

IRINA KHOMCHENKOVA (Moscow), POLINA PLESHAK (College Park)

## RUSSIAN NUMERALS IN MOKSHA AND HILL MARI

**Abstract.** This paper deals with the use of Russian numerals in spontaneous speech of Moksha-Russian and Hill Mari-Russian bilinguals. Based on a comparative corpus study of numeral phrases with code-switching in Moksha and Hill Mari, we propose an analysis in terms of the Matrix Language Frame (MLF) model. We discuss the factors that might influence the choice of language: the numerical value, the type of the context, and the syntactic type of the numeral. Although the numerical systems are maintained in each of these Uralic languages, there is a strong tendency to use Russian numerals to express larger quantities. We argue that ordinal numerals are switched more frequently for structural reasons: they do not occupy the same position as cardinal numerals in the noun phrase. We also argue that the formation of embedded language islands is influenced by the types of dependencies established in a construction (quantifier-like cardinal numerals vs. adjective-like ordinal numerals), which supports Muysken's model of categorical equivalence as a condition for code-switching.

**Keywords:** Moksha, Hill Mari, code-switching, numerals, Matrix Language Frame model, categorical equivalence.

### 1. Introduction

In this paper we discuss the status of the occurrences of Russian numerals in the spontaneous speech of Moksha-Russian and Hill Mari-Russian bilinguals. Moksha and Hill Mari both have a long history of contact with Russian (Bereczki 1968; Johanson 2000), which has led to numerous borrowings from Russian and frequent spontaneous switches between the two languages. There are some works on contact phenomena in the related languages Erzya (Janurik 2017; Лемов 2018) and Meadow Mari (Гаврилова 2012; 2013). For Moksha, there is only a brief overview of Moksha-Russian code-switching (Сааринен 2014), and, as far as we know, there are no works on code-switching in Hill Mari. In contrast to previous works, our study is a comparative analysis of code-switching patterns and the structural restrictions on them in two different but closely-related languages, based on the same criteria.

In the spontaneous speech of Moksha-Russian and Hill Mari-Russian bilinguals, some numerals or nouns (or both) in numerical construc-

tions<sup>1</sup> are in Russian. For example, in (1) the first numerical construction is in Hill Mari, while the second one is in Russian.

- (1) *mǎńǎ mar-lan ke-n-äm t ä ž e m ä n d e k š š ü d ä*  
 I man-DAT go-PRET-1SG thousand nine hundred  
*k ä n d ä k š l u - š ä i - n s e m n a d c a t y j m a j - ä n*  
 eight ten-ORD year-GEN seventeen.ORD may-GEN  
 'I got married in 1980, May 17'

The data come from two spoken corpora, which mainly consist of recordings of oral speech transcribed and translated together with consultants. The Moksha corpus was collected during the Lomonosov Moscow State University field trips in 2013–2016 in the villages of Lesnoje Tsibajevo and Lesnoje Ardaševo in the Republic of Mordovia (98 texts / 20103 tokens). It is available upon request. The texts are mostly life stories, interesting events, tales and descriptions of pictures and videos. The Hill Mari corpus was collected during the Lomonosov Moscow State University field trips in 2014–2018 in the village of Kuznetsovo in the Mari El Republic (167 texts / 63522 tokens; <http://hillmari-exp.tilda.ws/en/corpus>). The main genres are the following: stories from everyday life, stories on the history of villages, the description of some procedures (games, recipes, getting to some location), fairy tales, and experiments (descriptions of visual stimuli). The text authors are bilingual, and they speak both the indigenous language (Moksha or Hill Mari) and Russian. The median of speakers' age in both corpora are similar: in the Moksha corpus it is 62 and in the Hill Mari corpus it is 60. Most of the Moksha speakers have secondary/vocational education, some people have incomplete secondary education, and only few people have higher education. Higher education is much more widespread for Hill Mari speakers: half of them have higher education, and half of them have secondary/vocational education. For most people, Moksha or Hill Mari is the main language of their everyday communication. The sociolinguistic situation, as was studied by the research groups during the fieldwork, is much worse in Moksha than in Hill Mari: the language is not being transmitted to children. Most Hill Mari children, in contrast, learn Hill Mari as a native language. However, Hill Mari has been in contact with Russian for a long time, and many younger speakers choose to use Russian more. The sociolinguistic situation of the Mordvin languages (Moksha and Erzya) is described in more detail in Аришкин 1993; Пуссинен 2010, and that of the Mari languages (Hill Mari and Meadow Mari) in Кушлин 2010 and in Шабыков, Кудрявцева 2017.

We annotated all the constructions that include numerals in these two corpora (367 in Moksha and 1409 in Hill Mari). The annotation includes the numerical value of a numeral (large/small), its type (ordinal/cardinal), the type of the constituent (in terms of Myers-Scotton's (1993) Matrix Language Frame model (MLF model)), the language, and the context in which it was uttered.

In our study of Russian numerals in Moksha and Hill Mari speech, we leave aside the problem of distinguishing borrowings from switched fragments, discussed, e.g., in Poplack 1988. Based on the frequency criterion proposed by Myers-Scotton (1992 : 35–36), we would assume that these are instances of

<sup>1</sup>With the term *numerical construction*, we refer both to numeral phrases with cardinal numerals and noun phrases with ordinal numerals.

code-switching rather than borrowing, since indigenous numerals prevail in the texts (33% and 13% Russian numerals among all the numerals in Moksha and Hill Mari, respectively). For comparison, Myers-Scotton (1992 : 36) showed that in Shona (< Bantu) corpus, 86% of all the numbers were in English and not in Shona, which means that English numerals have been borrowed. However, according to Treffers-Daller (1991), no criteria is reliable enough to distinguish between code-switching and borrowing. Since the status of Russian numerals is not crucial in our study, we follow Treffers-Daller (1991) and treat the two phenomena uniformly, instead concentrating on their functional distribution in the discourse and formal restrictions on their distribution.

The current study has two main goals. First, we determine the preferred contexts for Russian numerals in Moksha and Hill Mari speech; and second, we uncover and characterize syntactic restrictions on switching inside a numerical construction. Both sets of results speak in favor of code-switching, not borrowing. Given that numerical constructions with cardinal numerals are structurally different in Finno-Ugric and Russian, whereas noun phrases with ordinal numerals are similar, we hypothesize that a) the ordinals may be switched more readily, and b) there will be no switches between cardinal numerals and head nouns. We show that these hypotheses are borne out. The proposed structural analysis is couched in the MLF model (Myers-Scotton 1993) as well as Muysken's (2000) classification of types of code-switching.

The paper is organized as follows. Section 2 describes the background information: we discuss the syntax of numeral constructions in Russian, Moksha and Hill Mari (2.1), typologically relevant factors that might influence the choice of the language for a numerical construction (2.2), and the two models we used to model code-switching (2.3). Section 3 presents our data: we discuss the aforementioned factors (3.1–3.3) with a special emphasis on the structural restrictions on the switched fragments (3.3), and interactions among the different factors (3.4). Section 4 is the conclusion.

## **2. Background**

### **2.1. Numerical constructions in Russian, Moksha and Hill Mari**

Understanding the syntactic structure of numerical constructions with different types of numerals is crucial for our analysis of the structural restrictions on code-switching (see section 3.3). In this section, we offer a short description of the general syntactic patterns of numerical constructions in the three languages involved in our investigation.

#### **2.1.1. Cardinal numerals**

Constructions with cardinal numerals in Moksha and Hill Mari differ from their Russian counterparts in terms of case and number morphology. In this section, we address each of these issues separately. The structure of Russian numerical constructions with cardinal numerals was broadly discussed in the formal literature (Мельчук 1985; Babby 1987; Corbett 1993; Pesetsky 2013, among others). In Russian, cardinal numerals from two to four (and numerals containing them) require a special form of the noun if the entire numerical construction is in subject position (2a). The same case is received by inani-

mate nouns in numerical constructions in direct object (2b) position. The special form is nearly identical to the singular genitive form of the noun. However, due to differences in stress in some nouns, it was treated as a distinct paucal form (Зализняк 2002; Ionin, Matushansky 2018). In the same positions, larger numerals require a plural genitive form (2c–2d). Notice also that, in addition to case differences, we also see number differences: numberless (or paucal, singular-like) with "small" numerals, and plural with "large" numerals.

- (2) a. *т р и к о т - а и д - у т*  
 three.NOM cat-SG.GEN/PAUC go-NPST.3PL  
 'Three cats are walking'
- b. *В а с я с - р у б и - л т р и д е р е в - а*  
 Vasja.NOM PV-cut-PST.M.SG three tree-SG.GEN/PAUC  
 'Vasja cut three trees'
- c. *п я т ь к о т - о в и д - у т*  
 five.NOM cat-PL.GEN go-NPST.3PL  
 'Five cats are walking'
- d. *В а с я с - р у б и - л п я т ь д е р е в ь - е в*  
 Vasja.NOM PV-cut-PST.M.SG five tree-PL.GEN  
 'Vasja cut five trees'

Simple noun phrases (i.e., ones that are not numerically modified) in subject and object positions would surface as plural nominative and accusative respectively (3a–3b).

- (3) a. *к о т - ы и д - у т*  
 cat-PL.NOM go-NPST.PL  
 'Cats are walking'
- b. *В а с я р у б и - л д е р е в ь - я*  
 Vasja.NOM cut-PST.M.SG tree-PL.ACC  
 'Vasja cut trees'

In numerical constructions with animate nouns in object positions (4a), (4c) and numerical constructions in oblique positions (4b), (4d), the case on the noun is the one that is required by the external syntax (i.e., the position that the noun phrase occupies relative to other elements in the clause), and it is uniform with both "small" (4a), (4b) and "large" numerals (4c), (4d). The cardinal numeral always receives the case which is required by the syntactic position of the entire numerical construction. The resulting pattern in (4), unlike the one in (2) looks like a case concord of cardinal numeral and head noun. The number is always plural.

- (4) a. *К а т я п о - г л а д и - л - а т р е х к о т - о в*  
 Kate.NOM PV-pat-PST-F.SG three.ACC cat-PL.ACC  
 'Kate patted three cats'
- b. *К а т я д а - л - а т р е м к о т - а м м о л о к - а*  
 Kate.NOM give-PST-F.SG three.DAT cat-PL.DAT milk-SG.GEN  
 'Kate gave some milk to three cats'
- c. *К а т я п о - г л а д и - л - а п я т ь к о т - о в*  
 Kate.NOM PV-pat-PST-F.SG five.ACC cat-PL.ACC  
 'Kate patted five cats'

- d. *Катя да-л-а пя-т-и кот-а м молоко-а*  
 Kate.NOM give-PST-F.SG five-DAT cat-PL.DAT milk-SG.GEN  
 'Kate gave some milk to five cats'

In these cases nouns in simple noun phrases would have the same form as they do in numerical constructions (5a–5b).

- (5) a. *Катя по-глади-л-а кот-о в*  
 Kate.NOM PV-pat-PST-F.SG cat-PL.ACC  
 'Kate patted the cats'  
 b. *Катя да-л-а кот-а м молоко-а*  
 Kate.NOM give-PST-F.SG cat-PL.DAT milk-SG.GEN  
 'Kate gave some milk to the cats'

In Moksha and Hill Mari, the case on the head noun in numerical constructions with cardinal numerals (numerals themselves do not bear case markers) is always the one that is required in the context where the numeral is not present (i.e., the one required by the external syntax), and there is no difference across the syntactic positions of numerical constructions, see Hill Mari examples (6)–(8). Cardinal numerals are never case-marked in presence of a noun, so they neither assign genitive case (6a), nor do they exhibit case-concord with the noun.

- (6) a. *k ɔ̂ m k o t' i / \*k o t' i - n a š k e d - ä t*  
 three cat cat-GEN walk-NPST.3PL  
 'Three cats are walking'  
 b. *k a t a k ɔ̂ m k o t' i - m n i ä l t - e n*  
 Kate three cat-ACC pat-PRET  
 'Kate patted three cats'  
 c. *k a t a k ɔ̂ m k o t' i - l ä n š ä š e r - ä m p u - e n*  
 Kate three cat-DAT milk-ACC give-PRET  
 'Kate gave some milk to (three) cats'

Due to number restrictions (see discussion below), in Hill Mari, the nouns in noun phrases appear in their singular form (compare to (7) with singular noun phrases without numerals). In plural contexts without numerals, the case marking would be the same, but they would bear a plural marker (8).

- (7) a. *k o t' i a š k e d - e š*  
 cat go-NPST.3SG  
 'A cat is walking'  
 b. *k a t a k o t' i - m n i ä l t - e n*  
 Kate cat-ACC pat-PRET  
 'Kate patted a cat'  
 c. *k a t a k o t' i - l ä n š ä š e r - ä m p u - e n*  
 Kate cat-DAT milk-ACC give-PRET  
 'Kate gave some milk to a cat'
- (8) a. *k o t' i - v l ä a š k e d - ä t*  
 cat-PL go-NPST.3PL  
 'Cats are walking'  
 b. *k a t a k o t' i - v l ä - m n i ä l t - e n*  
 Kate cat-PL-ACC pat-PRET  
 'Kate patted cats'

- c. *katá k o t' i - v i l ä - l ä n šäšer-äm pu-en*  
 Kate cat-PL-DAT milk-ACC give-PRET  
 'Kate gave some milk to cats'

Compare also Moksha examples with numerals (9) and without numerals (10). As in Hill Mari, the case marking in numerical constructions is the same as it would be for a simple noun phrase in the same position (but see the discussion about number below).

- (9) a. *k o l m ə k a t ə - t / \*katə-t' / \*katə-t'ńə-ń jaka-J-t'*  
 three cat-PL cat-DEF.SG.GEN cat-DEF.PL-GEN walk-NPST.3-PL  
 'Three cats are walking'
- b. *kat'é sud'ə'é-žəń k o l m ə k a t ə - t'ń ə - ń*  
 Kate caress-PST.S.3SG.O.3PL three cat-DEF.PL-GEN  
 'Kate patted the three cats'
- c. *kat'é maksə-žə lof'ć-t' k o l m ə k a t ə - t'ń ə - ń d' i*  
 Kate give-PST.S.3SG.O.3 milk-DEF.SG.GEN three cat-DEF.PL-DAT  
 'Kate gave the milk to the three cats'
- (10) a. *k a t ə - t jaka-J-t'*  
 cat-PL walk-NPST.3-PL  
 'Cats are walking'
- b. *kat'é sud'ə'é-žəń k a t ə - t'ń ə - ń*  
 Kate caress-PST.S.3SG.O.3PL cat-DEF.PL-GEN  
 'Kate patted the cats'
- c. *kat'é maksə-žə lof'ć-t' k a t ə - t'ń ə - ń d' i*  
 Kate give-PST.S.3SG.O.3 milk-DEF.SG.GEN cat-DEF.PL-DAT  
 'Kate gave the milk to the cats'

In Hill Mari, as reported in (Сидорова 2018a), in numerical constructions plural marking is determined primarily by the presence of dependents between the numeral and the noun, by the number of such dependents, and by their type. She claims that the more the linear distance between a numeral and a noun is, the more acceptable plural marking is. In Moksha, "small" (10 and smaller) and "large" (larger than 10) numerals behave differently with respect to plural marking. Small numerals require a plural marker on the head: *kaftə ćora-t* [two boy-PL] / *\*kaftə ćora* [two boy[SG]] 'two boys', whereas large numerals prohibit it in the absence of other nominal dependents: *šisgeməń ćora* [seventy boy[SG]] / *šisgeməń ćora-t* [seventy boy-PL] 'seventy boys'. Large numerals are also sensitive to the definite/indefinite and possessive/nonpossessive marking on the head, as well as to the number and type of nominal dependents (Сидорова 2018b), which is very similar to the Hill Mari system.

We would like to emphasize that both Moksha and Russian have a distinction between small and large numerals with respect to number marking, but they split the numeral scale differently: in Moksha, "small" numerals are below 10, and in Russian — below 4 (and those that contain numerals below 4, e.g. 54). Another crucial fact is that it is a matter of lively debate which element of the numerical construction in Russian is the head (Мельчук 1985; Babby 1987). For example, Pesetsky (1982) and Bowers (1982) claim that it is the cardinal numeral, not the noun, that heads the Russian numerical construction. Following this approach, Ionin and Matushansky (2018 : 180) treat CNs as heads which take NP complements as their sisters, and number-gender agreement between

the two results in case assignment on the noun. In Finno-Ugric numerical constructions, however, the noun is the head (Плешак 2017). In a situation of code-switching, this can create a conflict in the process of determination of the matrix language of the entire noun phrase (see section 2.2, on the theoretical model being used here). The fact that different elements are heads in Russian and Finno-Ugric is relevant, as we assume that syntactic constraints on code-switching reveal syntactic properties of the languages in question.

### 2.1.2. Ordinal numerals

Ordinal numbers are adjective-like nominal modifiers in all the three languages. In Russian, ordinal numerals exhibit concord with noun in case, gender and number (11a), as adjectives do (11b).

- (11) a. *третья*            *девочка* / *третьих*            *девочек*  
 three.ORD.F.NOM girl.NOM    three.ORD.PL.ACC girl.PL.ACC  
 'the third girl / [I see] the third girls'
- b. *красивая*        *девочка* / *красивых*        *девочек*  
 beautiful        girl.NOM    beautiful.PL.ACC girl.PL.ACC  
 'beautiful girl / [I see] beautiful girls'

In Moksha and Hill Mari, ordinal numerals are derived from cardinal numerals using a special affix. The distribution of ordinal numerals is also adjective-like. However, in contrast to Russian, adnominal modifiers in Moksha (12) and Hill Mari (13) show no case concord.

- (12) a. *kolmǎ-ćǎ* *štǎr-ś*        / *kolmǎ-ćǎ* *štǎR-ńǎ-ń*  
 three-ORD girl-DEF.SG    three-ORD girl-DEF.PL-GEN  
 'the third girl / [I see] the third girls'
- b. *mazi* *štǎr-ś*        / *mazi* *štǎR-ńǎ-ń*  
 nice girl-DEF.SG    nice girl-DEF.PL-GEN  
 'a nice girl / [I see] nice girls'
- (13) a. *kǎm-šǎ* *ǎdǎrǎš* / *kǎm-šǎ* *ǎdǎrǎs-vlǎ-m*  
 three-ORD girl        three-ORD girl-PL-ACC  
 'the third girl / [I see] the third girls'
- b. *cever* *ǎdǎrǎš* / *cever* *ǎdǎrǎs-vlǎ-m*  
 nice girl        nice girl-PL-ACC  
 'a nice girl / [I see] nice girls'

Morphosyntactically, ordinal numerals in Moksha and Hill Mari pattern with adjectives. They do not affect the number or the case marking on the noun and show no concord themselves. Therefore, we treat them as nominal modifiers that are integrated into the syntactic structure as other modifiers would be. Thus, the structures of numerical constructions with cardinal numerals are very different in Russian than they are in Finno-Ugric languages. Russian and Finno-Ugric languages exhibit a high degree of syntactic interaction between the cardinal numeral and the noun, but in different ways: Russian cardinal numerals influence both number and case marking of the noun, and Moksha and Hill Mari numerals influence only number marking. Noun phrases with ordinal numerals can be considered as structurally similar in Russian and in the two Finno-Ugric languages, with ordinal numerals showing a higher

degree of structural independence. This plays an important role in code-switching in the light of the principle of congruence (see Section 2.3.2).

## **2.2. Typologically relevant factors**

Matras (2007) showed that the borrowability of numerals in the languages of the world can be affected by such factors as numerical value (e.g. over 10 > below 10), context (more formal contexts: dates, addresses, transactions involving money, etc. > less formal contexts) and numeral type (ordinals and cardinals have different hierarchies of numerical value: "higher numerals 1000, 100 > above 20 > above 10 > above 5 > below 5" for cardinals and "lower ordinals > higher ordinals" for ordinals (Matras 2007 : 51–52). Given that we adopt a uniform approach to borrowing and code-switching, these hierarchies are also applicable to code-switching. This is attested in the use of Russian numerals in many Finno-Ugric languages, despite the presence and preservation of their own numeric system.

First, the *n u m e r i c a l v a l u e* of a numeral can influence the choice of the language. For example, in Komi-Permyak speech, Russian numerals are usually used to refer to larger numerals (10+) (Максимов 2017 : 12), as well as in Votic speech (Turunen 1997 : 218).

Second, the *c o n t e x t* may also play a part. For instance, this factor is relevant for the Northern dialects of Udmurt. Russian numerals may be used in some special contexts, such as expressing time, date, age, monetary value, or building-numbering (Максимов 2017 : 12–13). The use of Russian numerals in Mordvin languages (Moksha and Erzya) was already discussed as well. Saari-nen (Сааринен 2014 : 542) points out that some Moksha speakers use mostly Moksha for numerals, and only years get switched to Russian. According to Janurik (2017 : 117), in Erzya typical contexts for Russian numerals are temporal expressions, school grades, class numbers, distances (e.g. kilometers), and money. In section 3.2, we compare these findings to our own data.

The combination of both of these factors is attested in Karelian speech. Russian numerals are used with large (complex) numerals (Sarhimaa 1999 : 234) and while referring to time, age, date of birth, school grades, class numbers, and money (Pyöli 1996 : 295). In Kildin Saami, which, according to Auer (1999) is a merged Kildin-Russian variety, there are similar restrictions: Russian numerals are used with 7+ numerals, especially in contexts of the date of birth and age (Пинеда 2009 : 31–34).

As already noted by Gumperz and Hernandez-Chavez (1971), there are also particular *g r a m m a t i c a l c o n s t r a i n t s* on code-switching. For example, code-switching is possible between a subject and a predicate, but not between an auxiliary and a main verb. Given that *d i f f e r e n t t y p e s o f n u m e r a l s* occupy different structural positions, this can be relevant in our discussion of code-switching with numerals. This is addressed in section 3.3.

## **2.3. Theoretical models of code switching**

In order to explain the structural restrictions on code-switching that we find in our data, we will take into account the types of constituents in the MLF model proposed by Myers-Scotton (1993). Based on this, Muysken devel-



oped his model of constraints on code-switching and possible conditions where they can be avoided, as well as a classification of types of code-switching (Muysken 2000). These models have become the most influential ones, having more explanatory adequacy than previous proposals (Poplack 1980; 1981; Joshi 1985; Belazi, Rubin, Toribio 1994, among others). Recently, the Generative approach proposed by MacSwan (1999) has been claimed to have more accurate predictions (MacSwan 2005). Due to the unavailability of acceptability judgements for our materials, which are crucial for the Generative approach (Wyngaerd 2017), we chose the MLF model, which operates with corpus data.

### 2.3.1. Matrix Language Frame model (Myers-Scotton 1993)

The Matrix Language Frame (MLF) model is based on the distinction between a more activated matrix language (ML) and an embedded language (EL). A bilingual sentence can consist of a) ML islands, which contain only ML morphemes and are well-formed according to the grammar of ML; b) EL islands, which have only EL morphemes and are well-formed according to the grammar of EL; and c) mixed ML + EL constituents, which have both ML and EL morphemes.

The central hypothesis of the MLF model is that the matrix language forms the morphosyntactic frame of the clause. This hypothesis is realized by the following two principles (Myers-Scotton, 1993 : 83): the *Morpheme-Order Principle*, according to which the morpheme order in ML + EL constituents should be that of ML, and the *System Morpheme Principle*, according to which all system morphemes in ML + EL constituents, which have grammatical relations external to their head, should be that of ML.

MLF models operates with the terms *system* and *content morphemes*.<sup>2</sup> System morphemes (e.g. articles, inflections) are morphemes that do not express any semantic or pragmatic aspects of meaning as opposed to content morphemes (e.g. nominal, verbal, and adjectival roots). System morphemes express the relation between content morphemes, while content morphemes assign or receive thematic roles. The ML provides both content and system morphemes, while the EL mostly provides content morphemes. The only possibility for the EL system morpheme to appear in the ML is to build an EL island.

The example which includes ML + EL constituents is provided in (14). Two English words are inserted into the Swahili frame. First, the word *certificate* and its modifiers follow Swahili word order. Second, the verb *depend* agrees with the subject using Swahili morpheme *i-*.

- (14)  $\emptyset$ -saa hi-yo i-na-depend na  $\emptyset$ -certificate  
 c.9-time dem-c.9 c.9-NONPAST-depend with c.10-certificate  
 z-ako z-a  $\emptyset$ -shule  
 c.10-your c.10-ASSOC c.10-school  
 'At this time, it depends on your school certificates' (Myers-Scotton 2004 : 108)

<sup>2</sup> In the later version of MLF-model (4-M model) there is a more fine-grained distinction between system morpheme types: early system morpheme, late bridge system morpheme, late outsider system morpheme, see (Myers-Scotton, Jake 2000). However, since both numerals and nouns are content morphemes, the specification of system morphemes is irrelevant in the discussion of code-switching in numerical constructions.

Example (15) illustrates an EL island (French) which occurs in an Arabic morpho-syntactic frame and which meets well-formedness condition: the object is a well-formed French noun phrase.

- (15) *yə-t-haka-w wahed les histoires*  
 3.M-P-tell-PL INDEF DEF stories  
 'They tell each other some [fantastic] stories' (Boumans, Caubet 2000 : 152; cit. from Myers-Scotton 2002 : 116)

EL islands may appear in those circumstances when the structures of the two languages are incongruent, and switching between morphemes in accordance with the principles above is impossible. This is discussed in more detail in section 2.3.2.

### 2.3.2. Muysken's (2000) approach

In his research, Muysken takes as a base the idea that there can be no switches between the elements if one of them is a lexical item that selects the other (Shaffer 1978). This prohibition can be formalized in terms of the government model, where one of the elements governs the other (DiSciullo, Muysken, Singh 1986):

- (16) \*[ Xp Yq ], where X governs Y, and p and q are language indices  
 (Muysken 2000 : 21)

This model predicts that there can be no switches between such configurations as a verb and its complement or an adposition and its complement unless there is a neutralizing element (e.g. an article, as functional elements are exceptions to the government constraint). In (17), we see a Spanish verb which selects a determiner phrase. If such a determiner phrase is formed by a Spanish article, the structure is grammatical (17a). If the article is in English, the Spanish verb does not recognize it as a determiner phrase that it selects. This element is called *language index carrier* and is defined as "the highest (non-lexical) node in a tree" (DiSciullo, Muysken, Singh 1986 : 4).

- (17) a. *ve-o la-s house-s*  
 see-PRES.1SG DEF.F-PL house-PL  
 'I see the houses'.  
 b. *\*ve-o the house-s*  
 see-PRES.1SG DEF house-PL  
 'I see the houses' (Muysken 2000 : 22)

Nevertheless, there are violations of this government constraint, which are due to congruence (Sebba 1998). In other words, if the categories of a governing head and its complement are equivalent in the two languages respectively, such configurations are not subject to the government constraint. Pandit (1990 : 43) formulates that as follows: "Code-switching must not violate the grammar of the head of the maximal projection within which it takes place". This principle is also shared by the Generative approach (McSwan 2005). Based on these assumptions, Muysken (2000 : 30–31) proposes a violable constraint against mixed strings of type "A B", where A and B are elements in different languages. There are three strategies that allow this constraint to be violated: insertion, alternation and congruent lexicalization. Here, we concentrate on the first one,

where categorical equivalence makes it possible to insert elements that have at least one of the following properties: a) they have the same status in the two languages, b) they are morphologically encapsulated, c) they are shielded off by a functional element from the matrix language, or d) they could belong to either language (Muysken 2000 : 31). We focus our discussion on this phenomenon, because noun phrases (both numerical constructions and non-numerical constructions) are claimed to be a prototypical case of insertion (Muysken 2000 : 62). Consider the following properties which support the view of these switches as instances of insertion:

1. they form a single constituent (*viz.* noun and its modifiers) (18);
2. they occur in nested structures, so that the elements that surround them are in the ML (18);
3. they are content words (nouns, adjectives) rather than functional particles/ words (18);
4. they are complements, not adjuncts (18); and
5. they can undergo morphological integration (19).

Moksha (18) and (19):

(18) *i ul'-ś jota-ft-f ba l ma s ka pa d klup-sə*  
 and be-PST.3SG conduct-CAUS-PTCP.PASS ball masquerade club-IN  
 'And there was a masqued ball in the club'

(19) *soń-ć-əń o dno komnаты й квартира-с*  
 he-EMPH-GEN one.room.ADJ flat-3SG.POSS.SG  
 'He has a one-room flat'

Muysken (2000) points out that insertions in his classification "correspond to mixed ML + EL constituents" in the MLF model, and "alternations to EL islands combined with ML islands" (Muysken 2000 : 17). In this study, we consider noun phrases. Some of them we treat as ML + EL constituents, but some others we consider to be EL islands inserted into ML sentences. The cases of alternation that are exemplified in our case by noun phrases that serve as temporal adverbial adjuncts are also considered to be EL islands (see examples and discussion in section 3.3). Within these models, the constraints on code-switching inside a noun phrase can shed light on the internal structure of different types of noun phrases.

### 3. Data and analysis

In this section, we consider different factors that can influence the choice of language in code-switching scenarios.<sup>3</sup> See, e.g., the following examples (Hill Mari (20), Moksha (21)):

(20) *näl-lä i juže li-eš, vot, sem' d e s j a t p j a t y ŷ*  
 four-ten year already be-NPST.3SG PTCL seventy five.ORD  
*z o d, v t o r o ŷ o k t j a b p r ь, vot, texeňä-vlä*  
 year second october PTCL such-PL  
 'It will be forty years, well, the year seventy five, October second, well, so'

<sup>3</sup> Our examples are only of the type "Russian numeral + Russian head". Constructions like "Russian numeral + Moksha / Hill Mari head" were not attested (see more about structural types of code-switching in 3.3.2).

- (21) *Федя за-кончил восемь классов,*  
 Fedja.NOM PV-finish-PST.M.SG eight grade-PL.GEN  
*mol'-i d'e v a t a j - s*  
 go-NPST.3SG ninth-ILL  
 'Fedja finished eight grades, he goes to the ninth'

The interaction of the factors is not trivial and varies across text collections. First, we discuss every factor separately (3.1–3.3), then, we show how they interact and provide decision trees that model a probabilistic algorithm of language choice (3.4).

### 3.1. Numerical values

According to Matras (2007 : 50), there is a hierarchy of numerical values of numerals, where the leftmost members are more readily borrowed or switched: "over 10 > below 10". Our corpus data show that there is indeed an asymmetry between these two groups (we call them "large" and "small" respectively). In both language communities the hierarchy "large > small" holds (see Figure 1), and the difference is statistically significant (the Fisher exact test p-value is < 0.00001).

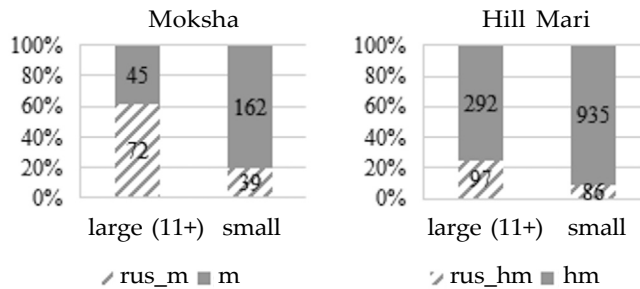


Figure 1. Numerical value dependence. m – Moksha, hm – Hill Mari, rus\_m, rus\_hm – Russian.

However, there is also a difference between the two collections. In Hill Mari there are more native numerals than Russian ones among both large and small numerals, even though the number of switched large numerals is significantly larger than the number of switched small numerals. In Moksha, on the other hand, the percentage of switched numerals within the group of large numerals is much higher than the percentage of non-switched ones.

### 3.2. Semantic context

All occurrences of numerals in the corpora were annotated for the type of context, see Table 1. The specific tags were chosen on empirical grounds.

Table 1

Context types		Context types	
Type	Example	Type	Example
year	1973, 19th century (20)	measure	<i>x grams, x liters</i> (24)
date	July, 21st (20)	time	60 years, first day, for two days (25)
education	5th grade; number of lessons (21)	age	6 years old (26)
number	N <sup>o</sup> , phone number (22)	other	
money	<i>x rubles</i> (23)		

Number (Hill Mari):

- (22) *mämnä-m* часть-*ǰš* *kand-evä*, часть *man-alt-eš* *сорок*  
 we-ACC part-ILL bring-AOR.3PL part say-DETR-NPST.3SG forty  
*пять сто пять*  
 five hundred five  
 'We were brought to the military unit, the unit's name was 45105'

Money (Moksha):

- (23) *триста* *рубл-е ü-də* *бодяга* *sev-at*  
 three.hundred ruble-PL.GEN-ABL alcohol.drink take-NPST.2SG  
 'You take alcohol drink for three hundred rubles'

Measure (Hill Mari):

- (24) *kužǰ* *pandǰ-m* *ške-žǰ* *näl-ät*, *ну* *iktä*  
 long stick-ACC REFL-POSS.3SG take-NPST.2SG PTCL INDEF  
*сантиметр-ов* *пятьдесят-шестьдесят*  
 centimeter-PL.GEN fifty-sixty  
 'You take the long stick, well, around fifty or sixty centimeters long'

Time (Moksha):

- (25) *i* *rabota-ń* *двадцать* *два* *год-а*  
 and work-PST.1SG twenty two year-SG.GEN  
 'And I've been working for twenty two years'

Age (Hill Mari):

- (26) *сорок* *пять* *лет*, *nällä vǰc i* *totá-n*  
 forty five year.PL.GEN forty five year grandfather-GEN  
 'The grandfather is forty five years old'

We calculated the percentage of Russian numerals in each context. The hierarchies of contexts for Moksha and Hill Mari are in (27a) and (27b) respectively.

- (27) a. number (100%) > money (90%) > year (77%) > date (38%) > time (18%)  
 b. number (75%) > date (59%) > year (40%) > money (26%) > education (25%)

This is very close to what is reported by Janurik (2017) for Erzya. The variation between Moksha and Hill Mari may be due to a difference in text collections, as the texts are not balanced with respect to genres and topics. But the general pattern is clear enough. There is a strong tendency to use Russian numerals in contexts of "pure" numbers and in conversations about exact dates and years in general. Money is also a frequent context for Russian, as well as education.

### 3.3. Structural restrictions on code-switching

#### 3.3.1. Cardinal numerals vs. ordinal numerals

In contrast to cardinal numerals, ordinal numerals tend to behave more like adjectives: they are closer to the (head) noun, and do not influence its number or its case, as already shown in section 2.1. As mentioned in section 2.1, we consider ordinal numerals to exhibit a lower degree of interaction with the noun, which facilitates the process of switching. For example, ordinal numerals do not require any specific case or number features from the noun. Although



constituents was made based on the following criterion: if the combination of two Russian words in Moksha / Hill Mari speech is grammatical in Russian, then it is the EL island; otherwise, it is the ML + EL constituent (see below the more detailed description of EL islands and ML + EL constituents). The results are presented in Figure 3. Numbers for ordinal numerals and cardinal numerals were calculated separately. ML(r) are Russian numerical constructions (NumC) that occur in Russian sentences, where the ML is Russian.

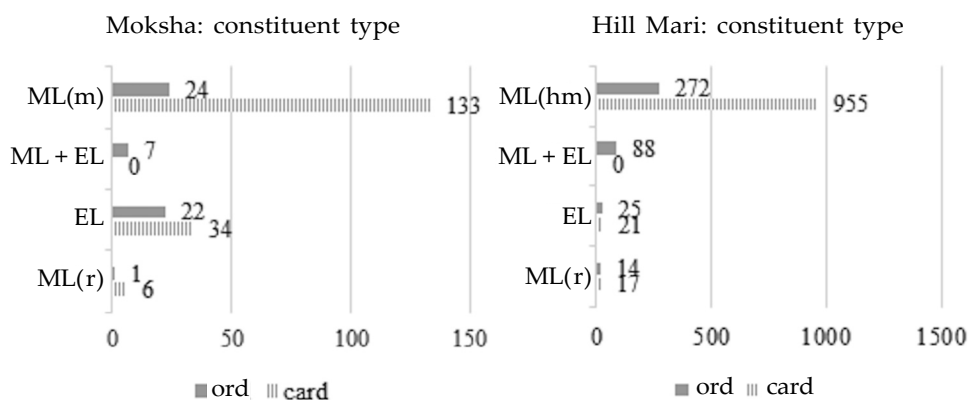


Figure 3. Constituent types. ML(hm) / ML(m) / ML(r) – Hill Mari / Moksha / Russian is ML of the clause and the language of the NumC, ML + EL – mixed constituents, EL – Russian is the language of the NumC but not ML of the clause.

In the charts we can see that despite a general prevalence of numerical constructions with cardinal numerals in the corpora, ML + EL constituents are attested only with ordinal numerals. They all represent the combination (28d), where both the numeral and the noun are switched. No instances of combination (28b) or (28c), where only the noun or only the numeral is switched, were attested. This perfectly fits Muysken’s theory, as it is adjuncts that are switched more readily (in comparison to complements). Another tendency is that although there are more ordinal numerals in code-switching contexts, EL islands in Moksha mostly contain cardinal numerals, in contrast to ML + EL constituents, which contain ordinal numerals exclusively. We argue that as cardinal numerals have different syntactic behavior in Finno-Ugric and Russian, there can be no switches between them and the rest of the numerical construction (mainly noun). Therefore, they only can occur in EL islands, whose internal structure is quite strict. In contrast to cardinal numerals, ordinal numerals do not suffer from this mismatch between their syntax in Russian and in Finno-Ugric, and so they may occur in ML + EL constituents.

Consider now the structure of EL islands. A syntactic pattern adhering to Russian is obligatory throughout the entire constituent (Hill Mari (29) and Moksha (30)). Cardinal numerals influence the case of the noun; therefore, we can conclude that the noun receives case inside the EL island. As shown in (30), the genitive case on the noun cannot be omitted, as it is required by the Russian cardinal numeral. The entire numerical construction receives the case from the ML, so there can be additional morphological markers of ML on top of the EL, which help to integrate the inserted EL into the ML sentence structure.

- (29) *stípendij-žä*                    *д в е н а д ц а т ь р у б л - е ю žl-ä*                    *в м е с я ц*  
 scholarship-POSS.3SG                    twelve                    ruble-GEN.PL                    be-AOR.3SG                    in month  
 'The scholarship was twelve rubles per month'
- (30) *téjə-st*                    *é'áv-ə'l-ø*                    *sa-m-s*                    *п я т ь ч а с - о в - с ə /*  
 PRON.DAT-3PL.POSS                    need-PQP-SG                    come-INF-ILL                    five                    hour-GEN.PL-IN  
 \**п я т ь ч а с - с ə*  
 five                    hour-IN  
 'They had to come at 5' (example courtesy Maria Kholodilova)

In contrast to cardinal numerals, ordinal numerals receive case together with the head noun within the entire determiner phrase (see subsection 2.1.2). Ordinal numerals themselves do not influence the case on the noun. In (31) (Moksha), we see a Russian numerical construction in genitive case, and the ordinal numeral shows concord in gender, number and case with the noun, as required by the grammar of Russian.

- (31) *al'e-žə*                    *д е в я т ь с о т ч е т в е р т - о г о г о д - а - l<sup>4</sup>*  
 father-1SG.POSS.SG                    nine.hundred                    four-ORD-M.SG.GEN                    year-SG.GEN-PQP  
 'My father was from the year nine hundred five (was born)'

Now, consider the structure of ML + EL constituents. These are mainly constructions with ordinal numerals. In these constituents, Russian words are inserted into the ML frame. Therefore, even though it is taken from Russian, the ordinal numeral does not have to show concord in gender, number and case with the head noun, and we see a bare adjective-like form (Hill Mari (32) and Moksha (33)).<sup>5</sup>

- (32) *п е р в ы ю smenä-m=ät*                    *nänge-ä*  
 first                    shift-ACC=ADD                    lead-NPST.3SG  
 'And he is leading the first shift'
- (33) *т р и д ц а т ь п я т а ю šekcije-š*  
 thirty                    five.ORD                    section-DEF.SG  
 '(Do you hear or not,) the thirty fifth section'

Auer and Muhamedova (2006) analyze Kazakh (Turkic) sentences with Russian insertions analogous to (32) and (33) as instances of EL islands, and not as two single word insertions, since "there is a relationship of dependency between the two words" (Auer, Muhamedova 2006 : 44). They argue that the matrix language (in this case having no gender system) can have an impact on the embedded language (in this case having a gender system). On this view, even though neither *pervâj smenä* 'first shift' in (32) nor *tricat' pataj šekcije* 'thirty fifth section' in (33) are grammatical in Russian due to gender mismatches, they are still EL islands. We do not accept this analysis, at least not for the data under discussion here. We argue that these are ML + EL constituents, since all members of the constituent are taken from the EL in their bare form (or in the "default" masculine gender), and yet no island is formed: the internal syntax is that of the ML.

<sup>4</sup> Nominal predicates in Moksha take verbal inflection, so the PQP marker on the predicative noun phrase in (31) is natural in Moksha (Холодилова 2018).

<sup>5</sup> The reviewer pointed out that the absence of agreement means that these numerals are rather borrowings than switched fragments. However, we would like to emphasize once more that there are no good tests that could unambiguously discriminate between these two phenomena, so we treat them uniformly, see the discussion in Section 1.



As we see, ordinal numerals are switched more readily than cardinal numerals. The latter are more in conflict with nouns as loci of morphological encoding of case: in Russian, they are analyzed as heads of numerical constructions, and in Finno-Ugric languages the head is the noun. Therefore, it would be unclear what the language of the numerical construction is, if a cardinal numeral and a noun were taken from different languages. This results in cardinal numerals occurring only in EL islands. Ordinal numerals, on the other hand, can occur either in EL islands or in ML + EL constituents.

### 3.4. Interaction of the factors

A question arises regarding to what extent the aforementioned factors are independent of one another. Larger numerals can be less frequent in colloquial speech and more frequent in those topics that are mostly uttered in Russian, which is tightly connected to factor of context. For example, years normally have large numerical values.

#### 3.4.1. Numerical value and context

We looked at the interdependence between context and numerical value, as shown in Table 2. In both languages the context "date" is balanced with respect to numerical values of numerals. The context "age" in Moksha and the context "measure" in Hill Mari also have equal proportions of small and large numerals. Year and money are inclined to have large numerals in both languages, while education and time tend to consist of small numerals.

Table 2

The percentage of small and large numerals in different contexts

a)									
Moksha	date	year	education	money	age	measure	time	other	Sparkline
small	43%	13%	72%	19%	50%	29%	78%	83%	
large	57%	87%	28%	81%	50%	71%	22%	17%	
b)									
Hill Mari	date	year	education	money	age	measure	time	other	Sparkline
small	51%	8%	90%	34%	17%	56%	86%	13%	
large	49%	92%	10%	66%	83%	44%	14%	87%	

Therefore, while some contexts are independent from the numerical value, others show a high correlation with larger or smaller numbers.

#### 3.4.2. Context and type of a numeral

The types of numerals are also sensitive to context. For example, years are expressed with ordinals, not cardinals. We counted cardinal numerals and ordinal numerals for each context separately, and our results are in Table 3. Money and education contexts are less uniform and may contain both small and large numbers. The variability of different types of numerals in such a context as education, for example, can be quite high. At the same time, it is quite logical that money and education are among those contexts where the numerals are switched more frequently: at school, in college, in shops, because of the domi-

nance of Russian in formal contexts. The figures show that year, date and numeration contexts are indeed those where mostly ordinal numerals are used.

Table 3

**Context dependence**  
(ord – ordinal numeral, card – cardinal numeral)

	date	year	educaton	number	money	age	measure	time	other
a)									
Moksha	6	21	4	5	0	0	0	1	2
(ord/card)	/0	/1	/9	/0	/19	/8	/7	/8	/15
Moksha	7	6	2	0	0	0	0	8	21
(ord/card)	/1	/1	/4	/0	/2	/13	/7	/35	/100
b)									
Hill Mari	41	42	18	1	0	0	0	0	23
(ord/card)	/0	/0	/2	/5	/15	/7	/12	/3	/14
Hill Mari	29	63	43	2	0	0	0	19	116
(ord/card)	/1	/0	/17	/0	/43	/38	/112	/153	/592

The interaction of context and type of numeral is highlighted in Table 4. We see a prevalence of ordinal numerals in date and year, which are switched most frequently. Age and measure, which show fewer switches, are expressed with numerical constructions with cardinal numerals, not ordinal numerals. Money, numerals and time show less interaction with the type of numeral. Finally, education is a context for more switches in Hill Mari and fewer switches in Moksha. As we have seen, it is exactly this context where we find more ordinal numerals in Hill Mari and more cardinal numerals in Moksha.

Table 4

**The percentage of ordinals and cardinals in different contexts**

	date	year	educaton	number	money	age	measure	time	other	Sparkline
a)										
Moksha										
ord	92%	93%	30%	100%	0%	0%	0%	24%	17%	
card	8%	7%	70%	0%	100%	100%	100%	76%	83%	---
b)										
Hill Mari										
ord	100%	100%	76%	38%	0%	0%	0%	11%	19%	
card	0%	0%	24%	63%	100%	100%	100%	89%	81%	---

**3.4.3. Numerical value and type of numeral**

The numeral type and the numerical value are statistically significant factors taken independently (see Section 3.1 and 3.3.1). It is important to check whether this significance is preserved when they are taken together. Using the ctree() function in the party package in R (Hothorn, Hornik, Zeileis 2006), we plotted the trees based on numerical value and numeral type, see Figure 4.

Figure 4 shows that in Moksha speech, small numerals are usually in Moksha, while large numerals are usually in Russian. Among these large numerals, ordinal numerals are more likely to be in Russian than cardinal numerals. In Hill Mari speech, cardinals are more often in Hill Mari than ordinals. In both cases, large numerals are more likely to be in Russian. The

two languages differ in their preferences: numerical value is crucial for Moksha, while for Hill Mari the type of the numeral is more important.

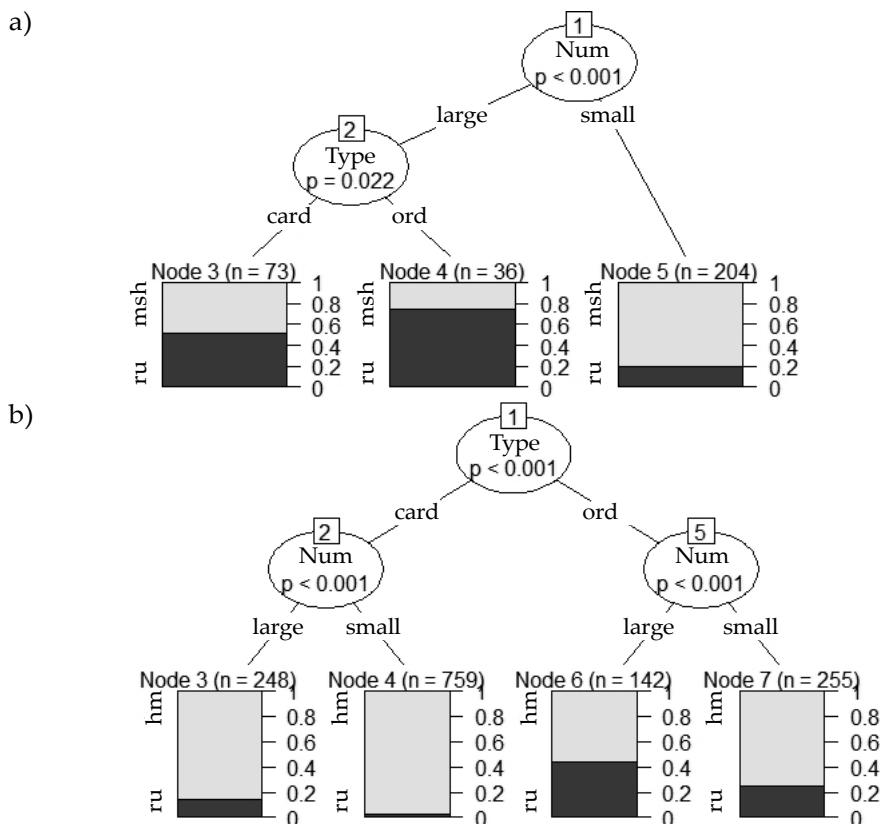


Figure 4. Decision trees: the dependence of language on numeral type and numerical value. msh – Moksha, hm – Hill Mari, ru – Russian, card – cardinal numeral, ord – ordinal numeral, Num – numerical value, p – p-value, n – number of numerical phrases.

We also checked whether the result will be different if we take into account the numerical value without the distinction into "large" and "small" numerals, see Figure 5.

Figure 5 shows that the hierarchy of the factors did not change. However, there are some differences. For Moksha the picture becomes a bit more complicated: apart from the first two choice-points that did not change (the first is basically the distinction between "small" and "large" numerals, and the second one is the distinction between ordinals and cardinals), there is the third choice-point, which shows that cardinals further depend on the numerical value: if they are more than 91, then they are almost always in Russian, if they are less than 91 or equal to it, then they are more likely to be in Moksha, but still there are about 40% that they can be in Russian.

On the contrary, for Hill Mari the picture becomes simpler: Russian is the most probable with ordinals (but still they are often in Hill Mari, as can be seen from the tree). Cardinals are usually in Hill Mari, but large cardinals (14+) are more likely to be in Russian than small cardinals.

Thus, both numerical value and numeral type are important for both languages. However, the hierarchies of factors are different. Basically, this tree

is the reverse version of the Moksha tree in Figure 5a. The hierarchy of the factors in Moksha is numerical value > type of numeral, while in Hill Mari it is the reverse.

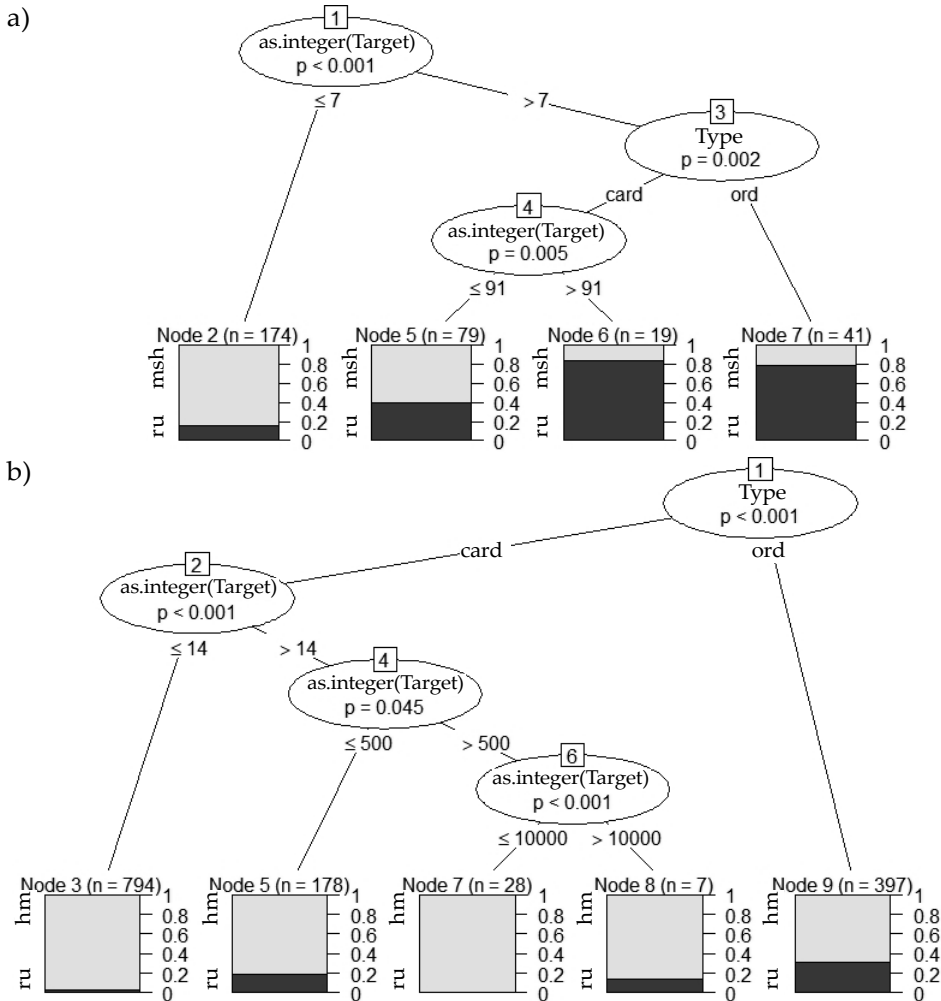


Figure 5. Decision trees: the dependence of language on numeral type and numerical value as integer. msh – Moksha, hm – Hill Mari, ru – Russian, card – cardinal numeral, ord – ordinal numeral, as.integer – numerical value as integer, not the distinction "large" vs. "small", p – p-value, n – number of numerical phrases.

### 3.4.4. Decision trees for the context, the numeral type and the numerical value

In addition to the interdependence between the numeral type and the numerical value, we studied the interdependence of all the three factors (including the context). In other words, it is important to understand whether a date is in Russian because of the context or purely because of its numerical value and numeral type. Also, this study reveals which factor plays the most crucial role in each language and what place the context occupies in each of our hierarchies of factors.

We made decision trees in order to trace the interdependence between all the three factors, as shown in Figure 6.

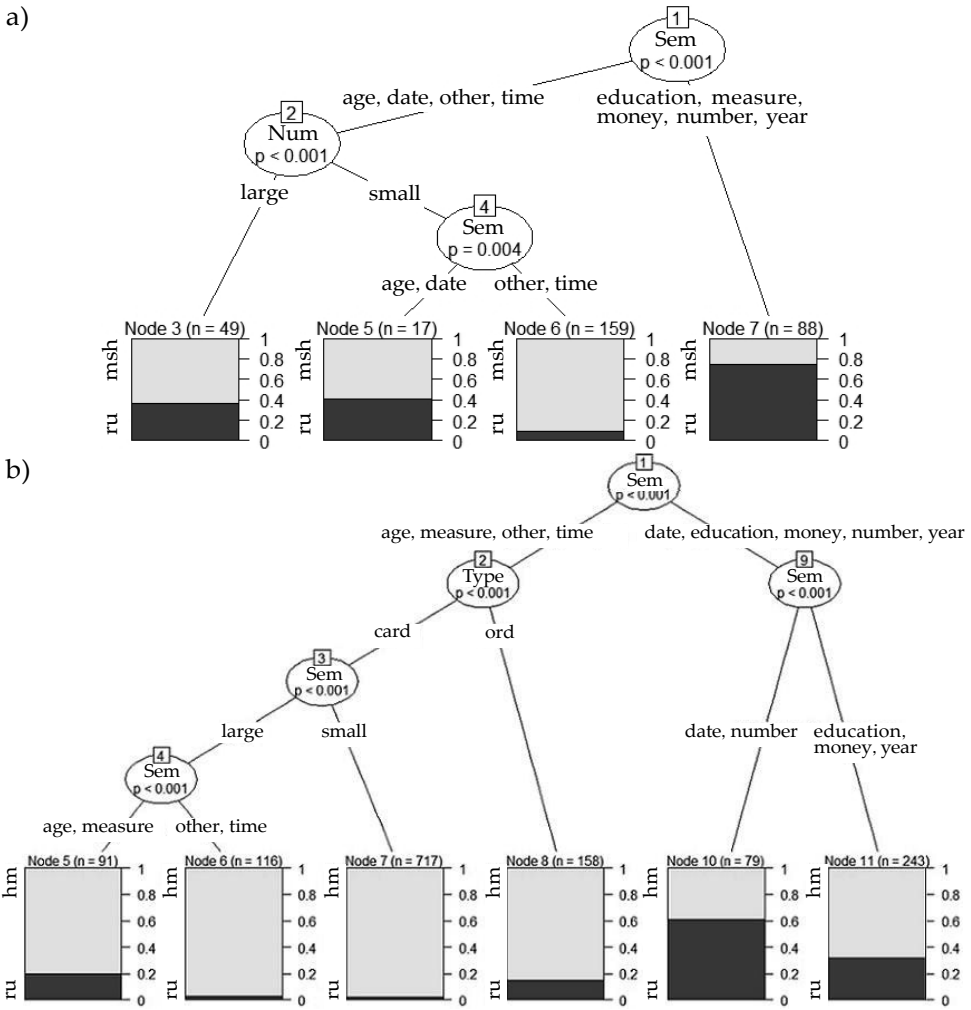


Figure 6. Decision trees: dependence of language on context, numeral type and numerical value. msh – Moksha, hm – Hill Mari, ru – Russian, card – cardinal numeral, ord – ordinal numeral, Sem – context type, Num – numerical value, p – p-value, n – number of numerical phrases.

In Moksha, the context is the most prominent choice-point: if the numerical construction is about money, year, education, measure and numeration, then it will be in Russian. The next choice-point is the numerical value. If the numeral is large, it will be in Russian. The final stage is the context again: if it is about date and age, then it will be in Russian, otherwise it will be in Moksha.

Hill Mari has a more complicated tree. The first choice-point is the context: date and numeration are more likely to be in Russian; education, money and year are also often in Russian, but with lower relative probability. The next choice-point is numeral type: some ordinals are in Russian, even though they are still more likely to be in Hill Mari. The numerical value is the next choice-point: small numerals are usually in Hill Mari. Finally, there is also some probability for Russian in the contexts "age" and "measure".

Figure 6 shows that in both languages, the context is the most influential factor. When all factors are considered together, the numerical value turns out

to be ineffectual. If we make the clusterization using actual numerical values instead of the distinction between large and small numerals (as we did in section 3.1), the influence of numerical value completely disappears. Thus, the context plays a crucial role in Moksha, and the type of the numeral is also important in Hill Mari.

#### **4. Conclusion**

In this paper we showed that the choice of the language of numerals in numerical constructions and entire numerical constructions is a result of a complex interaction of several factors: the numerical value of the numeral, its type (cardinal vs. ordinal), and the context where it is used. The weight of each parameter varies between the two corpora. Some contexts (e.g. "year", "date" and "money") trigger the usage of Russian numerals more than others (e.g. "measure", "age"). The type of the numeral also plays a role in Hill Mari: ordinal numerals are switched more frequently than cardinal numerals. Despite the seeming importance of numerical value, as seen in the fact that larger numerals are expressed in Russian with a higher frequency than smaller ones, this seems to be nothing more than the dependence of numerical value on the context.

Although the type of the numeral is irrelevant for the language choice in Moksha, and, while somewhat relevant in Hill Mari, is still less relevant than the context there, it plays a crucial role in the choice of the language within the noun phrase. Cardinal numerals are not only switched less frequently than ordinal numerals, but there can be no switches between a cardinal numeral and a noun. We argue that this is because of structural reasons: cardinal numerals and ordinal numerals occupy different positions in the noun phrase. We adopt the idea that there can be no switches between languages except the cases of categorical equivalence (following Muysken 2000), and this is exactly what prevents speakers from having a cardinal numeral in one language and a noun in the other. The head/modifier status of each part of a numerical construction is different in Russian and Finno-Ugric, which makes the determination of the ML in a numerical construction problematic. This results in the absence of mixed numerical constructions with cardinal numerals, and the formation of EL islands instead. These constraints shed light on differences in the types of syntactic relations found in a given construction cross-linguistically, and on the congruence of the structures across different languages.

#### **Addresses**

Irina Khomchenkova  
Vinogradov Russian Language Institute,  
Russian Academy of Sciences (Moscow)  
Lomonosov Moscow State University  
E-mail: irina.khomchenkova@yandex.ru

Polina Pleshak  
University of Maryland, College Park  
E-mail: pleshak@umd.edu

#### **Abbreviations**

**1, 3** — 1<sup>st</sup>, 2<sup>nd</sup> person, **ABL** — ablative, **ACC** — accusative, **ADJ** — adjective, **AOR** — aorist, **ASSOC** — associative, **c.1—c.10** — classes, **CAUS** — causative, **DAT** — dative, **DEF** — definite, **DETR** — ditransitive, **EMPH** — emphatic, **F** — feminine, **GEN** — genitive, **ILL** — illative, **IN** — inessive, **INDEF** — indefinite, **INF** — infinitive, **O** — object, **ORD** — ordinal, **M** — masculine, **NOM** — nominative, **NPST** — non-past tense, **PASS** — passive, **PAUC** —

paucal form, PL — plural, POSS — possessive, PQP — plusquamperfect, PRES — present, PRET — preterite, PRON — pronoun, PST — past tense, PTCL — particle, PTCP — participle, PV — preverb, REFL — reflexive, s — subject, SG — singular.

REFERENCES

- Auer, P. 1999, From Code-Switching via Language Mixing to Fused Lects: Toward a Dynamic Typology of Bilingual Speech. — *International Journal of Bilingualism*. Vol. 3, 309–332.
- Auer, P., Muhamedova, R. 2006, 'Embedded Language' and 'Matrix Language' in Insertional Language Mixing: Some Problematic Cases. — *Italian Journal of Linguistics*. Vol. 17. Code Switching, 35–54.
- Babby, L. H. 1987, Case, Prequantifiers, and Discontinuous Agreement in Russian. — *Natural Language & Linguistic Theory*. Vol. 5, 91–138.
- Belazi, H. M., Rubin, E. J., Toribio, A. J. 1994, Code Switching and X-bar Theory: The Functional Head Constraint. — *Linguistic Inquiry*. Vol. 25, 221–237.
- Bereczki, G. 1968, Wichtigere lautgeschichtliche Lehren der russischen Lehnwörter im Tscheremissischen. — *CIFU II. Pars I*, 70–76.
- Boumans, L., Caubet, D. 2000, Modelling Intrasentential Codeswitching: A Comparative Study of Algerian/French in Algeria and Moroccan/Dutch in the Netherlands. — *Arabic as a Minority Language*, Berlin (Contributions to the Sociology of Language [CSL] 83), 113–180.
- Bowers, J. 1984, On the Autonomy of Inflectional Morphology. — *Cornell University Working Papers in Linguistics* 6. Special Issue: Papers from the Second Cornell Conference on Government and Binding Theory, June 17–18, 1983, Ithaca NY, 23–41.
- Corbett, G. G. 1993, The Head of Russian Numeral Expressions. — *Heads in Grammatical Theory*, Cambridge, 11–35.
- Disciullo, A., Muysken, P., Singh, R. 1986, Government and Code-Mixing. — *Journal of Linguistics*. Vol. 22, 1–24.
- Gumperz, J. J., Hernandez-Chavez, E. 1971, Cognitive Aspects in Bilingual Communication. Problems of Multilingualism with Special Reference to Eastern Africa. — *Language Use and Social Change*, Oxford (Linguistic Surveys of Africa), 111–125.
- Hothorn, T., Hornik, K., Zeileis, A. 2006, Unbiased Recursive Partitioning: A Conditional Inference Framework. — *Journal of Computational and Graphical Statistics*. Vol. 15, 651–674.
- Ionin, T., Matushansky, O. 2018, Cardinals. The Syntax and Semantics of Cardinal-Containing Expressions, Cambridge MA—London (Linguistic Inquiry Monograph Seventy-Nine).
- Janurik, B. 2017, Erzya-Russian Bilingual Discourse: A Structural Analysis of Intrasentential Code-Switching Patterns. PhD Dissertation, Szeged.
- Johanson, L. 2000, Linguistic Convergence in the Volga Area. — *Languages in Contact*, Amsterdam—Atlanta (Studies in Slavic and General Linguistics 28), 165–178.
- Joshi, A. 1985, Processing of Sentences with Intrasentential Code Switching. — *Natural Language Parsing. Psychological, Computational and Theoretical Perspectives*, Cambridge (Studies in Natural Language Processing).
- MacSwan, J. 1999, A Minimalist Approach to Intra-Sentential Code Switching, New York.
- — 2005, Précis of a *Minimalist Approach to Intrasentential Code Switching*. — *Italian Journal of Linguistics*. Vol. 17. Code Switching, 55–92.
- Matras, Y. 2007, The Borrowability of Structural Categories. — *Grammatical Borrowing in Cross-Linguistic Perspective*, Berlin (Empirical Approaches to Language Typology [EALT] 38), 31–73.
- Muysken, P. 2000, Bilingual Speech. A Typology of Code-Mixing, Cambridge.
- Myers-Scott, C. 1992, Comparing Codeswitching and Borrowing. — *Journal of Multilingual and Multicultural Development*. Vol. 13, 19–39. <https://doi.org/10.1080/01434632.1992.9994481>.

- 1993, *Duelling Languages. Grammatical Structure in Code-switching*, Oxford.
- 2002, *Contact Linguistics. Bilingual Encounters and Grammatical Outcomes*, Oxford.
- 2004, Precision Tuning of the Matrix Language Frame (MLF) Model of Codeswitching. — *Sociolinguistica* 18, 106–117.
- Myers-Scott, C., Jake, J. L. 2000, Testing the 4-M Model: An Introduction. — *International Journal of Bilingualism*. Vol. 4, 1–8.
- Pandit, I. 1990, Grammaticality in Code-switching. — *Code-Switching as a World-Wide Phenomenon*, New-York, 33–69.
- Pesetsky, D. 1982, Paths and Categories. <http://www.ai.mit.edu/projects/dm/theses/pesetsky82.pdf>.
- 2013, *Russian Case Morphology and the Syntactic Categories*, Cambridge MA (Linguistic Inquiry Monograph Sixty Six).
- Poplack, S. 1980, Sometimes I'll Start a Sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a Typology of Code-Switching. — *Linguistics* 18, 581–618.
- 1981, The Syntactic Structure and Social Function of Codeswitching. — *Latino Language and Communicative Behavior*, Norwood, 69–92.
- 1988, Contrasting Patterns of Code-Switching in Two Communities. — *Codeswitching. Anthropological and Sociolinguistic Perspectives*, Berlin—New York—Amsterdam (Contributions to the Sociology of Language 48), 215–244.
- Pyöli, R. 1996, Venäläistyvä aunuksenkarjala: kielenulkoiset ja -sisäiset indikaattorit kielenvaihtolanteessa, Joensuu (Joensuun yliopiston humanistisia julkaisuja 18).
- Sarhima, A. 1999, Syntactic Transfer, Contact-Induced Change, and the Evolution of Mixed Codes. Focus on Karelian-Russian Language Alternation, Helsinki (Studia Fennica Linguistica 9).
- Sebban, M. 1998, A Congruence Approach to the Syntax of Code-Switching. — *International Journal of Bilingualism* 2, 1–20.
- Shaffer, D. 1978, The Place of Code-Switching in Linguistic Contacts. — *Aspects of Bilingualism*, Columbia SC, 265–274.
- Treffers-Daller, J. 1991, Towards a Uniform Approach to Codeswitching and Borrowing. — *ESF Network on Codeswitching and Language Contact. Papers for the Workshop on Constraints, Conditions and Models*, Strasbourg, 259–279. <http://centaur.reading.ac.uk/29333/>.
- Turunen, M. 1997, Nykyvatjan koodinvaihdoista. — *Vir.*, 208–232.
- Vanden Wyngaerd, E. 2017, The Adjective in Dutch-French Codeswitching: Word Order and Agreement. — *International Journal of Bilingualism* 21, 454–473.
- Арискин Н. И. 1993, Социальный аспект функционирования мордовских языков. — *Вестник Мордовского университета. Инженерные технологии и системы* 2, 20–22.
- Гаврилова В. Г. 2012, Дублирование как одно из проявлений переключения. — *Финно-угорский мир*, № 3/4, 56–59.
- 2013, Марийско-русское переключение и смешение кодов. — *Вестник Удмуртского университета. Вып. 2. История и филология*, 16–22.
- Зализняк А. А. 2002, «Русское именное словоизменение» с приложением избранных работ по современному русскому языку и общему языкознанию, Москва.
- Ку克林 А. Н. 2010, Особенности функционирования марийского языка в Урало-Поволжской историко-этнографической области. — *Финно-угорский мир*, № 1, 17–23.
- Лемов А. В. 2018, Особенности функционирования русского и эрзянского языков среди коренных жителей мордовской деревни. — *Историческая и социально-образовательная мысль*. Том 10, № 2/1, 92–96.
- Максимов С. А. 2017, Числительные в пермских языках: прошлое и настоящее. — *Ежегодник финно-угорских исследований*. Том 11, № 3, 7–13.
- Мельчук И. А. 1985, Поверхностный синтаксис русских числовых выражений, Wien (Wiener slawistischer Almanach. Sonderband 16).



- Пинеда Д. 2007, К проблеме «парадокса наблюдателя» и переключение кодов. — Полярный вестник 10, 40—52. <https://doi.org/10.7557/6.1309>.
- Плешак П. С. 2017, Морфосинтаксис именной группы в мокшанском и горномарийском языках [Manuscript].
- Пуссине О. 2010, Особенности языковой ситуации и русского языка в Мордовии. — Instrumentarium of Linguistics. Sociolinguistic Approaches to Non-Standard Russian, Helsinki (Slavica Helsingiensia 40), 106—126.
- Сааринен С. 2014, Переключение кодов в мокшанских диалектных текстах. — V Всероссийская конференция финно-угроведов «Финно-угорские языки и культуры в социокультурном ландшафте России». Материалы, Петрозаводск, 540—543.
- Сидорова М. А. 2018a, Числовое маркирование существительного в горномарийских количественных конструкциях. — Acta Linguistica Petropolitana XIV 2, 400—427.
- — 2018b, Количественные конструкции. — Элементы мокшанского языка в типологическом освещении, Москва, 328—341.
- Холодова М. А. 2018, Именная предикация. — Элементы мокшанского языка в типологическом освещении, Москва, 661—632.
- Шабьков В. И., Кудрявцева Р. А. 2017, Языковая ситуация в Республике Марий Эл в начале 2010-х годов: социолингвистический анализ. — Урало-алтайские исследования, № 3 (26), 208—229.

*ИРИНА ХОМЧЕНКОВА* (Москва), *ПОЛИНА ПЛЕШАК* (Колледж-Парк)

## **РУССКИЕ ЧИСЛИТЕЛЬНЫЕ В МОКШАНСКОМ И ГОРНОМАРИЙСКОМ**

В статье рассматривается употребление русских числительных в спонтанном устном дискурсе двуязычных носителей мокшанского/горномарийского и русского языков. Основываясь на сравнительном анализе конструкций с числительными, содержащими переключение кодов, в мокшанском и горномарийском корпусах, мы предлагаем анализ в рамках модели матричного языка Майерс-Скоттон. Обсуждаются факторы, которые могут влиять на выбор языка: арифметическое значение числительного, тип контекста, а также синтаксический тип числительного (количественное vs. порядковое). Несмотря на сохранность исконных систем числительных в каждом из исследуемых уральских языков, нельзя не отметить тенденцию использовать русские числительные для обозначения больших количеств. Кроме того, русские порядковые числительные встречаются чаще, чем русские количественные. Наконец, важным является фактор контекста: при обозначении года, речи об образовании или деньгах более вероятно появление русского числительного, чем при рассказе о повседневной жизни, количестве детей, бытовых проблемах. То, какой фактор оказывает большее влияние, различается для разных языков. Так, фактор арифметического значения более заметен в мокшанском корпусе. Для горномарийского важнее оказывается тип числительного. Для обеих текстовых коллекций самым важным является контекст.

Мы также показываем, что более частое употребление русских порядковых числительных по сравнению с количественными связано в том числе со структурным фактором: порядковые и количественные числительные занимают различные позиции в именной группе (ИГ). Последние имеют более тесные синтаксические связи с существительным, поскольку влияют на его числовое значение, а также (в русском языке) на падеж. Порядковые числительные в свою очередь никак не влияют на другие элементы в ИГ. Мы обсуждаем проблемы формирования островов вложенного языка, принимая во внимание различия в структурных отношениях разных типов числительных: количественные числительные сходны с квантификаторами, а порядковые — с адъективными модификаторами имени. Взаимодействие количественных числительных с именем различно в русском и в уральских языках, что создает конфликт при определении матрично-

го языка в составляющей и затрудняет процесс переключения кодов внутри ИГ. Порядковые числительные, не влияя на морфосинтаксические признаки других членов ИГ, ведут себя в русском и уральских языках более схожим образом и не создают структурных конфликтов. Рассматриваемые в статье ограничения на переключение внутри групп с числительными говорят в поддержку модели Мэйскена об эквивалентности категорий как условия для переключения кодов.

*IRINA HOMTŠENKOVA* (Moskva), *POLINA PLEŠAK* (College Park)

## **VEENE ARVSÕNADE KASUTAMISEST MOKŠA JA MÄEMARI KEELES**

Artiklis vaadeldakse mokša ja mäemari keelekorpusse tuginedes, kuidas kakskeelsed mokšalased ja mäemarilased spontaanses kõnes vene arvsõnu kasutavad. Keele valikut (koodivahetust) mõjutavad põhiliselt arvu suurus, kontekst ja arvsõna liik (põhiarvsõna või järgarvsõna). Üldiselt esineb tendents kasutada suuremate arvude tähistamiseks vene arvsõnu ning vene järgarvsõnad on tavalisemad kui vene põhiarvsõnad. Konteksti puhul saab esile tuua, et vene arvsõna tarvitamine on aastaarvu märkimisel, haridusest või rahast rääkides tõenäolisem kui argielust, laste arvust, olme probleemidest kõneldes. Koodivahetust põhjustavate tegurite osa on keeleli erinev. Kõige olulisem on mõlema keele puhul kontekst. Arvu suuruse mõju on märgatavam mokša keeles ja arvsõna liik on olulisem mäemari keele puhul. Vene järgarvsõnade sagedasem kasutamine võrreldes põhiarvsõnadega on muu hulgas tingitud struktuuritegurist. Põhiarvsõnadel on nimisõnaga tihedamad süntaktilised seosed, mis sõltuvad arvu suurusest ja mõjutavad ka käännet (vene keele puhul). Järgarvsõnad teisi nimisõnafraasi osi ei mõjuta. Põhiarvsõnad sarnanevad kvantooritega ja järgarvsõnad omadussõnadega. Põhiarvsõna ja nimisõna seos on uurali keeltes ja vene keeles erinev, mis koodivahetust nimisõnafraasi puhul raskemaks teeb. Järgarvsõnad käituvad uurali keeltes ja vene keeles sarnasemalt ega tekita struktuurseid konflikte.