ESUKA – JEFUL 2022, 13–2: 347–381

LEXICAL OUTCOMES OF KARELIAN-RUSSIAN BILINGUALISM IN TVER KARELIAN

Susanna Tavi
University of Eastern Finland, FI
susanna.tavi@uef.fi

Abstract. This study investigates the language contact between Tver Karelian and Russian, attempting to provide a comprehensive overview of the lexicon of bilingual code. The methodology includes a combination of statistical analyses and handling contact-induced change in terms of the Code-Copying Framework (=CCF). Nine interviews with nine people were conducted using the memory walk method. In code copying, correlations were found between different word classes and contact-relatedness. In code alternation, few differences were found between different speakers and one commonality was the use of complex numerals as Russian phrases without adapting them into the Tver Karelian code. The findings confirm that the copies are of a certain kind and appear in certain word classes. Code alternation sequences suggest that, according to the CCF, the discourse rather than the language is mixed. The findings within CCF have implications on minority language policies, as the findings support the use of bilingual terminology.

Keywords: code-copying, bilingualism, lexicon, contact-induced language change, Karelian, Russian

DOI: https://doi.org/10.12697/jeful.2022.13.2.11

1. Introduction

Karelian in Tver Oblast, Russia is an endangered Finnic language variety belonging to the Finno-Ugric language group of Uralic languages. In this article, Tver Karelian is studied from a lexical point of view. The lexicon of Tver Karelian has been studied along with other Karelian varieties, for example in the dictionaries of Karelian (KKS & Bubrih et al. 1997) and in Sarhimaa and Siilin (1994). The Tver Karelian variety has been of interest to linguists since the 18th century as D. E. D. Europaeus did his first expeditions to study the enclave's variety. In addition, Juho Kujola, A. A. Beljakov, G. N. Makarov and

A. V. Punžina collected data and studied Tver Karelian since the turn of the 19th century. (Virtaranta 1980: 100.) Pertti and Helmi Virtaranta did the most extensive work on Tver Karelian in the second half of the 20th century. After the work of the Virtarantas, Tver Karelian has been studied in the 1990s (Sarhimaa & Siilin 1994) and it has attracted new interest starting in 2010 and onwards (e.g. Novak 2014, 2016, 2019, Oranen 2019, Uusitupa 2020, 2021, Kehayov et al. 2021).

Tver Karelian descends from the Southern Karelian dialect and was formed in the 17th century due to wars on the north-western border of Russia causing population shifts from the border of Sweden (currently Finland) and Russia (Virtaranta 1980: 94–97). Karelian persisted, surrounded by the Russian language until the Soviet period, when strict language and assimilation policies resulted in growing bilingualism and language shift. In Russia's 2010 census, the number of ethnic Karelians in Tver Oblast was 7,294 and only 2,750 of them reported Karelian as their mother tongue. By contrast, according to the Soviet census of 1926, 140,567 speakers of Karelian lived in Tver Oblast at that time. (Laakso et al. 2016: 97.) Today, the remaining speakers of Tver Karelian are bilingual, also speaking Russian, and are middle-aged or older. In this article, bilingualism is defined according to Grosjean (2008: 13–14) as using two languages without assessing the speakers' competence in the languages.

Avoiding bilingual speech in data collection used to be common in the research of Finno-Ugric languages and in European linguistics in general due to monolingual bias (Laakso et al. 2016: 1–6). Monolingual bias is a term describing the general view in linguistics that monolingualism is a norm, whereas bi- or multilingualism is always marked (Verschik 2008: 2). Because of monolingual bias, Russian influence, especially code alternation, was seen as the unwanted mixing of two languages. In this research, in order to avoid monolingual bias in the data collection, the memory walk method was employed. This method allows the informants to use their bilingual resources freely, without any further instructions or restrictions.

The data consist of nine interviews conducted in the autumn of 2018 in six different villages in Tver Oblast, Russia: Klyčevoi, Seltsa, Stuanu, Miikšinä, Zaluazina and Maksuatiha; the latter is a small town and the administrative centre of the district of Maksuatiha. The villages are in two different districts, Maksuatiha and Lihoslavlja. The areas were

chosen to get a view on spoken Karelian in different places in the main areas of the Tver Karelian variety. Only the Karelian names of the villages and the districts are given in this article. Figure 1 illustrates the locations of the villages as well as the location of Tver Oblast in the Northern European context. Maksuatiha and Seltsa are situated northeast from the area and do not show on the map.

Figure 1. Map of the Tver Karelian villages and the location of Tver Oblast in the Northern European context.

Another attempt to avoid monolingual bias is to view Tver Karelian as a bilingual code. In order to study bilingual code and speech, the Code-Copying Framework (CCF) was employed to analyse all kinds of contact-induced changes, i.e., in traditional terms, borrowing and code-switching of lexicon within the same framework. Among all Finno-Ugric languages, traditional lexical borrowing and code-switching (CS) have rarely been researched together within the same theoretical framework. The study of lexical borrowing has concentrated on etymology (e.g. Kalima 1952, Kallio 2006, Saarikivi 2009, Junttila 2015). Studies on code-switching have, on the other hand, focused on the majority languages of the Finno-Ugric language group (e.g. Finnish-English CS: Halmari 2005, Estonian-Finnish CS: Praakli 2014), although some studies on minority languages do exist (e.g. Karelian-Russian CS: Sarhimaa 1999, Erzya-Russian CS: Janurik 2017). More recently, borrowing and code-switching in Finno-Ugric minority languages have been studied from the point of view of multilingualism (e.g. Kovács & Janurik 2018). In addition, CCF has been utilized in Finno-Ugric contexts (e.g. Verschik 2008, 2014, Kask 2019, 2021, Verschik & Kask 2019, Tavi 2018, Tavi & Tavi 2019, 2021).

The aim of this study was to investigate all contact-induced phenomena of lexicon together by using the CCF in order to provide an overview of the outcome of bilingualism in the lexicon of the Tver Karelian code. The research questions were the following:

1) Which lexical units are copied from Russian, and which are originally Karelian in the data?
2) In addition to code-copying, what kind of code alternation appears in the data?
3) Are contact-induced features common for the community or are they rather individual?

The article is structured as follows: Section 2 presents the CCF and its relation to the data of the study. Section 3 presents the data and the methods of the study. The analyses include statistical tests, which are presented in Section 4, and qualitative analysis according to the CCF, which is presented in Section 5. Finally, Section 6 presents the results and discusses their theoretical implications.

2. Code-copying and code alternation

Many studies have argued that borrowing and codeswitching are situated at different diachronic ends of the same phenomenon (e.g. Kovács & Janurik 2018: 38–40). For example, one-word codeswitches may later conventionalize into loanwords (for an extensive overview of theories of borrowing and other contact-induced changes, see e.g. Verschik 2008: 49–58; for more information about code-switching in the context of borrowing and language change, see Backus 2005, 2010). In the context of Karelian-Russian bilingualism, some Russian lexical units are adapted to Karelian without any morphophonological clues. For example, Karelian has Russian particles such as *a* 'well', *i* 'and, well', *da* 'and', *hot* ~ *hos* 'although', for which alternatives of Karelian origin do not exist. These particles are part of an old, conventionalized vocabulary in Karelian (Tavi & Tavi 2019: 406–408). Traditionally, morpho-phonological adaption is seen as a sign of conventionalization, i.e., distinguishing between a loanword and a codeswitch. In this article, it is demonstrated that adaption is not a sign of conventionalization by defining Russian global copies as nativized or non-nativized referring to their morpho-phonological adaption.

Many studies have suggested that it may not even be necessary to distinguish between borrowing and codeswitching (e.g. Frick, Grünthal & Praakli 2018: 168). The Code-Copying Framework (Johanson 1993, 1998, 1999, 2002a, 2002b; in a Finno-Ugric context Verschik 2008) even discards the terms borrowing and codeswitching, replacing them with the term "copy". Within this framework, both synchronic (emergent) and diachronic (conventionalized) changes can be studied (Verschik 2008: 54–55). Thus, the traditional terms borrowing and code-switching are discarded in this study and the terms of CCF are utilized to describe contact-induced phenomena of lexicon.

According to the CCF, language varieties in contact are called codes. The model code is the language from which the elements are copied (here Russian), and the basic code is the language into which the copies are accommodated (here Karelian). Within the framework, there are three types of copies: global, selective, and mixed (for further descriptions of the types of copies, see Johanson 1998: 327–331, 2002a: 291–292, Verschik 2008: 54, 61–62). Global copying means accommodation of the foreign blocks into the frame of basic code. Global copies include

all properties, i.e., material, semantic, combinational and frequential properties of the copied item inserted into the basic code. They may be simple, i.e., morphologically simple word stems such as the particles mentioned above, or derived or complex words, such as compounds or several words that together form an expression, i.e., multiword units. An example of this is *dačnik-susieda* 'a cottage-neighbour, a neighbour that uses his/her house only as a cottage' (Stuanu1). In this study, the division of simple and complex global copies is extended to include mixed copies that are lexical and, as lexical units, similar to lexical global copies. The extension is made to cover all data in the statistical analyses.

Selective copying refers to copying onto units of the basic code. Selective copying includes copying only some of the properties of the foreign blocks. For example, changes in word order are selectively copied from the Russian model into Karelian (see Tavi & Tavi 2022: 695 and its references). Selective copying concerns often grammar in Karelian context and is thus not dealt here. Mixed copying, in turn, combines the processes of global and selective copying. For example, Karelian prowords include mixed copies, as the Russian negative particle *ni* is copied globally to original words such as *midä* 'what' and *kedä* 'who', forming the negative prowords *nimidä* 'nothing' and *nikedä* 'no-one'. The meanings are also copied from Russian: *nečego* 'nothing', *nekogo* 'no-one'. These types of pronouns are old, conventionalized copies in Karelian.

In CCF, the non-monolingual discourse is called code interaction. Code interaction can be further divided into code-copying, which is always insertional as described above, and code alternation (CA), which is switching the code during the discourse, i.e., a sequence spoken according to the model code that emerges during the conversation (Johanson 1998: 39) as in Example 1.[1] Furthermore, CA can be divided into intraclausal and extraclausal sequences. Intraclausal alternation is phrase-like, representing sequences often without predication

1 The examples of the data are transcribed roughly according to FU transcription. In the transcription of Russian sequences, the model of the standard Russian has been utilised as the dialect spoken in Tver Oblast is Central Russian dialect on which the standard Russian is based. The examples are glossed according to the Leipzig glossing rules. For instance, sequences spoken in other codes than the basic code are capitalized. In addition, bolding has been used to refer to a certain part of the example if needed for the analysis.

(Example 1), whereas extraclausal alternation is sentence-like, representing-sequences often with predication (Example 2).

(1) *Pertti* *on* *miu-la* *šuuri* *I* *NOV-YJ*
home be.3SG I-ADESS big and new-M
NOV-YIJ *DOM.* *BOL'Š-OJ* *NOV-YIJ* *DOM.*
new-M house big-M new-M house
I have a big house and a new, new house. A big, new house. [Miiksinä1.]

(2) *EXCUSE* *ME.* *Myö* *randaze-s* *kača-mma.*
excuse me we door step-ELA look-1PL
IZ KRAJOŠK-U *POSMOTR-IM.*
from border-ACC look-1PL
'Excuse me. We will take a look from the doorstep.' [Maksuatiha1.]

According to Johanson, CA is "a switching between two basic codes yielding a 'mixed discourse'" (1999: 55). Therefore, it is suggested that the language is not mixed and that the copies are not external elements in the language. This view may be useful to guide the ideologies of the researchers and speakers of minority languages towards accepting bilingualism and its resources in the speech instead of the restrictive views of monolingually biased language ideologies.

3. Memory walk, lexical data, and methods

A memory walk is a data collection method used in cultural studies to research especially sensory memory, i.e., remembering the smells and sounds of childhood or youth (see Järviluoma 2016). The method aims to focus on the personal past of the informant rather than on the common narrative's point of view in the past (ibid. 200). The memory walk method can briefly be summarized in the following three points (ibid. 191):

1) Ask informant(s) to choose a personal, meaningful path from the past.
2) Instruct the informant(s) to walk along the chosen path and to talk about the smells and sounds of the past.
3) Record the discussion.

In this paper, the basic idea of this method is copied from cultural studies and accommodated into a linguistic fieldwork. For lexical research on the bilingualism of an endangered language, the memory walk method is, to the author's knowledge, a novel approach. The method is used with the aim of avoiding monolingual bias corresponding to common narratives in cultural studies.

The nine informants of the study were chosen randomly from a group of Karelian speaking volunteers. They were informed that they would be interviewed in Karelian about the life in the villages while walking around. This information was given to them in Russian and in Karelian. The informants were asked to guide the interviewers in the village and to describe it while walking. Also, the informants' personal past and present, as well as the Karelian language, were topics. The interviews were recorded with a GoPro camera which the informant was wearing on a vest. The camera records speech as well as what the informant is looking and pointing at when speaking, which is illustrated in Figure 2, a still image of one of the interviews with an informant presenting old, displayed tools in her home.

Figure 2. A still of one the interview video recordings conducted with the memory walk method.

Data from the nine interviews differ from each other slightly, as the informants were free to choose their own paths. Table 1 illustrates the interview data, the informant and his/her village, district and gender as well as the duration of the interview. Table 1 also illustrates the lexical data, i.e., how many word tokens occur in each interview and how many lexical units and CA sequences appear in the speech of each informant.

The collected video data were transcribed with the ELAN program (ELAN 2019) to text files that were transferred to the Atlas.ti program (Atlas.ti 2020), where the entire 17,915-word data set was coded based on four criteria for nouns and three for other word classes. The first criterion was to code the data into word classes. For example, the word *krolikat* 'rabbits' (Zaluazina1) is a noun. The word class division of this study is based on descriptive Finnish grammar (= ISK 2004), which is possible because Finnish is a close cognate language of Karelian and they share most grammatical features. In addition, Karelian has less detailed grammars compared to ISK, which further supports the use of Finnish grammar.

The second criterion was to code the units as either simple or complex lexical units. In this study, each lexical item is called a lexical unit (LU) and each LU refers to one referent or meaning. Simple units are basic words such as the example word *krolikat* 'rabbits' that was given above, whereas complex units may be compounds such as the noun *dačnik-susieda* 'a cottage-neighbour' or multiword expressions, for example *odinastoi janvarja* 'the eleventh of January' (Klyčevoi1). In addition, CA sequences spoken in Russian were coded according to the sequence type, i.e., sentence- or phrase-like speech (see Examples 1 and 2 in Section 2).

The third criterion is to code the LUs into groups describing contact-relatedness: original Karelian, for example, *lehmä*, 'cow', nativized Russian, i.e., units that show morpho-phonological accommodation to the basic code, such as *krolikat*, or non-nativized Russian LUs, i.e., units that do not show any morpho-phonological adaption but are syntactically or grammatically accommodated into the basic code, such as *mešali* 'messed, mixed' (see Example 8 in Section 5.1). Defining Russian global copies either nativized or non-nativized is used to discuss the degree conventionalization of the lexical items.

Table 1. The interview data.

Informant	Village	District	Gender	Duration (min)	Word tokens	Lexical units	Code alternation sequences
Klyčevoi1	Klyčevoi	Maksuatiha	Woman	10:33	816	764	16
Klyčevoi2	Klyčevoi	Maksuatiha	Woman	14:46	896	835	11
Maksuatiha1	Maksuatiha	Maksuatiha	Woman	13:46	1,079	1,038	3
Miiksinä1	Miiksinä	Lihoslavlja	Woman	31:46	1,354	835	115
Miiksinä2	Miiksinä	Lihoslavlja	Woman	24:39	1,882	1,740	24
Seltsa1	Seltsa	Maksuatiha	Woman	43:01	3,102	2,983	3
Stuanu1	Stuanu	Lihoslavlja	Man	32:59	2,760	2,663	8
Stuanu2	Stuanu	Lihoslavlja	Man	35:22	2,936	2,745	34
Zaluazina1	Zaluazina	Lihoslavlja	Woman	45:53	3,090	3,000	8
9	6	2	2	4:12:45	17,915	16,603	222

The fourth criterion includes only nouns that are also coded according to the semantic group of the word; for example, the nouns *krolikat* and *lehmä* are coded as animal. The semantic groups consist of 14 occurring topics in interviews regarding the environment and people in the Karelian villages. The coding strategies resulted in 97 different codes; for example, the code for *krolikat* is a simple nativized Russian noun animal. From these codes, three are excluded from this paper: the names of the interviewers, stutters and other small sounds during the interview, and one phrase produced in English (*excuse me* in Example 2).

The analysis of the lexical data consists of, firstly, statistical analyses and, secondly, analyses of the statistical findings within the CCF. First, two different statistical analyses were conducted with the data: The Chi Square test on all word classes and Fisher's exact test of independences on the CA sequences. All the statistical tests were conducted, and plots were drawn with the R program (R Core Team 2017). The Chi Square test was used to test the independencies between word classes and the groups of contact-relatedness of the LUs. The Fisher's exact test of independencies tests the independencies of the CA types between the speakers. The test aims to conclude whether the CA sequences are common for the code or are individual occurrences. Secondly, the independencies between word classes and the contact-relatedness of the LUs in addition to interspeaker variation of CA were analysed qualitatively within the CCF.

4. Statistical analyses

This Section introduces the results of the statistical analyses of the lexical units (Section 4.1) and the code alternation sequences (Section 4.2).

4.1. The relationship between word classes and contact-related groups

The analysis was constructed as follows: All codes created with Atlas.ti (Atlas.ti 2020) were exported as two contingency tables. First, all speakers' codes referring to simple and complex LUs were united and divided according to the word classes. In the contingency tables,

the word classes are presented in relation to the groups describing the contact-related features of the LUs, i.e., Karelian, nativized Russian and non-nativized Russian groups. The contingency tables with the frequencies of the LUs were imported to the R program (R Core Team 2017). Differences between the speakers were checked by comparing the percentages that each word class formed in the contact categories. No major differences were detected. Therefore, all the speakers' LUs in each category were summed. The resulting data are presented in Figure 3, which is a balloon plot of a contingency table, in which the size of each balloon describes the relative magnitude of the corresponding component. For example, verbs and Karelian having a large dot refers to the fact that most verbs are of Karelian origin, not Russian origin.

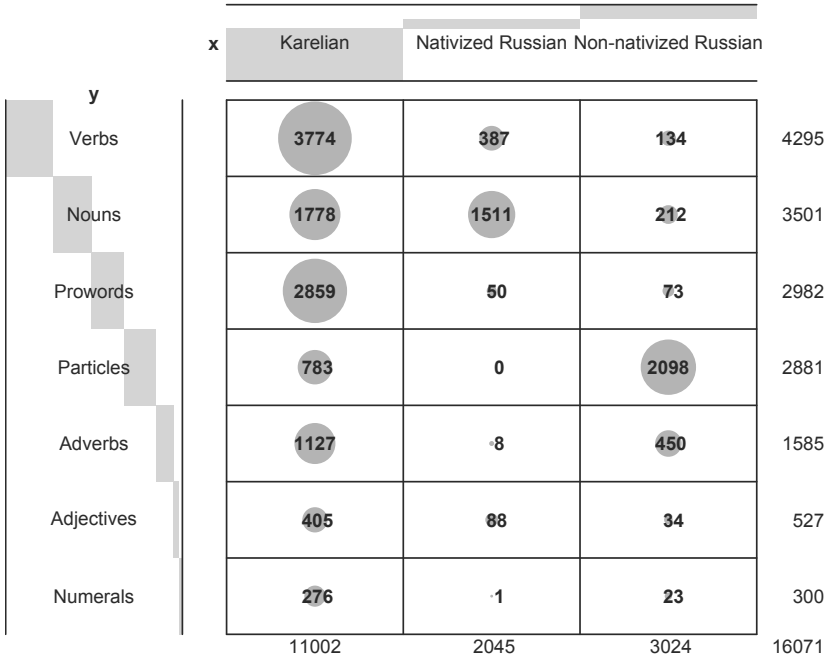

Figure 3. A balloon plot of the frequencies of all lexical units by word classes in the contact-related groups.

The independences of word classes between the three contact-related groups were tested with the Chi Square test of independence. The null hypothesis was that the proportions of the word classes are equal in all

three groups. The test showed that there is a statistically significant difference between some or all of the proportions of the word classes in the contact-related groups (X-squared = 11,003, df = 12, p-value < 0.001). Thus, the null hypothesis was rejected. In order to compare which proportions of word classes in the contact-related groups are statistically significant, the residuals of the Chi Square test were compared. Figure 4 shows the residuals and the degree of correlation between all the cells demonstrated in Figure 3. In Figure 4, a white dot shows a positive and a black dot a negative correlation. The larger the dot, the stronger the correlation. Correlation is expressed in the right side of the figure. 66.83 is the highest positive value of a residual and –26.73 is the lowest value of a residual. Positive values show positive correlation and, correspondingly, negative values show negative correlation. Residual values between 2 and -2 show no correlation between the tested word classes and contact-related groups.

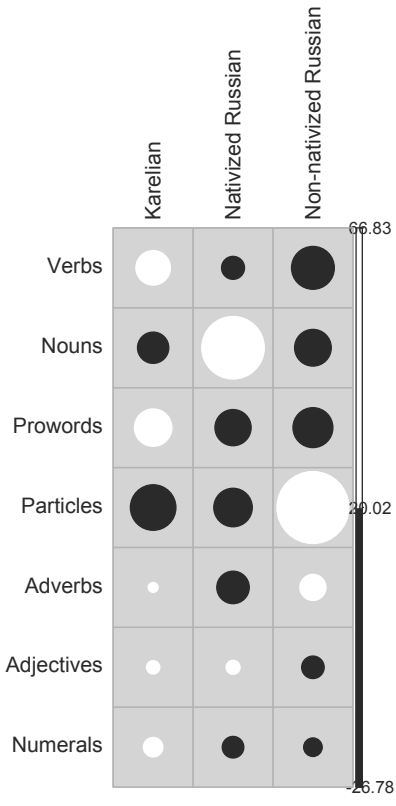

Figure 4. A correlation plot describing the correlation of residual values between the groups tested with Chi Square.

As Figure 4 shows, there is a strong positive correlation between verbs and the Karelian group, between nouns and the nativized Russian group, between prowords and the Karelian group, and between particles and the non-nativized Russian group. Correspondingly, these word classes show strong negative correlation in other contact-related categories. The results indicate that the word classes are dependent on the contact-related groups, which indicates that Tver Karelian has coherence with what is copied from Russian, rather than arbitrary copying. In other words, the verbs are likely to be original Karelian. The word class of nouns is more open to Russian global copies that are nativized or accommodated into the basic code. Some prowords have developed via mixed copying, which is an indication of the high intensity of the contacts between the two languages. Tver Karelian particles, in turn, are likely to be non-nativized Russian global copies, which is due to old bilingualism as particles often are copied in bilingual speech from the pragmatically dominant code (e.g. Matras 2000). These results are further analysed in Section 5.

4.2. Code alternation sequences and interspeaker variation

The model code's sentence- and phrase-like sequences are not incorporated morphologically or syntactically into the basic code. In the Code-Copying Framework these sequences are classified as code alternation. Code alternation also appears in the speech of Tver Karelians. All codes created with Atlas.ti marking sentence- or phrase-like CA sequences of each speaker were exported as a contingency table to the R program. Figure 5 illustrates the absolute frequencies of the two CA types as used by the speakers.

Due to the low frequencies of the CA sequences for each speaker, Fisher's Exact Test of independence (for Count Data hybrid in the R program) was applied to test the relationship between the speakers and the two types of CA sequences. The null hypothesis was that the speakers' proportions of CA were equal between the two types. The p-value of the test was statistically significant (exp=5, perc=80, Emin=1, $p < 0.001$). Thus, the null hypothesis was rejected. To see which pairs of the speakers and CA types have statistically significant differences, Fisher's multicomparison was conducted as a post hoc test (R program, RVAideMemoire package, p.method=Bonferroni). The multicomparison test

gives separate p-values for all speaker comparisons. Speaker Miiksinä1's proportions of the two types of CA differ statistically from every speaker except Maksuatiha1 (p-value < 0.05). Furthermore, speaker Maksuatiha1's proportions of the two types of CA differ statistically from speakers Klyčevoi1 and Stuanu2 (p-value < 0.05).

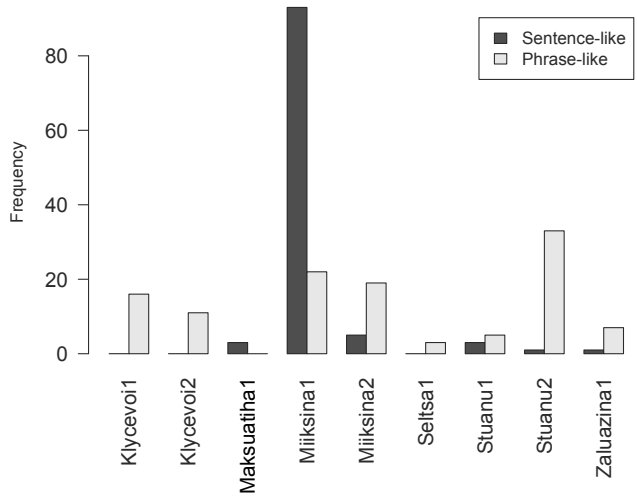

Figure 5. A bar plot of code alternation by the speakers.

The speaker Miiksinä1 differs in her use of the two CA types from most of the speakers because she talked to a shopkeeper in Russian during the memory walk and had difficulties speaking in Karelian to the interviewers (see Example 3). The sequences Miiksinä1 used are sentence-like. The speakers often differentiated these code alternation sequences from other parts of speech. An example of using Russian sentence-like sequences is from the end of one memory walk: the interviewers thanked the informant who in turn commented on her use of the Karelian language (Example 3).

(3) *IZVINJA-JU-S'* *NE* *OČEN'* *HOROŠO* *KONEČNO*
apologize-1SG-REFL not very well indeed
VLADE-JU *JAZYK-OM.* *Nu* *maltt-ua* *malta-n,* *da.*
master-1SG language-INSTR well understand-INF can-1SG yes.
A *pagiš-ša* *ka* *ošibka-t,* *ošibka-t.* *Tak.*
but speak-INF well mistake-PL mistake-PL well
'I apologize, I do not speak the language very well. Well, I understand, yes. But to speak, well, mistakes, mistakes. Well.' [Miiksinä1.]

These kinds of code alternation sequences can be defined as meta sequences which reflect the fact that Russian is a pragmatically stronger code for the bilingual informant. The speaker Maksuatiha1 is the only one in addition to Miiksinä1 who used sentence-like CA more than phrase-like CA. Maksuatiha1 works in education in a Karelian context and may therefore be more conscious of using CA in her speech, as she uses both types very rarely in contrast to Miiksinä1. Otherwise, phrase-like CA is more common in the data.

Studies on codeswitching in Uralic languages report interspeaker variation in the appearance of the codeswitches (Kovács & Janurik 2018: 23). Contingently, when studying CA, interspeaker variation occurs less than when studying CS. This is probably due to most insertional codeswitches being classified as global copies. There seems to be some sort of convention for example on using numerals and their modificands as CA. In the data, variation arises from meta speech and side conversations with outsiders during the interview.

5. The analysis of contact-related lexical units and code alternation in Tver Karelian within the CCF

This Section discusses the contact phenomena of the lexicon on the basis of the results of the statistical analyses presented in Section 4. As is already mentioned, the calculations of the word classes include simple and complex lexical units as global copies, and they are divided in such a manner in the Code-Copying Framework (see Section 2). The contact phenomena in word classes, i.e., global copying, are presented in Sections 5.1–5.5 and, in addition, some cases of mixed copies are discussed in Section 5.3. Interspeaker variation did not exist between the contact-related groups of word classes which indicates that the memory walk method of data collection provided quite a cohesive data set from the nine speakers. On the contrary, interspeaker variation exists to some extent in the types of code alternation. The code alternation consists of sentence- and phrase-like Russian sequences that are not accommodated into the Karelian code. There are, however, similarities in the contents and functions of the CA sequences and the interspeaker variation concerned mostly two of the speakers. Consequently, cohesiveness is also expected to appear among CA. CA sequences are discussed in Section 5.6.

5.1. Verbs

According to the data, the most common word class of the LUs is verbs. Nominals have more occurring types of words, but they are used less frequently in the data than the common verbs such as Karelian's copula and negation. Due to negation being a form of the verb (Example 4) and Karelian having a copula in the present tense (Example 5), unlike Russian, this word class has mostly original Karelian LUs.

(4) *Tiijä-tte* *šuarnas-ta?* ***Et-te*** *tii-jä* *šuarna-.*
know-2PL story-PART **not-2PL** know-CNG story
Et-te *tii-jä,* *da.*
not-2PL know-CNG yes
'Do you know the story? You don't know the story-. You don't know, ok.' [Maksuatiha1.]

(5) *Tämä* ***on*** *šoppi,* *da.*
this **be.3SG** kitchen yes.
'This is the kitchen, yes.' [Klyčevoi1.]

In addition, the informants use nativized and non-nativized Russian verbs. For example, nativized Russian verbs often describe actions that originate from Russian society. The verbs are related to, for example, building (Example 6) and administration (Example 7). In Example 7, the nativized Russian verb describes abstract movement and has a non-nativized Russian profession as a supplement.

(6) *Tämä* *on* *kaik* *ka* *kolhoza-s* ***stroit-tu.***
this be.3SG all well kolkhoz-INESS **build-PASS.PST.PTCP**
'Well all this is built in the Kolkhoz.' [Stuanu2.]

(7) *A* *jälgehen* *öö* *ol-dih* *vybora-t* *vybora-t*
well after mm be-PASS.PST.PL election-PL election-PL
i ***uij-i-n*** *glavn-* *NA GLAV-U.*
and **leave-PST-1SG** head- to head-ACC
'Well, after mm there were the elections, and I became the head.' [Seltsa1.]

Non-nativized Russian verbs mostly occur alone in speech and are syntactically and semantically adapted but in a Russian form (Example 8). The verb in Example 8 could be defined as an insertional codeswitch in traditional terms, but here it is not differentiated from other global copies. It does not modify the basic code, i.e., the complement of the verb is in genitive case as if the verb would have been originally Karelian or accommodated into the Karelian inflectional morphology.

(8) *Ei* *ole* ***meša-li*** *miu-n* *ver-da.*
not.3SG be.CNG **mess-PST.PL** I-GEN blood-PART
Mie *ole-n* *čisto* *karielane.*
I be-1SG purely Karelian.
'My blood has not been mixed. I am purely Karelian.' [Stuanu1.]

The verbs seem to have a certain consistency with respect to what are original Karelian and what are nativized Russian: Original Karelian verbs are either verbs having the basic syntactic functions or concrete semantics and nativized Russian verbs have often abstract semantics or they are culture-specific. On the other hand, the non-nativized Russian verbs may have concrete and abstract semantics. The non-nativized Russian verbs, in turn, belong to the bilinguals' repertoire and are syntactically accommodable to the basic code as they do not modify the basic code, rather they seem elements that the basic code can be embedded with. This supports the idea of global copying rather than switching a code mid-sentence. The CCF offers explanations on the accommodations of Russian verbs into the Karelian code. In this case, the division into the contact-related groups reflects mostly the history of the language contacts: nativized Russian verbs reflect the innovations of the more prestigious society, whereas the non-nativized Russian verbs reflect the bilingualism of individuals and their resources in a discourse.

5.2. Nouns

Nouns are the second most common word class according to the occurrence of the LUs of the data. A total of 14 different semantic groups of nouns (see Section 3 for the formation of these groups) occur in the memory walk data. The groups refer to referents that are named

during the memory walks in the village. The groups are constructions (houses, fences, etc.), people (family members, neighbours, etc.), village and yard (names of the villages and their parts, natural objects such as the river found in the village, etc.), edibles (food of the people and animals, plants), utensils (concrete tools and more abstract things such as paperwork), names (people and place names, not including the interviewers' names), animals, abstract (words with meanings such as 'life' or 'work'), date, etc. (words describing date, time, weather and measures), organization and profession (names of these referents), plants (non-edibles), cars (includes for instance tractors), body parts and events (parties, etc.). In Figure 6, a balloon plot illustrates the distribution of the 14 semantic groups of nouns into the contact-related groups.

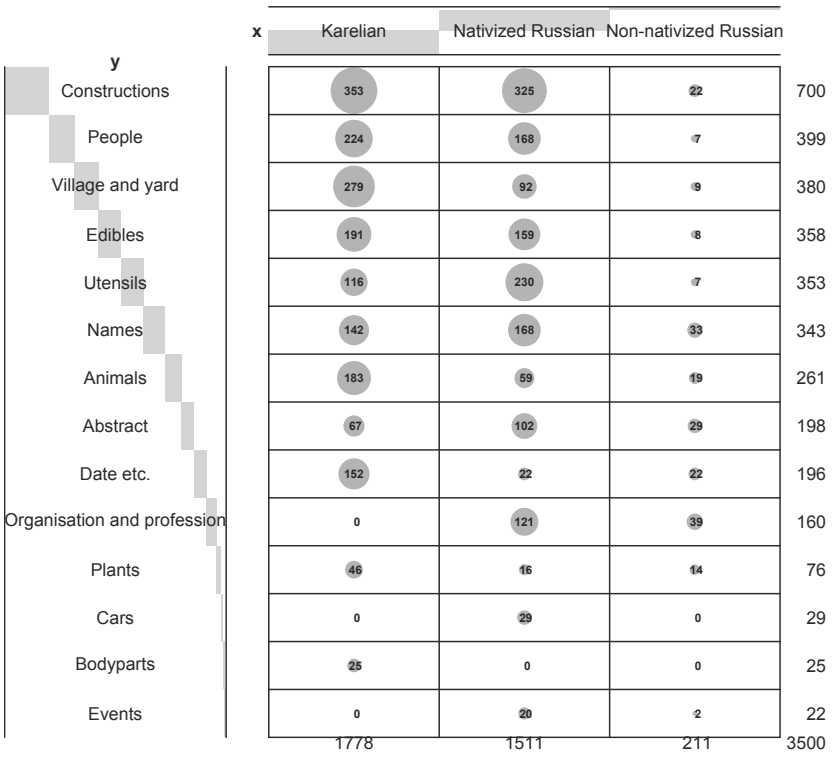

Figure 6. A balloon plot illustrating the distribution of the semantic groups of nouns in relation to the contact-related groups.

As can be seen from Figure 6, in the distribution of the nouns into the contact-related groups, Karelian and nativized Russian are more common groups than non-nativized Russian as the bigger dots suggest higher correlation. This may be because of the environment of the memory walks: the villages are the primary area where the Karelian language is used, which is reflected in the vocabulary used by the informants when describing the environment and the Karelian village, as the environment of the interview may have enhanced the use of the original Karelian words. Only the semantic group of organization and profession has more words in the groups of nativized and non-nativized Russian words than in the Karelian group (see Example 6: *vyborat* 'elections' as a nativized Russian LU and *na glavu* '[become] the head' as non-nativized Russian LU). Pragmatics explains this feature, as Tver Karelians are a part of Russian society but the use of Karelian is limited in professional life. However, recently, the use of Karelian in professional life has increased (Karjalainen et al. 2013: 37, 59, see also Kovaleva 2006).

Words referring to buildings, houses, rooms and other referents constructed by people are almost equally either original Karelian or nativized Russian LUs. For example, the name for a living room *šuuri puoli* (Seltsa1) is an original Karelian compound ('large+side'). The part of a Karelian house where animals were kept, *liävä* ('a barn, a cowshed, a sheepfold', Klyčevoi1) is an old Russian global copy (< *hlev*) that is established in all Karelian dialects and in other Finnic languages such as the Finnish *läävä*. Words meaning home and cottage are the original Karelian *kodi* and *pertti*. All informants use both words repeatedly.

There seems to be some sort of coherence with regard to which referents are expressed using an original Karelian word and which are expressed using the nativized Russian word. Many basic meanings, i.e., simple everyday referents of the Karelian environment have Karelian words. Conversely, nouns with culture-specific, more complex referents are nativized Russian global copies. Often the appearance of nativized Russian lexical units can be explained by the sociohistorical factors of the language contact: the innovations have also been introduced to Karelians via Russian culture. In addition, the culture of Karelians and Russian is largely common which is a result of long and intense contacts (see Sarhimaa 1995).

5.3. Prowords, particles and adverbs

In ISK, all pronouns, question words, etc., with human referents are defined as prowords (ISK 2004: § 713). According to the statistical analysis, prowords have a strong dependence on the original Karelian group. In the data, for example, the first and second plural pronouns in Examples 9 and 10, and the prowords with meanings 'this', 'there' and 'what' in Example 10 are originally Karelian, and in fact common for the Finnish language as well.

(9) *Ka* *täššä* ***mei-jän*** *jogi.* *Jogi.* *Ka,* *on.*
well here **we-GEN** river river well be.3SG
Ozuta-n ***tei-llä*** *joje-n,* *aštu-kkua.*
show-1SG **you-ADESS** river-GEN step-IMP.2PL
'Well here our river, river, well, is. I show you the river, come.' [Stuanu1.]

(10) *Veräjäine* *da.* ***Tämä*** *pikkaraine* ***šielä*** *veräjäine.*
gate yes **this** small **there** gate
Vot, *nu* ***midä*** *vielä* ***tei-le*** *šan-uo.*
well well **what** else **you-ALLAT** say-INF
'A gate yes. This little gate there. Well, well, what else to say to you?' [Zaluazina1.]

Moreover, Russian has been the model for prowords with negation, as Karelian varieties have developed mixed copies with the Russian model and the Russian participle expressing the negation. An example of a mixed proword is provided in Section 2 (*nimidä* 'nothing' < Rus. *nečego* 'nothing'). In this study, prowords that are mixed copies are defined in the division of the contact-relatedness as nativized Russian LUs, because mixing indicates more complex accommodation of the LUs to the basic code and for the purpose of including them in the statistical analyses.

As Figures 3 and 4 (Section 4.1) illustrate, most particles are non-nativized Russian global copies. Examples of these kinds of particles are given in Examples 11 (*nu* 'well', *I* 'and') and 12 (*da* 'yes', *i* 'and'). Only a few particles such as *ka* 'well' (see Example 11) are originally Karelian in the data. The particles in Examples 11 and 12 have the syntactic functions of conjunctions and discourse particles. Especially the function of discourse particles explains why these LUs are copied: in the

speech of bilingual individuals, discourse particles often come from the pragmatically dominant language (Matras 2000: 506–511). In Karelian, the conventionalization of the Russian particles indicates that Karelian-Russian bilingualism is an old, conventionalized phenomenon (see also Tavi & Tavi 2019).

(11) ***Ka*** *käzipaika-t.* *Käzipaika-t.* ***Nu*** ***i*** *sundukka.*
well towel-PL towel-PL **well** **and** chest
'Well, towels. Towels. Well and a chest.' [Seltsa1.]

(12) *Sygyžy,* ***da,*** *sygyžy* *on* *šoma,* ***da.*** ***I,*** ***i***
autumn **yes** autumn be.3SG pretty **yes** **and** **and**
lämmin, ***da.***
warm **yes**
'Autumn, yes, autumn is pretty, yes. And, and warm, yes.' [Miiksinä1.]

Like particles, adverbs are mostly non-nativized Russian global copies. In the statistical analysis, adverbs showed negative dependence on nativized Russian and positive dependence on Karelian and non-nativized Russian, which supports the fact that adverbs behave similarly to particles. Examples 13 and 14 demonstrate the use of the non-nativized Russian adverbs *dal'š'e* 'further' and *nagole* (northern dialects of Russian, see Sarhimaa & Siilin 1994: 273) 'always'. In Example 13, original Karelian adverbs also occur: *jo* 'already' and *ylen* 'very' are typical for all informants. In Example 15, adverbs referring to time are used. The word *čas* 'now' is a non-nativized Russian global copy and *jällest* 'after' is an original Karelian word. An original Karelian word for 'now' *nyt* also exists and was used along with *čas* by most of the informants. The use of these adverbs could also be described as congruent lexicalization (Muysken 1997), as they can be retrieved from either of the codes. This indicates the codes have similar structures at least from the bilingual individual's point of view. The similarities have developed during long and intense contacts between the languages (Sarhimaa 1995).

(13) ***Dal'š'e,*** *käzipaika-t.*
further towel-PL
Nämä *käzipaika-t* *olla-h* ***jo*** ***ylen*** *vahna-t.*
these towel-PL be-3PL **already** **very** old-PL
'Furthermore, towels. These towels are already very old.' [Seltsa1.]

(14) *Nu nämä käzipaika-t tože olla-h* ***vesma*** *šoma-t i*
well these towel-PL also be-3PL **very** pretty-PL and
hei-dä piett-i-h ***nagole*** *i ruškie-š ugla-s*
they-PART keep-PST-3PL **always** well red-INESS corner-INESS
'Well, these towels also are very pretty and they were always kept, well, in the red corner.' [Seltsa1.]

(15) ***Čas*** *läht-öy, mei-jän* ***jälle-st*** *läht-öy.*
now leave-3SG we-GEN **after-ELAT** leave-3SG
'[It] leaves now, [it] leaves after us.' [Klyčevoi2.]

In contrast to nouns and verbs, in the word classes of particles and adverbs, non-nativized Russian lexical units are more common and the words are conventionalized in the Karelian language which is an indication that the morpho-phonological nativization does not itself reveal anything about the status of conventionalization of the LUs. Interestingly, despite the mixed copies, most prowords are originally Karelian and the word class seems not to be as prone to copying as particles and adverbs. These findings are in accordance with the statistical analysis, which showed strong dependencies between prowords and original Karelian and particles and non-nativized Russian. Adverbs showed negative dependency with nativized Russian, which means most adverbs are either original or non-nativized Russian LUs. Thus, as a word class, they in some cases behave similarly to particles that have global copies from Russian used as bilingual resources.

5.4. Adjectives

Within the word class of adjectives, the statistical analysis did not show any strong correlation with the contact-related groups. The observations on the data showed that often the original Karelian words for colours are used, for example when describing the autumn colours of a certain tree (Example 16). Also, basic adjectives meaning young and old, including the comparative forms (Example 17), and small and large are original Karelian LUs. Thus, adjectives with basic meanings are often of Karelian origin.

(16) *Hyö olla-h ka v osen-ju ylen šoma-t.*
they be-3PL well in autumn-ACC very beautiful-PL
Ka mečä-s hiän on ***keldaine*** *sitten* ***ruskie.***
well forest-INESS (s)he be-3SG **yellow** then **red**.
'They [the certain kind of trees] are very beautiful in the autumn. Well, in the forest, it is yellow, then red.' [Klyčevoi2.]

(17) *Kirikkö. da, on* ***vahna.*** *Nu ka on i* ***uuzi.***
church yes be.3SG **old** well so be.3SG and **new**
Kirikkö da. ***Vahne-mbi*** *tämä. a tua* ***nuore-mbi.***
church yes **older-COMPV** this but that **younger-COMPV**
'The church, yes, is old. Well, also there is a new one. A church yes. This one is older and that one younger.' [Miiksinä1.]

Some adjectives are copied from Russian as well. The main nativization strategy of Russian adjectives is use of the adjectives in the form ending with *-oi* (see Example 18). In Russian, *-oi* is in some cases the ending of the masculine form of the nominative case, but it is also a common form in many cases of the conjugation of adjectives. Thus, the reason for the copiability of the form may be frequential.

(18) *Heinä.* ***Iskusstvennoi*** *heinä.*
hay **artificial** hay
'[That is] hay, artificial hay.' [Maksuatiha1.]

(19) *Ka vesi* ***čist-aja*** *mittuine da.*
well water **pure-F** what kind yes
'Well, the water is pure, that kind, yes.' [Miiksinä2.]

Non-nativized Russian adjectives occur in the data more rarely than nativized Russian and Karelian adjectives. In Example 19, the non-nativized Russian adjective *čistaja* is in the nominative singular form. The adjective has the feminine gender because of its referent *vesi* 'water', which is in Russian *vodá* and has the feminine gender. This example illustrates well the bilingual resources of the speakers: even original Karelian words may be declined according to the Russian gender categories.

To conclude, the word class of adjectives seems to consist mostly of Karelian LUs. This may be because many of the adjectives used to

describe the environment are simple or basic according to their semantics, such as colours, size and age. When the semantics of adjectives reflect new innovations such as in Example 18, the adjective is copied from Russian. Traces of the bilingual cognition of the speakers can also be found among adjectives as sometimes even the Russian gender category of a noun may be indicated in the global copy even though the noun is a Karelian word as in Example 19. This is an indication of structural change or selective copying of gender categories at least on the individual level.

5.5. Numerals

Generally, in a bilingual situation of a Finno-Ugric language and Russian (e.g. Kovács & Janurik 2018: 47–49), the first numerals are original and the higher the numeral, the more likely it is to be copied from the pragmatically dominant code, i.e. Russian. This is also true in the case of Tver Karelian. In Example 20, the informant uses Karelian numerals from one to three to describe the number of her grandchildren. The modificands of the numerals are conjugated correctly in partitive singular form.

(20) *Vunuka-t.* *miu-l* *on* *ka* *tyttäre-llä*
grandchild-PL I-ADESS be.3SG well daughter-ADESS
kakši *tyttös-tä* *i* *poijja-la* *tyttöne.*
two girl-PART and son-ADESS girl
Kolm *vunukku-a* *on.* *Da,* *da.*
three grandchild-PART be.3SG yes yes
'Grandchildren. I have, well, [my] daughter has two girls and son has a girl. There are three grandchildren. Yeah, yeah.' [Klyčevoi1.]

Pragmatics explain the use of original words for lower numerals: the lower numerals are basic and in everyday use in the villages, whereas higher numerals and dates are learned in Russian schools and used in mathematics and other domains of life where Karelian language is not used. Thus, the difficulties in expressing original numerals emerge after the number five, depending on the speaker. In Example 21, the hesitations associated with using the correct numeral (six or five) are bolded. Because of the difficulties, many speakers also use simple

Russian numerals, which attract other Russian LUs, and the resulting expression is a phrase-like CA sequence which are analysed in Section 5.6.

(21) *A* *miu-la* *on* *čikko,*
well I-ADESS be.3SG sister
hiän *mladš'e* *milma* *NA ŠEST'* *LET,* *kuin* *tämä*
she younger I.PART prep six year.PL how this
lie-u, ***viizi,*** ***kuuzi,*** ***kuu-*** ***vot,*** ***let,*** *da.*
become-3SG **five** **six** **six-** **well** **year.PL** yes
'Well, I have a sister, she [is] younger than me, by six, how this will be, five, six, six well, years, yes.' [Miiksinä2.]

To conclude, the distribution of the numerals to the contact-related groups depends on whether the LUs are simple or complex, and how high the numerals are on the numerical scale. These factors do not show in the statistical analysis, and as numerals they are rarely used in the data and complex numerals are produced in Russian in phrase-like units as they attract more linguistic units from the model code along with them. Thus, most occasions of complex numerals are classified as CA in this study.

5.6. Code alternation

In the data, two types of CA exist: extra- and intra-clausal CA sequences. The first CA type is sentence-like sequences in Russian spoken within the Karelian discourse. The speakers often differentiate these CA sequences from other parts of speech and the sequences often reflect meta text function, for example speaking to oneself. In Example 22, the informant proposes that the interviewers take a photograph of her and her husband at the end of the interview. Her husband seems to be reluctant, which causes the informant to say to herself in Russian that the host does not want [to be photographed]. These kinds of code alternation sequences can be defined as meta sequences which reflect the fact that Russian is a pragmatically stronger code (see e.g. Matras 2000) for the bilingual informant. Sequences with meta text function are common among code alternation in the data.

(22) *Snimi-kkiä* *minu-u.* *ID-I.* *NE* *HOČE-T* *HOZJAJN.*
film-IMP.2PL I-PART come-IMP.2SG not want-3SG host
'Film me. Come. The host does not want to.' [Zaluazina1.]

The second type of CA is a Russian phrase-like sequence with no single referent produced in the middle of a Karelian sentence or clause. As in the case of the first type, the second one also emerges from the pragmatic needs of the speaker. Many of the CA sequences in this type are complex expressions containing high or complex numerals (see Example 23). The bilingual production of numerals is often described as codeswitching and is explained by pragmatic reasons in the context of Finno-Ugric minority languages spoken in Russia (Kovács & Janurik 2018: 47). The speakers of these languages have received education mostly in Russian, which has led to the fact that many bilinguals have not even learned the high or complex numerals in the minority language. In addition, these kinds of numerals are not common in the daily domains where the minority languages are used. Thus, expressing the date, for example, is often produced in Russian as in Example 23.

(23) *Mie* *ole-n* *šyndy-n* *talve-lla.* *ODINAST-OI*
I be-1SG born-ACT.PST.PTCP winter-ADESS eleven-th
JANVAR-JA *PJAT'DESJAT'* *VOS'M-OM* *GOD-A.*
January-GEN fifty eigh-th year-GEN
Miu-l, *ka,* *lie-u*
I-ADESS well become-3SG
V JANVAR-E *jo* *ŠESTDESJAT' ODIN* *GOD.*
in January-PREP already sixty-one year
'I was born in winter. The eleventh of January in the year fifty-eight. In January, I will, well, already be a sixty-one year-old.' [Klyčevoi1.]

To conclude, CA suggests which of the languages of the informant is pragmatically stronger. As is known also for other Finno-Ugric minority languages, the use of the Russian numerals appears in similar pragmatic conditions. Possibly, these tendencies may even indicate that the CA of numerals is conventionalized or about to be conventionalized as a means of discourse. CA can be seen as discourse mixing and the conventions of that may conventionalize in bilingual speech.

6. Results and conclusion

This study has focused on code-copying and code alternation in the lexicon of Tver Karelian based on nine interviews conducted using the memory walk method. The interview data were classified as simple and complex lexical units and CA sequences. LUs were analysed according to their word class and to their distribution to the contact-related groups, i.e., original Karelian, nativized Russian and non-nativized Russian. The use of CA sequences that are either sentence- or phrase-like was analysed by the speakers.

Two statistical analyses were conducted, and the LUs and CA sequences were analysed within the Code-Copying Framework. As a result, a strong positive correlation was detected between verbs and Karelian LUs, nouns and nativized Russian LUs, prowords and Karelian LUs, and particles and non-nativized Russian LUs. The reason for the correlations between these groups can be found by inspecting the history of Karelian-Russian language contact. As Figure 7 illustrates, original Karelian lexical units are in use for nouns, adjectives and verbs of concrete and basic meanings. This also applies to the lower, simple numerals. In addition, prowords are original Karelian lexical units, with the exception of certain compound prowords developed via mixed copying. Nativized Russian lexical units are very common among nouns and verbs, but their meanings are more complex and culture-specific: these conventionalized global copies reflect the fact that Karelian and Russian have had a long and intense contact. Russian has been the source for new innovations of culture, society, and inventions. Non-nativized Russian LUs form a rather frequently occurring part of lexicon in speech, as many conventionalized particles are non-nativized Russian LUs. This in turn indicates that Karelian-Russian bilingualism is an old phenomenon as the Russian particle system has been adopted into the Karelian code. Furthermore, in the use of LUs, interspeaker variation did not occur.

In the use of the two types of CA, i.e., sentence- and phrase-like sequences, differences between speakers were found, as suggested by previous studies. However, on closer inspection, the frequency of these sequences was rather low and only two speakers differed from others: Miiksinä1 differed from everyone except Maksutiha1 and Maksuatiha1 differed from two other speakers. These two speakers differed from others as Miiksinä1 talked to a shopkeeper in Russian during the

interview and Maksuatiha1 used Karelian very mindfully as she is aware of the linguistic issues concerning Karelian. The remaining speakers did not differ from each other, which indicates that there is very little inter-speaker differences. The types of the CA are common for all speakers: the phrase-like use of high, complex expressions of numerals indicating, for example date and time, and the use of sentences that express meta speech. Phrase-like units expressing complex numerals may conventionalize or have already conventionalized into the bilingual Tver Karelian code as a means of discourse. Thus, in some cases, differentiating CA from CC might be difficult. Sentence-like units, on the other hand, reflect the nature of bilingualism: these meta speech sequences show that Russian is the pragmatically dominant code for many, if not all, speakers. CA is not a part of CCF that goes deep with details, rather, it is a helpful notion to explain the pragmatics of bilingual discourse and CA has been utilized in such a manner in this study as well. CA is also presented as a part of the lexical resources in Tver Karelian in Figure 7.

The goal of the study was to describe the contact-induced language change in the lexicon and the Tver Karelian variety in terms of the CCF. The framework is well compatible with the data because the data can be seen as bilingual and Tver Karelian appears as a high-copying code (Johanson 2002b). The term *high-copying code* has many advantages in the Karelian context as it suggests voluminous copying being typical for the code instead of more ideologically marked terms such as a mixed language. The term *mixed* may be deceptive because copying seems to have certain patterns, i.e., the bilingual resources seem to have common characteristics among speakers and using Russian resources during the discourse is not arbitrary. These theoretical findings indicate that Karelian-Russian bilingualism and the bilingual code itself seem to be conventionalized.

The study has illustrated the explanations of the CCF for spontaneous bilingual speech data of an endangered language variety. The strength of the framework is in accepting a constraintless perspective on contact-induced language change by combining the examination of borrowing and code-switching as global copying. As Figure 7 illustrates, when it comes to the bilingual lexicon, these resources behave similarly: all shorter sequences of Russian origin are global copies that are accommodated into the syntax of the basic code and thus have become part of the code whether they are nativized or non-nativized. They behave

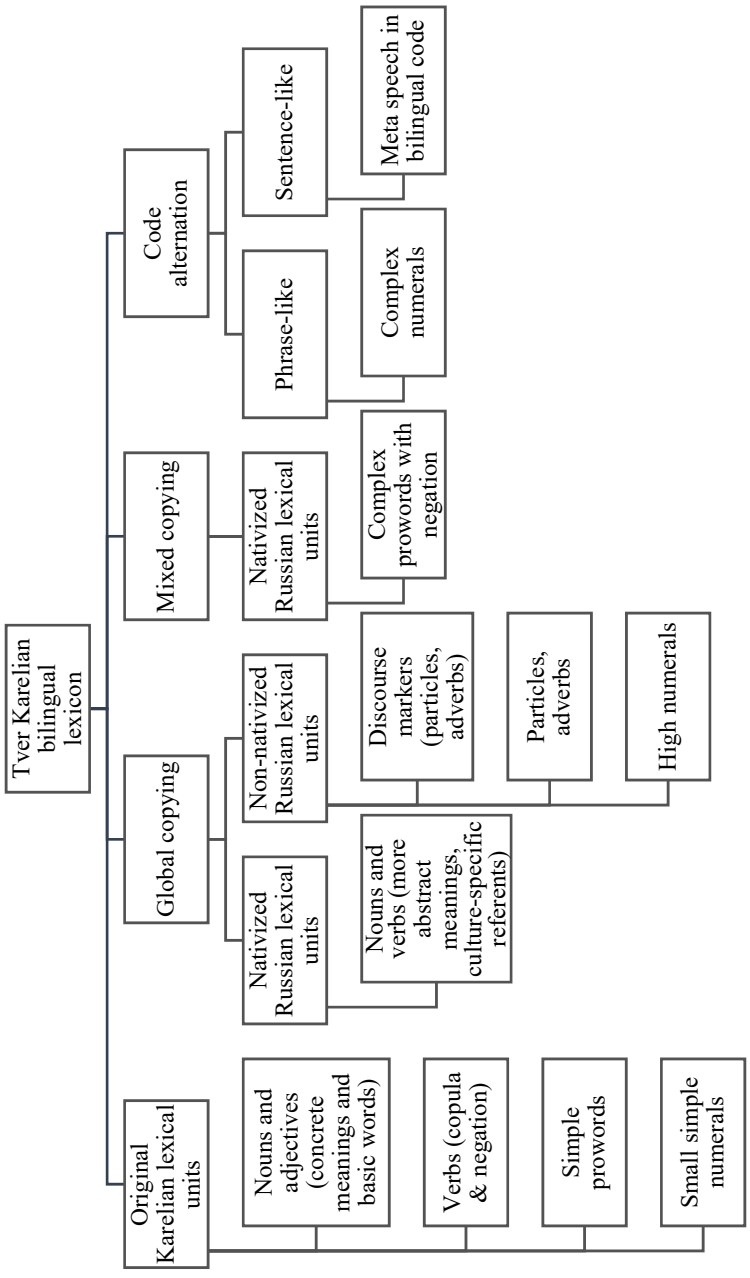

Figure 7. The structure and resources of the lexicon of the Tver Karelian bilingual code.

similarly to Karelian lexical units. Defining Russian global copies either nativized or non-nativized was used to discuss the degree conventionalization of the lexical items: in Karelian bilingual data, Russian LUs may be nativized or non-nativized despite of the degree of conventionalization. This demonstrates the usefulness of the term global copy as it can be used of any copied item regardless of the diachronic aspect of the copies. This, in turn, supports the fact that it is not possible to distinguish between borrowing and code-switching in traditional terms.

Sequences spoken in Russian during the discourse often have a certain function or they may indicate the fact that Russian is the stronger language for the speaker. In these cases, the sequences appear as mixing of the discourse rather than the language, which supports the idea presented in CCF. Concludingly, as is stated in other studies as well, it is not necessary or even possible to differentiate the two phenomena of borrowing and code-switching in the bilingual resources of the Tver Karelian lexicon. Thus, the CCF succeeds in explaining contact-induced language change from the point of view of lexicon. More studies are needed, however, to study the contact-induced language change from a grammatical perspective and to examine selective copying in general in Tver Karelian to describe the variety and its contacts to Russian in more detail.

Acknowledgements

This article is connected to the project Migration and Linguistic Diversification: Karelian in Tver and Finland (SA 314848). The author would like to thank MA Laura Tuomainen for excellent work and cooperation in data collection during fieldwork in 2018. The author would also like to thank the anonymous reviewers for the suggestions that have improved the article.

References

ATLAS.ti. (Version 8) [Computer software]. 2020. Scientific Software Development GmbH. https://atlasti.com/.

Backus, Ad. 2005. Codeswitching and language change. One thing leads to another? *International Journal of Bilingualism* 9(3–4). 307–340. https://doi.org/10.1177/13670069050090030101.

Backus, Ad. 2010. The role of codeswitching, loan translation and interference in the emergence of an immigrant variety of Turkish. *Working papers in corpusbased linguistics and language education* 5. 225–241.

Bubrih, Dmitrij B., Aleksandr A. Beljakov & Aleksandra A. Punžina. 1997. *Karjalan kielen murrekartasto*. Helsinki: Suomalais-Ugrilainen Seura.

ELAN (Version 5.3) [Computer software]. 2019. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan.

Frick, Maria, Riho Grünthal & Kristiina Praakli. 2018. A linguistic encounter between neighbors and relatives: reconnecting Estonian and Finnish in terms of contemporary multilingualism. In Outi Tánczos, Magdolna Kovács, Magdolna & Ulriikka Puura (eds.), *Multilingual practices in Finno-Ugric communities*. Uralica Helsingiensia 13, 163–196. Helsinki: Finno-Ugrian Society.

Grosjean, Francois. 2008. *Studying Bilinguals*. Oxford: Oxford University Press.

Halmari, Helena. 2005. "I'm forgetting both": L1 maintenance and code-switching in Finnish–English language contact. *International Journal of Bilingualism* 9. 397–433. https://doi.org/10.1177/13670069050090030601.

ISK = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Janurik, Boglárka. 2017. *Erzya–Russian bilingual discourse: A structural analysis of intrasentential code-switching patterns*. Szeged: Szegedi Tudományegyetem PhD dissertation.

Johanson, Lars. 1993. Codecopying in immigrant Turkish. In Guus Extra & Ludo Th Verhoeven (eds.), *Immigrant languages in Europe*, 197–221. Clevedon: Multilingual Matters.

Johanson, Lars. 1998. Code-copying in Irano-Turkic. *Language Sciences* 20. 325–337. https://doi.org/10.1016/S0388-0001(98)00007-2.

Johanson, Lars. 1999. Dynamics of code-copying in language encounters. In: Bernt Brendemoen, Elizabeth Lanza & Else Ryen (eds.), *Language encounters across time and space*, 37–62. Oslo: Novus Press.

Johanson, Lars. 2002a. Contactinduced change in a codecopying framework. In Mari C. Jones & Edith Esch (eds.), *Language change. The interplay of internal, external and extralinguistic factor*s, 286–313. Berlin: Mouton de Gruyter.

Johanson, Lars. 2002b. Do languages die of 'structuritis'? On the role of codecopying in language endangerment. *Rivista di Linguistica* 14. 249–270.

Junttila, Santeri. 2015. *Tiedon kumuloituminen ja trendit lainasanatutkimuksessa. Kanta suomen balttilaislainojen tutkimushistoria*. Helsinki: Helsingin yliopisto. Väitöskirja.

Järviluoma, Helmi. 2016. The art and science of sensory memory walking. In Marcel Cobussen, Vincent Meelberg & Barry Truax (eds.), *The Routledge Companion to Sounding Art*, 191–204. Oxford: Routledge.

Kalima, Jalo. 1952. *Slaavilaisperäinen sanastomme. Tutkimus itämerensuomalaisten kielten slaavilaisista lainoista.* Suomalaisen Kirjallisuuden Seuran Toimituksia 234. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kallio, Petri. 2006: On the earliest Slavic loanwords in Finnic. In Juhani Nuorluoto (ed), *Slavicization of the Russian north. Mechanisms and chronology*. Slavica Helsingiensia 27, 154–166. Helsinki: University of Helsinki.

Karjalainen, Heini, Puura, Ulriikka, Grünthal, Riho and Kovaleva, Svetlana. 2013. *Karelian in Russia*. *ELDIA Case-Specific Report.* Mainz: ELDIA.

Kask, Helin. 2019. To agree or not to agree? English adjectives in Estonian-English bilingual blogs and vlogs. *Journal of Estonian and Finno-Ugric Linguistics* 10 (2), 85–123. https://doi.org/10.12697/jeful.2019.10.2.06.

Kask, Helin. 2021. *English-Estonian code-copying in Estonian blogs and vlogs*. PhD dissertation. Tallinn University. Tallinn: TLU Press.

Kehayov, Petar, Denis Kuzmin & Rogier Blokland. 2021. Reflections of Russian dialect geography in Djorža Karelian. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 98, 279–319. https://doi.org/10.33340/susa.91524.

KKS = *Karjalan kielen sanakirja*. (Kotimaisten kielten tutkimuskeskuksen verkkojulkaisuja 18.) Helsinki: Kotimaisten kielten tutkimuskeskus 2009. https://kaino.kotus.fi/cgi-bin/kks/kks_etusivu.cgi.

Kovács, Magdolna & Boglárka Janurik. 2018. Grammatical and sociolinguistic aspects of code-switching in Finno-Ugric languages. In Outi Tánczos, Magdolna Kovács, Magdolna & Ulriikka Puura (eds.), *Multilingual practices in Finno-Ugric communities*. Uralica Helsingiensia 13, 13–76. Helsinki: Finno-Ugrian Society.

Kovaleva, Svetlana Viktorovna. 2006. *Pis'mennaja tradicija i leksikičeskij potencijal jazyka (na materiale karel'skogo jazyka).* Avtoreferat dissertacii na soiskanie učenoj stepeni kandidata filologičeskih nauk. Petrozavodsk: Petrozavod, gos, un-t.

Laakso, Johanna, Anneli Sarhimaa, Anastassia Spiliopoulou Åkermark & Reetta Toivanen. 2016. *Towards openly multilingual policies and practices assessing minority language maintenance across Europe*. Bristol: Multilingual Matters.

Matras, Yaron. 2000. Fusion and the cognitive basis for bilingual discourse markers. *International Journal of Bilingualism* 4, 505–528. https://doi.org/10.1177/13670069000040040701.

Muysken, Pieter. 1997. Codeswitching processes. Alternation, insertion, congruent lexicalization. In Martin Pütz (ed*.), Language choices. Conditions, constraints and consequences*, 361–380. London: Routledge. https://doi.org/10.1075/impact.1.25muy.

Novak, Irina Petrovna. 2014: Fonetičeskie markery tverskih dialektov karelskogo jazyka – Finno-ugorskie jazyki i kultury v sotsiokulturnom landšafte Rossii. Materialy V Vserossijskoi konferentsii finnougrovedov. Petrozavodsk: Karelski naučnyj tsentr RAN. 47–49.

Novak, Irina Petrovna. 2016: Tverskie dialekty karelskogo jazyka: Fonetika. Fonologija. Petrozavodsk: Karelski naučnyj tsentr RAN.

Novak, Irina. 2019. Venäjän kielen vaikutus tverinkarjalan murteiden äännejärjestelmään. *Multilingual Finnic. Language contact and change*. Uralica Helsingiensia 14, 227–248. Helsinki: Helsingin yliopisto. https://doi.org/10.33341/uh.85037.

Oranen, Nicklas. 2019. On the use of the adessive case in Tver Karelian. *Lähivõrdlusi. Lähivertailuja* 29. 204–227. https://doi.org/10.5128/LV29.07.

Praakli, Kristiina. 2014. On the bilingual language use of the Estonian-speakers in Finland. *Sociolinguistic Studies* 8 (3). 383–407. https://doi.org/10.1558/sols.v8i3.14087.

R Core Team. 2017. *R. A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.Rproject.org/.

Saarikivi, Janne. 2009. Itämerensuomalaisslaavilaisten kontaktien tutkimuksen nykytilasta – Jussi Ylikoski (ed.), *The quasquicentennial of the FinnoUgrian Society*. (SuomalaisUgrilaisen Seuran Toimituksia 258), 109–160. Helsinki: Suomalais-Ugrilainen Seura. https://doi.org/10.33340/susa.82021

Sarhimaa, Anneli. 1995. Karjalan kansat ja kielet kontakteissa. Asutushistoriallista taustaa ja lingvistisiä seurauksia. *Virittäjä* 99. 191–223.

Sarhimaa, Anneli. 1999. *Syntactic transfer, contact-induced change, and the evolution of bilingual mixed codes*. Helsinki: The Finnish Literature Society.

Sarhimaa, Anneli & Lea Siilin. 1994. ”Pasportas on kirjutettu mie olen karjalan’e”. *Studia Carelica Humanistica* 5, 267−278. Joensuu: University of Joensuu, Faculty of Humanities.

Tavi, Susanna. 2018. Rajakarjalaismurteiden kielikontaktit venäläiskopioiden taajuuden ja fonologian valossa. *Lähivõrdlusi. Lähivertailuja* 28. 316–356. https://doi.org/10.5128/LV28.10.

Tavi, Susanna & Lauri Tavi. 2019. ’Ja’-, ’tai’- ja ’vaikka’-merkityksisten lekseemien kontaktilähtöinen variaatio rajakarjalaismurteissa. *Virittäjä* 123. 401–429. https://doi.org/10.23982/vir.70346.

Tavi, Susanna & Lauri Tavi. 2021. Vocabulary-Based Classification and Contact-Induced Formation of Neologisms in Two Standard Varieties of Karelian. *Journal of Language Contact* 14. 684–719. https://doi.org/10.1163/19552629-14030006.

Uusitupa, Milla. 2020. On the relationship between subject expression and modal meaning in Tver Karelian necessity and possibility constructions. *SKY Journal of Linguistics* 33. 165–198.

Uusitupa, Milla. 2021. Subjektin merkintä tverinkarjalan pidiä-, šuaha- ja voija-konstruktioissa. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 98. 399–436. https://doi.org/10.33340/susa.95404.

Verschik, Anna. 2008. *Emerging bilingual speech. From monolingualism to code-copying*. London: Continuum International Publishing Company.

Verschik, Anna. 2014. Estonian-Russian Code-Copying in Russian-Language Blogs: Language Change and a New Kind of Linguistic Awareness. In V.-A. Vihman and K. Praakli (eds.), *Negotiating linguistic identity, Nationalism Across the Globe*, 59–87. Oxford: Peter Lang Europäischer Verlag der Wissenschaften.

Verschik, Anna & Kask, Helin. 2019. English-Estonian code-copying: Comparing blogs and vlogs. In Antje Quick & Anna Verschik (eds.), *Applied Linguistics Review*. De Gruyter Mouton. https://doi.org/10.1515/applirev-2019-0028.

Virtaranta, Pertti. 1980. Karjalaisista entisessä Tverin läänissä. Suomalais-ruotsalainen historian tutkijain symposio (Suomalais-ruotsalaisen kulttuurirahaston julkaisuja), 93–112. Helsinki: Helsingin yliopisto.

Kokkuvõte. Susanna Tavi: Karjala-vene kakskeelsuse mõju tverikarjala keele sõnavarale. Käesolevas uurimistöös uuritakse tverikarjala ja vene keele kontakte. See uuring püüab anda tervikliku ülevaate kakskeelse koodi sõnavarast. Metoodika sisaldab kombinatsiooni statistilistest analüüsidest ja kontaktidest põhjustatud muutuste käsitlemisest koodikopeerimise raamistiku (*Code-Copying Framework* = CCF) osas. Üheksa intervjuud üheksa inimesega viidi läbi mälukõnni meetodil. Leiti seoseid erinevate sõnaklasside ja kontaktidega seotuse vahel. Koodivahelduses leiti eri kõnelejate vahel vähe erinevusi ja üheks ühiseks jooneks oli keerukate arvsõnade kasutamine venekeelsete fraasidena, ilma neid tverikarjala koodi sobitamata. Leiud kinnitavad, et koopiad on teatud liiki ja esinevad teatud sõnaklassides. Koodi vaheldumise jadad viitavad sellele, et CCF-i kohaselt on segatud eelkõige diskursus, ja mitte keel. CCF-i leiud avaldavad mõju vähemuskeelte poliitikale, kuna leiud toetavad kakskeelse terminoloogia kasutamist.

Märksõnad: koodikopeerimine, kakskeelsus, leksikon, kontaktist tingitud keelemuutus, karjala keel, vene keel