

SYNTAX OF THE NUMERAL PHRASE IN SOIKKOLA INGRIAN

Fedor Rozhanskiy

University of Tartu, EE

handarey@yahoo.com

Abstract. This article describes the numeral phrase in Soikkola Ingrian. It focuses on the external syntax of the numeral phrase (first of all, case marking of its components) and on the number agreement between the numeral phrase in the subject position and the predicate. The sources of data are (a) a collection of spontaneous speech samples recorded in 2006–2013, (b) samples of spontaneous speech published by previous researchers, (c) elicited material recorded in 2006–2023. Though the numeral phrase in Soikkola Ingrian preserves most of the common Finnic traits, it has some less common features, e.g. agreement in all numeral phrases with a numeral ending in ‘one’ or expressing approximate quantity through a reversed word order. Most likely these features arose due to the contact influence of Russian.

Keywords: Ingrian, syntax, numeral phrase, agreement, case marking, language contact

DOI: <https://doi.org/10.12697/jeful.2024.15.1.06>

1. Introduction

Ingrian is a minor Finnic language closely related to Karelian and Finnish.¹ It is a moribund language – according to my estimation there are no more than 10 fluent speakers left. Most of them live in the Kingisepp district of the Leningrad province of Russia, not far from the Estonian border. For many decades, Russian has been their main language of communication (see Rožanskij & Markus 2013; Kuznetsova, Markus & Muslimov 2015 for more detailed information on language shift and other sociolinguistic issues).

Ingrian syntax has never been described. There are monographs and dissertations focused on Ingrian phonetics/phonology (Sovijärvi 1944;

¹ More information on Ingrian can be found in Markus & Rozhanskiy (2022).

Kuznecova 2009) and morphology/morphophonology (Saar 2017) but not on syntax. None of the existing descriptions of Ingrian – dialectal (Porkka 1885), historical and dialectal (Laanest 1978, 1986), or normative (Junus 1936) – has a chapter dedicated to syntax.²

The current article examines the Ingrian numeral phrase (NumP). Its main goal is to describe the syntax of the NumP in contemporary³ Soikkola Ingrian.⁴

In general, a description of a numeral phrase in any language should address three levels (see Table 1). On the first level, the internal structure of numeral forms is analysed. For simple numerals containing one numeral root (e.g. ‘three’ or ‘five’),⁵ it is primarily a question of their inflection (if they can be inflected in the studied language). Thus, it is more a morphological issue than a syntactic one. Such analysis has been done for Ingrian to some extent: declension of numerals is described in Porkka (1885) and Saar (2017), and the etymology of Uralic numerals is presented in Blažek (1999). Complex numerals, which consist of several numeral roots (e.g. *fifteen* or *five hundred*), require an analysis of their constituents and relations between them (see, for example, de la Villa 2010 on “atoms” and “bases”). Some complex numerals are compound words and demonstrate features of both a single word and a word combination, cf. Ingrian *üks-toižeešt* ~ *ühešt-toižeešt* ‘eleven.ELA’ (from *üks-toišt* ‘eleven’), where the whole numeral or both its structural parts can decline. Thus, in the case of complex numerals the researcher deals with both morphology and “internal syntax”; Greenberg (1989: 105) defines the latter as “syntactic relations within a complex numerical expression such as *two hundred* in English”. In the present article, I skip the analysis of this level, because it has already been largely described

2 There are several pages describing Ingrian syntax in the short description by Markus & Rozhanskiy (2022: 324–328). Fragmentary information on Ingrian syntactic features can be also found in other publications, e.g. Rožanskij & Markus 2014; Rožanskij 2017; Rozhanskiy & Markus 2017.

3 By “contemporary Ingrian” I mean the language of Ingrian native speakers in the 21st century.

4 See Laanest (1966b, 1998) on Ingrian dialects.

5 See, for example, Žoha, Wągiel & Caha (2022: 200–201) for definitions of simple and complex numerals.

for Ingrian, and the morphological issues are beyond the scope of this study.⁶

The second level concerns the structure of the NumP. It requires the analysis of relations between the numeral and the quantified noun and other possible elements of a NumP (e.g. in *two big houses* the relation with the adjective should also be specified). In the case of Ingrian, which is a morphologically rich language, these relations are expressed primarily by the distribution of case and other nominal categories between the elements of the NumP. This part of the description is “external syntax” of numerals in terms of Greenberg (1989).

The third level deals with the syntactic relations outside the NumP. It concerns mostly the agreement of the NumP with the predicate (both its verbal and nominal parts).

Table 1. Levels of a syntactic description of numerals.

N	Level	Sphere
1	Internal structure of numerals	Morphology for simple numerals Syntax for complex numerals (internal syntax)
2	Structure of NumP – relations between the numeral and the quantified noun	Syntax (external syntax)
3	Relations between NumP and other constituents of a sentence	Syntax

The current study is focused on the second and third levels. It means that it (a) describes case and number marking of constituents in the numeral phrase and (b) analyses how a numeral phrase is integrated in the clause – first of all, from the point of view of agreement. Since this is the first study dedicated to the Ingrian NumP, it addresses the most important features of the NumP and does not purport to contain a thorough analysis of all details. It is a descriptive rather than a comparative study; nevertheless, in passing, I make some comments concerning the general Finnic context. The article analyses exclusively the cardinal

6 The principles of typological analysis of numerals on this level are formulated in Comrie (2021).

numerals. An analysis of the ordinal numerals is beyond the scope of this paper. First, the syntactic behaviour of ordinal numerals is very close to that of adjectives. Second, a representative corpus containing many occurrences of various ordinal numerals is required for a detailed analysis. There is no such data at my disposal – in the corpus (see section 2), ordinal numerals are relatively rare if compared with cardinal ones.

This article has the following structure: section 2 describes the data; section 3 analyses case marking in the numeral phrase, section 4 studies the agreement of the numeral phrase with the predicate; section 5 presents the conclusions.

2. Data

This article aims primarily to provide a synchronic description of the Soikkola Ingrian NumP and therefore requires a data set that is homogeneous in terms of time period. For this reason, the main source of data is the corpus of recordings collected by me and my colleagues during fieldwork on Ingrian in 2006–2023. This corpus includes a collection of spontaneous speech samples (2006–2013) and a corpus of elicited materials (2006–2023). The collection of spontaneous speech samples includes 4 hours of recordings (approximately 20,000 words) by 24 Soikkola Ingrian speakers. Below I will refer to this collection as the “Spontaneous Speech Corpus”. The corpus of elicited materials includes about 700 hours of recordings. I will refer to it as the “Corpus of Elicitation”. I primarily use the data from the spontaneous speech corpus because the elicited material is prone to calque some constructions from the intermediary language (here, Russian). However, in cases where such influence of the intermediary language is not likely, the elicited data are also used in the article, because some constructions are rare and not attested in the relatively small corpus of spontaneous speech.

For every example from the Spontaneous Speech Corpus, the name of the text is indicated. Two final capital letters in this name are the index of the native speaker, e.g. Püüdämääž_talveel_OM. For examples from the Corpus of Elicitation, the alphanumeric index of the recording is indicated (the first three characters stand for the index of the file, and the two final letters denote the native speaker). The index of recording

is followed by the letter “E” in parentheses indicating that this example comes from the Corpus of Elicitation, e.g. D12AL(E).

It should be noted that, even though for contemporary Ingrian speakers Russian has been the main language of communication for many decades, all speakers represented in the abovementioned corpora are fluent in Ingrian, spoke Ingrian from their childhood and many of them did not know Russian before starting school. The next generation (born in 1940–1950s) spoke only Russian and among them fluent Ingrian speakers were an exception. This means that Ingrian did not undergo a long period of decline when one might typically observe attrition processes, and which usually significantly affects the syntax of the language.

Ingrian speech samples published by previous researchers are rather short, and it is thus impossible to carry out a thorough comparative study which might reveal recent changes in the syntax of the NumP. However, I decided not to ignore this material and so I use examples from earlier publications as additional illustrations of the analysed phenomena. I suppose that information about constructions attested in earlier data can be valuable.

The following published Ingrian texts were analysed while working on the current paper:

- two tales *Der goldene Vogel* and *Die goldene Feder* published in Porkka (1885); they are recorded by two speakers from the villages of Tarinai and Taatsoi, respectively;
- a collection of spontaneous speech samples and some pieces of folklore in Ariste (1960); mainly recorded by one speaker who was born in Saarove and lived in Tarinai. Additionally, this publication contains two songs and a few sentences from 3 more speakers; altogether approximately 7,500 words;
- a collection of spontaneous speech samples recorded by one native speaker from the village of Voloitsa (Laanest 1966a); approximately 5,300 words.

The original transcription of the Ingrian data from earlier sources is preserved in this article.⁷ The main differences in transcription systems used by different scholars are summarized in Table 2.

⁷ However, symbols denoting pauses, stress and liaison are omitted.

Table 2. Main differences in transcription of Ingrian data.

	Porkka 1885	Ariste 1960	Laanest 1966a	Current article
Long vowels (e.g. <i>õ</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>tt̥</i>)	<i>t ~ tt</i>	<i>t</i>	<i>tt̥ ~ t</i>	<i>tt̥</i>
Short geminates of sonorants (e.g. <i>nn̥</i>)	<i>nn</i>	<i>nn̥</i>	<i>nn̥</i>	<i>nn̥</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/ž</i>	<i>s/z</i>	<i>š/ž</i>

As the current study is based primarily on corpus data and published speech samples, I can report only the presence or absence of certain constructions; a detailed analysis of variation and limits of grammaticality is not possible.

Examples from Standard Finnish and Standard Estonian, which I give in the article for comparison with Ingrian, are recorded by native speakers of these languages.⁸ The standard orthography is used in these examples.

3. Case marking in the numeral phrase

This section consists of three parts. First, I will discuss the general principles of case marking in the Ingrian NumP, which are attested in other Finnic languages (for example, in Finnish and Estonian) as well (section 3.1). After that some features observed in Ingrian but not in most Finnic languages will be discussed (section 3.2). These sections focus on the NumP serving as a verbal argument. The NumP in an adverbial function is not examined in detail in this paper but section 3.3 briefly addresses this issue.

⁸ Contemporary Ingrian does not have a written variety and all materials used in this article come from the spoken language. In this situation, it would be reasonable to use data from other languages which represent the spoken rather than the written variety. For example, Karlsson (1959: 358) mentions that in Finnish the difference between spoken and written language can be significant for the agreement of NumP with the predicate.

3.1. General principles

The general principles of case marking in the NumP follow a common pattern of the Finnic languages in the sense that there are two types of NumP.⁹ In the first type, the numeral has a nominative singular form and governs the partitive singular form of the quantified noun.¹⁰ As the cases of both the numeral and noun are fixed, I call this NumP “strong”. In the second type, the numeral has the same marking as the noun, and the case depends on the syntactic position of the NumP in the clause. As there is no fixed case form for both constituents of the NumP, I call it “weak”.¹¹ Example (1) contains a strong NumP – the nominative form of the numeral *kakš* ‘two’ and the partitive form *kañnaa* from *kana* ‘hen’. The two examples which follow it contain weak NumPs: in (2), both numeral and noun are marked with the genitive and in (3), with the partitive.

(1) D12AL(E)

<i>kakš</i>	<i>kañna-a</i>	<i>lovajaa-d</i>	<i>tanvaas-t</i>	<i>mõõd</i>
two(NOM) ¹²	hen.PART	walk.PRS-3PL	yard.PART	along

‘Two hens walk around the yard.’

(2) Metsääž_EN

<i>metsää</i>	<i>eñtäälle</i>	<i>ve-i</i>	<i>taiñtaa</i>	
forest.ILL	far	lead-PST.3SG	perhaps	
<i>üli</i>	<i>kolme-n</i>	<i>vai</i>	<i>neljä-n</i>	<i>oja-n</i>
across	three-GEN	or	four-GEN	stream-GEN

‘(She) took me far into the forest, perhaps, across three or four streams.’

9 See Hurford (2003: 579) and Hakulinen et al. (2004: §772, §789) about Finnish, and Metslang (2019: 463) about Estonian. Note that comparison of syntactic structures in the Finnic languages can be complicated due to the different terminological traditions, primarily because of a different understanding of the term “accusative” (see Ovsjannikova & Rozhanskiy 2024: 18).

10 It should be noted that typologically a numeral that governs the case of a quantified noun is a relatively rare phenomenon. For example, it exists in Slavic languages (Corbett 1978) and in Old English (von Mengden 2010: 210–227). It is not clear if it developed in the Finnic languages independently or was borrowed from Indo-European.

11 “Strong” and “weak” are technical terms, which I use for the ease of reference. Cf. notions of heterogeneous vs homogeneous case distribution (see, for example, Babby 1985: 2).

12 Usually, I do not gloss the nominative forms. However, I provide the gloss NOM for numerals to show the structure of the NumP more explicitly.

- (3) 334IV(E)
- | | | | | |
|------------|------------|------------|--------------|---------------|
| <i>miä</i> | <i>e-n</i> | <i>nää</i> | <i>kah-t</i> | <i>puu-da</i> |
| 1SG | NEG-1SG | see.CNG | two-PART | tree-PART |
- ‘I do not see two trees.’

In fact, the syntactic interpretation of (3) is ambiguous concerning the grounds for the case of the noun (the case of the numeral follows from the syntactic context: the object acquires the partitive case in negative sentences). Though I interpret the partitive form of the noun as a result of agreement with the numeral, theoretically, it can be considered as dependent on the numeral (as in the case of a strong NumP). However, there are no obvious advantages to such an interpretation. First, it increases the number of possible constructions (besides Num_{NOM} + N_{PART} one should introduce Num_{PART} + N_{PART}) and therefore makes the description more complicated. Second, it is unclear why there are no further constructions where the numeral governs the noun. In other words, if a numeral has a dependent partitive noun, why is it that this numeral can have the nominative or partitive but not some other marking, such as the genitive, which is also a syntactic case (see 9).¹³

The distribution of strong vs weak NumPs depends on the syntactic position of the NumP in a sentence. Sections 3.1.1 and 3.1.2 contain examples of the most typical contexts for strong and weak NumPs, respectively.

3.1.1. Strong numeral phrase

The two most typical syntactic contexts for a strong NumP are the positions of the subject (1, 4, 5) and of the total object (6, 7).¹⁴

13 This problem is discussed for the Finnish data in Dolbey (1998: 20). The author concludes that there is no agreement in constructions where both the numeral and the noun have the partitive marking and argues that the difference between the two types of numeral constructions correlates with grammatical vs non-grammatical case of both constituents. However, this approach is only possible if the accusative case is understood in terms of the Finnish linguistic tradition (see Ovsjannikova & Rozhanskiy 2024: 18), when all occurrences of the morphological genitive are distributed between “grammatical accusative” (like *haugin* ‘pike.GEN’ in (9)) and “non-grammatical genitive” (like *ojan* ‘stream.GEN’ in (2)).

14 “Total object” refers to an object marked with a total case (i.e. nominative, genitive or accusative), as opposed to the partial object marked with the partitive. The distribution of objects depends on a number of grammatical and semantic/pragmatic factors. See details in Ovsjannikova & Rozhanskiy (2024: 12–18).

In a strong NumP, the numeral is the head, and the quantified noun is a dependent.¹⁵ This is obvious when the NumP is a subject (1, 4, 5) because the case of the numeral is defined by the syntactic position of the NumP (subject usually has the nominative form) and the case of the noun (partitive) is defined by the numeral. The analysis of the NumP in the position of a total object (6, 7) is more complicated. I consider these examples as representing a strong NumP where the numeral in the nominative case is the head. However, one might suggest that the NumP in examples (6, 7) takes the position of the partial object, and the head of the NumP is the noun in the partitive rather than the numeral in the nominative. From my point of view, this interpretation is implausible. First, the partial object is used in specific semantic conditions (Ovsjannikova & Rozhanskiy 2024: 12–16) and is not expected in (6) and (7). Second, the nominative singular form of the numeral looks very logical in the position of the total object. Semantically, the NumP expresses plurality of the object and is therefore comparable with a noun in the plural. A total object expressed with a plural noun is marked with the nominative plural, but in the case of a NumP there is no need for grammatical marking of the plural and the nominative singular is used. Note that the head defining criteria may yield controversial results in the case of NumPs, and a more profound analysis is required, cf. with the discussion of headedness in the Russian numeral expressions in Corbett (1993) and the analysis of Estonian numeral phrase in Rutkowski (2001).

- (4) Püüdämaaž_talveel_OM
šiž kakštoišť henkki-ü brigadaa-ž ol-i
 then twelve(NOM) soul-PART brigade-INE be-PST.3SG
 ‘At that time there were twelve people in the brigade.’
- (5) A61AG(E)
viiž ihmiiš-t jä-i-väd metsää
 five(NOM) person-PART remain-PST-3PL forest.ILL
 ‘Five people remained in the forest.’

15 The same is claimed for a similar construction in Finnish (see, for example, Hakulinen et al. 2004: §789 and Karttunen 2006: 411). While analyzing the similarity of Polish and Estonian numeral phrases, Rutkowski (2001: 181) notes that “the noun which is the semantic nucleus of the phrase becomes syntactically subordinated to the numeral” (for so-called structural cases).

- (6) Püüdämääž_talveel_OM
ühe-n kerra-n mõõ ša-i-mma kakš tonni-a
 one-GEN time-GEN 1PL get-PST-1PL two(NOM) ton-PART
kalla-a
 fish-PART
 ‘Once we caught two tons of fish.’
- (7) Porraž_EN
miä od-i-n kakš mogoma-a suur-d nigu
 1SG take-PST-1SG two(NOM) such-PART big-PART kind.of
riučku-a
 pole-PART
 ‘I took two such big kind of poles.’

In (7), three forms – the pronoun (*mogomaa*), adjective (*suurd*) and noun (*riučkua*) – follow the numeral and are marked with the partitive singular, i.e. the pronoun and the adjective agree with the noun in case and number. Compare with (8) where the pronoun *nämäd* precedes the numeral and, therefore, is not within the scope of the NumP and preserves the total case.¹⁶

- (8) A69ST(E), B82EI(E)
hää ošt-i nämäd kakš kotti-a
 3SG buy-PST.3SG this.PL.NOM kakš(NOM) house-PART
 ‘He bought these two houses.’

As already mentioned, the nominative singular marking of numerals in the position of the total object distinguishes them from nouns. In this position, the latter are marked with the genitive singular if the object is singular (9) or with the nominative plural if the object is plural (10).

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

¹⁶ Agreement depends on the position of adjectives and demonstratives with respect to the numeral, see Kaiser (2022a: 423–424) for an analysis of the Finnish data.

(10) Maa-muna(B)_ZD

<i>miä</i>	<i>ša-i-n</i>	<i>muna-d</i>	<i>iššutta-a</i>
1SG	get-PST-1SG	potato-PL.NOM	plant-INF

‘I managed to plant (these) potatoes.’

While analysing Finnish data, Ionin & Matushansky (2018: 63) compare the construction under discussion to the construction with measure nouns (*aski tupakkaa* packet.NOM tobacco.PART ‘a packet of tobacco’).

3.1.2. Weak numeral phrase

In a weak NumP, the case of the numeral and the noun is defined by the syntactic position of the NumP in the sentence, in the same way as in “adjective + noun” sequences where the agreement of both constituents in case and number is observed (e.g. *vanha nain* old.NOM woman.NOM ‘old woman’, *vanhaašt naižeešt* old.ELA woman.ELA ‘about the old woman’, etc.). A weak NumP usually appears as an oblique object (11, 12), in constructions with the negation of a direct object (3), in prepositional (2) or postpositional (13) constructions, and when a quantified noun is a plurale tantum. Unlike with strong NumPs, the noun is the head here, as usually the elimination of the noun renders the sentence ungrammatical, while the elimination of the numeral does not.

(11) Marjad(B)_ST

<i>a</i>	<i>obett-i</i>	<i>ain</i>	<i>miñnu-a</i>	<i>kahe-l</i>	<i>käe-l</i>
and	teach-PST.3SG	always	1SG-PART	two-ADE	hand-ADE
<i>vod</i>	<i>näin-ikkee</i>	<i>korjaa-maa</i>			
PTCL	such	gather-SPN			

‘And she was teaching me to pick berries with two hands like this.’

(12) Püüdämääž_GI

<i>kahe-ž</i>	<i>putka-ž</i>	<i>eļl-ii-d</i>
two-INE	cabin-INE	live-PST-3PL

‘(They) lived in two cabins.’

(13) Porkka 1885: 144

<i>kolme-n</i>	<i>tunni-n</i>	<i>peräst</i>	<i>ol-i</i>	<i>lindu</i>	<i>sîn</i>
three-GEN	hour-GEN	later	be-PST.3SG	bird	here

‘Three hours later the bird was here.’

In constructions with pluralia tantum nouns, both the agreeing numeral and the noun are in the plural form¹⁷ (14–16), cf. with (17), where the context is the same as in (14) but the noun is not a plurale tantum, so the strong NumP is used.

(14) A60ST(E)

<i>hää</i>	<i>ošt-i</i>	<i>kolme-d</i>	<i>atška-d</i>
3SG	buy-PST.3SG	three-PL.NOM	eyeglasses-PL.NOM

‘She bought three pairs of glasses.’

(15) Laanest 1966a: 139

<i>miu-l</i>	<i>ol-i</i>	<i>tää-l</i>	<i>nõre-mma-l</i>	<i>Boja-l</i>
1SG-ADE	be-PST.3SG	this-ADE	young-CPR-ADE	son-ADE
<i>vā</i>	<i>ühe-t</i>	<i>sütka-D</i>		
only	one-PL.NOM	day-PL.NOM ¹⁸		

‘This younger son of mine was only one day old (when he was baptized).’

(16) A60ST(E)

<i>miul</i>	<i>ei-oo</i>	<i>kakš-i-a</i>	<i>atško-j-a</i>
1SG.ADE	NEG.3SG-be	two-PL-PART	eyeglasses-PL-PART

‘I do not have two glasses.’

(17) 929EN(E)

<i>hää</i>	<i>ošt-i</i>	<i>kold</i>	<i>kañna-a</i>
3SG	buy-PST.3SG	three(NOM)	hen-PART

‘She bought three hens.’

When analysing the Finnish data, Ionin & Matushansky (2018: 80–81) offer two explanations for the fact that a singular numeral requires the partitive form of the quantified noun while a plural numeral does not. The first explanation is that the plural morphology of the numeral

17 See similar examples for Estonian (Metslang 2019: 463) and Finnish (Dolbey 1998: 20; Hurford 2003: 585–586; Hakulinen et al. 2004: §774).

18 The plural form appears here due to a calque from Russian. In contemporary Russian, the word *sutki* ‘(calendar) day’ only has a plural form. In the Ingrian example, the corresponding form has the Ingrian plural suffix, which is added to a constructed stem *sütka-*.

removes its ability to assign case.¹⁹ However, except for the nominative, there are no forms of a singular numeral able to assign case, so it would be more correct to say that the ability to assign case is a specific feature of the nominative singular form.

The second explanation is that “all cardinals assign adnumerative case to their complement, which is syncretic with the partitive case in the singular and with the nominative case in the plural. The underlying adnumerative case is different from the true partitive, as evidenced by the fact that <...> it is overridden by the oblique case in nondirect-case positions” (Ionin & Matushansky 2018: 81). This explanation does not look plausible, especially in the Finnish tradition, where objective nominative and genitive are already merged under the label “accusative”. If we accept that the same morphological form is sometimes considered the accusative and sometimes is merged with the partitive, the correspondence between the forms and their labels becomes unreasonably complicated.

In Finnish, the plural forms of numerals can appear not only with pluralia tantum nouns but also with other nouns, if the meaning ‘*n* groups of’ is expressed, cf. *kolme saapasta* three.NOM boot.PART ‘three boots’ vs *kolmet saappaat* three.PL.NOM boot.PL.NOM ‘three groups (pairs) of boots’ (see Hurford 2003: 584–590; Hakulinen et al. 2004: §774; Nenonen & Niemi 2010: 108; Ionin & Matushansky 2018: 79). In Ingrian, the plural forms of numerals can be used in the same function. This construction is similar to NumPs with pluralia tantum nouns, cf. (14) and (18).

(18) D38GI(E)

<i>miä</i>	<i>ošš-i-n</i>	<i>kolme-d</i>	<i>kengä-d</i>
1SG	buy-PST-1SG	tree-PL.NOM	boot-PL.NOM

‘I bought three pairs of boots.’

A NumP with the numeral *ükš* ‘one’ is always weak (19, 20).²⁰

19 Cf. Danon (2012: 1304), where the change from partitive government to the agreement strategy in Finnish is explained by the pluralization of the numeral.

20 See Ionin & Matushansky (2018: 62–63) about the adjectival behaviour of the numeral *yksi* ‘one’ in Finnish.

(19) Munad(B)_AI

i *siž* *emä* *ando-i* *mei-lle* ***ühe-n***
 and then mother give-PST.3SG 1PL-ALL one-GEN
tšuguniga-n
 cast.iron.pot-GEN

‘And then my mother gave us one cast-iron pot.’

(20) 439MB(E)

kaigin *ellää-d* ***ühe-ž*** ***koi-ž***
 all live.PRS-3PL one-INE house-INE

‘They all live in one house.’

3.2. Deviations from the main principles

As mentioned above, the principles of case marking described in 3.1 are applicable to other Finnic languages. However, Ingrian demonstrates several deviations from this system.

1. Strong NumPs spread to the contexts where the partial object is expected (21).

(21) D23AL(E)

miä *tihti* *lue-n* ***kakš*** *kirja-a*
 1SG often read.PRS-1SG two(NOM) book-PART

‘I often read two books.’

In Standard Finnish and Estonian, a similar construction requires the partitive marking of the NumP (22, 23).

(22) Finnish

lue-n *usein* ***kah-ta*** *kirja-a*
 read.PRS-1SG often two-PART book-PART

‘I often read two books.’

(23) Estonian

ma *loe-n* *sageli* ***kahte*** *raamatu-t*
 1SG read.PRS-1SG often two.PART book-PART

‘I often read two books.’

It is not clear whether the Ingrian construction with the strong numeral phrase might have appeared due to Russian influence. On the one hand, in the Russian examples (24) and (25), the form *dve* appears in the syntactic positions of both the subject and object as a result of case syncretism. Thus, copying of the subject case marking to the direct object does not look completely unjustified.

(24) Russian

Dv-e *knig-i* *lež-at* *na* *stol-e*
 two-NOM.F book-GEN lie-PRS.3PL on table-LOC
 ‘Two books are on the table.’

(25) Russian

Ja *čita-ju* *dv-e* *knig-i*
 1SG read-PRS.1SG two-ACC(=NOM).F book-GEN
 ‘I am reading two books.’

On the other hand, case forms of Russian numerals coincide in subject and object positions only if the noun is inanimate. For animate nouns there is no such syncretism, so there are contexts where Russian does not use the subject-like marking of the object (cf. 26 and 27), whereas Ingrian does (28).

(26) Russian

Ja *proda-l* *dv-ux* *kuric*
 1SG sell-PST.SG.M two-ACC.ANM hen.ACC.PL
 ‘I sold two hens.’

(27) Russian

Dv-e *kuric-y* *ubeža-l-i*
 two-NOM.F hen-GEN.SG run.away-PST-PL
 ‘Two hens run away.’

(28) 376AL(E)

miä *mö-i-n* *kakš* *kañnaa*
 1SG sell-PST-1SG two(NOM) hen.PART
 ‘I sold two hens.’

The fact that there are no unambiguous examples with the nominative form of the numeral instead of the expected partitive in the Spontaneous Speech Corpus nor in earlier text collections (Porkka 1885; Ariste 1960; Laanest 1966a) suggests that this phenomenon is not very frequent.

2. Though numerals in constructions with pluralia tantum nouns are usually in the plural form, some examples with the singular form of the numeral are also attested in my material (29). The variation between singular and plural forms of numerals is usually speaker dependent, cf. (29) with (30), where the expected plural form of the numeral is used.²¹ Sometimes such variation is observed in data from the same speaker. It is not clear whether it is an original Ingrian construction or a calque from Russian in the elicited material – in the Spontaneous Speech Corpus there are no examples that could answer this question. In Standard Finnish and Estonian, this construction requires the plural form of the numeral. It is possible that in Ingrian it is the result of attrition.

(29) B33EN(E)

<i>kümme<i>n</i>ü<i>-l</i></i>	<i>kaad'o-i-l</i>	<i>ei-oo</i>	<i>nüblä-ä</i>
ten-ADE	trousers-PL-ADE	NEG.3SG-be	button-PART

‘Ten trousers do not have buttons.’

(30) A60ST(E)

<i>kolm-i-l</i>	<i>pökšü-löi-l</i>	<i>ei-oo</i>	<i>nübl-i-ä</i>
three-PL-ADE	trousers-PL-ADE	NEG.3SG-be	button-PL-PART

‘Three trousers do not have buttons.’

3. The adjective-like syntactic behaviour of the numeral *üks* ‘one’ has spread to compound numerals ending with *üks*. NumPs with these

21 Note that, in examples (29) and (30), different words for trousers are used. Both of them exists in Soikkola Ingrian though the word *kaat'ĩ'aad* ‘trousers’ is more widespread than *pöksiid* ‘trousers’. The latter was borrowed from Finnish dialects (Nirvi 1971: 457). It is likely that the singular (29) vs plural (30) number of the word *nüblä* ‘button’ reflects some difference in meaning: one button of every pair of trousers in the case of the singular form and all buttons in the case of the plural form. However, in this particular instance the difference in meaning was not checked with the native speakers.

numerals also became weak (31, 32), which is not common for the major Finnic languages, cf. the Finnish example in (33).²²

(31) A60ST(E)

<i>hää</i>	<i>ošt-i</i>	<i>kakškümmend ühe-n</i>	<i>kirja-n</i>
3SG	buy-PST.3SG	twenty.one-GEN	book-GEN

‘He bought twenty-one books.’

(32) B82EI(E), (Markus & Rozhanskiy 2022: 327)

<i>naap̄puri</i>	<i>šah̄ha-iž</i>	<i>vižküm̄mend ühe-n</i>	<i>puu-n</i>
neighbour	saw-PST.3SG	fifty.one-GEN	tree-GEN

‘The neighbour sawed fifty-one trees.’

(33) Finnish

<i>naapuri-ni</i>	<i>saha-si</i>	<i>viisikymmentäyksi</i>	<i>puu-ta</i>
neighbour-P1SG	saw-PST.3SG	fifty.one(NOM)	tree-PART

‘My neighbour sawed fifty-one trees.’

It is likely that this change was induced by contact with the Russian language, where the last part in complex numerals has the same form as in the respective simple numeral and defines the syntactic relation with the quantified noun (Hurford 2003: 615), cf. (34, 36) with the agreement and the accusative form of the noun and (35, 37) with the numeral governing the genitive form of the noun.

(34) Russian

<i>Ja</i>	<i>vzja-l</i>	<i>odn-u</i>	<i>knig-u</i>
1SG	take-PST.SG.M	one-ACC.F	book-ACC.SG

‘I took one book.’

(35) Russian

<i>Ja</i>	<i>vzja-l</i>	<i>dv-e</i>	<i>knig-i</i>
1SG	take-PST.SG.M	two-ACC.F	book-GEN.SG

‘I took two books.’

(36) Russian

<i>Ja</i>	<i>vzja-l</i>	<i>dvadcat’</i>	<i>odn-u</i>	<i>knig-u</i>
1SG	take-PST.SG.M	twenty	one-ACC.F	book-ACC.SG

‘I took twenty-one books.’

²² See similar Finnish examples in Ionin & Matushansky (2018: 127–128).

(37) Russian

Ja vzja-l dvadcat' dv-e knig-i
 1SG take-PST.SG.M twenty two-ACC.F book-GEN.SG
 'I took twenty-two books.'

It is not clear when this change happened – earlier texts do not contain compound numerals ending in *ükš* in the position of the total object.

The same phenomenon was attested in Vaipooli Votic in the 21st century (see Rožanskij 2017: 63).

4. As shown by the examples above, the word order in the NumP is fixed: the numeral precedes the dependent noun. However, the reversed order, when the numeral follows the noun, is also possible: it expresses an approximate quantity (38, 39).

(38) Skvortsad_EN

taiĭtaa nüd jo on viikko-a kold
 probably now already be.PRS.3SG week-PART three(NOM)
ku hõõ männ-ii-d poiž
 when 3PL go-PST-3PL away
 'It is now probably about three weeks since they went away.'

(39) Püüdämääž_GI

hää šiz šeiššo-o šęęl tunni-a kakš
 3SG then stand.PRS-3SG there hour-PART two(NOM)
kold
 three(NOM)
 'Then it stands there about two or three hours.'

The same construction exists in Russian, so it is likely that in Ingrian it has been borrowed.²³ This is not a recent borrowing – this construction is also attested in earlier texts (40, 41).

(40) Ariste 1960: 19

päivä ***kold*** *ol-i lapši, šiz rišši-ttī.*
 day.PART three be-PST.3SG child then baptize-IPS.PST
 'The child was about three days old, then (the child) was baptized.'

23 This Russian construction was widely discussed in literature, see, for example, Matushansky (2015).

- (41) Laanest 1966a: 149

ja sīs pie-ttī säGi-s sīl päivā kaks
 and then hold-IPS.PST sack-INE there day.PART two(NOM)
 ‘And then it was kept there in a sack for about two days.’

The reversed word order is not the only means to express approximate quantity. Ingrian also has a crosslinguistically widespread construction with two different numerals. This construction is attested in earlier texts (42). Example (39) shows that a succession of numerals can be combined with the reversed order.

- (42) Ariste 1960: 17

kakš kolt päivā oļl-ī-D hullu-D
 two three day.PART be-PST-3PL crazy-PL
 ‘Two or three days they were crazy.’

3.3. Numeral phrase in adverbial function

A NumP in adverbial function can be either strong or weak. A strong NumP (with the nominative singular form of the numeral) usually expresses the most general indication of time period or distance (43, 44).²⁴

- (43) Püüdämääž_GI

hä kä-i va kakštoiš voot-ta mere-l
 3SG go-PST.3SG only twelve(NOM) year-PART sea-ADE
 ‘He had been going to the sea for just twelve years.’

- (44) Elo(B)_AI

pid-i män-nä viištoišt kilometri-a täšt
 have.to-PST.3SG go-INF fifteen(NOM) kilometre-PART from.here
 ‘It was necessary to go fifteen kilometres from here.’

In more specific constructions, a weak NumP is used. In (45), a particular year is indicated and the adessive marking is used; in (46) a particular time interval for which breads were prepared is marked with the translative, and in (47) a particular time interval when the work was fulfilled is specified.

²⁴ See, for example, Hakulinen et al. (2004: §972) about adverbials of quantity in Finnish.

- (45) Pedro_OM
neljütoizēe-l vōōvee-l hää män-i šoitaa
 fourteen-ADE year-ADE 3SG go-PST.3SG war.ILL
 ‘In 1914, he went to the war.’
- (46) Taikina_RI
mokkooma-d leivä-d šuure-d leivä-d
 such-PL.NOM bread-PL.NOM big-PL.NOM bread-PL.NOM
paišto-i taṗpa-iž kahe-kš viigoo-kš
 bake-PST.3SG be.enough-PST.3SG two-TRSL week-TRSL
 ‘Such breads, big breads she baked, it was enough for two weeks.’
- (47) 895OM(E)
kahe-ž kuu-ž hää teg-i koi-n
 two-INE month-INE 3SG do-PST.3SG house-GEN
 ‘He built the house in two months.’

4. Agreement of numeral phrase with the predicate

As mentioned in section 3.1, a NumP in the position of the subject is strong, and both the numeral and dependent noun have a singular form (except NumPs with pluralia tantum nouns and constructions with the meaning ‘*n* groups of’). Usually, the NumP expresses plurality and this causes a conflict between syntactic and semantic agreement with the predicate: the syntactic construction assumes the singular form of the predicate while from the semantic point of view we would expect the plural form.²⁵ In Finnic, this ambivalence often results in the variation of forms, depending on a particular language and particular construction. For example, Finnish allows both singular and plural predicates in cases of the direct word order²⁶ but only the singular form in cases of

25 I discuss only agreement in number because there is no morphological gender in Uralic languages and a NumP cannot be anything other than the 3rd person, so the only parameter for agreement is number.

26 Kaiser (2022b) notes that “singular verbs are the default option with Finnish numeral-noun constructions; plural verbs need special semantic/pragmatic support.”

the reversed word order.²⁷ Unlike Finnish, Estonian does not have such a restriction, cf. (48) and (49).

(48) a. Finnish

<i>Kaksi</i>	<i>mies-tä</i>	<i>tul-i / tul-i-vat</i>
two(NOM)	man-PART	come-PST.3SG/PST-3PL
‘Two men came.’		

b. Finnish

<i>Tul-i / *tul-i-vat</i>	<i>kaksi</i>	<i>mies-tä</i>
come-PST.3SG/PST-3PL	two(NOM)	man-PART
‘Two men came.’		

(49) a. Estonian

<i>Kaks</i>	<i>mees-t</i>	<i>tul-i / tul-i-d</i>
two(NOM)	man-PART	come-PST.3SG/PST-3PL
‘Two men came.’		

b. Estonian

<i>Tul-i / tul-i-d</i>	<i>kaks</i>	<i>mees-t</i>
come-PST.3SG/PST-3PL	two(NOM)	man-PART
‘Two men came.’		

In this section, I will first consider agreement with the verb ‘to be’ (4.1) and then agreement with other verbs (4.2). The final subsection analyses agreement in predications containing nominals (4.3).

4.1. Agreement with the verb ‘to be’

In the Spontaneous Speech Corpus, there are about 60 occurrences of a NumP in the subject position, combining with both singular and plural forms of the predicate. The main factor influencing the agreement is lexical: the verb ‘to be’ is in the singular form (50) in 50 of 52 occurrences (for other verbs this is not the case, see 4.2).

27 In fact, the Finnish system of agreement of a subject expressed by a numeral phrase with the predicate is rather complicated and depends on various factors (type of subject, word order, type of sentence, etc.), see Karlsson (1959).

- (50) Šoomeež_OM
šeitse *henkki-ä* *laps-i-a* *ol-i*
 seven(NOM) soul-PART child-PL-PART be-PST.3SG
 ‘There were seven children.’

One of the two remaining occurrences looks irregular: the native speaker used the form *olivat* ‘be.PST.3PL’, which is a Finnish form (the correct Soikkola Ingrian form is *ölliid*).²⁸ In the second case, the impersonal form of the predicate was used (51). Using an impersonal form instead of a personal one (i.e. in a sentence with a subject) is not very typical for Soikkola Ingrian but it is not rare, cf. examples (56, 59). No other constructions with a NumP and the impersonal form of the verb ‘to be’ are attested in the Spontaneous Speech Corpus so it is not clear how natural this construction is (example (56) from Laanest (1966a) has a plural form of the numeral so it should be considered a different construction).

- (51) Hirvi_GI
kakš *męęš-t* *kerä* *ol-ittäi* *obogaa-ž*
 two man-PART also be-IPS.PST mushroom-INE
 ‘Two men were gathering mushrooms too.’

Neither word order nor the tense of the verb²⁹ influences the agreement of the NumP subject with the verb ‘to be’. Examples (46–49) illustrate that these factors are not significant for agreement: word order distinguishes (52, 54) from (53, 55), the verbal tense distinguishes (52, 53) from (54, 55), but in all these examples the singular form of the verb is attested.

28 It is not uncommon for a Soikkola Ingrian speaker to have some knowledge of Finnish. Usually, this is a result of the deportation to Finland in the course of the Second World War (see, for example, Kuznetsova, Markus & Muslimov 2015: 157–158) but it can also be due to contacts with Ingrian Finns or with visitors from Finland. In the current case, the latter is the most probable reason. The form *olivat* ‘be.PST.3PL’ also exists in Lower Luga Ingrian. Though there are no obvious reasons why this particular speaker uses the Lower Luga Ingrian form, this explanation is also possible.

29 An asymmetry between different tense forms of the verb ‘to be’ is not rare. For example, in Estonian, the verb ‘to be’ lost the difference between 3SG and 3PL in the present tense. In Votic and Ingrian, the present tense copula *on* ‘be.PRS.3SG’ can be omitted but the past tense copula *oli* ‘be.PST.3SG’ cannot.

- (52) Šoomeež_ OM
kakš tüttöj-ä häne-l ol-i
 two(NOM) girl-PART 3SG-ADE be-PST.3SG
 ‘She had two daughters.’
- (53) Pahhain_elo_MM
miu-l ol-i kold tükkü-ä laps-i-a
 1SG-ADE be-PST.3SG three(NOM) item-PART child-PL-PART
 ‘I had three children.’
- (54) Huvä_elo_AI
kakš henkki-ä täž on mogomb-i-a
 two(NOM) soul-PART here be.PRS.3SG such-PL-PART
naaburi-loj-a
 neighbour-PL-PART
 ‘There are two such neighbours here.’
- (55) Šakšalaiššiin_aiga_MM
meije-n pid-i šiin hõõnee-ž eällä-ä
 1PL-GEN have.to-PST.3SG that.INE room-INE live-INF
kuž on kakš-šaitaa henkki-ä
 where be.PRS.3SG two.hundred(NOM) soul-PART
 ‘We had to live in that room with two hundred people in it.’

In Porkka (1885), Ariste (1960), and Laanest (1966a), there are no constructions with the direct word order and the verb ‘to be’ in the present tense, but in all other cases, the agreement with the verb ‘to be’ is the same as in the contemporary Ingrian language.

It is no longer possible to check how strict the ban on the plural form is in the examples discussed above. However, a considerable number of examples with the verb ‘to be’ in the Spontaneous Speech Corpus makes it possible to claim that the plural form of the verb ‘to be’ is at least highly atypical (if not impossible). In Finnish, the variation in the number of the predicate depends on the type of clause (locative clauses are prone to such variation, unlike existential ones). However, in the Ingrian data, there are not many unambiguous locative clauses – in most cases only the existential clauses or sentences which allow for different

interpretations³⁰ are observed. In any case, examples where the subject is a NumP and the predicate is a regular plural form of the verb *olla* ‘to be’ are not attested.

There are not enough relevant examples in the Spontaneous Speech Corpus or in earlier publications to describe the agreement of NumPs with pluralia tantum nouns. In Laanest (1966a: 122), there is an example with the impersonal form of the verb (56), which sometimes serves as the 3Pl personal form, cf. (51) and (59).

(56) Laanest 1966a: 122

<i>kahe-t</i>	<i>kusseGa-D</i>	<i>ol-tī</i>	<i>ukko-loi-l</i>
two-PL.NOM	sash-PL.NOM	be-IPS.PST	old.man-PL-ADE

‘The old men had two sashes.’

4.2. Agreement with other verbs

In the Corpus of Elicitation, there are only nine occurrences of verbs other than ‘to be’ agreeing with a subject NumP. Five of these examples contain verbs in the plural, for example (57, 58).

(57) Šoomeež_OM

<i>ühekšä</i>	<i>henkki-ä</i>	<i>kanne-n</i>	<i>takkaakš</i>	<i>jü-i-vüd</i>
nine(NOM)	soul-PART	table-GEN	behind	remain-PST-3PL

‘Nine people remained at the table.’

(58) Hirvi_GI

<i>kakš</i>	<i>hirvi-ä</i>	<i>jokš-ii-d</i>
two(NOM)	elk-PART	run-PST-3PL

‘Two elks ran.’

Two more examples have impersonal forms of the verb, e.g. (59). Both examples were recorded by the same native speaker who is the youngest and whose speech is affected by Russian to a greater extent than the speech of other Ingrians.³¹

30 Word order, which is often considered a criterion to distinguish existential and locative clauses (see, for example, Etxebarria 2012: 87 on Finnish), does not always work.

31 One of the hypotheses suggests that the replacement of 3Pl personal forms by impersonal forms appeared due to the Russian influence (Nirvi 1947).

(59) Lastotškad_LK

vot kakš lastotška-a teh-tii viiž laš-t
 PTCL two swallow³²-PART do-IPS.PST five(NOM) child-PART
 ‘Two swallows produced five children.’

Two remaining occurrences differ from the others because they contain the singular form of the verb. In example (60), the verb is in the pluperfect form. The pluperfect form consists of two components – the personal form of the verb *olla* ‘to be’ and the participle of the lexical verb. One can hypothesize that the singular number of the predicate in the perfect tenses (i.e. the perfect or pluperfect) is because of the auxiliary ‘to be’, which takes a singular form with a NumP subject (see 4.1), while the lexical verb does not determine the agreement.

(60) Pedro_OM

a neljä ol-i kōqol-d
 and four be-PST.3SG die-PTCP.ACT.SG
 ‘And four (children) died.’

Since there are no other examples with a NumP as the subject and the perfect form of the predicate in the Spontaneous Speech Corpus, I cannot make more definite conclusions, but in Ariste (1960: 9, 24) one also finds examples of pluperfect (61) and pluperfect impersonal (62, 63)³³ where the auxiliary is in the singular form. Thus, the older recordings support the proposed hypothesis.

(61) Ariste 1960: 24

mei-l ol-i kahekšav³⁴-varšā hävi-nnēD
 1PL-ADE be-PST.3SG eight foal.PART get.lost-PTCP.ACT.PL
 ‘We had eight foals lost.’

32 Ingrian has its own word for ‘swallow’ – *pääsköi*. However, this native speaker always uses the Russian word *lastotška*.

33 Theoretically these examples could also be interpreted as a stative passive. The difference between impersonal perfect forms and passive forms in the Finnic languages is a complicated theoretical question that I will not discuss in this article. See Oskolskaia (2024) on these forms.

34 The final consonant *-n* is assimilated here by the initial consonant of the following word.

(62) Ariste 1960: 9

<i>neljä šaittä</i>	<i>ol-i</i>	<i>kuþpu-a</i>	<i>teh-tü</i>
four hundred.PART	be-PST.3SG	sheaf-PART	do-PTCP.PASS

‘Four hundred sheaves were made.’

(63) Ariste 1960: 42

<i>kaks</i>	<i>naiš-t</i>	<i>ol-i</i>	<i>pan-du</i>	<i>korjā-mā</i>
two	woman-PART	be-PST.3SG	put-PTCP.PASS	gather-SPN

‘Two women were assigned to gather.’

The last example (64) contains the word *tuhatta* ‘one thousand’. It is not clear why the predicate has the singular form here. However, I can hypothesize that the cardinal *tuhatta* ‘one thousand’ might possess some features of a noun. Crosslinguistically, numerals such as *thousand*, *million*, etc. often demonstrate differences from other numerals (see the discussion on semilexical cardinals in Ionin & Matushansky 2018: 216–218). Greenberg (1989: 111) notes that “the other basic principle is that the lower a numeral is, the more it is treated as an adjective and the higher it is, especially for bases, the more it is treated as a noun, morphologically and syntactically”. Zabbal (2005) distinguishes *low numerals*, that are adjectival, and *high numerals*, that are nominal. In Finnish, a syntactic feature of nouns (the genitive case in the position of total object) was attested for large numerals (*miljoona* ‘million’, etc.) but not for *tuhatta* ‘thousand’ (Hakulinen et al. 2004: §776). However, it would not be surprising if the syntactic behaviour of the Ingrian *tuhatta* ‘one thousand’ acquired some nominal features due to the influence of Russian, cf. (65) where the predicate can be either in the plural form or the singular form of the neutral gender and (66) where the feminine singular form of the predicate appears due to the syntactic agreement of the nominal type. See also Corbett (2001: 215) on the singular agreement of ‘thousand’ in Slavonic.

(64) Elo(B)_AI

<i>tuhatta</i>	<i>rubl'a-a</i>	<i>kuu-ž</i>	<i>mahtaa</i>
thousand(NOM)	rouble-PART	month-INE	be.able.PRS.3SG

šinne tuļloo
there come.PRS.3SG

‘One thousand roubles per month, maybe, comes there.’

- (65) Russian
Dvadcat' čelovek ostalis' / ostalos'
 twenty person.PL.GEN leave.PST.PL/leave.PST.SG.N
bez rabot-y
 without work-GEN
 'Twenty people left without the job.'
- (66) Russian
Tysjača čelovek ostalas' / ostalis' / ostalos'
 thousand person.PL.GEN leave.PST.SG.F/leave.PST.PL/¹leave.PST.SG.N
bez rabot-y
 without work-GEN
 'One thousand people left without the job.'

The number of the predicate does not correlate with its tense, cf. (57) and (67).

- (67) Lastotškad_LK
kaig viiž tükkü-ä nii kovašt mätti-jüü-d peššää
 all five item-PART so very climb-PRS-3PL nest.ILL
 'All five (nestlings) climb so actively into the nest.'

All mentioned occurrences of verbs other than 'to be' agreeing with the NumP have a direct word order: the NumP precedes the verb. There are no examples with the reversed word order in the Spontaneous Speech Corpus.

It is interesting that all such examples in earlier texts – one from Porkka (1885) and four from Ariste (1960)³⁵ – have the reversed word order (see examples 68–70). In all cases, the predicate has the singular form.

- (68) Porkka 1885: 130
i tul-i kolD tēnhārā
 and come-PST.3SG three(NOM) road.branch.PART
 'And (they) came to a triple road fork (lit. a fork of three roads came).'

35 I failed to find such examples in Laanest (1966a).

(69) Ariste 1960: 19

šündü *kakš* *poikā* *i* *ükš* *tüttöi*
 be.born.PST.3SG two(NOM) son.PART and one daughter
 ‘Two sons and one daughter were born.’

(70) Ariste 1960: 21

nāppapā-le *tuļļō* *jo* *kakš* *jalkā,*
 tadpole-ADE come.PRS.3SG already two(NOM) leg.PART
ja *händä* *hävijā*
 and tail be.lost.PRS.3SG
 ‘The tadpole already acquires two legs, and the tail is lost.’

If we do not assume that a crucial change in the syntactic distribution happened in recent decades, we should conclude that the word order is a significant factor for the agreement of verbs other than ‘to be’: the direct word order requires the plural form of the predicate, and the reversed word order requires the singular form.

In the Corpus of Elicitation, I have examples of the reversed word order both with the singular and with the plural form of the verb. The plural forms usually appear in cases when the Russian stimulus contains the plural form as well. I evaluate such examples as a direct influence of the stimulus and do not take them into account. However, one cannot exclude the possibility that such variation became more widespread in recent decades. Sometimes it was observed in the examples from the same speaker.

One should not dismiss the idea that there are some semantic features of the subject expressed by a NumP which influence the number of the predicate, e.g. its definiteness, cf. with Finnish (Hakulinen et al. 2004: §790). However, in my data I failed to find reliable evidence that such a correlation exists.

4.3. Agreement of the nominals

Agreement in number also concerns the nominal parts of predication: nouns, adjectives, and participles in the perfect tenses. In (60), there is a singular form of the participle. However, in Ariste (1960) the plural form of the participle is attested in the pluperfect form (61).

Hence, while it is likely that there is variation in the number form of the participles in the perfect tenses, there is not enough data to conclude what governs this variation.

In the case of the nominal predicate, the noun or adjective takes the plural form even though the copula has the singular form (71, 72).

(71) Elo(B)_AI

<i>kakš</i>	<i>vunukka-a</i>	<i>kovašt</i>	<i>hüvä-d</i>	<i>on</i>
two(NOM)	grandson-PART	very	good-PL.NOM	be.PRS.3SG

‘Two grandsons are very good.’

(72) C06AL(E)

<i>kakš</i>	<i>ihmiiš-t</i>	<i>on</i>	<i>šüitiä-d</i>
two(NOM)	person-PART	be.PRS.3SG	sober-PL.NOM

‘Two people are sober.’

In example (73), the dependent noun is repeated after the copula and the adjective agrees with this noun – both are in the partitive singular (without the repeated noun we would expect the nominative plural form, cf. 71 and 72).

(73) Kurvi_rokka2_AG

<i>kold</i>	<i>šäkki-ä</i>	<i>ol-i</i>	<i>šuur-d</i>	<i>šäkki-ä</i>
three(NOM)	sack-PART	be-PST.3SG	big-PART	sack-PART.NOM

‘Three sacks were big sacks.’

Conclusions

The syntax of the Ingrian numeral phrase does not offer many surprises – it preserves most of the Finnic traits and generally looks similar to that of Finnish. However, there are a number of features that distinguish the Ingrian numeral phrase from the Finnish one. The influence of Russian seems to be the main cause of these differences. It is not a recent influence as these features are also found in the collections of texts recorded in the middle of the 20th century.

Table 3 lists the main features of the Ingrian NumP. The features that Ingrian shares or does not share with Standard Finnish and Standard Estonian are marked with “+” or “–”, respectively. In the rightmost column, the potential Russian influence is evaluated: “+” means that this influence is highly likely, “?” means that this influence is possible but there is no solid evidence for it, “–” marks the features that distinguish Ingrian from Russian, and N/A means that this feature is not

applicable or relevant (e.g. it is common for Russian and all three Finnic languages under consideration).

Table 3. Main features of the Ingrian NumP.

Phenomenon	Same as in Standard Finnish and Estonian	Russian influence
Structure of NumP		
Opposition of the strong NumP (Num _{NOM} + N _{PART}) vs the weak NumP (agreement of Num and N in case)	+	–
The strong NumP can appear in the position of the partial object	–	?
The weak NumP is used with pluralia tantum nouns	+	–
Singular form of a numeral occasionally appears with pluralia tantum nouns	–	N/A
Numerals ending in “one” require the weak NumP: 21, 31, etc.	–	+
Reversed word order in NumP expressing approximate quantity	–	+
Agreement of subject NumP with predicate		
The verb ‘to be’ usually requires the singular form of the predicate	+	–
The number of verbs other than ‘to be’ is defined primarily by word order (plural for NumP-V and singular for V-NumP)	–	–
In perfect tenses, the auxiliary is in the singular if the subject is a NumP	+	N/A
The nominal part of the predicate takes the plural form	+	N/A

A moribund language which has had sustained contact with a major language of higher prestige and which does not have any monolingual speakers left typically borrows many features from the major language. The analysis carried out in this paper has shown that the syntax of numeral phrase in Ingrian is sensitive to the contact influence only to

a limited extent and should be considered as a relatively stable part of the language.

Acknowledgements

I am very grateful to Éva Dékány, Ekaterina Gruzdeva, Riho Grünthal, Juha Janhunen, Mehmed Muslimov and Eva Saar for their comments on the article and/or consultations concerning Finnish, Estonian and Ingrian data.

This research was supported by the University of Tartu, grant PHVEE18904.

Abbreviations

ACC – accusative, ACT – active, ADE – adessive, ALL – allative, ANM – animate, CNG – connegative, CPR – comparative form, ELA – elative, F – feminine, GEN – genitive, ILL – illative, INE – inessive, INF – infinitive, IPS – impersonal, LOC – locative, M – masculine, N – neutral, NEG – negation marker, NOM – nominative, P1 – 1st person possessive marker, PART – partitive, PASS – passive, PL – plural, PRS – present tense, PST – past tense, PTCP – participle, PTCL – particle, SG – singular, SPN – supine, TRSL – translative, 1 – 1st person, 3 – 3rd person.

References

- Ariste, Paul. 1960. Isuri keelenäiteid. *Keel ja kirjanduse instituudi uurimused* 5. 7–116.
- Babby, Leonard H. 1985. Noun phrase internal case agreement in Russian. *Russian Linguistics* 9(1). 1–15. <https://doi.org/10.1007/BF00161063>.
- Blažek, Vaclav. 1999. *Numerals. Comparative-etymological analyses of numeral systems and their implications (Saharan, Nubian, Egyptian, Berber, Kartvelian, Uralic, Altaic and Indo-European languages)*. Brno: Masarykova Univerzita.
- Comrie, Bernard. 2021. *Typology of numeral systems*. Version of 2021 May 26. https://lingweb.eva.mpg.de/channumerals/TypNum_Latest_21ho.pdf (12 August, 2022).
- Corbett, Greville G. 1978. Problems in the Syntax of Slavonic Numerals. *The Slavonic and East European Review* 56(1). 1–12.
- Corbett, Greville G. 1993. The head of Russian numeral expressions. In Greville G. Corbet, Norman M. Fraser & Scott McGlashan (eds.), *Heads in grammatical theory*, 11–35. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511659454.002>.

- Corbett, Greville G. 2001. *Number*. Cambridge: Cambridge University Press.
- Danon, Gabi. 2012. Two structures for numeral-noun constructions. *Lingua* 122. 1282–1307. <https://doi.org/10.1016/j.lingua.2012.07.003>.
- de la Villa, Jesús. 2010. Numerals. In Philip Baldi & Pierluigi Cuzzolin (eds.), *New Perspectives on Historical Latin Syntax. Volume 3: Constituent Syntax: Quantification, Numerals, Possession, Anaphora*, 175–238. Berlin/New York: De Gruyter.
- Dolbey, Andrew. 1998. Constructional inheritance and case assignment in Finnish numeral expressions. *Nordic Journal of Linguistics* 21(1). 17–45. <https://doi.org/10.1017/S0332586500004133>.
- Etxeberria, Urtzi. 2012. Quantification in Basque. In Edward L. Keenan & Denis Paperno (eds.), *Handbook of Quantifiers in Natural Language* (Studies in Linguistics and Philosophy 90), 83–164. Dordrecht/Heidelberg/London/New York: Springer. https://doi.org/10.1007/978-94-007-2681-9_3.
- Greenberg, Joseph H. 1989. The internal and external syntax of numerical expressions: Explaining language specific rules. In Michel Kefer & Johan van Auwera (eds.), *Universals of Language* (Belgian Journal of Linguistics 4), 175–238. Brussels: Editions de l'Université de Bruxelles. <https://doi.org/10.1075/bjl.4.07gre>.
- Hakulinen, Aili, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho. 2004. *Iso suomen kielioppi*. Online version. <http://scripta.kotus.fi/visk> (16 May, 2023).
- Hurford, James R. 2003. The interaction between numeral and nouns. In Frans Plank (ed.), *Noun Phrase Structure in the Languages of Europe* (Empirical Approaches to Language Typology 20–7), 561–620. Berlin/New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110197075.4.561>.
- Ionin, Tania & Ora Matushansky. 2018. *Cardinals. The syntax and semantics of cardinal-containing expressions*. Cambridge London: The MIT Press. <https://doi.org/10.7551/mitpress/8703.001.0001>.
- Junus, Väino I. 1936. *Izoran keelen grammatikka: morfologia opettajaa vart*. Leningrad/Moskova: Ucpedgiz.
- Kaiser, Elsi. 2022a. Referential effects on verb agreement: Finnish numeral-noun constructions. In John R. Starr, Juhyae Kim & Burak Öney (eds.), *Proceedings of the 32nd Semantics and Linguistic Theory Conference*. <https://doi.org/10.3765/salt.v1i0.5388>.
- Kaiser, Elsi. 2022b. Unexpected verb agreement: Finnish numeral-noun constructions. Paper presented at the online conference “Syntax of Uralic Languages 4 (SOUL-4)” on 14–16 June 2022.
- Karlsson, Göran. 1959. Suomen kielen predikaatin numeruksesta kardinaalilukusanan ollessa subjektina. *Virtittäjä* 63(3). 356–368.
- Karttunen, Lauri. 2006. Numbers and Finnish numerals. In Mickael Suominen, Antti Arppe, Anu Airola, Orvokki Heinämäki, Matti Miestamo, Urho Määttä, Jussi Niemi, Kari K Pitkänen & Kaius Sinnemäki (eds.), *A Man of Measure: Festschrift in Honour of Fred Karlsson on his 60th Birthday* (SKY Journal of Linguistics 19), 407–421. Turku: The Linguistic Association of Finland.
- Kuznecova, Natalija V. 2009. *Fonologičeskie sistemy iżorskix dialektov*. Institut lingvističeskix issledovanij RAN, Sankt-Peterburg. Ph.D. thesis.

- Kuznetsova, Natalia, Elena Markus & Mehmet Muslimov. 2015. Finnic minorities of Ingria. In Heiko F. Marten, Michael Rießler, Janne Saarikivi & Reetta Toivanen (eds.), *Cultural and Linguistic Minorities in the Russian Federation and the European Union. Comparative Studies on Equality and Diversity* (Multilingual Education 13), 127–167. Cham: Springer. https://doi.org/10.1007/978-3-319-10455-3_6.
- Laanest, Arvo. 1966a. *Isuri murdetekste*. Tallinn: Eesti NSV Teaduste Akadeemia, Keele ja Kirjanduse Instituut.
- Laanest, Arvo. 1966b. *Ižorskie dialekty. Lingvogeografičeskoe issledovanie*. Tallin: Akademiija nauk Ėstonskoj SSR, Institut jazyka i literatury.
- Laanest, Arvo. 1978. *Istoričeskaja fonetika i morfoloģija ižorskogo jazyka*. Tallin: Institut jazyka i literatury Akademii nauk Ėstonskoj SSSR, Tallin. Ph.D. thesis.
- Laanest, Arvo. 1986. *Isuri keele ajalooline foneetika ja morfoloģia*. Tallinn: Valgus.
- Laanest, Arvo. 1998. Isurid ja isuri keel. In Jaan Ōispuu (compiler), Marje Joalaid (ed.), *Kaheksa keelt, kaheksa rahvast*, 30–36. Tallinn: Tallinna Pedagoogikaülikooli Kirjastus.
- Markus, Elena & Fedor Rozhanskiy. 2022. Ingrian. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*, 308–329. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198767664.003.0018>.
- Matushansky, Ora. 2015. On Russian approximative inversion. In Gerhild Zybatoŵ, Petr Biskup, Marcel Guhl, Claudia Hurtig, Olav Mueller-Reichau & Maria Yastrebova (eds.), *Slavic Grammar from a Formal Perspective: The 10th Anniversary FDSL Conference*, 303–316. Frankfurt: Peter Lang.
- Metslang, Helle. 2019. Kvantorifraas. In Erelt, Mati & Helle Metslang (eds.), *Eesti keele süntaks*, 463–478. Tartu: Tartu Ülikool kirjastus.
- Nenonen, Marja & Jussi Niemi. 2010. Mismatches between grammatical number and conceptual numerosity: A number–decision experiment on collective nouns, number neutralization, pluralia tantum, and idiomatic plurals. *Folia Linguistica* 44(1). 103–125. <https://doi.org/10.1515/flin.2010.004>.
- Nirvi, Ruben E. 1947. *Passiivimuotojen aktiivistumisesta*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Nirvi, Ruben E. 1971. *Inkeröismurteiden sanakirja* (Lexica societatis Fenno-Ugricae 18). Helsinki: Suomalais-Ugrilainen Seura.
- Oskolskaia, Polina. 2024. Impersonal or passive? Some approaches to analysis (based on the Veps and Estonian languages). *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 15(1). 221–253. <https://doi.org/10.12697/jeful.2024.15.1.07>.
- Ovsjannikova, Maria & Fedor Rozhanskiy. 2024. Syntax of the Uralic languages: main 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.2024.15.1.04>.
- Porkka, Volmari. 1885. *Ueber den Ingrischen Dialekt: mit Berücksichtigung der übrigen finnisch-ingermanländischen Dialekte*. Helsingfors: J. C. Frenckell & Sohn.
- Rozhanskiy, Fedor & Elena Markus. 2017. Negation in Soikkola Ingrian. *Finnisch-Ugrische Mitteilungen* 41. 189–219.

- Rožanskiy, Fedor. 2017. Differencirovannoe markirovanie ob'ekta v pribaltijsko-finskix jazykax: osnovnye principy i deviacii. *Acta Linguistica Petropolitana* 13(3). 41–75.
- Rožanskiy, Fedor & Elena Markus. 2013. Ižora Sojkinskogo poluoostrova: fragmenty sociolingvističeskogo analiza. *Acta Linguistica Petropolitana* 9(3). 261–298.
- Rožanskiy, Fedor & Elena Markus. 2014. Padežnoe oformlenie ob'ekta v ižorskom jazyke s točki zrenija sintaksičeskoj tipologii. In Mixail A. Daniël', Ekaterina A. Ljutikova, Vladimir A. Plungjan, Sergej G. Tatevosov & Ol'ga V. Fedorova (eds.), *Jazyk. Konstanty. Peremennye. Pamjati Aleksandra Evgen'eviča Kibrika*, 564–578. Sankt-Pererburg: Aletejja.
- Rutkowski, Paweł. 2001. Numeral phrases in Polish and Estonian. In Arthur Holmer, Jan-Olof Svantesson & Åke Viberg (eds.), *Proceedings of the 18th Scandinavian Conference of Linguistics*, vol. 2 (Travaux de l'Institut de Linguistique de Lund 39:2), 181–190. Lund: Lund University Press.
- Saar, Eva. 2017. *Isuri keele Soikkola murde sõnamuutmissüsteem* (Dissertationes philologiae uralicae Universitatis Tartuensis 19). Tartu: University of Tartu.
- Sovijärvi, Antti. 1944. *Foneettis-äännehistoriallinen tutkimus Soikkolan inkeröismurteesta* (Suomi-kirjan eripainoksia 5). Helsinki: Suomalaisen Kirjallisuuden Seura.
- von Mengden, Ferdinand. 2010. *Cardinal Numerals. Old English from a Cross-Linguistic Perspective* (Topics in English Linguistics 67). Berlin/New York: Walter de Gruyter. <https://doi.org/10.1515/9783110220353>.
- Zabbal, Youri. 2005. *The syntax of numeral expressions*. Unpublished manuscript. University of Massachusetts at Amherst.
- Žoha, Lukáš, Marcin Wągiel & Pavel Čaha. 2022. The morphology of complex numerals: A cross-linguistic study. *Linguistics Beyond and Within* 8. 200–217. <https://doi.org/10.31743/lingbaw.14966>.

Kokkuvõte. Fedor Rozhanskiy: Numeraalfraasi süntaks isuri keele Soikkola murdes. See artikkel kirjeldab numeraalfraasi isuri keele Soikkola murdes. Artikli keskmes on numeraalfraasi välissüntaks (eeskätt selle komponentide käändemärgistus) ja arvu ühildumine numeraalfraasi ning predikaadi vahel. Andmeallikad on (a) 2006.–2013. aastal salvestatud spontaanse kõneüksuste kogu, (b) varasemate uurijate avaldatud spontaanse kõne näited, (c) 2006.–2023. aastal salvestatud küsitletud materjal. Kuigi numeraalfraas Soikkola murdes säilitab enamiku tavalistest läänemeresoome keele tunnustest, on sellel mõningaid vähemlevinud omadusi, näiteks ühildumine kõigis numeraalfraasides numbriga 'üks' lõppeva numeraali või ligikaudse koguse väljendamise kaudu pööratud sõnajärgjega. Need omadused on tekkinud tõenäoliselt vene keele kontakti tulemusena.

Märksõnad: isuri keel, süntaks, numeraalfraas, ühildumine, käändemärgistus, keelekontakt