Tallinna].
there walk-3SG man.NOM REL move-3SG tomorrow Tallinn.ILL
'A/the man who moves to Tallinn tomorrow is walking there.'

(16) Standard Estonian

Näg-i-n täna mees-t, [kes koli-b homme Tallinna].
see-PST-1SG today man-PART REL move-3SG tomorrow Tallinn.ILL
'Today I saw a/the man who moves to Tallinn tomorrow.'

In this work, the part of the relative clause construction that the relative clause is attached to or that it modifies is called the **head of the relative clause**. As pointed out by Dixon (2010: 317), the term *head* has varying uses in different areas of linguistics and using it may sometimes cause confusion. Still, in lack of a better term, it will be adopted here.

Languages use different means to mark the syntactic role of the common argument in the relative clause. Comrie (1989: 147–153) distinguishes four main **relativization strategies**: non-reduction, pronoun-retention, relative pronoun and the gap strategy. In languages applying the non-reduction strategy, the head noun is in full form in the relative clause. In pronoun-retention, the head is represented in the relative clause by a pronoun. Languages which use clause-initial relative pronouns for marking the role of the common argument in the relative clause employ the relative pronoun strategy. If the gap strategy is used, the head is not explicitly

represented in the relative clause by any element. Each relativization strategy is illustrated in examples (17) through (20) below.

- (17) Non-reduction strategy, Gooniyandi, McGregor 1990: 438

[*ginharndi yoowooloo jijaggiddaa-nhi* *wambiggoowaari*
 you:know man we:are:speaking-of:him he:is:going:inside
 'The man who we're talking about is going outside.'

- (18) Pronoun-retention strategy, Persian, Comrie 1998: 63

mardhâi [*ke ketâbhâ-râ be ânhâ dâde bud-id*]
 men that books-ACC to them given were-2SG
 'the men that you had given the books to'
 (lit. 'the men that you had given the books to them')

- (19) Relative pronoun strategy, Russian, Comrie and Kuteva 2013b

Ja poterjal nož [*korotym ja narezal xleb*].
 I lose.PST knife.ACC REL.INSTR I cut.PST bread
 'I lost the knife with which I cut the bread.'

- (20) Gap strategy, Turkish, Comrie 1998: 82

[*kitab-ı al-an*] *öğrenci*
 book-ACC buy-PTCP student
 'the student who bought the book'

3.3.3. Headedness of the relative clause

As Shagal (2023: 939–941) describes, if the common argument is realized as a full noun phrase in either of the clauses or both, it is considered a **headed** relative clause. In other words, if the relative clause has a head that is overtly expressed as a common noun or a proper noun, it is a headed relative clause. Only constructions where the head is a full noun are considered canonical. Headed relative clauses can be divided into **externally headed** ones, if the head is located in the main clause and **internally headed** ones, if it is found inside the relative clause. If, in addition to the full noun head in one of the clauses, the common argument is represented by an anaphoric pronoun in the other one, the construction is called **correlative**. If the head is not realized in either of the clauses, the relative clause is considered **headless** (also called **free**). An example of an externally headed relative clause can be seen above in (16). Sentences (21) and (22) provide examples of an internally headed relative clause and a headless one, respectively. A correlative relative clause is presented in (23), where the head *ärvezäš* is inside

the relative clause and the common argument is represented by the pronoun *tädä* in the main clause.

- (21) Moksha Mordvin, Privizentseva 2025: 29

[*Kona kata-t t'ejə-n kaz'-əzʹ*] *mon*
 REL cat-DEF.SG.GEN PRON.DAT-1SG.POSS gift-PST.3.O.3PL.S I
ul'-an kən'er'd'-f kədə _____ karma-j kunc'-əmə
 be-NPST.1SG happy-PTCP.RES if become-NPST.3[SG] catch-FREQ.INF
šejəʹ-t'
 mouse-PL
 'I will be happy if the cat that was gifted to me starts catching mice.'

- (22) Standard Finnish, Manninen 2003: 669

Sirkku teki [mitä Pulmu käski]
 Sirkku did REL.PART Pulmu told
 'Sirkku did what Pulmu told her to do.'

- (23) Hill Mari, Kaškin 2023: 715

[*kädä ärvezüš urok-əm a-k äštä*],
 REL boy lesson-ACC NEG.NPST-3 do.CNG
tädä jalaxaj əl-eš
 3SG lazy be-NPST.3SG
 'The boy who is not doing his homework is lazy,'

In free relative clauses, the relativizer is the only surface level representation of the common argument in the whole relative clause construction, meaning that it does not have a head. For example, in sentence (22) above, the relative pronoun *mitä* functions as the sole representation of the common argument and also as the marker of the relative clause. Due to this double role, this type of construction is sometimes called a **fused relative clause** and it is considered here a non-canonical relative clause construction (Dixon 2010: 359). Other non-canonical relative clause constructions include ones where the head is some other element than a full noun phrase. See, for example, sentence (24) where the head is the demonstrative pronoun *azt* and its pair in the relative clause is the relative pronoun *amit*.

- (24) Standard Hungarian, Lipták 2004: 289

János megette azt [amit Mari tegnap főzött].
 János ate that.ACC REL.ACC Mari yesterday cooked
 'John ate up what Mari cooked yesterday.'

3.3.4. Restrictiveness

Syntactically, the relative clause modifies its head and semantically, it restricts the referent of the common argument or provides additional information about it. Depending on the type of semantic effect, relative clauses can be divided into *restrictive* and *non-restrictive* ones (Shagal 2023: 941). Sentences (15) through (24) are all examples of restrictive relative clauses. For example, in (16), the relative clause *who moves to Tallinn tomorrow* narrows down the range of the referent, *a/the man*, from all possible men to only those who perform the action described in the relative clause. Example (25) below shows a case of a non-restrictive relative clause, where the relative clause semantically adds information to a referent already specified in the context. As the referent of the common argument in the main clause, *my mother*, is already unambiguous and uniquely identifiable, the relative clause cannot delimit the range of reference, but it rather provides additional information about the referent.

(25) Standard Finnish

Äiti-ni, [joka tulee huomenna kylä-än], on nyt Helsingi-ssä.
mother-1SG REL come.3SG tomorrow village-ILL be.3SG now Helsinki-INE
'My mother, who's coming for a visit tomorrow, is now in Helsinki.'

As Shagal (2023: 941) reports, there are no descriptions of the differences between restrictive and non-restrictive relative clauses available for any Uralic language so far. Taking this into account and considering that the relation between these two types of relative clauses is not in the focus in this work, restrictive and non-restrictive relative clauses will not be differentiated by default further on. Nevertheless, both types of relative clauses will appear in the presented examples later.

3.3.5. Internal syntax

Depending on the syntactic position of the relative clause in terms of its head, canonical relative clauses can be divided into **postnominal** and **prenominal** ones. If the relative clause comes after the head, it is considered a postnominal relative clause. As reported by Shagal (2023: 947), postnominal relative clauses are the most common in the Uralic languages. In prenominal relative clauses, the relative clause precedes the head. Examples of a postnominal and a prenominal relative clause are presented in sentences (26) and (27) below.

(26) Pite Saami, Wilbur 2014: 257

ja dä manje-mus skoteri-jd [ma-jd ini-ga]
 and then after-SUPERL snowmobile-ACC.PL REL-ACC.PL have-3DU.PST
 ‘...and the last snowmobiles which they had.’

(27) Tundra Nenets, Nikolaeva 2014: 324

[səwa-w^ona ləx^ona-qma] wada-da
 good-PROL speak-PERF.AN word-3SG
 ‘the language he spoke well’

If the relative clause comes immediately before or after the head, it is considered to be embedded in the main clause. In all the relative clause constructions presented above, the relative clauses are embedded. If any constituent separates relative clause from its head, the relative clause is not embedded but rather adjoined to the main clause. An example where the relative clause is adjoined to the main clause is given in (28) below.

(28) German, Eisenberg & Wöllstein 2016: 1040

Fritzchen leerte alles aus, [was im Aschenbecher lag].
 Fritzchen emptied everything out REL in ashtray laid
 ‘Fritzchen emptied out everything that was in the ashtray.’

Typologically, the position of the relativizer can vary. It is considered to be an areal feature of European languages that the relative clause is marked with a clause-initial relativizer (Shagal 2023: 948). The relative pronoun strategy, that is, using clause-initial relative pronouns for marking relative clauses, is rare in languages outside Europe (even in cases where the languages belong to the same language family as European languages), and this strategy usually arises due to contact-induced developments with European languages (Comrie 1998: 61). Ingrian is not expected to differ from other European languages in this sense, but in any case, the position of the relativizer will be analyzed for each relative clause construction found in the source materials.

3.3.6. Noun Phrase Accessibility Hierarchy

Already in the late 1970s, linguists observed that the relativization strategies in a language follow certain universal principles. Most notable works at the time was published by Keenan and Comrie (1977, 1979), who analyzed which syntactic functions can be relativized cross-linguistically. Based on a sample of 50 languages they proposed an implicational hierarchy (29) which describes how relativizable different syntactic roles in relative clause constructions are.

(29) Noun Phrase Accessibility Hierarchy

Subject > Direct Object > Indirect Object > Oblique > Possessor > Object of Comparison

The hierarchy suggests that subjects are the most suitable participants for relativization universally.⁹ Objects of comparison, on the other hand, are the least susceptible to it. Any relativization strategy must be applicable to a continuous section on the hierarchy. If a relativization strategy allows relativizing indirect objects, for example, then the same strategy must be applicable to relativizing every other type of participant to the left of it on the hierarchy. It is not possible to skip positions on the hierarchy. For example, a language cannot allow the relativization of indirect objects and possessors, but not oblique participants. Note that not all languages make distinctions between all the different positions in the hierarchy (for example if there is no syntactic distinction between a direct and an indirect object).¹⁰ In such cases, the hierarchy still holds, but neighboring positions will just be treated as one. In this work, the analysis of the Noun Phrase Accessibility Hierarchy will be restricted to the relative pronoun strategy, meaning that relative clause constructions where the relativizer is a relative pronoun will be examined.

Examples of subject and direct object relativization are presented in (30) and (31), respectively. Note that *subject relativization* means that the common argument is functioning as the subject in the relative clause. As discussed in section 3.3.2, this does not necessarily mean that the common argument is a subject in the main clause, although it is also possible.

(30) Ancient Greek, Whaley 2011: 468

<i>ho</i>	<i>Kuros</i>	[<i>hos</i>	<i>basileu-ei</i>]
ART.NOM.SG	Cyrus	REL.NOM.SG	reign-3SG.PRS
‘Cyrus, who is reigning’			

(31) Ancient Greek, Whaley 2011: 470

<i>ho</i>	<i>hipp-os</i>	[<i>hon</i>	<i>ho</i>	<i>Kuros</i>	<i>phil-ei</i>]
ART.NOM.SG	horse-NOM.SG	REL.ACC.SG	ART.NOM.SG	Cyrus	love-3SG.PRS
‘the horse which Cyrus loves’					

Note that the oblique participant on the Noun Phrase Accessibility Hierarchy only applies to so called **major oblique cases**, where the head of the relative clause is functioning as an argument of the predicate in the main clause (Keenan & Comrie 1977: 66). This means that,

⁹ In fact, one of the postulates stated by Keenan and Comrie is that all languages have a main strategy for relativization and that this strategy allows for subject relativization.

¹⁰ This is the case in the Bantu language of Kinyarwanda, for one. See Gary & Keenan (1977) for details.

for example, locatives that are adjuncts in the main clause (such as *here* in *He came here, where we live*) are not considered in the hierarchy.

3.4. Relative clause constructions in Finnic

3.4.1. Relative pronouns

Among Finnic languages, the syntax of complex sentences has been researched thoroughly only for Finnish and Estonian. When it comes to minor Finnic languages, many of the existing descriptions lack a section dedicated to relative clauses altogether. As in Uralic in general, Finnic relative pronouns also are usually identical in form with interrogative pronouns (Shagal 2023: 944). The interrogative pronouns have their roots in Proto-Uralic, and they usually based on the roots **ke-*, which has a human referent, **mi-*, which has a non-human referent and **ku-* ‘where, how’ (Grünthal 2023: 457). The relative pronouns in a selection of Finnic languages are presented in Table 2 below (for Ingrian relative pronouns and their inflectional forms, see Table 4 in section 3.5.).¹¹ In addition, each language has specialized relativizers for relativizing locatives, for example, but they are not in the focus in this work and are therefore missing from the table.

Table 2. Relative pronouns in Finnic.

Language	Relative pronouns	Source
Finnish	<i>joka, mikä, kuka</i>	Vilkuna et al. 2008: §735
Estonian	<i>kes, mis, kumb</i>	Metslang et al. 2024: 902
Ingrian Finnish	<i>kuka, mikä</i>	Kholodilova 2013: 98
Votic	<i>tšen, mikä, kump</i>	Markus & Rozhanskiy 2022b: 348
Seto South Estonian	<i>kiä ~ kua, miä</i>	Pajusalu 2022: 378
Livonian	<i>kis, mis</i>	Laakso 2022: 390
Veps	<i>kudam(b)</i>	Grünthal 2015: 302
Livvi Karelian	<i>ken, mi, kudai</i>	Pyöli 2021: 67

¹¹ Note that since relative clause constructions have not been thoroughly studied for minor Finnic languages, the list presented in the table may be incomplete.

3.4.2. Choice of relative pronoun

If a language has more than one relative pronoun, they are often used in different contexts. In Finnic languages in general, the semantics of the common argument affect the choice of the relative pronoun. In Estonian, the choice between the relative pronouns *kes* and *mis* depends on two hierarchies (32), that are governed by the animacy and individuality of the common argument (Metslang et al. 2024: 906).

(32) a. Animacy

human > higher animate > lower animate > inanimate

b. Individuality

individual > distributive set > collective

With individual humans, the relative pronoun is always *kes* and with inanimate referents it is *mis*. Other than that, the probability of choosing *kes* instead of *mis* is higher on the left side of the hierarchies. There are exceptions, such as the one that animals (higher animate) may receive the relative pronoun *mis*. Sometimes both options are possible. For examples, see Metslang et al. (2024: 906).

Individuality also plays a role in choosing the relative pronoun in Finnish. In Table 3, the different semantic aspects that affect the choice of the relative pronoun are presented (from Vilkkuna et al. 2008: §736).

Table 3. Semantic dimensions of individuality in Finnish.

MORE INDIVIDUALIZED	LESS INDIVIDUALIZED
human	non-human
animate	inanimate
individual	collective, plural
indivisible	divisible
concrete	abstract

Usually, referents that are more individualized receive the relative pronoun *joka*, whereas less individualized referents tend to occur with *mikä*. The relative pronoun *kuka* only appears in free relatives with a human referent, although *joka* is also possible in such cases (Vilkkuna et al. 2008: §737). In normative descriptions, such as Maamies (2011), the syntactic nature of the head of the relative clause is another factor that affects the choice of the relative pronoun. Namely, the choice depends on whether the head is a noun, an adjective or a pronoun. In actual

language use, variation is present and the choice is guided by the semantics (for examples, see Vilkuna et al. 2008: §736).

For Ingrian Finnish, animacy also affects the choice of relative pronoun. *Kuka* is used with people and sometimes with animals, whereas *mikä* is used with all other referents (Kholodilova 2013: 99). In Veps, the relative pronoun *kudam(b)* can be used with animate, inanimate, concrete and abstract referents (Grünthal 2015: 302). In terms of other Finnic languages, the factors affecting which relative pronoun is used have not been thoroughly studied and described yet, to my knowledge.

In this work, three of the aforementioned variables will be inspected: the TYPE OF HEAD, ANIMACY, and CONCRETENESS. The possible effects of individualization on the choice of relative pronoun in Ingrian are left for further study.

3.4.3. Number agreement

When the relative pronoun is marked for number, it generally takes on the number of the head. There are some deviations from this where the semantics of the common argument also have an effect. The nature of number agreement has not been studied for minor Finnic, but only for Finnish and Estonian, to my knowledge.

In Finnish, the common relative pronoun *joka* and the colloquial *kuka* have full inflectional paradigms, which means that their morphological forms are always marked for number. *Mikä* has separate forms for singular and plural only in the nominative, and in plural contexts, the same form is used as in singular. In general, the relative pronoun always agrees with its head in number, when the pronoun has distinct forms for singular and plural. However, the agreement is based on the semantics of the head, not the morphological form (Vilkuna et al. 2008: §735). This is why it is possible for the head to be formally singular, while the relative pronoun is in plural. An example of such a case is given in (33), where the head is singular morphologically, but its meaning is plural.

(33) Standard Finnish, Vilkkuna et al. 2008: §735

Sali-ssa hän ol-i näh-nyt kaksi ritari-a, [jotka
 hall-INE 3SG be-PST.3SG see-PTCP two knight.SG-PART REL.PL
pelas-i-vat lautapeli-ä kultais-i-lla nappulo-i-lla].
 play-PST-3PL board_game-PART golden-PL-ADE piece-PL-ADE
 ‘In the hall, he had seen two knights who were playing a board game with golden pieces.’

The relative pronouns *kes* and *mis* in Estonian do not have nominative plural forms, whereas *kumb* has a full inflectional paradigm. For *kes* and *mis*, the nominative singular forms are used in plural contexts. In other cases, the plural forms exist but usually singular forms are still used (Metslang et al. 2024: 907). Sentence (34) shows an example where the allative singular form *millele* of the relative pronoun *mis* is used to replace the existing plural form (*milledele*).

(34) Standard Estonian, Metslang et al. 2024: 907

elu-s on palju küsimusi, [millele on vahest
 life-INE be.3SG many question.PL.PART REL.SG.ALL be.3SG sometimes
väga raske vastata].
 really hard answer-INF
 ‘In life, there are a lot of questions which are sometimes hard to answer.’

Whether the presence of number agreement can be stated or not is directly connected to the set of inflectional forms the relative pronoun has in a language. Finnish and Estonian have relative pronouns with full inflectional paradigms, pronouns with distinct plural forms only in the nominative, and pronouns whose singular and plural forms differ in all cases except the nominative. Ingrian relative pronouns show similar patterns, and they are discussed in section 3.5.

3.4.4. Inverse attraction

In relative clause constructions that show inverse attraction, the syntactic marking of the common argument in the main clause is acquired from the relative clause (Kholodilova 2013: 96). In other words, the relativizer affects the case marking of the head, when usually the situation would be the other way around. An Ingrian Finnish example of a relative clause construction showing inverse attraction is given in (35) below. The head noun *lammas* would be expected to receive the nominative case as it is in the subject position in the main clause. Instead, it has genitive marking due to the inverse attraction effect (the relativizer *minkä* is in the genitive case in the relative clause).

(35) Ingrian Finnish, Kholodilova 2013: 97

lampà-n [minkä miä eilen ost-i-n] loikò koi-n
 sheep-GEN REL.GEN I.NOM yesterday buy-PST-1SG lie.PRS.3SG home-GEN
luon
 near
 ‘The sheep I bought yesterday is lying in front of the house’

Inverse attraction has been reported for some other Uralic languages, such as Moksha Mordvin and Beserman Udmurt (Shagal 2023: 951). One has to keep in mind, however, that the Ingrian Finnish examples where inverse attraction has been observed arise from elicited data. The constructions where inverse attraction appears are direct translations from Russian, and inverse attraction occurs in colloquial Russian (Kholodilova 2013: 97). To my knowledge, this phenomenon has not been attested in other Finnic languages.

3.5. Relativization strategies in Ingrian

The main relativization strategy employed in Ingrian is the relative pronoun strategy. Relative pronouns of Ingrian were first described by Porkka (1885: 86). He observed that the relative pronouns are identical in form with the interrogative pronouns *ken* ‘who’, *migä* ‘what’ and *kumba* ‘which’. Relative pronouns are also briefly mentioned in Junus (1936: 100), but only *kumba* is discussed. Examples (36) through (38) feature contemporary examples of constructions with the relative pronouns *ken*, *migä* and *kumba*, respectively.

- (36) Kala_ja_metsä_EN, glosses from Schwarz & Rozhanskiy 2024: 274)

täž nüd on uuž ... uuž no meğž [ken tökköö tōō-dä]
 here now be.PRS.3SG new new PRCL man who do.PRS.3SG work-PART
 ‘Here now there is a new ... new man who works’

- (37) Elo(B)_AI

oppi-tta tōō šiz šidä-gi keel-d [migä ono]
 study-2PL 2PL then that-PRCL language-PART REL be.3SG
laukkaha-n perild šiin]
 Luga-GEN from here
 ‘Do you study also that language that is there from the Luga river?’

- (38) Kala_ja_metsä_EN, glosses from Markus & Rozhanskiy 2022a: 328

i še haili ain vee-ťtii štarašta-lle [kumba šenen]
 and this sprat always bring-IPS.PST village.head-ALL REL this.GEN
haili-n jago-i]
 sprat-GEN divide-PST.3SG
 ‘And the sprats were always taken to the village head, who divided the sprats [between people].’

In addition to *ken*, *migä* and *kumba*, Markus & Rozhanskiy (2022a: 328) also categorize *millain* ‘which, what kind of’ as a relative pronoun in Ingrian. Schwarz and Rozhanskiy (2024: 274) note that it has five different variants attested in Soikkola Ingrian data: *millain*, *milläin*,

milltain, *milltäin* and *milliin*. Here, *millain* along with its variants is considered a relative proadjective, similarly as in Finnish grammar (Vilkuna et al. 2008: §739). The main property differentiating it from relative pronouns is that the relative proadjective indicates that the head and relativizer are similar in terms of their properties, but they are not necessarily referring to the same exact entity. An example where *millain* is used to relativize a full noun phrase is given below in (39).

(39) Kala_AI

[*milläin* ***poiga*** *kerkkii-ž* *ottaa* *oma-n* *tüdöi-n*]
REL boy manage-PST.3SG take.INF own-GEN girl-GEN
šiž *pääž-i-d* *tantsu-loi-le*
then get_in-PST-2SG dance-PL-ALL
‘The kind of boy who got themselves a girl got into the dances’

For relativizing locative participants, Ingrian uses lexicalized pronominal forms that are essentially identical in form with the relative pronoun *migä*. These forms include *miž* ‘where’, *mišt* ‘from where’ and *mihe* ‘to where’. For the corresponding inessive, elative and illative forms of *migä*, see Table 4. For expressing the same meaning, pronominal adverbs *kuž* ‘where’, *kušt* ‘from where’ and *kuhu* ‘to where’ can be used. These pairs of lexicalized pronominal forms and pronominal adverbs are likely in free variation and the choice of the form seems to depend on the idiolects of individual speakers (Schwarz & Rozhanskiy 2024: 277). Examples of the lexicalized pronominal form *miž* and the corresponding pronominal adverb *kuž* are given in (40) and (41) below.

(40) Porkka 1885: 132

Män-i ***sinne*** [*miz* *oll-ii-t* *tüttöi* *ja* *poiga*].
go-PST.3SG there REL.INE be-PST-3PL girl and boy
‘He went there where the boy and the girl were.’

(41) Kala_ja_metsä_EN

vod ***šəel*** [*kuž* *meijen* *vägi* *kä-i*]
so there REL.INE 1PL.GEN people go-PST.3SG
‘So there where our people went...’

In this work, relative pronouns are treated separately from the relative proadjective and the lexicalized proadverbials and corresponding pronominal adverbs. The different grammatical forms of Ingrian relative pronouns are showcased in Table 4 below (reproduced from Schwarz & Rozhanskiy 2024: 274).

Table 4. Paradigms of relative pronouns *ken* ‘who’, *migä* ‘what’ and *kumba* ‘which’

	who	what	which (one)	who	what	which (one)
	Singular			Plural		
NOM	<i>ken</i>	<i>migä</i>	<i>kumba</i>	<i>ked</i>	<i>mid</i>	<i>kummad</i>
GEN	<i>kenen</i>	<i>minen</i>	<i>kumman</i>			<i>kummiin</i>
PART	<i>kedä</i>	<i>midä</i>	<i>kumppaa</i>			<i>kumppia</i>
ILL	<i>kehe</i>	<i>mihe</i>	<i>kumppaa</i>			<i>kumppii</i>
INE	<i>kež</i>	<i>miž</i>	<i>kummaaž</i>			<i>kummiž</i>
ELA	<i>kešt</i>	<i>mišt</i>	<i>kummaašt</i>			<i>kummišt</i>
ALL	<i>kelle</i>	<i>mille</i>	<i>kummalle</i>			<i>kummille</i>
ADE	<i>kel</i>	<i>mil</i>	<i>kummaal</i>			<i>kummiil</i>
ABL	<i>keld</i>	<i>mild</i>	<i>kummaald</i>			<i>kummiild</i>
TRSL	<i>kekš</i> (~ ???kenekš)	<i>mikš</i> (~ ???mihekš ~ ???minekš)	<i>kummaakš</i>			<i>kummiikš</i>

As can be seen from the table, the relative pronouns *ken* and *migä* do not have full inflectional paradigms. However, this does not mean that these pronouns cannot be used in a plural setting. Instead, the pronouns show syncretism, where the singular forms are used also in plural contexts.¹² This structure of paradigm is similar to the relative pronoun *mikä* in Finnish (see section 3.4.3.). The forms of *ken* and *migä* attested in plural contexts could be categorized as plural forms, as in Saar (2017: 127). However, since these forms are formally identical to the singular, they are not treated here as separate plural forms, following Schwarz & Rozhanskiy (2024: 275). Nevertheless, a clear number distinction is possible only in instances where the relative pronoun takes the nominative case. Examples (42) through (45) show *ken* and *migä* used in both singular and plural in cases other than nominative.

(42) Nirvi et al. 1967: 148

se pāz-i mīhe-lle, [kel midägi hāsse-ttī]
 that get-PST.3SG man-ALL REL.ALL something speak-IPS.PST
 ‘That one got married (lit. got to a man) to whom something was said’

(43) Kala_ja_metsä_EN

i kan’ešna ken itse kä-i mere-l še välištää
 and surely who oneself go-PST.3SG sea-ADE that sometimes
ša-i veļel vāhemb kuin need [kel oļ-i-d šuure-d pere-d]
 get-PST.3SG even less than those REL.ADE be-PST-3PL big-PL family-PL
 ‘And surely that one who went fishing by themselves sometimes got even less than those who had large families’

¹² Note that the term *syncretism* is used in a very broad sense here, following Rozhanskiy & Markus (2024). Here, it simply means that the singular and plural forms are identical in form. It does not consider the historical developments (which have not been researched) that have led to the forms being identical.

(44) Kala_AI

nu šiz möö män-i-mmä toiššee poolee [miž eļlää
well then 1PL go-PST-1PL other.ILL side.ILL REL.INE live.3SG
nüd tämä tan'a]
now this T.
'Well then we went to the other side, where this Tanja is living now'

(45) Hirvi_GI

šēļ meil ennemäž oļ-ii-d need paiga-d [miž kažvo-i
there 1PL.ADE before be-PST-3PL these place-PL REL.INE grow-PST.3SG
paljo obokka-a]
many mushroom-PART
'There we used to have the places where many mushrooms were growing'

The relative pronoun *kumba* has a full inflectional paradigm, just like *kumb* in Estonian. As noted by Schwarz & Rozhanskiy (2024: 276), *kumba* is translated into the Finnish *kumpi* 'which of two' in Nirvi (1971: 218) when used as an interrogative pronoun, while contemporary data show the use of *kumba* also in cases where there are more than two alternatives, as exemplified in (46).¹³

(46) Schwarz & Rozhanskiy 2024: 276

tāž on kold koiti-a kumba on šiu-n oma
here be.PRS.3SG three house-PART which be.PRS.3SG 2SG-GEN own
'Here are three houses. Which one is yours?'

It should be mentioned that even in the relative function, the number of referents is not limited to two. This can be confirmed by contemporary data, as can be seen in example (47). Furthermore, the oldest written Ingrian sources verify that *kumba* has been used to refer to more than two entities already in the 1880s. This is reflected in example (48). For context, the protagonist of the story encounters three creatures in the forest, and these creatures are fighting over who gets the things. According to Nirvi (1971: 483), the word *riista* translates into the Finnish *kapine*, *kalu* 'thing, object, tool', which leaves the interpretation rather ambiguous and the original story does not clarify what these things are more specifically. In any case, the sentence shows *kumba* referring to three entities.

¹³ There appears to be a typo in the original reference citing page 217, the correct page in Nirvi's dictionary is 218.

(47) Kolhožiiž_PE

mõõ kä-i-mmä kaig [kummad ol-ttii kolhožii anne-ttu
1PL go-PST-1PL all REL.PL be-PST.IMP kolkhoz.ILL give-PST.PTCP
talo-loi-št]
house-PL-ELA

‘All of us who had [horses] given from their farms to the kolkhoz, [we all] went...’

(48) Porkka 1885: 133

[Kumba saa-p senen lōi-n ja tô-b miulle],
REL get-3SG that.GEN bullet-GEN and bring-3SG 1SG.ALL
sille siit kaig ne riissa-t.
that.ALL then all they thing-PL

‘Whichever (of you) gets that bullet first and brings (it) to me, to that one all the things (go) then.’

All the examples shown above have featured relative clauses with finite verbs. However, non-finite relative clauses have also been attested in Ingrian. In Uralic in general, the gap strategy is employed for non-finite relativization, and in such constructions, the relative clauses are usually prenominal (Shagal 2023: 940). As in many related language, Ingrian non-finite relative clauses are formed with participles. Examples (49) and (50) show instances of Soikkola Ingrian non-finite relative clauses introduced by an active present participle and an active past participle (from Norvik et al. 2022).¹⁴

(49) Active present participle

tämä-g se on [sodaväkkem] männõ-vä?
it-Q this be.3SG army.ILL go-ACT.PRS.PTCP
‘Is s/he the one who is going to the army?’

(50) Active past participle

[k̄hu-nt] kukkoī ol-i keskikannè-l
bake-ACT.PST.PTCP cock be-PST.3SG mid_table-ADE
‘The baked cock-pastry was in the middle of the table’

In both examples, the participles are used to relativize the copula subject. Additionally, Markus and Rozhanskiy (2022a: 321) note that the active present participle *va/vä* is not used productively in the contemporary language anymore, and the instances where it appears should be considered lexicalized. Unfortunately, not much more can be said about non-finite relativization in Soikkola Ingrian at this point due to lack of data. Even in the existing sources

¹⁴ The transcriptions have been adapted to reflect the original sources of the examples. They are dictionary entries from Nirvi (1971: 325, 159).

they are very rare. In fact, there are no examples of participles used for relativization in the source materials used in this study. Therefore, this thesis focuses on Soikkola Ingrian relative clause constructions that employ the relative pronoun strategy.

3.6. Research questions

Other than the inflectional forms of the relative pronouns and other relativizers, not much more research has been done on the nature of relative clause constructions in Soikkola Ingrian. Using different variables and parameters described in sections 3.3. and 3.4., this thesis aims to expand the knowledge on aspects of Soikkola Ingrian relative clause constructions that have not been researched before. More specifically, the focus will be on the following: syntactic functions of the common argument, head type, animacy and concreteness of the referent, presence of number agreement, headedness of the relative clause, position of the relative clause and the relativizer, and the presence of inverse attraction.

By analyzing the aforementioned variables, the aim of this thesis is to answer the following research questions:

- 1) What are the possible syntactic functions of the common argument in Soikkola Ingrian relative clause constructions?
- 2) How do the type of head, animacy and concreteness of the referent affect the choice of relative pronoun in Soikkola Ingrian?
- 3) How does number agreement between the head and the relative pronoun function in Soikkola Ingrian relative clause constructions?
- 4) What is the internal syntactic structure of relative clause constructions in Soikkola Ingrian?

4. Data and method

4.1. Description of data sources

The data analysed in this thesis come from several sources, which can be divided roughly into three different time periods. The sources differ from one another in a number of other aspects as well, such as medium and means of collection. They also contain folk songs and poems, but they are left out from this work, as their language is distinct and best studied separately. In this chapter, each source is described and, in the end, they are compared to each other and critically evaluated in terms of their suitability for the purposes of this thesis.

4.1.1. Data from the 19th century

The oldest source used in this thesis is Volmari Porkka's doctoral dissertation *Über den Ingrischen Dialekt mit Berücksichtigung der übrigen finnisch-ingermanländischen Dialekte* (1885). It is a grammatical overview of the phonology and morphology of Ingrian in comparison with the Finnish Ingrian varieties spoken in Ingria. The appendix of the dissertation includes language samples from different dialects of Ingrian. Two of the language samples are Soikkola Ingrian: "Der goldene Vogel" ["The golden bird"] (pages 130–134), collected from the former village of Tarinaisi (Rus. *Andreevshchina*) and "Die goldene Feder" ["The golden feather"] (pages 143–145), collected from the village of Taatsoi (Rus. *Semeiskoe*). These samples are fictional tales. The first of them, "Der goldene Vogel", is a version of the Russian fairy tale "Tsarevitch Ivan, the Firebird and the Gray Wolf", although its plot does not always follow the canonical Russian one (Rožanskij & Markus 2012: 453). The storyline of the other one, "Die goldene Feder", bears resemblance to another Russian fairy tale, "The Firebird and Princess Vasilisa".

The exact years when the samples were collected are not mentioned in the work, but it is known that Porkka organized two expeditions to Ingria prior to the publishing of the dissertation, one in 1882 and another one in 1884 (Grünthal 2021: 124). As the dissertation is based on Porkka's notes from these linguistic and ethnographic expeditions, it can be deduced that the samples are from the early 1880s. Moreover, other than the places where the texts were collected, no information is given on the narrators of the texts.

The tales are written down in a writing system that is not described in the work itself. A couple of features in it are worth pointing out. First of all, Porkka uses the characters *ô*, *ô* and

ê to denote vowels that were historically long mid-vowels in the first syllable. Moreover, there is no distinction between short and long geminates, and Porkka does not use diacritic marking for reduction and prolongation. The last two remarks point towards the fact that the transcription used by Porkka is not entirely phonetic. There are also elements in the layout of the texts, such as the way the text follows the punctuation conventions of Finnish of that time, that would suggest that the transcription system Porkka uses is partly an adaptation of Finnish orthography. This is not very surprising considering that Ingrian did not have a written standard of its own at the time, and that Porkka was Finnish himself. Moreover, Porkka, and many scholars after him, categorized Ingrian as a dialect of Finnish. Therefore, the system in which the texts have been written down is best described as a mix of a phonetic transcription and an adapted late 19th century Finnish orthography.

4.1.2. Data from the 20th century

The data from this time period come from three publications that include Soikkola Ingrian language samples collected between the years 1944–1964. The first one of them is Paul Ariste’s *Isuri keelenäiteid* [“Language samples of Ingrian”] (1960), followed by Arvo Laanest’s *Isuri murdetekste* [“Ingrian dialect texts”] (1966a), Ruben Nirvi and others’ *Inkeröismurteet* [“Ingrian dialects”] (1967) and Seppo Suhonen and Pertti Virtaranta’s *Lähisukukielet* [“Closely related languages”] (1978). Note that the data are not presented chronologically here, but rather in order of the year the sources were published.

From Ariste’s work, texts 1–47 (pages 8–40) were collected in 1957 from an 80-year-old speaker who was born in Saarove (Rus. *Jugantovo*) and spent her life in Tarinaisi (Rus. *Andreevshchina*) for the most part. Texts 48–55 (pages 40–46) were narrated by another speaker in 1956. She was 74 years old at the time and also lived in Tarinaisi, although she was from the village of Viistina (Rus. *Vistino*) originally. Texts 56 and 57 contain only folk songs, so they are disregarded in this work. Text 58 (page 48) is a collection of sentences or short paragraphs that are unrelated to each other. They have been collected from several different speakers in Venakontsa (Rus. *Pakhomovka*) in 1956. The texts follow the traditional Finno-Ugric transcription.

Laanest has collected ten texts (numbered 22–31 on pages 111–154) from a 57-year-old speaker in Voloitsa (Rus. *Valyanitsy*). They have been written down using the traditional Finno-Ugric transcription with a few modifications concerning the quality of plosives, sibilants and laterals. The texts are interviews that cover various different topics, including past life,

work and traditions. Some of them feature folk songs mixed with narration, but the folk song parts are not analyzed in this work. Laanest has also translated the texts into Estonian and included them in the same publication.

The texts collected by Laanest were originally recorded on magnetophon tapes. Later on, the recordings have been digitized, and these audio files are the primary source from Laanest in this work. The transcriptions do not correspond to the recordings entirely, as Laanest has intentionally left out repetitions and parts that were unclear. The recorded interviews were set up in a way that allows for the capture of spontaneous speech. That is, the recording was started immediately after introducing the topic and usually there was no preparatory time before starting the recording.

Concerning Nirvi's materials, texts 1–4 (pages 138–141) have been written down based on dictation. Nirvi collected the texts in 1944 from a 75-year-old Soikkola Ingrian speaker who was living in a nursing home in Finland. She had been born in Oussimäki (Rus. *Evseeva Gora*) and lived in the village of Tammikontu (Rus. *Dubki*) after being married. Text 5 contains only poems, so it is disregarded. Text 6 (page 143) was narrated by a 43-year-old speaker who was from the village of Ruutsia (Rus. *Ruch'i*). The text was recorded in Sweden in 1951. Texts 7–12 (pages 144–149) were narrated by another speaker who was around 60 years old at the time. She was born in the village of Säätinä (Rus. *Slobodka*) and moved to Takaväljä (Rus. *Novaja Derevnja*), a former village in the Lower Luga speaking area. These texts were recorded in Sweden in 1951 as well. Texts 13 and 14 (pages 149–150) were narrated by a 72-year-old speaker from Saarove (Rus. *Jugantovo*). The recordings were made in Finland in 1954.

Suhonen and Virtaranta's publication contains one Soikkola Ingrian text (pages 54–57), narrated by a speaker born in Saarove. The text is a story recorded in Finland in 1964. In addition, the authors provide a Finnish translation of the text.

4.1.3. Data from the 21st century

The most recent data come from field recordings collected by Elena Markus, Fedor Rozhanskiy and their colleagues between 2006–2013. Most of the data have been collected between 2011–2013 as part of the project “Documentation of Ingrian: collecting and analysing fieldwork data and digitizing legacy materials” (for a detailed description of the project, see Rozhanskiy & Markus 2019). The processed and transcribed recordings are saved in the form of ELAN (2025) files, and the files are collectively called the Spontaneous Speech Corpus (Markus & Rozhanskiy 2023: 46).

The Spontaneous Speech Corpus consists of 4 hours of audio recordings, collected from 24 native speakers of Soikkola Ingrian, making the overall size of the corpus around 20,000 words (Rozhanskiy 2024: 190). For the most part, the texts are narratives about the personal life of the speakers, but approximately 40 minutes of dialogue is also included (Markus 2024: 159). The speakers come from different parts of the Soikkola Ingrian speaking area, and the texts represent different subdialects (northern, central, southern) of the language (Markus & Rozhanskiy 2023: 45–46).

The transcription system used in the Spontaneous Speech Corpus somewhat differs from the ones used by the authors of 19th and 20th century publications. The most distinct features are the marking of long vowels with a double character (such as *oo* pro *ō* or *ô*) and marking sibilants [š] and [ž] as *š* and *ž* instead of *s* and *z* (Rozhanskiy 2024: 192). For any example presented from the Spontaneous Speech Corpus, the name of the text along with the speaker identifier will be provided. Furthermore, wherever “contemporary Ingrian” is mentioned in this work, it refers to the language spoken in the 21st century, that is, language captured in the Spontaneous Speech Corpus.

4.1.4. Evaluation of the data sources

The source materials used in this study differ in terms of modality and include most of the available sources containing spontaneous Soikkola Ingrian speech that are available. There are a few other Soikkola Ingrian texts that are included in Sovijärvi (1944), but these are translations of Karelian speech samples into the Ingrian variant spoken in the village of Tarinaisi. As Ariste (1960: 7–8) points out, even though these texts have been translated by a native speaker, they are not ideal materials for researching Soikkola Ingrian. This is why these texts are left out from this work.

The 19th century materials contain around 2,000 words from 2 speakers, 20th century materials around 16,000 words from 8 speakers and 21st century materials around 20,000 words from 24 speakers, adding up to a total of around 38,000 words from 34 different native speakers. Note that the actual size of the dataset used in this study is somewhat smaller, as folk songs are excluded.

Although the diachronic development of relative clause constructions is not the main focus of this thesis, it is reasonable to include the older materials in the sources as well, since they offer the possibility for confirming that observations from the contemporary data are not idiolectal. Moreover, in terms of word count, 19th and 20th century data make up a significant

part of the dataset, accounting for almost a half of it in size. As the amount of Soikkola Ingrian data available is quite scarce altogether, the 19th and 20th century examples are a valuable addition to the contemporary data.

Other than the data from the 19th century, the source materials are quite uniform in terms of the genre of the texts, as the rest are life stories and personal narratives. The texts collected by Porkka are stories, so their genre is not too far from the other texts. However, it should be kept in mind that they are adaptations of Russian fairy tales, so they are not quite as authentic as the other texts.

The differences in the method the different texts have been collected should also be noted. Porkka's, Ariste's and Nirvi's materials have been written down on the spot, presumably with some help from the native speakers as well. However, Porkka bases his example texts only on his notes, which means that it is possible that he has edited the texts before publishing them, making them fit the Finnish norm better. But as the original notes are no longer available, this remains a question of suspicion.

Even among the materials that have been originally recorded (Laanest 1966a, texts from the 1950s in Nirvi et al. 1967, Suhonen & Virtaranta 1978 and the Spontaneous Speech Corpus), there is a difference in whether or not the audio files are available anymore. For Laanest and the Spontaneous Speech Corpus, the audio files are considered the primary source of data, and for the rest, the judgement of the original authors will be trusted by default, as there is no way to check the accuracy of the transcriptions since the audio is no longer available.

The different data sources from different times represent also varying systems of transcribing Ingrian. Even though these variations are not that significant for this thesis, the main differences are presented in Table 5 below for the sake of clarity (reproduced from Rozhanskiy 2024: 192, Nirvi et al. and Virtaranta & Suhonen added).

Table 5. Main differences in transcriptions of Ingrian data.

	Porkka 1885	Ariste 1960	Laanest 1966a, Nirvi et al. 1967, Virtaranta & Suhonen 1978	Spontaneous Speech Corpus
Long vowels (e.g. <i>ō</i>)	<i>ô</i>	<i>ō</i>	<i>ō</i>	<i>oo</i>
Assimilation of the final consonant by the following word	not marked	marked	marked	not marked
Half-voiced singletons (e.g. <i>D</i>)	<i>d</i>	<i>D</i>	<i>D</i>	<i>d</i>
Short geminates of plosives (e.g. <i>tt̃</i>)	<i>t ~ tt</i>	<i>t</i>	<i>tt̃ ~ t</i>	<i>tt̃</i>
Short geminates of sonorants (e.g. <i>nñ</i>)	<i>nn</i>	<i>nñ</i>	<i>nñ</i>	<i>nñ</i>
Velarized allophone of <i>l</i>	<i>l</i>	<i>l̥</i>	<i>l</i>	<i>l</i>
Sibilants <i>š/ž</i> developed from <i>s/z</i>	<i>s/z</i>	<i>š/z̥</i>	<i>s/z</i>	<i>š/ž</i>

The role of the translations is, on the other hand, is of more importance in this thesis. Wherever translations are available (i.e. Laanest 1966a and Suhonen & Virtaranta 1978, for a glossed Russian translation of Porkka’s “Der goldene Vogel”, see Rožanskij & Markus 2012), they can provide additional information about how the original authors have perceived the text. This can be useful especially when there is no audio available. For example, compare the punctuation in the original transcription and the Finnish translation of Virtaranta and Suhonen (1978: 54–55) in Table 6 below.

Table 6. Differences in punctuation between Ingrian transcription and Finnish translation.

Ingrian	Finnish	English
<i>hã siZ meille Gu läkkäis käBiöjä kãskoja ni, kui kaks sũrt poronã noiššĩt tappelomã, heil oltĩ sũret sarveD.</i>	Kun hän sitten meille kertoi kauniita satuja niin... Kun kaksi suurta pãssiã rupesi tappelemaan – niillä oli suuret sarvet.	When she then told us beautiful stories so... When two big rams started to fight – they had big horns.

As can be seen above, the commas in the transcription can translate to different characters, such as an ellipsis or a dash, marking that there is a change of sentence or thought perceived in the original Ingrian text according to the translator. This is not readily apparent from the transcription alone, which is why the translations function as a tool in helping to read the data.

To summarize, the sources used in this study are quite homogeneous. They form a dataset that is adequate in size and fairly uniform in genre. It contains mostly spontaneous speech samples that have been collected, transcribed and translated in various ways and to different degrees. These differences that the parts of the dataset showcase must be considered when extracting relative clause constructions from the dataset.

4.2. Method

According to Veismann, Muischnek and Padrik (2026: 20), a language corpus is a digital text collection with a uniform way of transcribing the texts and which is organized and usually includes metadata. As mentioned before, the data used in this study come from different types of sources and with varying transcriptions and metadata, so the sources do not form a corpus in this sense. However, as building a unified corpus for Soikkola Ingrian is still a work in progress, going through the different sources and annotating the data manually is the only option for researching relative clause constructions currently. This approach is suitable for making qualitative observations about Ingrian relative clause constructions and it also allows for presenting some descriptive statistics about them. The procedures for extracting and annotating the data points for subsequent analysis are described in sections 4.2.1. and 4.2.2.

4.2.1. Extracting the data

The process of extracting relative clause constructions from the data sources consists of four parts: finding all morphological forms that could potentially be relativizers, identifying them as relativizers, determining the borders of the main clause and the relative clause and moving the constructions to a separate database. Each of these steps is described in this section.

As mentioned in section 3.5., finite relative clauses in Ingrian are marked by a relativizer. In order to find all the relative clause constructions in the materials, every occurrence of a relative pronoun or another type of relativizer had to be identified. Therefore, the first step was to find all instances of *ken*, *migä* and *kumba*, the relative proadjective *millain* in any of its variants, lexicalized pronominal forms with *mi*-stem and the relative proadverbials with *ku*-stem.

Since all the relativizers are not unique in morphological form, they must be distinguished from interrogative pronouns which are identical to them. In written sources, interrogative

contexts are usually explicitly marked by a question mark. However, the presence of a question mark does not mean that the question itself could not include a relative clause construction. In audio sources, interrogatives can most often be distinguished either by the intonation or with the help of transcriptions.

There is another problem concerning the distinction between relativizers and interrogatives. Namely, interrogatives are used in indirect questions as well. Indirect questions are part of reported speech and therefore explicit marking, such as a question mark, is usually not present. An example of an Ingrian indirect question is given in (51) below.

(51) Porkka 1885: 131

Sao miulle, kuhu siä taho-t männä.
 say.IMP.2SG 1SG.ALL REL.ILL 2SG want-2SG go.INF
 ‘Say to me, where you want to go.’

Since the sources are not compiled together in digital format, the search had to be done manually. Only cases where the connector could be identified as a relativizer were included. If it was unclear whether the connector is a relativizer, it was dismissed. Then, the clause that includes the relativizer as well as the main clause were extracted. In some cases, the relative clause construction is part of a larger complex sentence. Other parts of the complex sentence, such as coordinated main clauses or adverbials were left out. For example, from sentence (52) only the underlined part was extracted.

(52) Porkka 1885: 132

Tüttöi i poiga istu-isii-t sue-lle selkää i
 girl and boy sit-PST-3PL wolf-ALL back.ILL and
männ-ii-t sinne riikkii, miz ol-i heppoin kuldaharja.
 fo-PST-3PL there kingdom.ILL REL.INE be-PST.3SG horse golden-maned
 ‘The girl and the boy sat on the wolf’s back and went to the kingdom where the golden-maned horse was.’

Note that in some cases the main clause is not realized as a full clause, since not all the utterances in spoken language make up full sentences. The common argument and its syntactic functions can still be analyzed, even if the main clause is not entirely complete. An example is given in (53) below, where the head *izännäd* can be identified as the subject of the main clause even though it is not fully complete.

(53) Kūlmä_kežä_IV_VF

<i>nęęd</i>	<i>įžännä-d</i>	[<i>mid</i>	<i>oļl-ii-d</i>	<i>ennemmāž</i>]
those owner-PL	REL.PL	be-PST-3PL	before	

‘Those, the owners, who were [here] before’

In written sources, sentence boundaries are usually marked with punctuation marks. Wherever available in written sources and in data extracted from the Spontaneous Speech Corpus, translations were used to help determine the sentence borders. There were a few files in the Spontaneous Speech Corpus and other audio sources where the translations and transcriptions were missing. In such cases, the sentence boundaries were determined by finding the intonational phares containing the relativizer.¹⁵

4.2.1. Annotating the data

After collecting all the relative clause constructions in the source materials, they were moved to a spreadsheet for annotation and further analysis. For each relative clause construction, the relativizer and the head were extracted into different columns in the form they appear in the original sources. In case of fused relatives, the head was marked empty. The next step was to annotate the type of the head, that is, whether the head is a full noun, a pronoun, an adjective or an adverb. This was done in order to separate canonical relative clause constructions from non-canonical ones.

The relativizers were lemmatized so that they could be easily filtered despite the differences in transcriptions and case forms. The nominative singular was used for lemmatizing the relative pronouns *ken*, *migä* and *kumba*. In terms of other relativizers, the variant *millain* was chosen for the relative proadverbial and the relative proadverbials and lexicalized pronominal forms were marked with either *mi-* or *ku-*, depending on their stem. The *mi-* forms were differentiated from the corresponding identical forms of *migä* on the basis of the type the head. If the head was an adverb, the relativizer was considered *mi-*, and if it was a full noun, it was understood as a form of *migä*.

The syntactic functions of the common argument were annotated for relative clause construction that have *ken*, *migä* or *kumba*. The values were SUBJECT, DIRECT OBJECT, INDIRECT OBJECT, MAJOR OBLIQUE CASE, POSSESSOR and OBJECT OF COMPARISON, as per the Noun Phrase

¹⁵ It is a cross-linguistic property of relative clauses that the relative clause and the main clause form an intonational unit (Dixon 2010: 339).

Accessibility Hierarchy. This was done separately for both the main clause and the relative clause with the help of the morphological form of the head and the relativizer.

When it comes to the choice of the relativizer, the type of head, animacy and the concreteness of the referent were analyzed, as these have an effect in closely related languages, such as Finnish and Estonian. The rest of the parameters of individualization are left for future study. The animacy of the entity behind the common argument was classified as either HUMAN, ANIMATE or INANIMATE. In terms of concreteness, each head was categorized as either CONCRETE or ABSTRACT.

Number agreement between the head and the relativizer was analyzed for all relative clause constructions with the relative pronouns *ken*, *migä* and *kumba*. The expectation is that the relativizer takes on the number of the head. If so, number agreement was marked with YES. If the relativizer did not agree in number with the head, it was annotated with the value NO. However, as shown in section 3.5., the singular and plural forms are distinct only in the nominative case for *ken* and *migä*. In other cases than the nominative, constructions with *ken* and *migä* received the value N/A.

Each relative clause construction was annotated also in terms of headedness, that is, whether it contains an EXTERNALLY HEADED, INTERNALLY HEADED, FREE or FUSED relative clause. For externally headed relative clauses, the position of the relative clause in terms of the head was also determined (POSTNOMINAL or PRENOMINAL). Moreover, if the head and the relative clause appeared next to each other, they were marked as EMBEDDED. If there was any constituent between them, their linking was marked as ADJOINED. For constructions including fused or free relative clauses, the position of the relative clause in terms of the main clause was annotated (whether it comes before, after, or in the middle of the main clause).

The position of the relativizer and the presence of inverse attraction were annotated for all relative clause construction. In terms of the first, the expectation was that the relativizer appears clause-initially, which is why the values were CLAUSE-INITIAL or OTHER. The presence of inverse attraction was also annotated for every relative clause construction.

Some of the variables used in the analysis apply to only part of the relative clause constructions, while others are annotated for every one of them. In Table 7 below, the variables, their corresponding values and the types of constructions (=domain) they are used in are presented.

Table 7. Variables, their domains and values in the analysis.

Variable	Domain	Values
Head type	all	FULL NOUN, PRONOUN, ADJECTIVE, ADVERB, NONE
Relativizer	all	KEN, MIGÄ, KUMBA, MILLAIN, MI-, KU-
Syntactic function (main clause)	relativizer = ken, migä, kumba	SUBJECT, DIRECT OBJECT, INDIRECT OBJECT, MAJOR OBLIQUE CASE, POSSESSOR, OBJECT OF COMPARISON
Syntactic function (relative clause)	relativizer = ken, migä, kumba	SUBJECT, DIRECT OBJECT, INDIRECT OBJECT, MAJOR OBLIQUE CASE, POSSESSOR, OBJECT OF COMPARISON
Animacy	relativizer = ken, migä, kumba	HUMAN, ANIMATE, INANIMATE
Concreteness	relativizer = ken, migä, kumba	CONCRETE, ABSTRACT
Number agreement	all	YES, NO, N/A
Headedness	canonical	EXTERNALLY HEADED, INTERNALLY HEADED
	non-canonical	FREE, FUSED
Position of relative clause	canonical	POSTNOMINAL, PRENOMINAL, N/A
	non-canonical	BEFORE, MIDDLE, AFTER
Position of relativizer	all	CLAUSE-INITIAL, OTHER
Inverse attraction	all	YES, NO

5. Results

In this chapter, the results of the annotation process are presented. The absolute frequencies of the annotated variable are summarized in forms of tables and figures, which are accompanied by texts describing the relative frequencies of the variables. First, a general overview of what kind of relative clause constructions were found in the materials is given in section 5.1. Section 5.2. presents the syntactic functions of the common argument in both relative clauses and main clauses. Next, section 5.3. focuses on the choice of relativizer, covering the type of head, animacy and concreteness. The presence and absence of number agreement are analyzed in section 5.4. Finally, internal syntax of the relative clause construction is discussed in 5.5, including the position of the relativizer and the position of the relative clause.

5.1. General overview

In this section, the total number of relative clause constructions, the types of heads and the frequencies of different relativizers are summarized. The numbers of relative clause constructions found from the source materials from different time periods are presented in Figure 2 below.

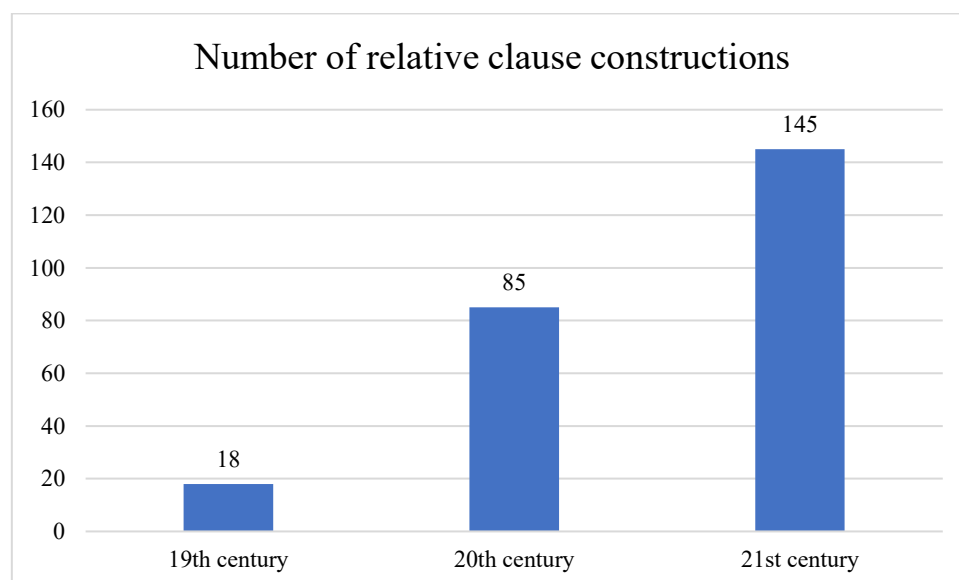


Figure 2. Relative clause constructions from different time periods.

Altogether, 248 relative clause constructions were identified in the source materials. 7 % of them came from the 19th century, 34 % from the 20th century and 58 % from contemporary

sources. This shows that the collected dataset is not balanced in terms of the different time frames the collected constructions come from. Since the share of 19th century examples is small compared to the 20th and 21st century materials, the 19th-century data are included in the analysis but are not used for direct quantitative comparison across time periods.

The *types of head* in the constructions are presented in Figure 3 below.

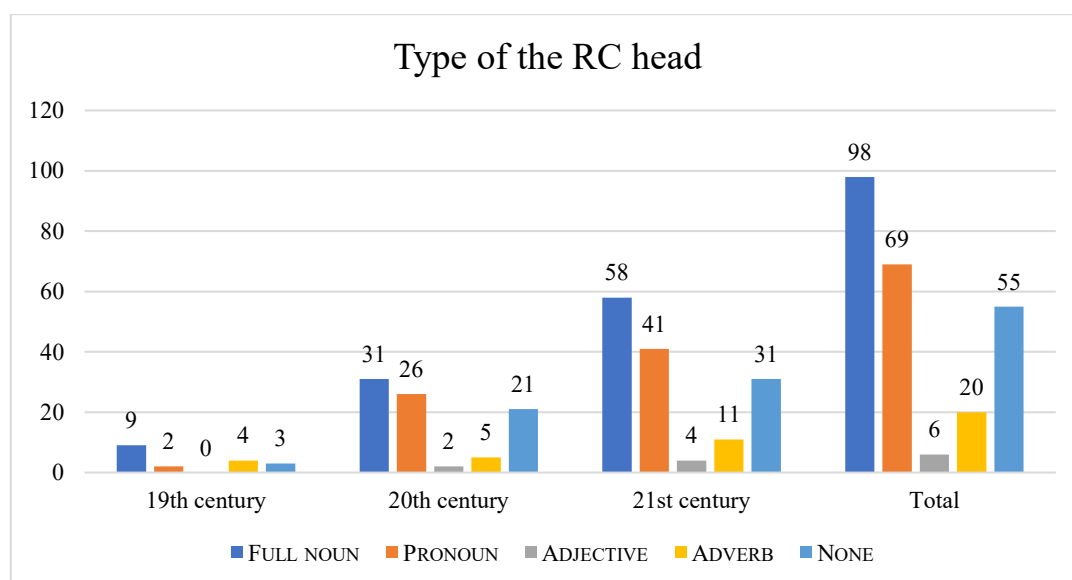


Figure 3. Part of speech of the relative clause head.

There were 98 canonical constructions found in the materials, meaning that in 40 % of the total number of constructions the head was a full noun. Out of them, 92 had an externally headed relative clause and 6 an internally headed one (for examples of internally headed relative clauses, see (93) and (94) in section 5.5.1.). Pronominal heads were the second most common, making up 28 % of the total. In 2 % of cases the head was an adjective and in 8 % an adverb. Counting together relative clauses with pronominal, adjectival and adverbial heads, there were 95 free relative clauses, which add up to 38 % of the total. The 55 other constructions did not have a head, meaning that fused relative clauses were present in 22 % of the constructions.

The distribution of different types of heads is similar across the 20th and the 21st century. Since there are only 18 constructions from the 19th century, it is not meaningful to compare that time period with the two others. Relative clauses with full noun heads make up 36 % of the total in the 20th century and 40 % in contemporary sources. Pronominal heads account for 31 % and 28 % of the totals, respectively. Free relative clauses are present in 25 % of constructions from the 20th century and in 21 % of the ones from the 21st century. Adverbs

make up 6 % of the total in the 20th century and 8 % in the 21st century. Adjectives are the rarest type of head in both: they appear in 2 % of the constructions from the 20th century and in 3 % from the 21st century.

In Figure 4 below, the overall frequencies of different *relativizers* are presented.

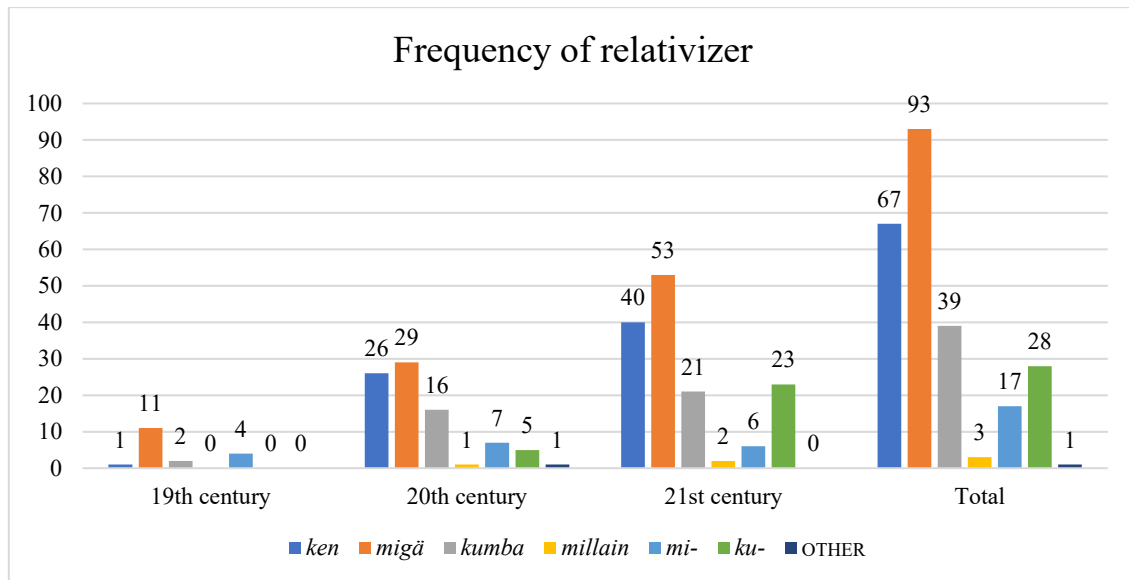


Figure 4. Overall frequencies of relativizers.

Altogether, *migä* is the most frequent relativizer in Ingrian, and it occurs in 38 % of relative clause constructions. After that, *ken* is the second most frequent relativizer with 27 %, followed by *kumba* with 16 %. *Millain* appears only in 3 constructions (1 %), and when it comes to the *mi-* and *ku-* forms, they appear in 7 % and 11 % of the constructions, respectively. One example was found where the relativizer is not any of the ones described in section 3.5., so it was categorized as OTHER. This example is presented in (54) below and it will be further discussed in chapter 6.

(54) Nirvi et al. 1967: 145

[*kuga vihastu-ĩ nāpuri-m pālle*], *sē noidu-ĩ*
 REL get_angry-PST.3SG neighbor-GEN on this curse-PST.3SG
 ‘That one who got angry with their neighbor, cursed her’

From the examples found from the 20th century, *migä* appears in 34 % of the constructions, *ken* in 31 %, *kumba* in 19 %, *mi-* in 8 %, *ku-* in 6 % and *millain* and OTHER in 1 % both. In contemporary materials, the corresponding relative frequencies are 37 % for *migä*, 28 % for *ken*, 14 % for *kumba*, 4 % for *mi-*, 16 % for *ku-*, 1 % for *millain* and 0 % for OTHER. The most

significant difference in terms of the frequencies of the relativizers between the 20th and 21st centuries is that it seems like pronominal adverbs with *ku-* stem are more frequent in contemporary materials than in the ones from the 20th century. Other than that, the distributions are similar over the two time frames.

5.2. Syntactic function of the common argument

This section focuses on relative clause constructions using the relative pronoun strategy, that is, constructions where the relativizer is *ken*, *migā* or *kumba*. The frequencies of syntactic functions of the common argument in the relative clause are summarized in Figure 5 below. In the figure, SU stands for subject, DO for direct object, IO for indirect object, OBL for major oblique case, POSS for possessor, and OCOMP for object of comparison.

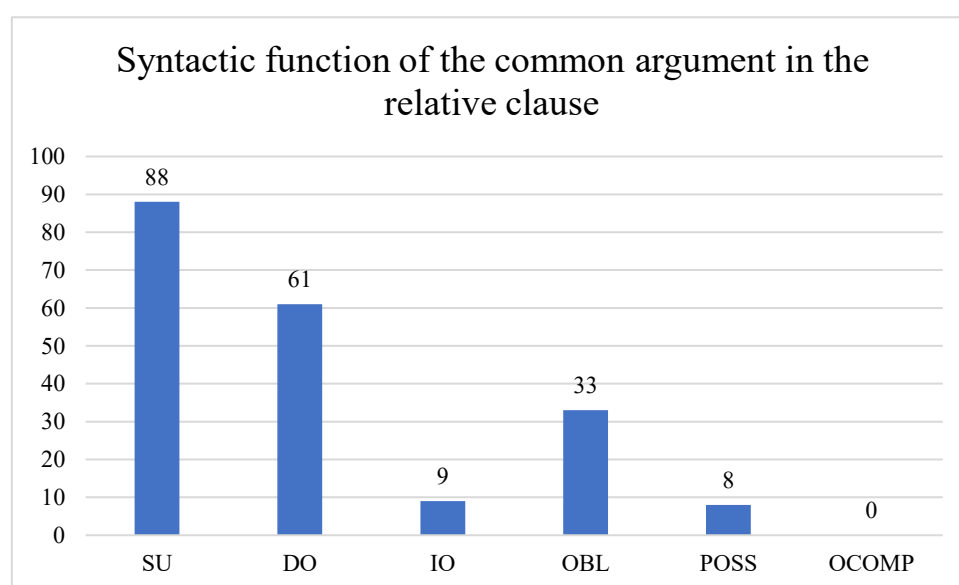


Figure 5. Syntactic functions in the relative clause.

Overall, subject relativization is the most frequent: in 44 % of the constructions the common argument is in the subject position in the relative clause. 31 % of constructions show direct object relativization and 17 % have oblique participant relativization. In addition, 5 % of the examples show indirect object relativization and 4 % possessor relativization. There are no examples in the data where the common argument is functioning as an indirect object or as an object of comparison. Examples of subject, direct object, indirect object, oblique participant and possessor relativization are given in (55) through (59) below.

- (55) Subject relativization, Püüdämääž_GI

ol-i mokkooma nemtsa [kumba miule teg-i ap'eratsija-n]
 be-PST.3SG such German REL 1SG.ALL do-PST.3SG operation-GEN
 'There was a German who did an operation on me'

- (56) Direct object relativization, Praažnikka_EN

ku ligemmäž tul-i še šuur pühä [midä pie-tti]
 when closer come-PST.3SG that big holiday REL.PART hold-PST.IPRS
 'When the big holiday that was celebrated came closer...'

- (57) Indirect object relativization, Nirvi et al. 1967: 148

se pāz-i mīhe-lle, [kel midägi hāsse-tti]
 that get-PST.3SG man-ALL who.ALL something speak-IPS.PST
 'That one got married (lit. got to a man) to whom something was said'

- (58) Oblique participant relativization, Skvortsad_EN

ja ain pan-i-mma pehko-loi-n alle [kuž ol-i vähä lun-da]
 and always put-PST-1PL bush-PL-GEN under REL.INE be-PST.3SG little snow-PART
 'And we always put [the white bread] under the bushes where there was little snow.'

- (59) Possessor relativization, Porkka 1885: 144

Kunigas käsk-i saa'a si-dä kuldalintu-a, [mill tämä sulga on].
 king order-PST.3SG get-INF that-PART golden_bird-PART REL.ADE this
 feather be.3SG
 'The king ordered to catch the golden bird that has this feather.'

The syntactic functions of the common argument in the main clause are presented in Figure 6 below.

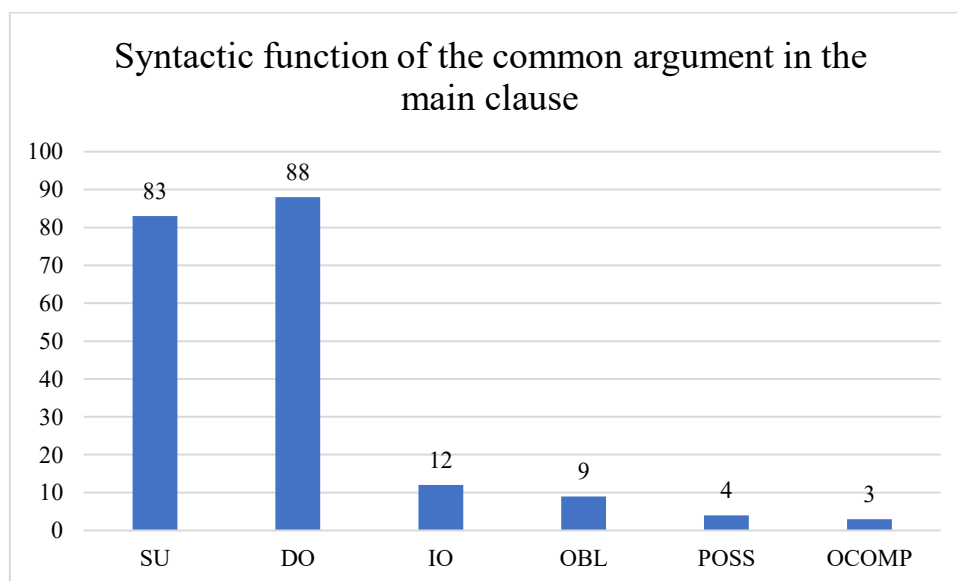


Figure 6. Syntactic functions in the main clause.

In main clauses, the common argument is the subject of the clause in 42 % of the constructions, direct object in 44 %, indirect object in 6 %, an oblique participant in 5 %, a possessor in 2 % and an object of comparison in 2 %. In example (55) above, the head *nemtsa* is a subject in the main clause. An example of the head, *kuldalintua*, as a direct object can be seen in (59). In (58), the phrase *pehkoloin alle* is an oblique participant. In (38), which is reproduced below, the head *štaraštalle* is functioning as an indirect object in the main clause. Examples where the common argument is filling the role of possessor and object of comparison are given in (60) and (61), respectively.

(38) Kala_ja_metsä_EN, glosses from Markus & Rozhankiy (2022a: 328)

i še haili ain vee-ĩtii štarašta-lle [kumba šene-n
 and this sprat always bring-IPS.PST village.head-ALL REL this-GEN
haili-n jago-i]
 sprat-GEN divide-PST.3SG
 ‘And the sprats were always taken to the village head, who divided the sprats
 [between people].’

(60) Common argument as a possessor in the main clause, Huvä_elo_AI

kaigille pittää anĩtaa šuure-d kiidokše-d
 everyone.ALL have_to.INF give.INF big-PL thank_you-PL
[ked käi-väd miñnua katso-maa-ž]
 REL.PL go-3PL 1SG.PART look-SUP-INE
 ‘[I] have to give great thanks to everyone who comes to visit me’

(61) Common argument as an object of comparison in the main clause, Praažnikka_EN

šiul on kakš kerĩta-a šuure-mb praažnikka tänäbää
 2sg.ADE be.3SG two time-PART big-COMP celebration today
pa sravneniju šenen kera [ken šö-i kaig]
 compared that.GEN with REL eat-PST.3SG everything
 ‘You have twice the celebration compared to the one who had eaten everything’

5.3. Choice of relativizer

The variables affecting the choice of relativizer that were analyzed in this work are the type of head, animacy and concreteness. These were annotated for constructions where the relativizer is a relative pronoun: *ken*, *migä* or *kumba*. This was the case in 199 constructions, accounting for 80 % of all the relative clause constructions found in the materials.

5.3.1. Type of head

The frequencies of relative pronouns appearing with different types of heads are presented in Table 8 below.

Table 8. Relative pronouns and types of heads.

	FULL NOUN	PRONOUN	ADJECTIVE	ADVERB	NONE
KEN	15	25	1	0	26
MIGÄ	43	29	3	0	18
KUMBA	24	9	2	1	3

The most common relative pronoun *migä* is used mostly with full noun heads (46 %) and pronouns (31 %), but it can also appear in fused constructions (19 %) and with adjectives (3 %). *Kumba* can be used with all types of heads, but in 62 % of cases, it appears in a construction with a full noun as the head. *Ken* is used most often in free relative clauses with a pronominal head (37 %) or fused relative clauses without a head (39 %). It can, however, appear within constructions with full nouns as the head. For canonical constructions where the head is a full noun, see examples (36) through (38). Examples of *ken*, *migä* and *kumba* with other types of heads are given in (62) through (71) below.

(62) *Ken*, pronoun, Ariste 1960: 44

se istu-i sēl [kel löü-zī]
 that sit-PST.3SG there REL.NOM find-PST.3SG
 ‘That one who found (it) was sitting there’

(63) *Ken*, adjective, Ariste 1960: 42¹⁶

ol-i pan-du mokkōma-d, [ken mahtā-d]
 be-PST.3SG put-PASS.PTCP such-PL REL be_able_to-3PL
tehä olu-zi-a]
 do.INF beer-PL-PART
 ‘Such [women] had been put [to work] who can make beer.’

(64) *Ken*, no head, Troitsa_ja_olud_VV

[ken mahto-i tehä], teh-tii olut-ta
 REL be_able_to-PST.3SG do.INF do-PST.IPRS beer-PART
 ‘Those who could, made beer’

¹⁶ Note that *mokkōma* also has pronominal use. In this example, it is clear from the context that it is used as an adjective. The head noun, *women*, is not explicitly mentioned but its presence is reflected in the English translation.

(65) *Migä*, pronoun, Hivä_elo_AI

meil on kaig [midä pittää]
 1PL.ADE be.3SG everything REL.PART have_to.INF
 ‘We have everything that is necessary’

(66) *Migä*, adjective, Šoomi_MM

muu-d ei ol-d [midä anttaa]
 other.PART NEG be-ACT.PTCP REL.PART give.INF
 ‘There was nothing else to give’

(67) *Migä*, no head, Püüdämaaž_GI

ei ol-d [miž töö-dä tehhä]
 NEG be-ACT.PTCP REL.INE work-PART do.INF
 ‘There was no place to work’

(68) *Kumba*, pronoun, Laanest 1966a: 129

senem piD-i arvaDa sīZ [kumBa istu-i]
 that.GEN have_to-PST.3SG guess.INF then REL sit-PST.3SG
sīl nurGā-Z
 there corner-INE
 ‘Then the one who was sitting in the corner had to guess.’

(69) *Kumba*, adjective, Laanest 1966a: 122

ol-ti argi päiväis-t [kummad mere-lle sio-ttī kīn]
 be-PST.IPRS everyday-PL REL.PL sea-ALL tie-PST.IPRS closed
 ‘We had everyday ones, which were tied [when going] to the sea.’

(70) *Kumba*, adverb, Elo(B)_AI

hää šiz oštaa šiid lekarstvoj [kummaž ovad...]
 3SG then buy.3SG there medicine REL.INE be.3PL
 ‘So she buys the medicine there, where they are... [cheaper for pensioners]

(71) *Kumba*, no head, Ariste 1960: 9

[*kumB ol-i riGaZ*], *naglo-i-l lö-i kīn*
 REL be-PST.3SG rich nail-PL-ADE hit-PST.3SG closed
 ‘That one who was rich, sealed it with nails.’

The frequency of the relative pronoun *ken* with different types of heads over time is shown in Figures 7.

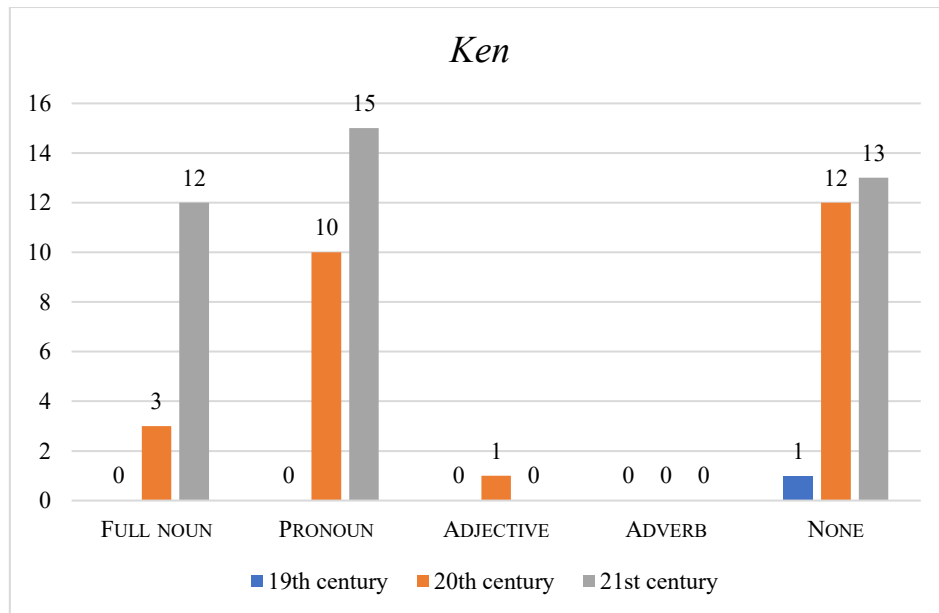


Figure 7. Frequency of *ken* over time.

In the 20th century, *ken* was used with fused relative clauses in 46 % of cases, with a free relative clause in 42 % of cases and with a full noun head in 12 % of cases. The most drastic change compared to contemporary materials concerns the last: *ken* appears with a full noun head in 30 % of the constructions in the 21st century. The relative frequency of *ken* in fused relatives in contemporary examples is 33 % and in free relatives it is 38 %. There is only one example from the 19th century where *ken* appears at all. In the construction, it is used with a fused relative clause. The example is presented in (72) below.

(72) Porkka 1885: 131

[*Ken senen linnu-n varastaa*], *niin siit sill kerra-l tabe-taa*.
 REL that.GEN bird-GEN steal.3SG so then that.ADE time-ADE kill-IPS
 ‘Whoever steals this bird will be killed.’

The frequency of the relative pronoun *migä* with different types of heads over time is shown in Figure 8.

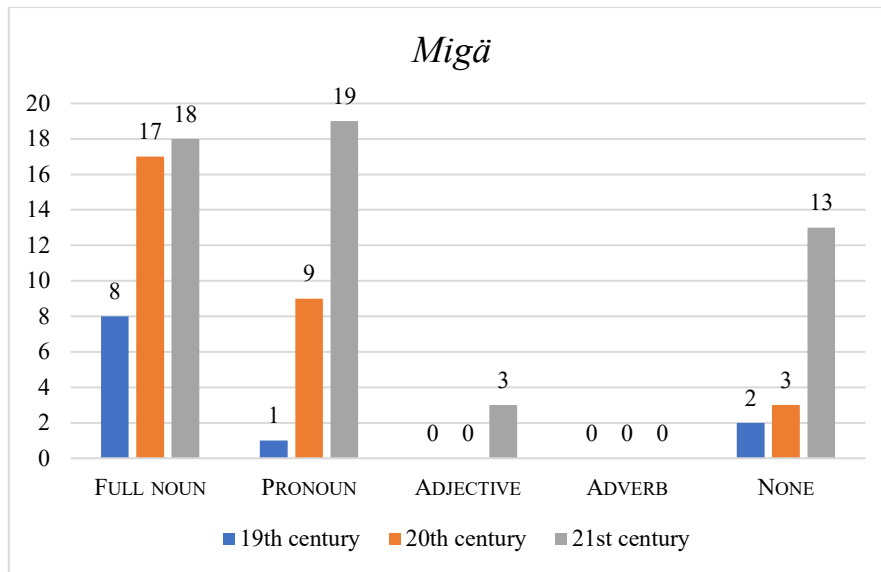


Figure 8. Frequency of *migä* over time.

In examples from the 20th century, *migä* appears with full noun heads in 59 % of cases, with pronominal heads in 31 % of cases and in 10 % of fused relative clauses. In contemporary materials, full noun heads and pronominal heads are almost as frequent, accounting for 34 % and 36 % of the total, respectively. Adjectival heads are present in 6 % of the constructions and the proportion of *migä* within fused relatives is 25 %. The most noticeable changes between the time periods are that the frequencies of *migä* have risen significantly with pronominal heads and within fused relatives in contemporary materials as compared to data from the 20th century.

The frequency of the relative pronoun *kumba* with different types of heads over time is shown in Figure 9.

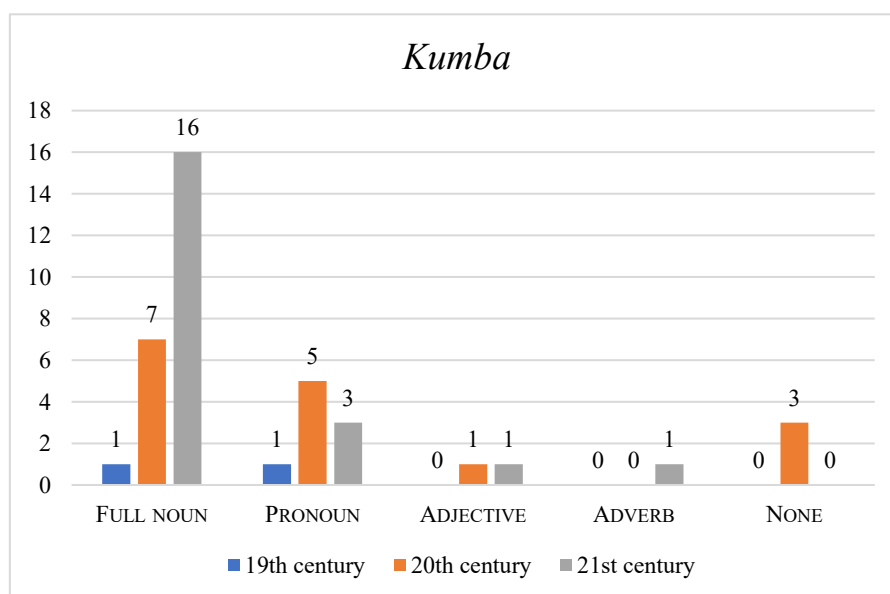


Figure 9. Frequency of *kumba* over time.

In the 20th century materials, *kumba* is used with a full noun head in 44 % of the cases, with pronominal heads in 31 %, within fused relative clauses in 19 % of the cases and with an adjectival head in 6 %. In contemporary materials, the share of *kumba* with full noun heads is 76 %, with pronominal heads 14 %, with adjectival heads 5 %, and with an adverbial head 5 %. The most drastic change has happened within the full noun heads, where the relative frequency has risen by 32 percentage points between the 20th and 21st centuries. Another significant observation is that *kumba* does not appear in fused relative clauses in the contemporary materials.

5.3.2. Animacy

The animacy of the referent was annotated for every relative clause construction where the relativizer was *ken*, *migä* or *kumba*. The animacy values of different relative pronouns are presented in Figure 10 below.

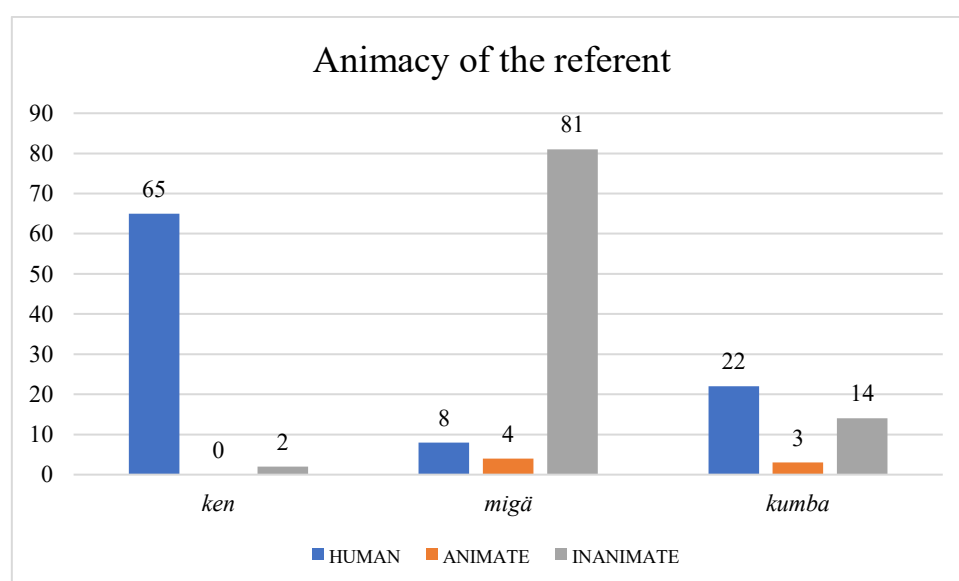


Figure 10. Animacy with different relative pronouns.

The relative pronoun *ken* is used with human referents in 97 % of the cases. In the two instances where *ken* is used with a referent tagged as inanimate, the head of the relative clause is *taloï* ‘house, building’. One such case is shown in (73) below. Note, however, that the word *taloï* is used in a particular context in this example. The (in)animacy of the head is discussed further in chapter 6.

(73) Lehmä_EI

vot mamoi-ld ain hilegaištaa varašš-i-n no sille
so mother-ABL always little_by_little steal-PST-1SG well that.ALL
talloi [*kel ei ol-d enžigää miittää*]
house.ILL REL.ADE NEG be-ACT.PTCP at_all nothing.PART
‘So I used to steal from my mother little by little [and give] to that house that had nothing at all.’

Migä is mostly used with inanimate referents (see examples (65) and (66) above). It can, however, be used with human and animate referents. Examples of these are given in (74) and (75).

(74) *Migä*, human referent, Ariste 1960: 42

paimen, [*migä katsō karjā*], *käü-b ümbär karja-n*
shepherd REL look.3SG herd.PART go-3SG around herd-GEN
‘The shepherd that is looking after the herd goes around the herd.’

(75) *Migä*, animate referent, Porkka 1885: 144

Kunigas käsk-i saa’a sidä kuldalintu-a, [*mill*
king order-PST.3SG get.INF that.PART golden_bird-PART REL.ADE
tämä sulga on].
this feather be.3SG
‘The king ordered to get the golden bird that has this feather.’

When it comes to *kumba*, it can be used with human, animate and inanimate referents (see examples (68) through (71)). In most cases (56 %), however, *kumba* appears with a human referent.

5.3.3. Concreteness

The concreteness values of the common arguments are presented in Table 9 below.

Table 9. Concreteness with relative pronouns.

	KEN	MIGÄ	KUMBA
CONCRETE	67	49	39
ABSTRACT	0	44	0

Whenever the common argument is referring to an abstract entity, *migä* is used as the relative pronoun. It is also frequently found with a concrete reference. *Ken* and *kumba*, on the

other hand, are used only to refer to concrete entities in the examined data. Examples of *migä* referring to concrete and abstract entities are given in (76) and (77) below.

(76) *Migä*, concrete referent, Kūlmä_kežä_IV+VF

nęd ižännä-d [*mid oļl-ii-d ennemmäž*]
 those owner-PL REL.PL be-PST-3PL before
 ‘Those, the owners, who were [here] before’

(77) *Migä*, abstract referent, Porkka 1885: 131

ku tê-t senen, [*minen miä käze-n*]
 if do-2SG that.GEN REL.GEN 1SG order-1SG
 ‘If you do that what I tell you to do...’

5.4. Number agreement

The presence of number agreement was annotated for all relative clause constructions with a relative pronoun as the relativizer. For *kumba*, the presence of number agreement can be unambiguously determined for all case forms, but in relative clauses with *ken* and *migä*, this can be done only in nominative contexts, because separate plural forms do not exist for other cases than the nominative (see section 3.5.). The presence of number agreement across different time periods for all relative pronouns combined is shown in Figure 11 below.

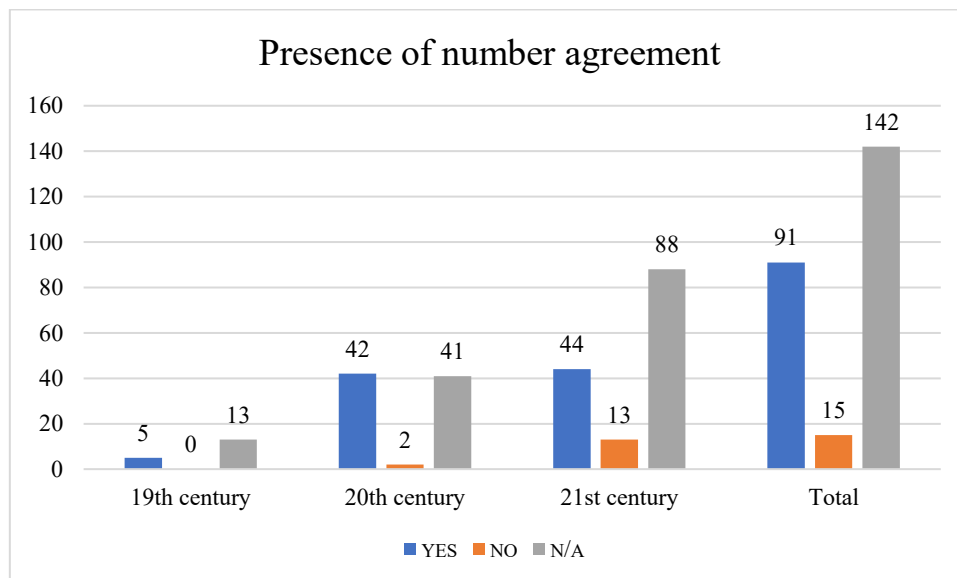


Figure 11. Presence of number agreement in Ingrian relative clause constructions.

In 57 % of cases, the presence of number agreement cannot be verified. Still, there are notably many (106 out of 248) instances where number agreement can be analyzed. In 14 % of

the total cases where the singular and plural forms of the relativizer are distinguishable, the relativizer does not agree with the head in number. Examples where number agreement is present, missing and not distinguishable are given in (78) through (80) below.

(78) Number agreement, Kolhožiiž_PE

i vergo-d [mid ol-ttii] kaig kolhožii
 and net-PL REL.PL be-PST.IPS everything kolkhoz.ILL
 ‘and the [fishing] nets one had, everything [was taken] to the kolkhoz (collective farm)’

(79) No number agreement, Selsavediiž_GI

a [ken ovad nqore-d] nii hqōq ovad
 but REL.SG be.3PL young-PL so 3PL be.3PL
 ‘And those who are young, they are...’

(80) No number distinction, Troitsa_ja_olud_VV

[kel ol-i ruiš-t], keide-ttii kovašt
 REL.ADE be-PST.3SG rye-PART cook-PST.IPS much
 ‘those who had rye, made a lot [of beer]’

In examples from the 20th century, the head and the relativizer agree in number in 95 % of the cases, whereas in contemporary examples, the corresponding proportion is 77 %. In other words, the number of cases where number agreement is not present grows from 5 % to 23 % between the two time periods. However, this result does not take into account whether the relativizer is *ken*, *kumba* or *migä*. The presence of number agreement for each relative pronoun separately is shown in Figure 12 below.

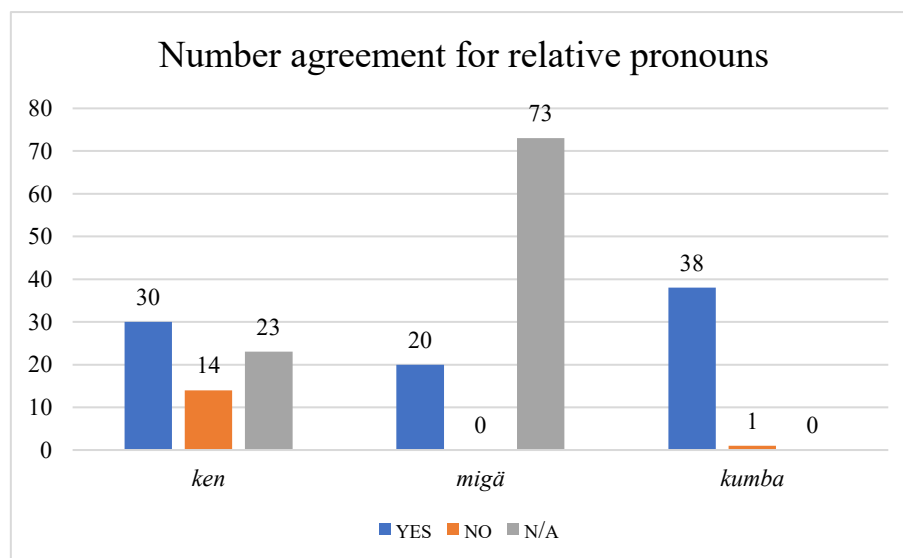


Figure 12. Number agreement of *ken*, *migä* and *kumba*.

In 14 out of the 15 cases where the relativizer does not agree in number with the head, the relativizer is *ken*. Overall, *ken* does not show number agreement in 32 % of cases where the singular and plural forms are distinguishable. The number agreement in relative clause constructions with *ken* as the relativizer in the nominative case over time is presented in Figure 13 below.

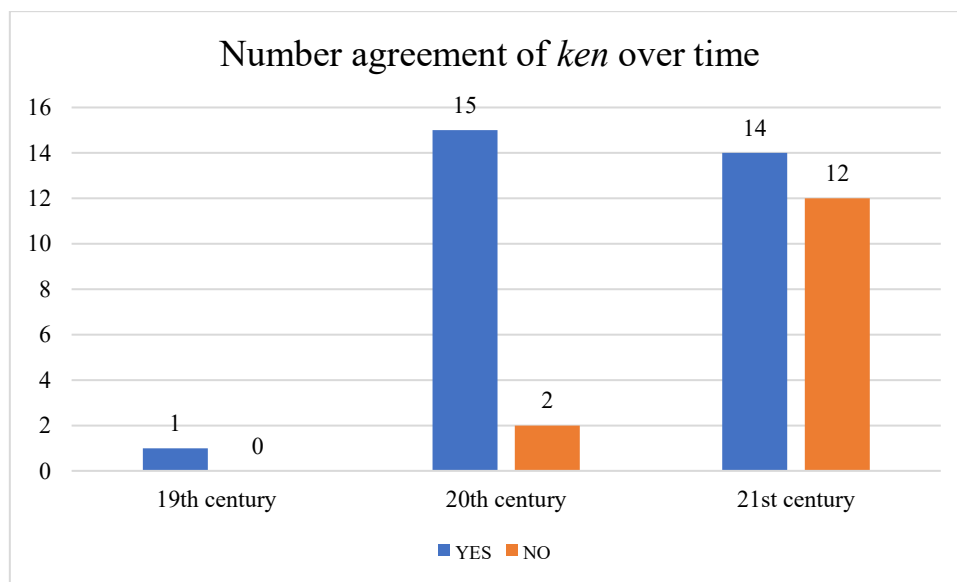


Figure 13. Number agreement of *ken* in nominative contexts in different time periods.

In 46 % of the cases *ken* does not agree in number with the head in contemporary examples. This is a radical change compared to the materials from the 20th century, where number agreement was not present in only around 12 % of the cases. In all of the 14 constructions with *ken* that do not show number agreement, it is always so that the singular form is used in a context where a plural would be expected. There are no instances where the nominative plural form *ked* would be used in a context where the nominative singular form *ken* is expected. Two examples where *ken* is used instead of *ked* are given in (81) and (82) below.

(81) Ariste 1960: 40

[*ken* *männ-ī-d* *sõmē* *kõi-st*], *ott-ī-d* *paljo keralē*
 REL.SG go-PST-3PL Finland.ILL home-ELA take-PST-3PL a.lot along
 ‘Those who went to Finland from home, took a lot (of things) along.’

(82) Püüdämääž_GI

a heil ei ol-d *męgh-i-ä* [*ken* *mere-l* *kä-i*]
 but 3PL.ADE NEG be.ACT.PTCP husband-PL-PART REL.SG sea-ADE go-PST.3SG
 ‘And they did not have husbands who went to the sea.’

An interesting case of number disagreement is presented in (83). The head is an indefinite pronoun in which the form *ked* is included. This is a plural form of *ken-ikkee* ‘somebody’. Despite this, the relative pronoun that follows immediately after the head is in the singular form, *ken*.

(83) Lastotškad_LK

tuumaa-n što vot [RUS] imena hõõ anttaa-d heid
 think-1SG that exactly 3PL give-3PL 3PL.PART
ked-ikkenä [*ken noiššoo heidä katso-maa obetta-maa*]
 somebody.PL REL.SG start.3SG 3PL.PART look-SUP.ILL teach-SUP.ILL
 ‘I think that exactly they give them to some people who will look after them and teach.’

Other than the cases where the relativizer is *ken*, there is only one instance where number agreement is not present. In example (84), the relativizer is *kumba*, and it is expected to be in a plural form, such as *kummad* or even *kumppia*. This example will be further discussed in chapter 6.

(84) Külvettii_EN

oļl-ii-d mokkooma-d spetsal’nije prispasabl’enija
 be-PST-3PL such-PL special equipment
 [*kumba vot niin lõõ-ttii lõõ-ttii*]
 REL.SG so so hit-PST.IPRS hit-PST.IPRS
 ‘There were such special devices, which were used to beat, beat this way.’

It seems that the number agreement is motivated by semantics more than the morphological form of the head. This is illustrated in example (85), where the head *väkkiä* is in morphological singular, but the meaning is plural. Therefore, the relativizer is also in the plural.

(85) Kala_ja_metsä_EN

no hää ain kutsuu väkki-ä sidä [ked
 but 3SG always invite.3SG people-PART that.PART REL.PL
lägäjää-d]
 speak-3PL
 ‘but he always invites the people who can speak [in Ingrian]’

5.5. Internal syntax

This section focuses on the position of the relativizer, and the position of the relative clause in relation to the main clause. These variables were annotated for all of the relative clause constructions found in the source materials. Another annotated parameter was inverse attraction; however, it was not attested in any of the constructions. For this reason, the phenomenon does not have a separate section in this chapter. However, it will be discussed in chapter 6.

5.5.1. Position of relativizer

The constructions were annotated in terms of whether the relativizer occurs at the beginning of the relative clause or in some other position. The results are presented in Figure 14 below.

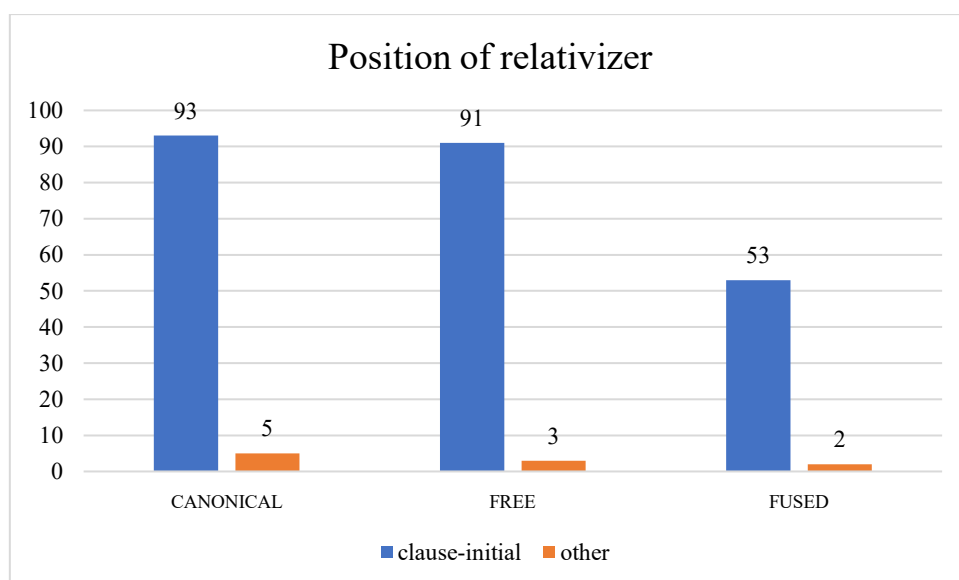


Figure 14. Position of the relativizer.

In 237 constructions (96 %), the relativizer occurs at the beginning of the relative clause. Examples of construction with a canonical relative clause, a free relative clause and a fused relative clause where the relativizer appears clause-initially are given in (86) through (88) below.

(86) Püüdämääž_GI

a miä šiare-n kera [*kumba nqɔrem miñnua vɛɣel ono*]
and 1SG sister-GEN with REL younger 1SG.PART yet be.3SG
'And me and my sister who is even younger than me [brought the fish]'

(87) Marjad(B)_ST

to miä ain tōrri-iži-n i tōrri-iži-n
that 1SG always talk_nonsense-PST-1SG and talk_nonsense-PST-1SG
kaigelaiš-t [*midä ei piä*]
all_sorts-PART REL.PART NEG have_to.CNG
‘Since I just talked and talked nonsense about all sorts of unnecessary [things]’

(88) Metsääž_EN

miul ei-qo [kenen kera läädä-gää]
1SG.ADE NEG-be.CNG REL.GEN with talk-PRCL
‘I have no one to talk to’

Deviations can be found both in canonical and non-canonical constructions. Examples of a canonical and a free relative clause where the relativizer is not in the first position are given in (89) and (90) below, respectively.

(89) Ariste 1960: 42

a illā-st opp-izi-n urakk-oj-a, [se obastaja
but evening-ELA study-PST-1SG homework-PART-PL that teacher
midä anto-i oppiš].
REL.PART give-PST.3SG study-INF
‘But in the evening, I did the homework the teacher gave me to study’

(90) Ariste 1960: 15

[*şū-şť midä laze-t*] *şe on şünni*
mouth-ELA REL.PART let-2SG that be.3SG sin
‘That what you let out of your mouth is a sin’

In three constructions, there is a complementizer before the relativizer. One such case, which will be further discussed in chapter 6, is shown in (91).

(91) Laanest 1966a: 113

no miä nüt sen verre-n laula-n [joD migä sīn laule-ttī
well 1SG now that.GEN song-GEN sing-1SG that REL here sing-PST.IPRS
Gu tul-ti]
when come-PST.IPRS
‘Well, I’ll now sing the song that was sung here when they came...’

In two non-canonical relative clause constructions, the relativizer is in a clause-final position. One of them is shown in (92) below.

(92) Kazahstaniž_RP

[šõõvvä [ei ol-d] midä]
eat.INF NEG be.ACT.PTCP REL.PART

‘There was nothing to eat (lit. There was not what to eat)’

In constructions containing an internally headed relative clause, the relativizer is clause-initial in 5 out of 6 cases. Two examples are given below: in (93), the relativizer is located at the beginning of the relative clause and in (94), in another position. Both of them are discussed further in chapter 6.

(93) Hirvi_GI

šiž perrää [ken tull-ii-d miihe-d]
then after REL come-PST-3PL man-PL

‘After that the men who came...’

(94) Püüdämääž_GI

tak što kogo [meil migä šugu ono]
so that whole 1PL.ADE REL family be.3SG

‘So the whole family that we have [went fishing]’

5.5.2. Position of the relative clause

Out of the 92 externally headed relative clauses, 91 are postnominal. The only exception to this is shown in (95). It will be discussed further in chapter 6.

(95) Laanest 1966a: 139

[kumBa jo laoDelDu kaiG] Ze kirkko-kkI
REL already destroyed all that church-PRCL

‘The church that has been completely destroyed’

When it comes to linking, 73 of the externally headed relative clauses are embedded in the head (79 %), and 19 are adjoined (21 %). In constructions where the relative clauses are adjoined to the main clause, the constituent between the head and the relative clause is usually a verb. It can be accompanied by an adverb, but usually there is only a verb separating the relative clause from its head. Examples of these are presented in (96) and (97) below.

(96) Porkka 1885: 144

Siä sen linnu-n sa-i-t, [midä miä käz-i-n].
2SG that.GEN bird-GEN get-PST-2SG REL.PART 1SG order-PST-1SG

‘You got the bird I told you to.’

(97) Laanest 1966a: 127

näiDä l̥johko-loj-a enne ei tant̥si-ttu [miDä nūt
 these.PART easy-PL-PART before NEG dance.PASS.PTCP REL.PART now
kaik tant̥si-Dā]
 all dance-IPRS
 ‘These easy ones that are all danced nowadays were not danced back then’

It is possible to have more than one relative clause referring to one head. In such a case, the relative clauses are combined either syndetically with conjunctions or asyndetically, without an explicit linking device. An example of asyndetically linked relative clauses is presented in (98) below. Additionally, there is one instance where an entire relative clause construction separates the relative clause from the head. That one is shown in (99).

(98) Praažnikka_EN

piitää ain muištaa om̥m-i-a vanhempp-i-a [ked
 have_to.INF always remember.INF own-PL-PART parent-PL-PART REL.PL
kool-ii-d] [ked šiun kažvatt-ii-d]
 die-PST-3PL REL.PL 2SG.GEN raise-PST-3PL
 ‘You always have to remember your parents who died, who raised you.’

(99) Nirvi et al. 1967: 148

tüdöi-n sugu tali poja-n sugu [kumbo ott-i nais-ta]
 girl-GEN family or boy-GEN family REL take-PST.3SG wife-PART
tali [kumbo m̥ihelle män-i], sis se sugu kutsu-ttī
 or REL married go-PST.3SG then that family invite-PST.IPRS
 ‘The girl’s family, or the boy’s family who was taking a wife or who got married
 (lit. went to a man), then that one was invited...’

The positions of the relative clauses in non-canonical relative clause constructions are presented in Figure 15 below.

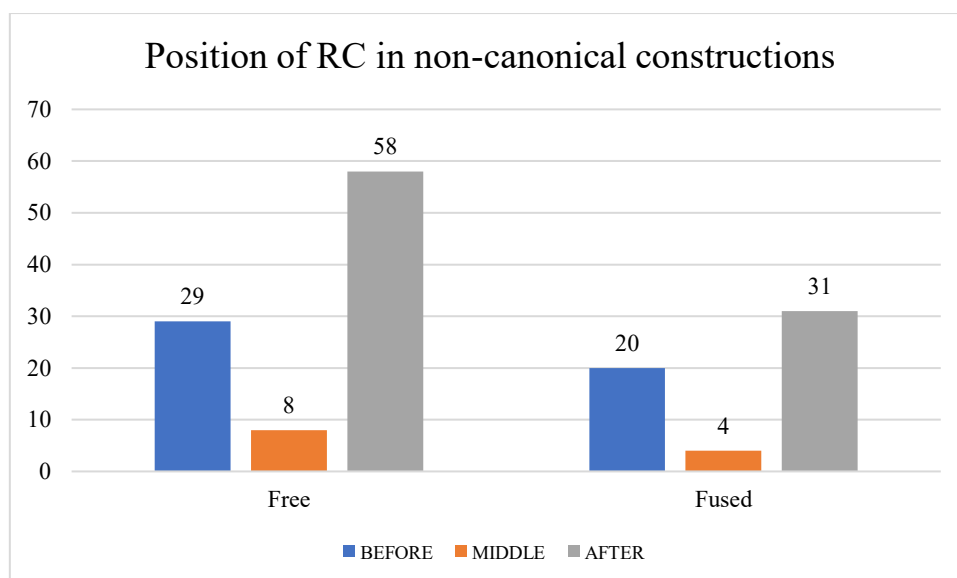


Figure 15. Position of the relative clause in relation to the main clause in non-canonical constructions.

Compared to canonical constructions, there are considerably more instances where the relative clause precedes the main clause among the non-canonical relative clause constructions (31 % of free relatives, 36 % of fused relatives). Still, in most constructions the relative clause comes after the main clause (61 % of free relatives, 56 % of fused relatives). It is possible for the relative clause to be in the middle of the main clause in both free and fused relatives. Examples of non-canonical relative clauses that come before in the middle and after the main clause are given in (100) through (102) below.

(100) Porraž_EN

[*kel on mehe-d*] *hõõ ošt-ii-d hüvä-d poritaha-d*
 REL.ADE be.3SG man-PL 3PL buy-PST-3PL good-PL ladder-PL
 ‘Those who have men [in the house], they bought good ladders’

(101) Nirvi et al. 1967: 140

a kalmoi-l [ken kehtajä], käü-p tihim_pälle
 but graveyard-ADE REL feel_like_doing.3SG go-3SG very_often
 ‘But those who feel like it go to the graveyard very often’

(102) Šoomi_MM

Šö-i-mmä [midä taho-i-mma]
 eat-PST-1PL REL.PART want-PST-1PL
 ‘We ate what we wanted’

6. Discussion

This chapter focuses on interpreting the results and especially finding explanations for deviations from the general trends presented in chapter 5. The topics covered in this chapter are a) the use of *kuga* as a relative pronoun, b) the missing syntactic role of the common argument in the relative clause, c) the use of *ken* with a non-human referent, d) the number disagreement of *kumba*, e) deviations of the position of the relativizer, f) internally headed relative clauses, g) deviation of the position of the relative clause, and h) inverse attraction.

a) In 5.1., one relativizer which has not traditionally been categorized as a relativizer in Soikkola Ingrian was found. *Kuga* appears as a relative pronoun in one sentence (54) in the dataset. The sentence came from a speaker born in the village of Säätinä (Rus. *Slobodka*) who later moved to the former village of Takaväljä (Rus. *Novaja Derevnja*) located in the area where Lower Luga Ingrian was spoken.

(54) Nirvi et al. 1967: 145

[*kuga vihastu-ĩ nāpuri-m pālle*], *sē noidu-ĩ*
REL get_angry-PST.3SG neighbor-GEN on this curse-PST.3SG
'That one who got angry with their neighbor, cursed her'

Kuga was described as a relative pronoun in the Hevaha dialect and Ingrian Finnish already by Porkka (1885: 85). In Nirvi's dictionary (1971: 211–212), in addition to the Hevaha dialect, two Soikkola Ingrian examples are given where the relative pronoun is *kuga*. These examples come from the villages of Saarove (Rus. *Jugantovo*) and Säätinä (Rus. *Slobodka*). It is possible that the use of *kuga* as a relative pronoun is a feature of the southern subdialect of Soikkola Ingrian, or a feature found in only these villages. Due to lack of data, this cannot be confirmed in this study. As there is only one construction from one speaker where the relativizer is *kuga*, it can be treated as an individual deviation.

b) As shown in section 5.2., examples of the common argument functioning as the subject, direct object, indirect object, oblique participant and possessor in the relative clause were found in the data. No example was found where the common argument acts as an object of comparison in the relative clause. However, this does not prove that objects of comparison cannot be relativized in Ingrian. The dataset used in this study is limited and the most likely explanation is that relativized objects of comparison just appear rarely in spontaneous speech.

Even if it were ungrammatical to relativize objects of comparison, it would not violate the Noun Phrase Accessibility Hierarchy. In fact, as noted by Dixon (2010: 321–322), the number of syntactic functions a language allows for the common argument in the main clause is either the same or higher than the number of allowed syntactic roles in the relative clause. Either Ingrian allows for the common argument to fill any syntactic function on the Noun Phrase Accessibility Hierarchy both in the main clause and in the relative clause, or it allows for all roles in the main clause and up to possessor in the relative clause. Out of these, the first seems more likely, but getting confirmation for this would require more research.

c) In (73), which is reproduced below, one of two examples was presented where the relative pronoun *ken* is used with an inanimate referent. In both of these examples the head *talo* ‘house, building’ is used. Although the concrete meaning of this word is inanimate, it can be understood with having the meaning ‘household’, that is, the word refers to the people living in the house. This explains the choice of relative pronoun, *ken*.

(73) Lehmä_EI

<i>vot</i>	<i>mamoi-ld</i>	<i>ain</i>	<i>hilegaištaa</i>	<i>varašš-i-n</i>	<i>no</i>	<i>sille</i>
so	mother-ABL	always	little_by_little	steal-PST-1SG	well	that.ALL
<i>talloi</i>	[<i>kel</i>	<i>ei</i>	<i>ol-d</i>	<i>enžigää mittää</i>]		
house.ILL	REL.ADE	NEG	be-ACT.PTCP.	at_all	nothing.PART	

‘So I used to steal from my mother little by little [and give] to that house that had nothing at all.’

Ken was not found to occur with any animate referents in the dataset used in this study. Note, however, that a referent annotated as ANIMATE only occurred in 7 constructions in total. In all of them, the relative pronoun was either *migä* or *kumba*. 5 of these are from folk stories from the 1880s and 2 from contemporary sources. On a dataset this small, it is not possible to make strong generalizations about animate referents. However, an example has been found where *ken* is used with an animate referent. It is presented in (103) below (from Norvik et al. 2022). What is clear is that *ken* is clearly preferred with human referents, and it does not appear with inanimate referents in the dataset.

(103) *Ken* with an animate referent

lehmä-D, [*ked...*]
 cow-PL, REL.PL
 ‘Cows, who...’

d) There was one example where the relative pronoun *kumba* did not agree in number with the head. This example (84) is reproduced below. However, it must be noted that the head is in Russian, which could affect the number agreement. The head, *spetsal'nije prispasabl'enija*, is plural in Russian, and so is the Ingrian determiner of the head, *mokkoomad*. Before the Russian part, the speaker holds a small pause, indicating that the speaker switches the language from Ingrian to Russian in between. Due to this code-switching effect and the fact that this appears only in one example, this example can be considered irregular.

(84) K÷lvettii_EN

<i>oñ-ii-d</i>	<i>mokkooma-d</i>	<i>spetsal'nije</i>	<i>prispasabl'enija</i>
be-PST-3PL	such-PL	special	equipment
[<i>kumba</i>	<i>vot</i>	<i>niin</i>	<i>lõõ-ttii</i>
			<i>lõõ-ttii</i>]
REL.SG	so	so	hit-PST.IPRS hit-PST.IPRS

‘There were such special devices, which were used to beat, beat this way.’

e) In section 5.5.1., two examples were shown where the position of the relativizer is not clause-initial. These examples, (89) and (90), are reproduced below. In both examples, the relativizer appears in the second position. In (89), the teacher had been mentioned in the preceding discourse, and in (90), the preceding sentence describes a person who speaks of bad things, meaning that the mouth is also present in the discourse, at least implicitly. The constituents before the relativizer are known from the context, which makes it seem like the unusual placement of the relativizer is somehow connected with the information structure of the sentence.

(89) Ariste 1960: 42

<i>a</i>	<i>illā-st</i>	<i>opp-izi-n</i>	<i>urakk-oj-a,</i>	[<i>se</i>	<i>obastaja</i>
but	evening-ELA	study-PST-1SG	homework-PL-PART	that	teacher
<i>midä</i>	<i>anto-i</i>	<i>oppĩš</i>].			
REL.PART	give-PST.3SG	study-INF			

‘But in the evening, I did the homework the teacher gave me to study’

(90) Ariste 1960: 15

[<i>şũ-şť</i>	<i>midä</i>	<i>laže-t</i>]	<i>şe</i>	<i>on</i>	<i>şünni</i>
Mouth-ELA	REL.PART	let-2SG	that	be.3SG	sin

‘That what you let out of your mouth is a sin’

In the data, there were three constructions where a complementizer, *jod* ‘that’, is used before the relativizer. No other sentences were found where a complementizer would precede the relativizer, and the three examples with *jod* were all produced by the same speaker. However,

there are examples in Nirvi (1971: 105) where *joD* is used before other linking devices, such as interrogative pronouns in indirect questions. It is likely that in (91) and similar examples *joD* functions as a discourse particle but this question requires further investigations.

(91) Laanest 1966a: 113

no miä nüt sen verre-n laula-n [joD migä sîn laule-ttī
 well 1SG now that.GEN song-GEN sing-1SG that REL here sing-PST.IPRS
Gu tul-ti]
 when come-PST.IPRS
 ‘Well, I’ll now sing the song that was sung here when they came...’

An example where the relativizer appears at the end of the clause was presented in (92). In the data, there is another similar example. Both sentences represent a fused relative clause, and they come from contemporary sources. In fact, the sentences are identical, but they are produced by different speakers. Since this is the only occurrence of a clause-final relativizer, the possibility of a fixed expression cannot be ruled out.

(92) Kazahstaniiz_RP

[šöǝvvä [ei ol-d] midä]
 eat.INF NEG be.ACT.PTCP REL.PART
 ‘There was nothing to eat (lit. There was not what to eat)’

f) Two interesting internally headed constructions, (93) and (94), were attested in the contemporary materials.¹⁷ In (93), the head appears in the final position of the relative clause. The head, *miihed*, is mentioned in previous discourse. In (94), a part of the relative clause ‘*that we have*’ appears before the relativizer, and the head *šugu* is located in the position where the pronoun *meil* would be expected. The preceding context is built around the speaker describing her relatives who went fishing. In both examples, it is likely that the unexpected positionings of the relativizer or the head are motivated by pragmatic factors and the placement of emphasis in the sentence.

(93) Hirvi_GI

šiž peřrää [ken tuļl-ii-d meǵhe-d]
 then after REL come-PST-3PL man-PL
 ‘After that the men who came...’

¹⁷ Note that also sentence (39) with *milläin* is also internally headed.

(94) Püüdämääž_GI

tak što kogo [meil migä šugu ono]
 so that whole 1PL.ADE REL family be.3SG
 ‘So the whole family that we have [went fishing]’

g) There was only one example, (95), where an externally headed relative clause is prenominal. The head, *the church*, has been mentioned in the preceding discourse. The speaker asked the interviewer whether he had seen the church in question. To this, the interviewer answered in the affirmative, after which the speaker responded with the sentence (95). It is likely that the unusual position of the head is motivated by focus placement. The church appears to be in focus in this sentence, as suggested by the use of the demonstrative pronoun *še* ‘that’ and the focus particle *(k)ki*.

(95) Laanest 1966: 139

[kumBa jo laoDelDu kaiG] Ze kirkko-kki
 REL already destroyed all that church-PRCL
 ‘The church that has been completely destroyed’

h) As mentioned in section 5.4., inverse attraction was not attested in any of the examples extracted from the examined Ingrian texts. However, inverse attraction is attested in elicited sentences recorded from contemporary speakers, see (104). The head of the relative clause, the apple tree, is expected to take the nominative case, as it is the subject of the main clause. Instead, it has genitive marking. The relativizer represents the common argument in the relative clause in the object position, which is why it is marked with the genitive. The relativizer affects the preceding nominal head, and as a result of inverse attraction, the head takes on the case form of the relativizer.

(104) Elicited inverse attraction, 154AI(E)

ommeenapuu-n, [kumman miä iššud-i-n],
 apple_tree-GEN REL.GEN 1SG plant-PST-1SG
hää on korgia-mb koitti-a
 3SG be.3SG tall-COMP house-PART
 ‘The apple tree I planted is taller than the house’

This example comes from the Soikkola Ingrian Corpus of Elicitation (for details, see Rozhanskiy 2024: 190–191). The speakers whose recordings are represented in the Corpus of Elicitation are partially the same as the ones in the Spontaneous Speech Corpus. The speakers

were asked to translate a Russian stimulus into Ingrian in the elicitation process. It is noteworthy that inverse attraction is present in 11 out of 20 elicited translations of the same sentence produced by the speakers. Out of the 16 speakers, 8 produced similar sentences as (104).

As with the Ingrian Finnish sentence (35), inverse attraction is only found in elicited Ingrian data. Although it is possible for inverse attraction to appear in colloquial Russian, the stimulus used in the Corpus of Elicitation does not have it (see (105)). A deeper understanding of the phenomenon would require further research, where the local Russian dialects and other Finnic languages of Ingria would be compared.

(105) Standard Russian

<i>Яблоня,</i>	[которую я	посадил],	уже	выше
<i>Jablonja,</i>	<i>kotoruju ja</i>	<i>posadil,</i>	<i>uže</i>	<i>vyše</i>
apple_tree.NOM	REL.ACC	1SG	plant.PST.3SG	already higher
<i>моего</i>	<i>дома</i>			
<i>moego</i>	<i>doma</i>			
1SG.GEN	house.GEN			

‘The apple tree that I planted is already taller than my house.’

7. Conclusions

In this study, 248 Soikkola Ingrian relative clause constructions from texts recorded in the 19th, 20th and 21st centuries were considered. Based on their manual annotation and analysis, the main findings of this thesis can be stated as follows:

- 1) The relative pronoun strategy allows for the common argument to be in any syntactic function on the Noun Phrase Accessibility Hierarchy in the main clause and at least up to the possessor function in the relative clause.
- 2) Most commonly, Soikkola Ingrian relative clauses are headed by full nouns, but free relative clauses with pronominal, adjectival and adverbial heads also occur.
- 3) The relative pronoun *kumba* is preferably used with a full noun head. The referent can be either human, animate or inanimate.
- 4) The relative pronoun *migä* tends to be used with an inanimate abstract referent. Its use with pronominal heads and within fused relative clauses has increased in the 21st century compared to the previous century.
- 5) The relative pronouns *ken* and *kumba* are always used with a concrete referent.
- 6) For relativizing locative adverbials, the use of pronominal adverbs with *ku-* stem has increased in the 21st century compared to the 20th century.
- 7) The relative pronouns *migä* and *kumba* always agree with the head in number.
- 8) The relative pronoun *ken* does not always agree with the head in number. The nominative singular form is used instead of the nominative plural in 46 % of contemporary examples, and only in 12 % of cases from the 20th century. This indicates that the singular form *ken* has started to replace the plural form *ked* in Soikkola Ingrian.
- 9) Externally headed relative clauses with a full noun head are predominantly postnominal and usually come immediately after the head. In constructions with free or fused relative clauses, it is more common for the relative clause to appear before the head, but in most cases, the relative clause comes after the head.
- 10) The relativizer appears almost always clause-initially.
- 11) Internally headed relative clauses are rare, but possible.
- 12) Inverse attraction is not attested in spontaneous speech, but it can be found in elicited materials.

The study also shows that deviations are possible in terms of the position of relativizer and relative clause. Such deviations seem to be motivated by information-structural factors, especially focus placement within the sentence, and require further research.

Future research on Ingrian relative clause constructions should be expanded to other dialects of the language. Concerning Soikkola Ingrian, the results of this study could be compared to elicited materials, the written variety of Ingrian, and dictionary examples, which were left outside the scope of the present study. Additionally, the parameters used in this work could be applied for studying relative clause constructions in other Finnic languages, especially those spoken in Ingria, in order to find out whether they exhibit similar patterns.

References

- Andrews, A. D. (2007). Relative clauses. In T. Shopen (Ed.), *Language typology and syntactic description 2: Complex constructions*, 206–236. Cambridge: Cambridge University Press.
- Ariste, P. (1960). Isuri keelenäiteid. In *Soome-ugri keelte küsimusi* (Keele ja Kirjanduse Instituudi uurimused V), 7–68. Tallinn: Eesti Riiklik Kirjastus.
- Borise, L. & Kiss, K. É. (2023). The emergence of conjunctions and phrasal coordination in Khanty. *Journal of Historical Linguistics* 13(2). 173–219. <https://doi.org/10.1075/jhl.21016.kis>
- Bradley, J. & Luutonen, J. (2023). Mari. In D. Abondolo & R.-L. Valijärvi (eds.), *The Uralic Languages* (2nd ed.), 527–575. London, New York: Routledge.
- Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology* (2nd ed.). Chicago: University of Chicago Press.
- Comrie, B. (1998). Rethinking the typology of relative clauses. *Language Design: Journal of Theoretical and Experimental Linguistics* 1(1), 59–86.
- Comrie, B. & Keenan, E. (1979). Noun Phrase Accessibility Revisited. *Language* 55(3), 649–664. <https://doi.org/10.2307/413321>
- Comrie, B. & Kuteva, T. (2013a). Relativization Strategies. In M. S. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>. (Accessed 20.05.2026).
- Comrie, B. & Kuteva, T. (2013b). Relativization on Obliques. In M. S. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online* Zenodo. <https://doi.org/10.5281/zenodo.13950591>. (Accessed 20.05.2026).
- Dixon, R. M. W. (2010). *Basic linguistic theory 2: Grammatical topics*. Oxford: Oxford University Press.
- Eisenberg, P. & Wöllstein, A. (2016). *Duden - die Grammatik: Unentbehrlich für richtiges Deutsch* (9th ed.). Berlin: Dudenverlag.
- ELAN. (2025). [Computer software] (Version 7.0). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>. (Accessed 05.02.2026).
- Ernits, E. (2007). Isurite asualadest ja algkodust. In J. Rahman & H. Koks, *Õdagumeresoomõ kodo. Läänemeresoome kodu: Konvõrents Võrol, 19-21. Rehekuul 2006* (Võro Instituudi toimõndusõq 20), 13–32. Võro: Võro instituut.

- Flink, T. (2010). *Kotiin karkotettavaksi: Inkeriläisen siirtoväen palautukset Suomesta Neuvostoliittoon vuosina 1944-1955* (Historiallisia Tutkimuksia 251). Helsinki: Suomalaisen Kirjallisuuden Seura.
- Gary, J. & Keenan, E. L. (1977). On Collapsing Grammatical Relations in Universal Grammar. In P. Cole & J. Sadock (eds.), *Grammatical Relations*, 83–120. New York: Academic Press.
- Gast, V. & Diessel, H. (2012). The typology of clause linkage: Status quo, challenges, prospects. In V. Gast & H. Diessel (eds.), *Clause linkage in cross-linguistic perspective: Data-driven approaches to cross-clausal syntax* (Trends in Linguistics. Studies and Monographs 249), 1–36. Berlin, Boston: De Gruyter Mouton.
- Glossary of Linguistic Terms. (2015). SIL Global. <https://glossary.sil.org/term/relativizer> (Accessed 16.03.2026)
- Grünthal, R. (2015). *Vepsän kielioppi* (Apuneuvoja suomalais-ugrilaisten kielten opintoja varten XVII = Hilfsmittel für das Studium der finnisch-ugrischen Sprachen XVII). Helsinki: Suomalais-Ugrilainen Seura.
- Grünthal, R. (2021). Inkeröisen kieli. In K. Kallio, R. Grünthal & L. Saressalo (Eds.), *Inkerikot, setot ja vatjalaiset: Kansankulttuuri, kieli ja uskomusperinteet* (Suomalaisen Kirjallisuuden Seuran Toimituksia 1467), 122–152. Helsinki: Suomalaisen Kirjallisuuden Seura. <https://doi.org/10.21435/skst.1467>
- Grünthal, R. (2023). The Finnic Languages. In D. Abondolo & R.-L. Valijärvi (eds.), *The Uralic Languages* (2nd ed.), 433–480. London, New York: Routledge. <https://doi.org/10.4324/9781315625096>
- Häikiö, K. (2022). *Paluumatkan mutkat: Inkeriläisten paluumuuton alkaminen ja laajeneminen Neuvostoliiton hajoamisen aikaan* [PhD thesis]. University of Helsinki.
- Hamari, A. & Ajanki, R. (2022). Mordvin (Erzya and Moksha). In M. Bakró-Nagy, J. Laakso, E. Skribnik, M. Bakró-Nagy, J. Laakso & E. Skribnik (eds.), *The Oxford Guide to the Uralic Languages* (Oxford Guides to the World's Languages), 392–431. Oxford, New York: Oxford University Press.
- Itkonen, T. (1983). Välikatsaus suomen kielen juuriin. *Virittäjä* 87(2), 190–229.
- Junus, V. I. (1936). *Izoran keelen grammatikka: Morfologia opettajaa vart.* Leningrad—Moskova: Ucpedgiz.
- Kaškin, J. (Ed.). (2023). *Elementy gornomariiskogo jazyka v tipologičeskom osveščeenii: Monografija.* Moscow: Buki Vedi.

- Keenan, E. L. & Comrie, B. (1977). Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry* 8(1), 63–99.
- Kholodilova, M. (2013). Inverse Attraction in Ingrian Finnish Relative Clauses. *Linguistica Uralica* 49(2), 96–116. <https://doi.org/10.3176/lu.2013.2.02>
- Klumpp, G., Mazzitelli, L. F. & Rozhanskiy, F. (2018). Typology of Uralic languages: Current views and new perspectives. Introduction to the special issue of ESUKA – JEFUL. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 9(1), 9–30. <https://doi.org/10.12697/jeful.2018.9.1.01>
- Kokko, O. (2007). *Inkerinsuomen pirstaleisuus: Eräiden sijojen kehitys murteen yksilöllistymisen kuvastajana* (Joensuun yliopiston humanistisia julkaisuja 48) [PhD thesis]. Joensuun yliopisto.
- Kuznetsova, N. (2009). *Fonologičeskie sistemy ižorskich dialektov* [PhD thesis]. St. Petersburg: Institut lingvističeskich issledovanij RAN.
- Kuznetsova, N. (2015). Two phonological rarities in Ingrian dialects. In M. Hilpert, J.-O. Östman, C. Mertzluff, M. Rießler & J. Duke (eds.), *New Trends in Nordic and General Linguistics*, 91–117. Berlin: De Gruyter. <https://doi.org/10.1515/9783110346978.91>
- Kuznetsova, N., Markus, E. & Muslimov, M. (2015). Finnic Minorities of Ingria: The Current Sociolinguistic Situation and Its Background. In H. F. Marten, M. Rießler, J. Saarikivi, & R. Toivanen (eds.), *Cultural and Linguistic Minorities in the Russian Federation and the European Union* (Multilingual Education 13), 127–167. Berlin: Springer. <https://doi.org/10.1007/978-3-319-10455-3>
- Laakso, J. (2022). Livonian. In M. Bakró-Nagy, J. Laakso, E. Skribnik, M. Bakró-Nagy, J. Laakso & E. Skribnik (Eds.), *The Oxford Guide to the Uralic Languages* (Oxford Guides to the World's Languages), 380–391. Oxford, New York: Oxford University Press.
- Laanest, A. (1966a). *Isuri murdetekste*. Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Laanest, A. (1966b). *Ižorskie dialekty: Lingvogeografičeskoe issledovanie*. Tallinn: Akademiya nauk Estonskoj SSR.
- Laanest, A. (1978). *Istoričeskaja fonetika i morfologija ižorskogo jazyka*. Tallinn: Institut jazyka i literatury.
- Laanest, A. (1986). *Isuri keele ajalooline foneetika ja morfoloogia*. Tallinn: Valgus.

- Laury, R., Etelämäki, M. & Couper-Kuhlen, E. (Eds.). (2017). *Linking Clauses and Actions in Social Interaction*. Helsinki: SKS Finnish Literature Society.
<https://doi.org/10.21435/sflin.20>
- Laury, R. & Suzuki, R. (Eds.). (2011). *Subordination in conversation: A cross-linguistic perspective* (Studies in Language and Social Interaction 24). Amsterdam: John Benjamins.
- Lipták, A. (2004). On the correlative nature of Hungarian left-peripheral relatives. *ZAS Papers in Linguistics* 35(2), 287–313. <https://doi.org/10.21248/zaspil.35.2004.231>
- Longacre, R. E. (2008). Sentences as combinations of clauses. In T. Shopen (Ed.), *Language typology and syntactic description 2: Complex constructions*, 372–420. Cambridge: Cambridge University Press.
- Maamies, S. (2011). Joka ja mikä: Relatiivipronominin käytöstä. *Kielikello* 4/2011.
<https://kielikello.fi/joka-ja-mika-relatiivipronominin-kaytosta/>. (Accessed 28.03.2026).
- Mägiste, J. (1925). *Rosona (Eesti Ingeri) murde pääjooned* (Acta et Commentationes Universitatis Tartuensis B VII 3). Tartu Ülikool.
- Manninen, S. (2003). To Raise or Not to Raise: The Case of Finnish Restrictive Relative Clauses. In A. Dahl & P. Svenonius (Eds.), *Proceedings of the 19th Scandinavian Conference of Linguistics* 31(4), 668–693. <https://doi.org/10.7557/12.32>
- Mantila, H. (2025). *Suomen nykymurteet*. Helsinki: Suomalaisen Kirjallisuuden Seura.
<https://doi.org/10.21435/tl.301>
- Markus, E. (2024). Syntax and functions of the Ingrian discourse particles *no* and *nu*. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics*, 15(1), 155–186. <https://doi.org/10.12697/jeful.2024.15.1.05>
- Markus, E. & Rozhanskiy, F. (2012). Votic or Ingrian: New evidence on the Kukkuzi variety. *Finnisch-Ugrische Mitteilungen* 35, 77–95.
- Markus, E. & Rozhanskiy, F. (2022a). Ingrian. In M. Bakró-Nagy, J. Laakso & E. Skribnik (eds.), *The Oxford Guide to the Uralic Languages* (Oxford Guides to the World's Languages), 308–329. Oxford, New York: Oxford University Press.
<https://doi.org/10.1093/oso/9780198767664.003.0053>
- Markus, E. & Rozhanskiy, F. (2022b). Votic. In M. Bakró-Nagy, J. Laakso & E. Skribnik (eds.), *The Oxford Guide to the Uralic Languages* (Oxford Guides to the World's Languages), 330–349. Oxford, New York: Oxford University Press.
<https://doi.org/10.1093/oso/9780198767664.003.0019>

- Markus, E. & Rozhanskiy, F. (2023). Functions of demonstrative pronouns in contemporary Soikkola Ingrian. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 14(2), 41–74. <https://doi.org/10.12697/jeful.2023.14.2.03>
- McGregor, W. B. (1990). *A Functional Grammar of Gooniyandi* (Studies in Language Companion Series 22). Amsterdam: John Benjamins.
- Metslang, H., Erelt, M., Habicht, K., Hennoste, T., Kasik, R., Teras, P., Viht, A., Asu, E. L., Lindström, L., Lippus, P., Pajusalu, R., Plado, H., Rääbis, A. & Veismann, A. (2024). *Eesti grammatika* (H. Metslang, Ed.). Tartu: Tartu Ülikooli Kirjastus.
- Nikolaeva, I. (2014). *A Grammar of Tundra Nenets* (Mouton Grammar Library 65). Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110320640>
- Nirvi, R. E. (1971). *Inkeröismurteiden sanakirja* (Lexica Societatis Fenno-Ugricae XVIII). Helsinki: Suomalais-Ugrilainen Seura.
- Nirvi, R. E., Laanest, A. & Virtaranta, P. (1967). Inkeröismurteet. In P. Virtaranta (Ed.), *Lähisukukielten lukemisto* (Suomalaisen Kirjallisuuden Seuran Toimituksia 280), 137–169. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Norvik, M. (2016). Research into Livonian syntax: The results of previous studies and the tasks ahead. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 7(1). <https://doi.org/10.12697/jeful.2016.7.1.08>
- Norvik, M., Yingqi Jing, Dunn, M., Forkel, R., Honkola, T., Klumpp, G., Kowalik, R., Metslang, H., Pajusalu, K., Piha, M., Saar, E., Saarinen, S. & Vesakoski, O. (2022). *Uralic Typological database—UraTyp* (Version v1.1) [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.5236365>. (Accessed 15.02.2026).
- Ovsjannikova, M. & Rozhanskiy, F. (2024). Syntax of the Uralic languages: Principal features and challenges. Introduction to the special issue of ESUKA – JEFUL. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 15(1), 5–53. <https://doi.org/10.12697/jeful.2024.15.1.01>
- Pajusalu, K. (2022). Seto South Estonian. In M. Bakró-Nagy, J. Laakso, E. Skribnik, M. Bakró-Nagy, J. Laakso & E. Skribnik (eds.), *The Oxford Guide to the Uralic Languages* (Oxford Guides to the World's Languages), 367–379. Oxford, New York: Oxford University Press.
- Patrakka, M. (2021). Soikkolan inkerikkojen 1900-luku. In K. Kallio, R. Grünthal & L. Saressalo (eds.), *Inkerikot, setot ja vatjalaiset: Kansankulttuuri, kieli ja uskomusperinteet* (Suomalaisen Kirjallisuuden Seuran Toimituksia 1467), 92–120. Helsinki: Suomalaisen Kirjallisuuden Seura. <https://doi.org/10.21435/skst.1467>

- Porkka, V. (1885). *Über den Ingrischen Dialekt mit Berücksichtigung der übrigen finnisch-germanländischen Dialekte* [PhD thesis]. Helsingfors: J. C. Frenckell & Sohn.
- Privizentseva, M. (2025). *Relative clauses in Moksha: Raising by reprojection* [Manuscript submitted for publication].
- Pyöli, R. (2021). *Livvinkarjalan kieliooppi* (2nd ed.). Helsinki: Karjalan Sivistysseura.
- Rantanen, T., Tolvanen, H., Roose, M., Ylikoski, J. & Vesakoski, O. (2022). Best practices for spatial language data harmonization, sharing and map creation—A case study of Uralic. *PLOS One* 17(6), e0269648. <https://doi.org/10.1371/journal.pone.0269648>
- Rantanen, T., Vesakoski, O., Ylikoski, J. & Tolvanen, H. (2021). *Geographical database of the Uralic languages* [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.4784188>. (Accessed 08.03.2026).
- Raudalainen, T. (2021). Näkymiä Inkerinmaan historiaan. In K. Kallio, R. Grünthal & L. Saressalo (eds.), *Inkerikot, setot ja vatjalaiset: Kansankulttuuri, kieli ja uskomusperinteet* (Suomalaisen Kirjallisuuden Seuran Toimituksia 1467), 36–58. Helsinki: Suomalaisen Kirjallisuuden Seura. <https://doi.org/10.21435/skst.1467>
- Rožanskij, F. I. & Markus, E. B. (2012). “Zolotaja ptica” (Publikacija izorskoj skazki, zapisannoj v XIX veke). In N. N. Kazanskij (Ed.), *Acta Linguistica Petropolitana. Transactions of the Institute for Linguistic Studies* 8(1), 448–503. St. Petersburg: Nauka.
- Rožanskij, F. I. & Markus, E. B. (2013). Ižora Sojkinskogo poluostrova: Fragmenty sociolingvističeskogo analiza. In N. N. Kazanskij (Ed.), *Acta Linguistica Petropolitana. Transactions of the Institute for Linguistic Studies* 9(3), 261–298. St. Petersburg: Nauka.
- Rozhanskiy, F. (2024). Syntax of the numeral phrase in Soikkola Ingrian. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 15(1), 187–220. <https://doi.org/10.12697/jeful.2024.15.1.06>
- Rozhanskiy, F. & Markus, E. (2019). A new resource for Finnic languages. *Uralica Helsingiensia* 14, 303–326. <https://doi.org/10.33341/uh.85039>
- Rozhanskiy, F. & Markus, E. (2024). Syncretism in Soikkola Ingrian verbal and nominal paradigms. In S. Björklöf, S. Jantunen, S. Junntila, J. Kuokkala, S. Laitinen & A. Pasanen (eds.), *Itämeren kieliapajilta Volgan verkoille: Pühendusteos Riho Grünthalile 22. mail 2024* (Suomalais-Ugri-laisen Seuran Toimituksia 278). Helsinki: Suomalais-Ugri-lainen Seura. <https://doi.org/10.33341/sus.965.1351>
- Rozhanskiy, F. & Markus, E. (2025a). Language Death or Murder? The Fate of Ingrians and their Language since 1944. *Studia Orientalia* 126, 449–471.

- Rozhanskiy, F. & Markus, E. (2025b). Negation in Soikkola Ingrian. In C. Hasselblatt & B. Wagner-Nagy (eds.), *Finnisch-Ugrische Mitteilungen* 39, 189–219. Hamburg: Buske. <https://doi.org/10.46771/978-3-96769-533-5>
- Saar, E. (2017). *Isuri keele Soikkola murde sõnamuutmissüsteem* (Dissertationes Philologiae Uralicae Universitatis Tartuensis 19) Tartu: Tartu Ülikooli Kirjastus. <http://hdl.handle.net/10062/58016>
- Schwarz, A. & Rozhanskiy, F. (2024). Pronominal System of Soikkola Ingrian: Interrogative/Relative, Indefinite, and Negative Pronouns and Their Variation. *Linguistica Uralica* 60(4), 271–295. <https://doi.org/10.3176/lu.2024.4.02>
- Shagal, K. (2023). Relative Clauses in Uralic. In D. Abondolo & R.-L. Valijärvi, *The Uralic Languages* (2nd ed.), 939–978. London, New York: Routledge. <https://doi.org/10.4324/9781315625096-21>
- Skribnik, E. (2022). Clause combining. In M. Bakró-Nagy, J. Laakso & E. Skribnik (eds.), *The Oxford Guide to the Uralic Languages* (Oxford Guides to the World's Languages), 996–1017. Oxford, New York: Oxford University Press. <https://doi.org/10.1093/oso/9780198767664.003.0053>
- Söder, T. (2023). South Saami. In D. Abondolo & R.-L. Valijärvi (eds.), *The Uralic Languages* (2nd ed.), 235–301. London, New York: Routledge. <https://doi.org/10.4324/9781315625096>
- Sovijärvi, A. (1944). *Foneettis-äännehistoriallinen tutkimus Soikkolan inkeröismurteesta*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Veismann, A., Muischnek, K. & Padrik, J. (2026). Sissejuhatus lingvistilisse uurimistöösse ja korpuslingvistikasse. In L. Lindström, M.-L. Pilvik, A. Veismann, K. Muischnek & J. Padrik (eds.), *Korpused lingvistilises uurimistöös. Praktiline käsiraamat*, 13–30. Tartu: Tartu Ülikooli Kirjastus. <https://hdl.handle.net/10062/119520>
- Vilkuna, M., Hakulinen, A., Korhonen, R., Koivisto, V., Heinonen, T.-R. & Alho, I. (2008). *Ison suomen kieliopin verkkoversio* (Kotimaisten kielten tutkimuskeskuksen verkkojulkaisuja 5). Helsinki: Suomalaisen Kirjallisuuden Seura. (Accessed 09.06.2025).
- Virtaranta, P. & Suhonen, S. (eds.). (1978). *Lähisukukielet. Finnic languages: Karelian, Ludic, Vepsian, Ingrian, Votic, Livonian*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Whaley, L. (2011). Syntactic typology. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology* (Oxford handbooks in linguistics), 465–486. Oxford, New York: Oxford University Press.

Wilbur, J. (2014). *A grammar of Pite Saami* (Studies in Diversity Linguistics 5). Berlin: Language Science Press. https://doi.org/10.26530/oopen_533870

Relatiivlausekonstruktsioonid isuri keele Soikkola murdes. Kokkuvõte

Käesolev lõputöö uurib relatiivlausekonstruktsioonide süntaksit isuri keele Soikkola murdes. Töös kasutatud andmed pärinevad mitmeist erinevaist tekstikogudest, millest vanimad on 19. sajandi lõpust ja uusimad 21. sajandi algusest. Allikaist eraldati käsitsi kokku 248 relatiivlausekonstruktsiooni, mis märgendati erinevate parameetrite suhtes. Töö eesmärgiks oli välja uurida, milliseid lauseliikmeid isuri keele Soikkola murdes on võimalik relativeerida, kuidas põhisõna sõnaliik ja referendi elusus ning konkreetsus mõjutavad relatiivpronoomeni valikut, kas relatiivpronoomenid ühilduvad põhisõnaga arvus ning milline relatiivlausekonstruktsioonide sisemine süntaks välja näeb.

Uurimuse kõige olulisemad tulemused on järgnevad:

- 1) Personaalpronoomeni strateegia lubab relativeerida kõiki relativeerumise hõlpsushierarhia lauseliikmeid pealauses ja vähemalt kõiki kuni omajalaiendini relatiivlauses.
- 2) Relatiivlause põhjaks on tavaliselt nimisõna aga esineb ka relatiivlauseid pronominaalsete, adjektiivsete ja adverbiaalsete põhjadega.
- 3) Relatiivpronoomenit *kumba* kasutatakse harilikult viidates nimisõnale. Referent saab olla inimene, loom või elutu.
- 4) Relatiivpronoomenit *migä* kasutatakse enamasti elutu abstraktse referendiga. Võrreldes eelmise sajandiga on selle kasutus pronominaalsete põhjade ja põhjata relatiivlausetega sagenenud.
- 5) Relatiivpronoomeneid *ken* ja *kumba* kasutatakse alati konkreetse referendiga.
- 6) Kohamääruste relativeerimiseks kasutatakse pronominaalseid *ku-* tüvega määrsõnu sagedamini 21. sajandil võrreldes 20. sajandiga.
- 7) Relatiivpronoomenid *migä* ja *kumba* ühilduvad alati oma peasõnaga arvus.
- 8) Relatiivpronoomen *ken* ei ühildu alati oma põhjaga arvus. Tänapäevastes näidetes kasutatakse ainsuse nimetavat vormi mitmuse nimetava vormi asemel 46 %-l juhtudest. 20. sajandi näidetes esineb seda vaid 12 %-l juhtudest. Sellest saab järeldada, et ainsuslik vorm *ken* on hakanud mitmuslikku vormi *ked* asendama isuri keele Soikkola murdes.
- 9) Konstruktsioonides, kus põhisõna asukoht on väljaspool relatiivlauset, paikneb relatiivlause enamasti vahetult oma põhja järel. Need relatiivlause, kus põhjaks on asesõna, omadussõna või määrsõna ning põhjata relatiivlause, võivad paikneda põhjast eespool, aga enamikul juhtudel asub relatiivlause siiski põhisõna järel.

- 10) Relativeerija esineb peaaegu alati lause alguses.
- 11) Konstruktsioonid, kus põhisõna asukoht on relatiivlause sees, on võimalikud, aga haruldased.
- 12) Põhisõna ei ühildu käändes relatiivpronoomeniga spontaanses kõnes, aga küsitletud materjali hulgast leidub näiteid, kus nii juhtub.

Uurimus näitab samuti, et hálbed relativeerija ja relatiivlause paiknemises on võimalikud. Säärased hálbed näivad olevat motiveeritud teatestruktuurilistest teguritest, eriti fookuse paigutusest lauses, aga vajavad edasist uurimist.

Tulevases isuri keele relatiivlausekonstruktsioonide uurimistöös tuleb käsitletust laiendada ka keele teistele murretele. Mis Soikkola murret puutub, siis on selle uurimuse tulemusi võimalik võrrelda küsitletud materjalide, isuri kirjakeele ning sõnastikunäidetega, mis jäid käesoleva töö käsitusalt välja. Samu parameetreid, mida siin kasutati, võib rakendada teiste läänemeresoome keelte relatiivlausekonstruktsioonide uurimisel, eriti Ingerimaal kõneldavate keelte puhul. Nii saaks välja uurida, kas lähisugukeeltes esineb sarnasusi isuri keelega.

Non-exclusive Licence for Reproduction and Public Availability of the Thesis

I, Riku Tapani Kuusijärvi,

1) grant the University of Tartu a free, non-exclusive licence to reproduce my thesis "Relative clause constructions in Soikkola Ingrian", supervised by Elena Markus and Fedor Rozhanskiy, for preservation, including adding it to the University of Tartu digital archive for the duration of copyright;

2) grant the University of Tartu permission to make the work mentioned in point 1 publicly available through the University of Tartu web environment, including the digital archive, under the Creative Commons license CC BY NC ND 4.0, which allows reproducing, distributing, and making the work available to the public with attribution to the author, and prohibits creating derivative works and using the work for commercial purposes, until the expiration of copyright;

3) am aware that the rights mentioned in points 1 and 2 remain with the author;

4) confirm that by granting this simple license, I do not violate the intellectual property rights of other persons or rights arising from personal data protection legislation.

Riku Tapani Kuusijärvi

21.05.2026