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.

Key words: code-switching, numerals, Matrix Language Frame model, Finno-Ugric languages, 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 are Finno-Ugric languages spoken in the Volga region of Russia. 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 (CS) (Сааринен 2014), and, as far as we know, there are no works on CS in Hill Mari.

In contrast to previous works, our study is a comparative analysis of CS 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 constructions (NumC)[1] are in Russian. For example, in (1) the first NumC is in Hill Mari, while the second one is in Russian.

HILL MARI

(1) *mǝn'ǝ* *mar-lan* *ke-n-äm* **tǝžem** **ǝndekš** **šüdǝ**
    I man-DAT go-PRET-1SG thousand nine hundred
    kändäkš **lu**-*šǝ* *i-n* **semnadcatyj** **maj**-*ǝn*
    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 Ardashevo 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). It is available at 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 texts authors are bilingual, and they speak both the indigenous language (Moksha or Hill Mari) and Russian. For most people, Moksha or Hill Mari is the main language of their everyday communication. The sociolinguistic situation 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

---

[1] With the term numerical construction, we refer both to numeral phrases with cardinal numerals and noun phrases with ordinal numerals.

languages (Moksha and Erzya) is described in more detail by Ariskin (1993), and that of the Mari languages (Hill Mari and Meadow Mari) by Kuklin (2010) and by Shabykov and Kudryavtseva (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 the MLF model), the language, and the context in which it was uttered.

Taking into consideration the large percentage of Russian numerals in the texts (33% and 13% Russian numerals among all the numerals in Moksha and Hill Mari, respectively), and the problem of distinguishing borrowings from switched fragments (Poplack 1988), the status of Russian numerals in Moksha and Hill Mari speech is not obvious. Based on the frequency criterion proposed by Myers-Scotton (1992: 35-36), we would assume that these are instances of CS rather than borrowing. However, no criteria ended up being reliable enough to distinguish between CS and borrowing (Treffers-Daller 1991). Therefore, following Treffers-Daller (1991), we treat the two phenomena uniformly, concentrating not on the status of Russian numerals but 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 NumC. Both sets of results speak in favor of CS, not borrowing. Given that NumCs with cardinal numerals (CN) are structurally different in Finno-Ugric and Russian, whereas noun phrases with ordinal numerals (ON) are similar, we hypothesize that (i) the ordinals may be switched more readily, and (ii) there will be no switches between CNs and head nouns. We show that these hypotheses are borne out. The proposed structural analysis is couched in the Matrix Language Frame (MLF) model (Myers-Scotton 1993) as well as Muysken's (2000) classification of types of CS.

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 NumC (2.2), and the two models we used to model CS (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. NumCs in Russian, Moksha and Hill Mari

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

2.1.1 Cardinal numerals

Constructions with CNs 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 NumCs with CNs was broadly discussed in the formal literature (Mel'čuk 1985; Babby 1987; Corbett 1993; Pesetsky 2013, among others). In Russian, CNs with numerals from two to four require a special form of the noun if the entire NumC is in subject position (2.a). The same case is received by inanimate nouns in NumCs in direct object (2.b) 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 (Zaliznjak 2002; Ionin & Matushansky 2018). In the same positions, larger numerals require a plural genitive form (2.c-2.d). 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. *три   кот-а   ид-ут*
     three.NOM   cat-SG.GEN/PAUC   go-NPST.3PL
     'Three cats are walking'.

   b. *Вася   с-руби-л   три дерев-а*
     vasja.NOM   PV-cut-PST.M.SG   three tree-SG.GEN/PAUC
     'Vasja cut three trees'.

   c. *пять кот-ов   ид-ут*
     five   cat-PL.GEN go-NPST.3PL
     'Five cats are walking'.

   d. *Вася   с- руби-л   пять деревь-ев*
     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 (3.a-3.b).

(3) a. *кот-ы   ид-ут*
     cat-PL.NOM   go-NPST.PL
     'Cats are walking'.

   b. *Вася   руби-л   деревь-я*
     Vasja.NOM   PV-cut-PST.M.SG   tree-PL.NOM/ACC
     'Vasja cut three trees'.

In NumCs with animate nouns in object positions (4.a), (4.c) and NumCs in oblique

positions (4.b), (4.d), 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" (4.a), (4.b) and "large" numerals (4.c), (4.d). The CN always receives the case

which is required by the syntactic position of the entire NumC. The resulting pattern in (4), unlike

the one in (2) looks like a case concord of CN and head noun. The number is always plural.

(4) a. *Катя   глад-ит   трех   кот-ов*
     Kate.NOM   pat-NPST.3SG   three.PL.ACC   cat-PL.ACC
     'Kate is patting three cats'.

   b. *Катя   да-л-а   трем   кот-ам   молок-а*
     Kate.NOM   give-PST-F.SG   three.PL.DAT   cat-PL.DAT   milk-SG.GEN
     'Kate gave some milk to three cats'.

   c. *Катя   глад-ит   пять   кот-ов*
     Kate.NOM   pat-NPST.3SG   five.ACC   cat-PL.ACC
     'Kate is patting five cats'.

   d. *Катя   да-л-а   пят-и   кот-ам   молок-а*
     Kate.NOM   give-PST-F.SG   five-DAT   cat-PL.DAT   milk-SG.GEN
     'Kate gave some milk to three cats'.

In these cases nouns in simple noun phrases would have the same form as they do in

NumCs (5.a-5.b).

(5) a. *Катя   глад-ит   кот-ов*
     Kate.NOM pat-NPST.3SG cat-PL.ACC
     'Kate is patting five cats'.

   b. *Катя   да-л-а   кот-ам   молок-а*
     Kate.NOM give-PST-F.SG cat-PL.DAT milk-SG.GEN
     'Kate gave some milk to three cats'.

In Moksha and Hill Mari, the case in NumCs with CNs 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 NumCs (6)-(9). CNs are never case-marked in presence of a noun, so they neither assign genitive case (6a)-(9a), nor do they exhibit case-concord with the noun.

(6) a. *kə̂m kot'i / *kot'i-n ašked-ə̂t*
     three cat   cat-GEN walk-NPST.3PL
     'Three cats are walking'.

   b. *kat'a kə̂m kot'i-m niält-en*
     Kate three cat-ACC pat-PRET
     'Kate is patting three cats'.

   c. *kat'a kə̂m kot'i-län   šə̂šer-ə̂m pu-en*
     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. *kot'i ašked-eš*
     cat   go-NPST.3SG
     'A cat is walking'.

   b. *kat'a kot'i-m niält-en*
     Kate cat-ACC pat-PRET
     'Kate is patting a cat'.

   c. *kat'a kot'i-län   šə̂šer-ə̂m   pu-en*
     Kate cat-DAT   milk-ACC   give-PRET
     'Kate some milk to a cat'.

(8) a. *kot'i-vlä   ašked-ə̂t*
     cat-PL   go-NPST.3PL
     'Cats are walking'.

   b. *kat'a kot'i-vlä-m niält-en*
     Kate cat-PL-ACC pat-PRET
     'Kate is patting a cat'.

c. *kat'a kot'i-vilä-län š∂šer-∂m pu-en*
   Kate cat-PL-DAT milk-ACC give-PRET
   'Kate some milk to a cat'.

Compare also Moksha examples with numerals (9) and without numerals (10). As in Hill Mari, the case marking in NumCs 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. *kolmə katə-t/ *katə-t'/ *katə-t'n'ə-n' jaka-j̊-t'*
       three cat-PL cat-DEF.SG.GEN cat-DEF.PL-GEN walk-NPST.3-PL
       'Three cats are walking'.
     b. *kat'ɛ sud'ər'ɛ-z'ən' kolmə katə-t'n'ə-n'*
       Kate caress-PST.S.3SG.O.3PL three cat-DEF.PL-GEN
       'Kate gave the three cats a pat'.
     c. *kat'ɛ maksə-z'ə lofc'-t' kolmə katə-t'n'ə-n'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. *katə-t jaka-j̊-t'*
       cat-PL walk-NPST.3-PL
       'Cats are walking'.
     b. *kat'ɛ sud'ər'ɛ-z'ən' katə-t'n'ə-n'*
       Kate caress-PST.S.3SG.O.3PL cat-DEF.PL-GEN
       'Kate gave the cats a pat'.
     c. *kat'ɛ maksə-z'ə lofc'-t' katə-t'n'ə-n'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 by Sidorova (2018a), 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 c'ora-*(t)* [two boy-PL] 'two boys', whereas large numerals prohibit it in absence of other nominal dependents: *s'is'gemən' c'ora-(*t)* [seventy boy-PL] 'seventy boys'. Large numerals are also sensitive to the definite/indefinite and possessive/non-possessive marking on the head, as well as to the number and type of nominal dependents (Sidorova 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. Another crucial fact is that it is a matter of lively debate which element of the NumC in Russian is the head (Mel'čuk 1985; Babby 1987). For example, Pesetsky (1982) and Bowers (1982) claim that it is the CN, not the noun, that heads the Russian NumC. Following this approach, Ionin and Matushansky (2018) 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 (p.180). In Finno-Ugric NumCs, however, the noun is the head (Плешак 2017). In a situation of CS, 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).

2.1.2 Ordinal numerals

Ordinal numbers (ON) are adjective-like nominal modifiers in all the three languages. In Russian, ONs 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.NOM   girl.NOM   beautiful.PL.ACC girl.PL.ACC
        'beautiful girl / [I see] beautiful girls'

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

HILL MARI
(12) a. *kə̂m-šə̂   ə̂də̂räš / kə̂m-šə̂   ə̂də̂räs-vlä-m*
        three-ORD girl   three-ORD girl-PL-ACC
        'the third girl / [I see] the third girls'
     b. *cever   ə̂də̂räš / cever   ə̂də̂räs-vlä-m*
        nice   girl   nice   girl-PL-ACC
        'a nice girl / [I see] nice girls'

MOKSHA

(13) a. *kolmə-c'ə   s't'ər'-s'* /   *kolmə-c'ə   s't'ə̣r-n'ə-n'*
     three-ORD   girl-DEF.SG   three-ORD   girl-DEF.PL-GEN
     'the third girl / [I see] the third girl'
   b. *mazi s't'ər'-s'* /   *mazi s't'ə̣r-n'ə-n'*
     nice   girl-DEF.SG   nice   girl-DEF.PL-GEN
     'a nice girl / [I see] nice girls'

Morphosyntactically, ONs in Moksha and Hill Mari pattern with adjectives. They do not affect the number nor 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 NumCs with CNs 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 CN and the noun, but in different ways. Noun phrases with ONs can be considered as structurally similar in Russian and in the two Finno-Ugric languages, with ONs showing a higher degree of structural "independence". This plays an important role in CS.

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 (p. 51) and "lower ordinals > higher ordinals" for ordinals (p.52)). Given that we adopt a uniform approach to borrowing and CS, these hierarchies are also applicable to CS. 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 **numerical value** 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). Similarly, Votic numerals are used only while counting from 1 to 7 (Turunen 1997: 219).

Second, the **context** 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. Saarinen (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 **grammatical constraints** on CS. For example, CS is possible between a subject and a predicate, but not between an auxiliary and a main verb. Given that **different types of numerals** occupy different structural positions, this can be relevant in our discussion of CS with numerals. This is addressed in section 3.3.

2.3. Theoretical models of code switching

In order to explain the structural restrictions on CS that we find in our data, we will take into account the types of constituents in the Matrix Language Frame (MLF) model proposed by Myers-Scotton (1993) (2.3.1). Based on this, Muysken (2000) develops his model of constraints on CS and possible conditions where they can be avoided, as well as a classification of types of CS (2.3.2). These models have become the most influential ones, having more explanatory adequacy than previous proposals (Poplack 1980, 1981; Joshi 1985; Belazi, Rubin and 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 MLF model (Myers-Scotton 1993)

The Matrix Language Frame model is based on the distinction between a more activated matrix language (ML) an embedded language (EL). A bilingual sentence can consist of (i) **ML islands**, which consist only of ML morphemes and are well-formed according to the grammar of ML; (ii) **EL islands,** which have only EL morphemes and are well-formed according to the grammar of EL; and (iii) **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" mophemes. System mophemes (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 only 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). First, the word *certificate* and its modifiers follow Swahili word order. Second, the verb *depend* agrees with the

subject using Swahili morpheme *i-*. In both these cases two English words are inserted into the Swahili frame:

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

Example (15) illustrates an EL island (French) which occurs in an Arabic morpho-syntactic frame and which meets well-formedness condition: the plural morphemes on both the article and the noun are French.

(15) *yə-t-haka-w   waħed **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 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 departs from 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, and 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 DP. If such a DP 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 DP that it selects. This element is called *language index carrier* and is defined as "the highest (non-lexical) node in a tree" (DiSciullo, Muysken, and 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 governing head and its complement are equivalent in the two languages respectively, such configurations are not subjects 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: (i) they have "the same status in the two languages", (ii) they are "morphologically encapsulated, (iii) [they] are shielded off by a functional element from the matrix language, or (iv) [they] could belong to either language" (ibid: 31). We focus our discussion on this phenomenon, because noun phrases (both NumCs and non-NumCs) are claimed to be a prototypical case of insertion (ibid: 62). Consider the following properties which support the view of these switches as instances of insertion:

   i.   they form a single constituent (*viz.* noun and its modifiers) (18);

   ii. they occur in nested structures, so that the elements that surround them are in the ML (18);

   iii. they are content words (nouns, adjectives) rather than functional particles/words (18);

   iv. they are complements, not adjuncts (18); and

   v. they can undergo morphological integration (19).

MOKSHA

(18) *i   ul'-s'   jota-ft-f*   **бал   маскарад**   *klup-sə*
and   be-PST.3SG conduct-CAUS-PTCP.PASS ball   masquerade   club-IN
'And there was a masqued ball in the club'.

MOKSHA

(19) *son'-c'-ən'*   **однокомнатный   квартира**-*c*
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 MLF model, and "alternations to EL islands combined with ML islands" (p.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 CS 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 CS scenarios. 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 statistic value is < 0.00001).
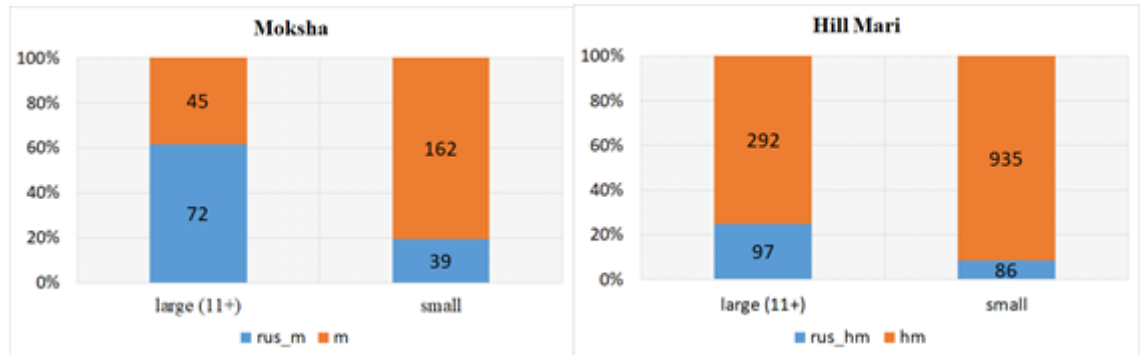
Figure 1. Numerical value dependence. m — Moksha, hm — Hill Mari, rus_m, rus_hm — Russian.

However, there is also a difference between the two collections. In Hill Mari there are more native numerals than Russian ones in both large and small conditions, 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.

Type	Example
year	1973, 19th century, (20)
date	July, 21st, (20)
education	5th grade; number of lessons, (21)
num(ber)	№, phone number, (22)
money	x rubles, (23)
measure	x grams, x liters, (24)
time	60 years, first day, for two days, (25)
age	6 years old, (26)
other	

Table 1: Context types

YEAR & DATE (HILL MARI)

(20) *näl-lə̈ i уже li-eš,*   *вот,* **семьдесят пятый год, второй**
four-ten year already be-NPST.3SG PTCL seventy five.ORD year second
октябрь, *вот, texen'ə̈-vlä*
october PTCL such-PL
'It will be forty years, well, the year seventy five, October second, well, so'.

EDUCATION (MOKSHA)

(21) *Федя за-кончи-л* **восемь класс-ов,** *mol'-i* **d'evataj-s**
Fedja.NOM PV-finish-PST.N.SG eight grade-PL.GEN go-NPST.3SG ninth-ILL
'Fedja finished eight grades, he goes to the ninth'.

NUM (HILL MARI)

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

MONEY (MOKSHA)

(23) **триста рубл-ей-*də̈*** *бодяга sɛv-at*
three.hundred ruble-PL.GEN-ABL alcohol.drink take-NPST.2SG
'You take alkohol drink for three hundred rubles'.

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

TIME (MOKSHA)

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

AGE (HILL MARI)

(26) **сорок пять лет,** *näl-lə̈ və̈c i t'ot'a-n*
fourty five year.PL.GEN four-ten 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 CS

3.3.1 Cardinals vs. ordinals

In contrast to CNs, ONs 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 ONs to exhibit a lower degree of interaction with the noun, which facilitates the process of switching. This predicts that there should be more switched ordinals than switched cardinals. This indeed is borne out, as shown in Figure 2, and the result is statistically significant (the Fisher exact test statistic value is 0.0004 and < 0.00001, respectively for Moksha and Hill Mari).
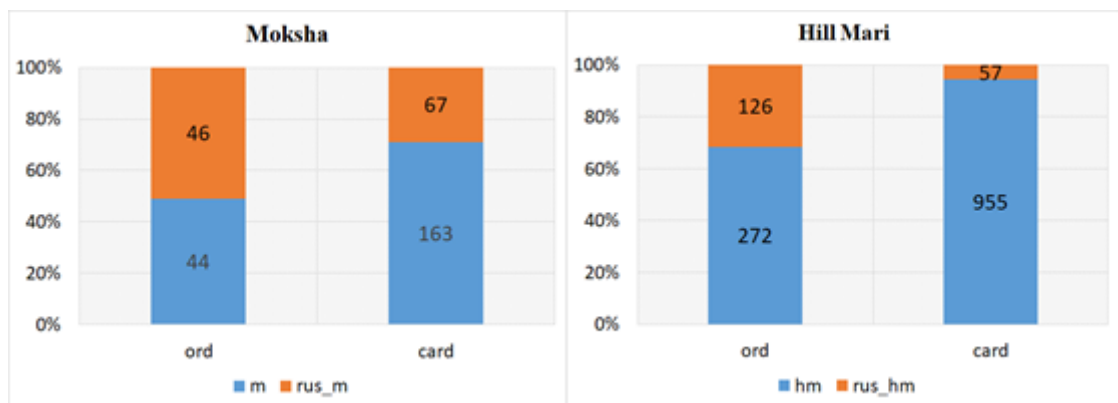
Figure 2. Numeral type dependence. ord - ordinal numerals, card - cardinal numerals, m - moksha, hm - Hill Mari, rus_m and rus_hm - Russian.

In the Hill Mari corpus, the number of Russian CNs is very small, whereas the number of ONs is quite large. In Moksha, there is a considerable number of switches of both types, but the percentage of Russian ONs is significantly larger that of Russian CNs.

3.3.2 EL islands vs. ML+EL constituents

In discussing the switching of numerals, one has to keep in mind that they are part of a noun phrase. This means that one has to distinguish switching of numerals themselves from switching of an entire noun phrase. There are four possible combinations:

(28) (i) Num(ML) + Noun(ML): no switching
     (ii) Num(ML) + Noun(EL): switching of a noun
     (iii) Num(EL) + Noun(ML): switching of a numeral
     (iv) Num(EL) + Noun(EL): switching of a noun phrase

The first combination is not considered as CS, as there are no elements in the EL (Russian). Combinations (ii) and (iii) contain elements of two different languages and are considered to be ML+EL constituents. They should be the most restricted, according to Muysken's model, and require categorical equivalence (see section 2.3.2). Combination (iv) can be analyzed differently. It can be treated as a classic case of an EL island, where all members of the constituent are taken from the EL. Another option is that it is an ML+EL constituent, where the syntactic structure is from the ML but the roots are taken from the EL.

We calculated the number of each type of constituent (ML; EL island; ML+EL) for both corpora. The distinction between EL islands and ML+EL 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 ONs and CNs were calculated separately. ML(r) are Russian NumCs that occur in Russian sentences, where the ML is Russian.
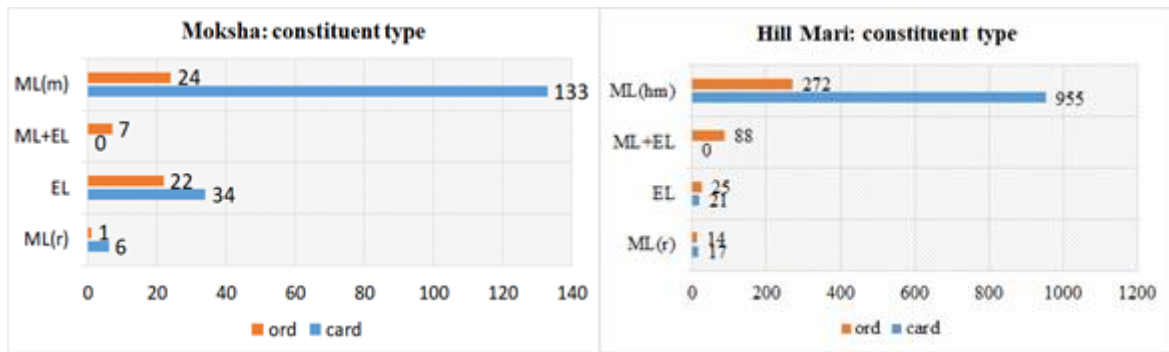
Figure 3. Constituent types. ML(hm) / ML(m) / ML(r) — Hill Mari / Moksha / Russian is ML of the clause and the language of the NumC, ML+EL — mixed constituents, EL — Russian is the language of the NumC but not ML of the clause.

In the charts we can see that despite a general prevalence of NumCs with CNs in the corpora, ML+EL constituents are attested only with ONs. They all represent the combination (ii), where the numeral but not the noun is switched. No instances of combination (iii), where only the noun is switched, were attested. This fits perfectly with Muysken's theory, as it is adjuncts that are switched more readily (in comparison to complements). Another tendency is that although there are more ONs in CS contexts, EL islands in Moksha mostly contain CNs, in contrast to ML+EL constituents, which contain ONs exclusively. We argue that as CNs have different syntactic behavior in Finno-Ugric and Russian, there can be no switches between them and the rest of the NumC (mainly noun). Therefore, they only can occur in EL islands, whose internal structure is quite strict. In contrast to CNs, ONs 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 (29)-(30). CNs 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 CN. The entire NumC 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.

HILL MARI
(29) *st'ip'end'ij-žə̑*   **двенадцать рубл-ей**   *ə̑l'-ə̂*   *в месяц*
scholarship-POSS.3SG   twelve   ruble-GEN.PL be-AOR.3SG in month
'The scholarship was twelve rubles per month'.

MOKSHA

(30) t'ejə-st er'av-əl'-ø sa-m-s **пять часов-sə** / ***пять**
PRON.DAT-3PL.POSS need-PQP-SG come-INF-ILL five hour-GEN.PL-IN five
час-сə
hour-IN
'They had to come at 5' (example courtesy Maria Kholodilova)

In contrast to CNs, ONs receive case together with the head noun within the entire DP (see subsection 2.1.2). ONs themselves do not influence the case on the noun. In (31), we see a Russian NumC in genitive case, and the ON shows concord in gender, number and case with the noun, as required by the grammar of Russian.

MOKSHA

(31) al'ɛ-z'ə **девятьсот четверт-ого год-а-l'.**
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 ONs. In these constituents, Russian words are inserted into the ML frame. Therefore, even though it is taken from Russian, the ON does not have to show concord in gender, number and case with the head noun, and we see a bare adjective-like form (32)-(33).

HILL MARI

(32) *первый смена-m=ät nänge-ä*
first shift-ACC=ADD lead-NPST.3SG
'And he is leading the first shift'.

MOKSHA

(33) *тридцать пятай s'ekcijɛ-s'*
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 *pervɔ̂j sm'enə* 'first shift' in (32) nor *tr'icat' p'ataj s'ekcijɛ* '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.

As we see, ONs are switched more readily than CNs. The latter are more in conflict with nouns as loci of morphological encoding of case: in Russian, they are analyzed as heads of NumCs, and in Finno-Ugric languages the head is the noun. Therefore, it would be unclear what the language of the NumC is, if a CN and a noun were taken from different languages. This results in CNs occuring only in EL islands. ONs, 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.

a)

Moksha	date	year	edu	mon	age	meas	time	oth
small	43%	13%	72%	19%	50%	29%	78%	83%
large	57%	87%	28%	81%	50%	71%	22%	17%

b)

Hill Mari	date	year	edu	mon	age	meas	time	oth
small	51%	8%	90%	34%	17%	56%	86%	13%
large	49%	92%	10%	66%	83%	44%	14%	87%

Table 2. The percentage of small and large numerals in different contexts.

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 CNs and ONs for each context separately, and our results are in Figure 4. 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 dominance of Russian in formal contexts. The figures show that year, date and numeration contexts are indeed those where mostly ONs are used.

Figure 4. Context dependence.

The interaction of context and type of numeral is highlighted in Table 3. We see a prevalence of ONs in date and year, which are switched most frequently. Age and measure, which show fewer switches, are expressed with NumCs with CNs, not ONs. 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 ONs in Hill Mari and more CNs in Moksha.

a)

Moksha	date	year	edu	num	mon	age	meas	time	oth
ord	92%	93%	30%	100%	0%	0%	0%	24%	17%
card	8%	7%	70%	0%	100%	100%	100%	76%	83%

b)

Hill Mari	date	year	edu	num	mon	age	meas	time	oth
ord	100%	100%	76%	38%	0%	0%	0%	11%	19%
card	0%	0%	24%	63%	100%	100%	100%	89%	81%

Table 3. The percentage of ordinals and cardinals in different contexts.

3.4.3. Numerical value and type of numeral

Let us now try to eliminate the interdependence of the context, the numeral type, and the numerical value. In order to check whether the date will be in Russian because of the context or purely because of the numerical value and numeral type, using the ctree() function in the *party* package in R (Hothorn et al., 2006), we plotted the trees based on numerical value and numeral type, see Figure 5.
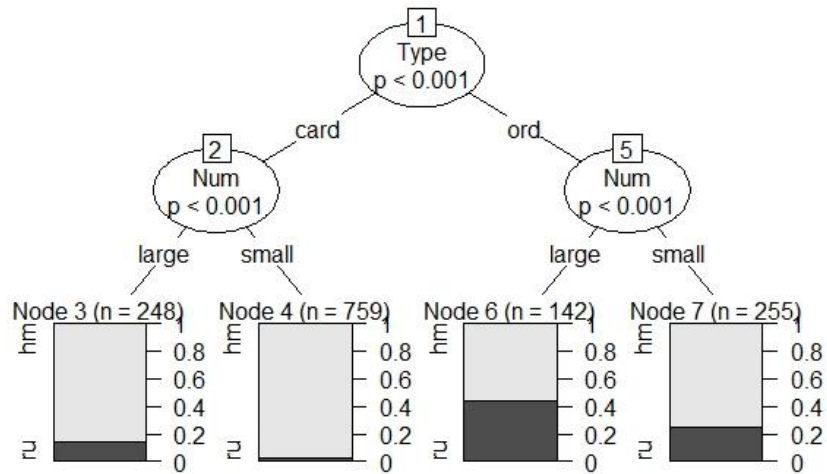
a)
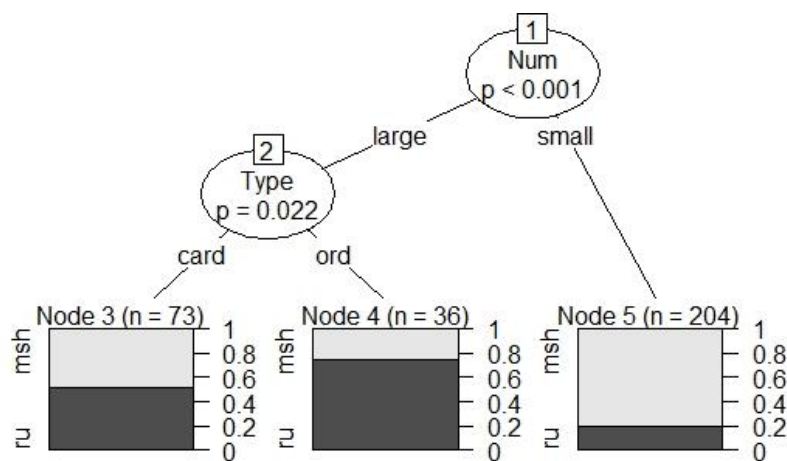
b)

Figure 5. Decision trees: the dependence of language on numeral type and numerical value.

The trees are very simple. In Moksha speech, small numerals are usually in Moksha, while large numerals are usually in Russian. Among these large numerals, ONs are more likely to be in Russian than CNs. 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.

Interestingly, the decision trees can change if we take into account the numerical value without the distinction into 'large' and 'small' numerals, see Figure 6.
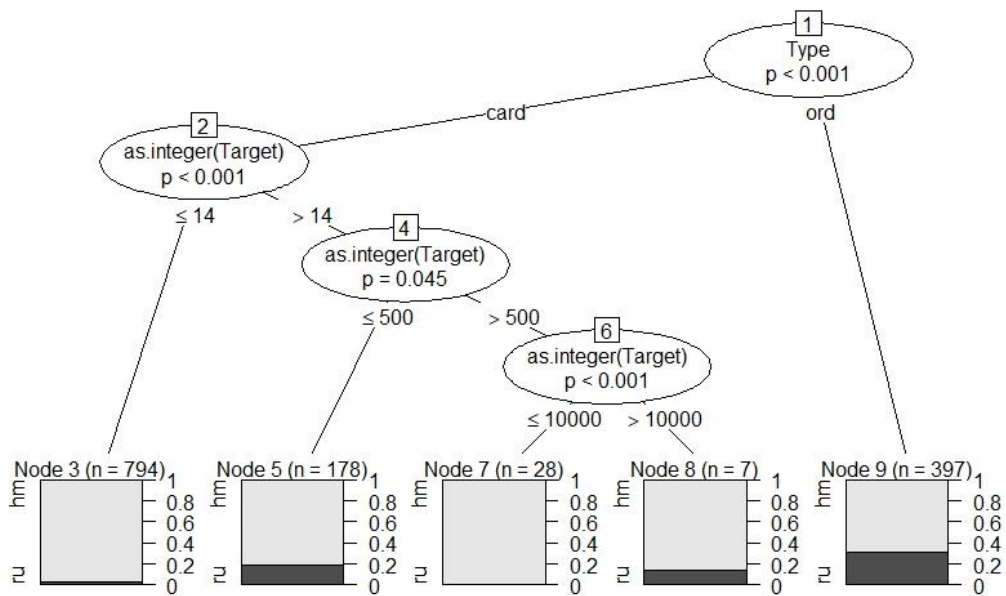
a)
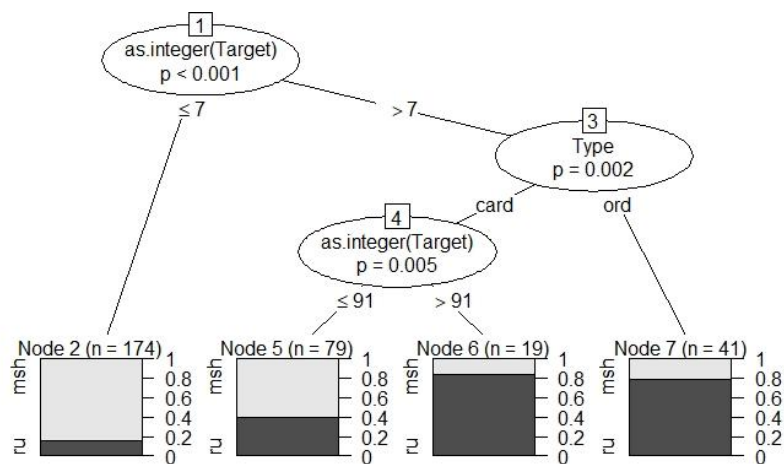
b)

Figure 6. Decision trees: the dependence of language on numeral type and numerical value as integer.

For Moksha the picture becomes a bit more difficult: numerals less than 7 or equal to it tend to be in Russian. Ordinals which are more than seven are usually in Russian. As for cardinals, 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. Basically, this tree is the reverse version of the Moksha tree in Figure 6a. The hierarchy of the factors in Moksha is numerical value > type of numeral, while in Hill Mari it is the reverse.

3.4.4. Decision trees for the three factors

We also made decision trees in order to trace the interdependence between all the three factors, as shown in Figure 7.
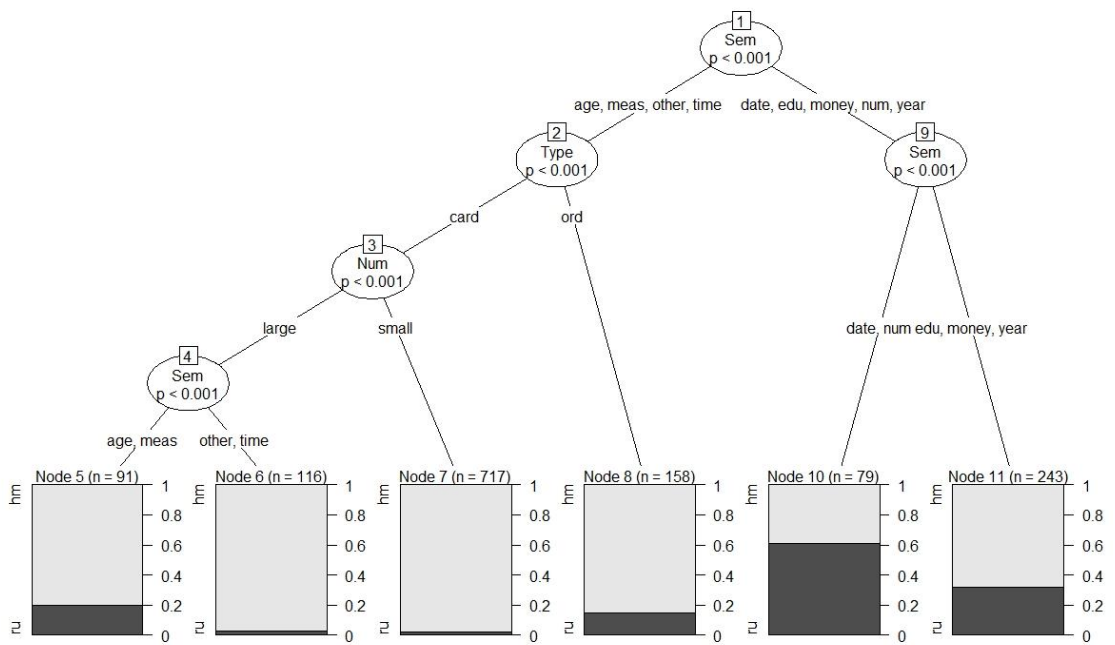
a)
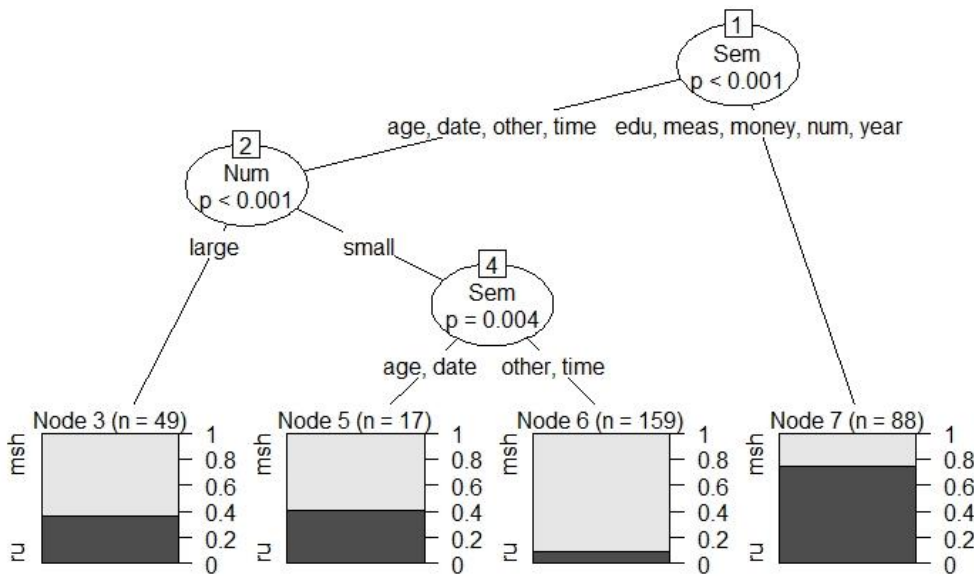
b)

Figure 7. Decision trees: dependence of language on context, numeral type and numerical value.

In Moksha the context is the most prominent choice-point: if the NumC 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'. 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 let the program cluster large and small numerals instead of using actual numerical values (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 NumCs and entire NumCs 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: ONs are switched more frequently than CNs. 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. CNs are not only switched less frequently than ONs, but there can be no switches between a CN and a noun. We argue that this is because of structural reasons: CNs and ONs occupy different positions in the NP. 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 CN in one language and a noun in the other. The head / modifier status of each part of a NumCs is different in Russian and Finno-Ugric, which makes the determination of the ML in a NumC problematic. This results in the absence of mixed NumCs with CNs, 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.

Abbreviations

1, 3 — 1st, 2nd 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, 3(4), 309–332.

Auer, P., Muhamedova, R. 2006, 'Embedded language' and 'matrix language' in insertional language mixing: Some problematic cases. – Italian Journal of Linguistics / Rivista di linguistica 17(1). [Special Issue: "Commutazione di codice e teoria linguistica"], 35–54.

Babby, L. H. 1987, Case, prequantifiers, and discontinuous agreement in Russian. – Natural Language & Linguistic Theory, 5(1), 91-138.

Belazi, H. M., Rubin, E. J., Toribio, A. J. 1994, Code switching and X-bar theory: The functional head constraint. – Linguistic Inquiry 25: 221-37.

Bereczki, G. 1968, Wichtigere lautgeschichtliche Lehren der russischen Lehnwoerter im Tscheremissischen. – Congressus Secundus Internationalis Fenno-Ugristarum. Pars I. Acta Linguistica, 70–76.

Boumans, L., Caubet, D. 2000, Modelling intrasentential codeswitching: A comparative study of Algerian/French in Algeria and Moroccan/Dutch in the Netherlands. – J. Owens, Arabic as a minority language, Berlin: Mouton de Gruyter, 113-18.

Bowers, J. 1984, On the Autonomy of Inflectional Morphology. – W. Harbert, Cornell Working Papers in Linguistics VI, Department of Modern Languages and Linguistics, Cornell University, Ithaca, New York, 23-41.

Corbett, G. G. 1993, The head of Russian numeral expressions. – Heads in grammatical theory, 11-35.

DiSciullo, A., Muysken, P., Singh, R. 1986, Code-mixing and government. – Journal of Linguistics 22, 1-24.

Gumperz, J. J., Hernandez-Chavez, E. 1971, Cognitive aspects in bilingual communication. – W. H. Whiteley, Language use and social change, Oxford University Press, 111-125.

Hothorn, T., Hornik, K., Zeileis, A. 2006, Unbiased Recursive Partitioning: A Conditional Inference Framework. – Journal of Computational and Graphical Statistics 15(3), 651-674.

Ionin, T., Matushansky, O. 2018, Cardinals: The syntax and semantics of cardinal-containing expressions. ⟨halshs-01679109⟩

Janurik, B. 2017, Erzya–Russian bilingual discourse: A structural analysis of intrasentential code-switching patterns, PhD dissertation, University of Szeged.

Johanson, L. 2000, Linguistic convergence in the Volga area. – Studies in Slavic and General Linguistics, 28, 165-178.

Joshi, A. 1985, Processing of sentences with intrasential code switching. – D. R. Dowty, L. Kattunen, A. M. Zwicky, Natural language parsing: Psychological, computational and theoretical perspectives, Cambridge: Cambridge University Press.

MacSwan, J. 1999, A Minimalist Approach to Intra-sentential Code-switching, New York: Garland.

MacSwan, J. 2005, Précis of a minimalist approach to intrasentential code switching. – Italian Journal of Linguistics, 17(1), 55-92.

Matras, Y. 2007, The borrowability of structural categories. – Y. Matras, J. Sakel, Empirical Approaches to Language Typology: Grammatical Borrowing in Cross-Linguistic Perspective, Berlin: Mouton de Gruyter, 31–73.

Muysken, P. 2000, Bilingual speech: A typology of code-mixing, Cambridge: Cambridge University Press.

Myers-Scotton, C. 1992, Comparing codeswitching and borrowing. – Journal of Multilingual and Multicultural Development, 13(1-2), 19-39. DOI: 10.1080/01434632.1992.9994481.

Myers-Scotton, C. 1993, Duelling languages: Grammatical structure in code-switching, Oxford: Clarendon Press.

Myers-Scotton, C. 2002, Contact linguistics: Bilingual encounters and grammatical outcomes, Oxford: Oxford University Press.

Myers-Scotton, C. 2004, Precision tuning of the Matrix Language Frame (MLF) Model of codeswitching. – Sociolinguistica, 18, 106–117.

Pandit, I. 1990, Grammaticality in code-switching. – R. Jacobson, Code-switching as a world-wide phenomenon, New-York: Peter Lang, 33-69.

Pesetsky, D. 1982, Paths and Categories, unpublished Ph.D. dissertation, MIT.

Pesetsky, D. 2013, Russian case morphology and the syntactic categories, MIT Press 66.

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.

Poplack, S. 1981, The syntactic structure and social function of codeswitching. – R. Durán, Latino language and communicative behavior, Norwood, Ablex, 69-92.

Poplack, S. 1988, Contrasting patterns of code-switching in two communities. – M. Heller, Codeswitching: Anthropological and sociolinguistic perspectives, Berlin and New York: Mouton de Gruyter, 215–244.

Pyöli, R. 1996, Venäläistyvä aunuksenkarjala: kielenulkoiset ja sisäiset indikaattorit kielenvaihtotilanteessa. Doct. diss. Joensuu.

Sarhimaa, A. 1999, Syntactic transfer, contact-induced change, and the evolution of mixed codes: Focus on Karelian–Russian language alternation, Helsinki: Finnish Literature Society.

Sebba, 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. – M. Paradis, Aspects of bilingualism, 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. European Science Foundation, Strasbourg, 259-279. Available at http://centaur.reading.ac.uk/29333/

Turunen, M. 1997, Nykyvatjan koodinvaihdosta. – Virittäjä, 101(2), 208-232.

Vanden Wyngaerd, E. 2017, The adjective in Dutch-French codeswitching: Word order and agreement. – International journal of bilingualism, 21(4), 454-473.

Арискин, Н. И. 1993, Социальный аспект функционирования мордовских языков. – Инженерные технологии и системы, (2), 20-22.

Гаврилова, В. Г. 2012, Дублирование как одно из проявлений переключения. – Финно-угорский мир, 3/4, 56–59.

Гаврилова, В. Г. 2013, Марийско-русское переключение и смешение кодов – Вестник Удмуртского университета, Серия «История и филология», 2. 16–22.

Зализняк, А. А. 2002, «Русское именное словоизменение» с приложением избранных работ по современному русскому языку и общему языкознанию, Москва: Языки славянской культуры.

Куклин, 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, 10(2-1), 92-96.

Максимов, С. А. 2017, Числительные в пермских языках: прошлое и настоящее. – Ежегодник финно-угорских исследований, 11(3), 7–13.

Мельчук, И. А. 1985, Поверхностный синтаксис русских числовых выражений, Institut für Slawistik der Universität Wien.

Пинеда, Д. 2007, К проблеме «парадокса наблюдателя» и переключение кодов. – Полярный вестник, 10, 40–52. doi.10.7557/6.1289

Плешак, П. С. 2017, Морфосинтаксис именной группы в мокшанском и горномарийском языках, диплом бакалавра, манускрипт.

Сааринен, С. 2014, Переключение кодов в мокшанских диалектных текстах. – Н. Г. Зайцева, И. И. Муллонен и др., V Всероссийская конференция финно-угроведов «Финно-угорские языки и культуры в социокультурном ландшафте России»: Материалы, Петрозаводск, 540–543.

Сидорова, М. А. 2018, Числовое маркирование существительного в горномарийских количественных конструкциях. – Acta Linguistica Petropolitana, 14(2), 400–427.

Сидорова, М. А. 2018, Количественные конструкции. С. Ю. Толдова, М. А. Холодилова, Элементы мокшанского языка в типологическом освещении, М.: Буки Веди, 328-341.

Шабыков, В. И., Кудрявцева, Р. А. 2017, Языковая ситуация в Республике Марий Эл в начале 2010-х годов: социолингвистический анализ. – Урало-алтайские исследования, 3, 208–229.

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 бо́льших 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 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 матричного 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 условии для переключения кодов.

Addresses

Irina Khomchenkova

Russian Language Institute, RAS

irina.khomchenkova@yandex.ru;

Polina Pleshak

University of Maryland, College Park

ppleshak@umd.edu