Uralic Studies PhD Program

Graduate School in Linguistics

University of Szeged

Boglárka Janurik

Erzya–Russian bilingual discourse:

A structural analysis of intrasentential code-switching patterns

PhD dissertation

Supervisor:

Magdolna Kovács, PhD, Adjunct Professor

Szeged, 2017

To my grandparents

Table of Contents

Acknowledgements

List of abbreviations

1	first person
2	second person
3	third person
1PL<3PL	first person plural object, third person plural subject
3PL<1SG	third person plural object, first person singular subject
3SG<1SG	third person singular object, first person singular subject
3SG<3SG	third person singular object, third person singular subject
ABESS	abessive
ABL	ablative
ACC	accusative
CCM	Code-Copying Model
COMP	comparative
COND	conditional
CONJ	conjunctive
CONNEG	connegative
CP	complementizer phrase
CS	code-switching
DAT	dative
DEF	definite
EL	embedded language
ELA	elative
EMP	emphatic
F	female
F1	fieldworker number one
F2	fieldworker number two
F3	fieldworker number three
FL	fused lect
FREQ	frequentative

GEN	genitive
GEN2	second genitive
GER	gerund
ILL	illative
IMP	imperative
INE	inessive
INF1	first infinitive
INF2	second infinitive
INS	instrumental
LAT	lative
LM	language mixing
M	masculine
ML	matrix language
MLF	Matrix Language Frame Model
N	neutral
OPT	optative
PART	partitive
PL	plural
POSS	possessive
PREP	prepositional
PROL	prolative
PST	past
PST2	second past
R1	reporter number one
R2	reporter number two
R3	reporter number three
REFL	reflexive
SG	singular
TRA	translative
VOC	vocative

Chapter 1. Introduction

This dissertation focuses on the structural description of code-switching (CS) patterns in Erzya–Russian bilingual discourse. I define CS as the use of elements from two languages in the same discourse or within the same utterance. Erzya (or Erzya-Mordvin) is a Finno-Ugric minority language spoken in the Russian Federation. According to the UNESCO Atlas of the World's Languages in Danger (2016) it is a *definitely endangered* language with approximately 400,000 speakers (I discuss the unreliability of the census data in Chapter 2 in detail). All the speakers are bilingual in Erzya and Russian, and their language use varies according to the extent of CS to Russian. While spoken discourse and informal written genres typically contain CS, contemporary formal written discourse lacks Russian elements, and CS is especially avoided in media products (Janurik 2016). This, however, was not always the case (cf. Hallap 1960). In the present paper, I refer to heavy CS publications from the mid-20th century which have possibly contributed to the widespread bilingualism and extensive CS of the Erzyas nowadays (cf. section 4.2).

Heavy CS is usually considered a transitory phase in contact situations (Bakker 2003). My dissertation aims to describe the characteristics of this in-between state and analyze the morphosyntactic structure of intrasentential code-switches in Erzya–Russian bilingual utterances (in two different data sets described in detail in Chapter 6). Although code-switching models (cf. Chapter 5) generally allow for hybrid constructions, they provide a framework that is more suited for the analysis of prototypical CS cases in which the dominance relations of the two languages involved can be identified unambiguously in all of the utterances. In Erzya–Russian CS data, there are hybrid or mixed constructions which defy the rules of both (standard) languages and show signs of an evolving contact variety. The study of this bilingual language use and the variation of mixed constructions can reveal ongoing (contact induced) changes and help us to understand what happens in general in contact situations involving typologically different languages which have converged during a prolonged period of contact. Zabrodskaja (2009: 32) describes the connection between convergence and CS as follows: "It should be clear that CS and convergence are related in the sense that both are linked to the use of another language. I state that CS helps to account for what morphosyntactic structures are more likely to be converged and why".

Apart from some sporadic articles (Hallap 1960, Mosin 2002, etc.) and my preliminary studies (Janurik 2011, 2013, and 2016), research on Erzya–Russian language contact has so far focused mainly on lexical borrowing. If the CS discourse was mentioned in these earlier papers, the authors regarded it corrupt Erzya language use (cf. section 4.2). It is paramount at this point in time to describe this phenomenon because, depending on sociolinguistic factors, the CS (language use) can give way either to language shift or to the strengthening of a more monolingual, "purified" Erzya variety. (In Chapter 2, I argue for the existence of a code-switching Low variety which is in a diglossic situation with the semi-monolingual High variety.) The heavy CS phase is generally short-lived in all contact situations, and the third possible outcome, the fossilization of a mixed language is the least typical one. A study of CS can yield useful insights into how typologically different languages "manage" CS. With this case study, I intend to provide information for contact linguists and typologists on the mechanisms of CS in the Erzya–Russian contact situation.

In my dissertation, I adopt a grammatical perspective although I am aware of the fact that CS is a multifaceted phenomenon, and agree with Backus (2015) and Muysken (2000) that only an interdisciplinary approach can yield an explanation of how CS really works. However, I also argue that a structural analysis as the first step can reveal the points that require further study involving quantitative research and not only naturally occurring data but also elicited sentences.

So, although I agree that there are many (structural, pragmatic, psycholinguistic, and sociolinguistic) factors responsible for variation of CS types in the Erzya–Russian data, my hypothesis is that structural (in)congruence between the constructions in the two languages is one of the main causes for the emergence of mixed utterances, and the variation partly results from speakers applying different coping strategies. My aim is to describe how these incongruent constructions are switched and which switching strategies of the speakers are prototypical and which are idiosyncratic.

There is only one exception to this predominantly structural approach, I devote a separate section (7.3) to the analysis of flagged switches. In these cases, the speaker signals the point where the switching takes place. Some of these examples of flagging are results of the incongruence between the structures of the two languages, and speakers cannot produce smooth

switches in these cases. However, flagging occurs in my data also at switch points which should not present any problems of processing to the speakers, for example, in cases involving lexical insertions and even established Russian loanwords. These instances are due to pragmatic and sociolinguistic reasons, partly to puristic tendencies and more conscious language use in formal situations.

A major aim of this dissertation is to represent the variation attested in the use of CS strategies. In order to give an account of this diversity, I use an Erzya–Russian continuum model based on earlier continuum models (Auer 1999, Kovács 2001). Different points on this continuum represent the various CS strategies attested in the Erzya–Russian data. I assign speakers to categories and to given points on the continuum on the basis of the types and amount of CS they use. With this qualitative study of the data, I intend to make provisions for later quantitative research of these CS types.

As a tertiary aim, I intend to show Erzya linguists how complex and rule-governed Erzya–Russian CS is, and that parallel existence of more varieties makes the language richer and more diverse. A further sociolinguistic study of the language choice preferences of Erzya–Russian bilinguals would be needed to reveal whether the stigmatizing of CS has contributed to language shift, and if yes, to what extent.

In order to achieve the above mentioned aims, my dissertation pursues the following research questions:

(1) What are the main types of insertional switches in the two Erzya–Russian data sets?

(2) Which CS constructions can be described with the binary model of the Matrix Language Frame model (MLF) (Myers-Scotton 2002)? In which cases does this binary model break down? Are these constructions instances of congruent lexicalization (Muysken 2000)? Can these cases be explained by the incongruence of the constructions in the two languages? What are the main strategies speakers use to realize CS in cases where the structures in the two languages are incongruent?

(3) How are CS and contact induced change connected? Are there any cases in which certain CS types might be indicators of ongoing change?

(4) Can we define different CS styles? Is there a correlation between the amount and types of CS in speakers' language use?

The dissertation is structured as follows. After this brief introduction, Chapter 2 presents the sociohistorical background of the contact situation. I discuss basic information and statistical data on the Erzya speech community (section 2.1), focus on the present situation of Erzyas (section 2.2), and give an overview of the main historical events that influenced Erzya–Russian relations (section 2.3).

In Chapter 3, I provide a brief typological description of the two language systems, focusing primarily on morphosyntactic features, especially on genitive constructions and verbal morphology because these are the structures showing the greatest level of mixture in Erzya–Russian CS.

Chapter 4 is connected both to Chapters 2 and 3, as it discusses CS research conducted among Finno-Ugric peoples living in the Russian Federation (section 4.1), and earlier studies of Erzya–Russian bilingualism and CS (section 4.2). As some of the contact phenomena seem to be present in all contact scenarios, I suggest conducting a comparative study which could reveal further similar CS patterns and mechanisms.

In Chapter 5, I provide an overview of the theoretical background of CS, and discuss a variety of questions connected to CS studies: the definition of CS and its relation to borrowing (section 5.1), the question of constraints and the existence of a matrix language (section 5.2). In section 5.3, I discuss the frameworks I use in my analysis. These include a binary model differentiating between a dominant and an embedded language (Myers-Scotton 2002), a typology applying an interdisciplinary approach (Muysken 2000) and frameworks connecting code-switching to contact induced change (Johanson 2002 and Backus 2015). After that, I discuss the role of equivalence and convergence in CS (Sebba 2009), and Gafaranga's 2000 model, which provides a framework for the analysis of the flagged switches. In section 5.4, I focus on two continuum models, Auer's 1999 and Kovács's 2001 models, which provide an apt tool for the demonstration of the variation attested in Erzya–Russian CS patterns.

In Chapter 6, the data and methodology are described. I discuss the characteristics of the two data sets used in the analysis, reflect on problems encountered during fieldwork, and explain

advantages and disadvantages of using this type of data for the analysis of CS. I also refer to the transcription principles I applied, and why I decided on using the Finno-Ugric Transcription (FUT) for both the Erzya and the Russian elements of the corpus.

Chapter 7 contains the main body of the paper, the analysis of the data. I use Muysken's typology for the categorization of the CS patterns, I distinguish between insertions, congruent lexicalizations, and alternations, but discuss instances of alternations separately (in section 7.1), as alternation involves CS mechanisms different from the ones at work in case of insertions and congruent lexicalization. In further subsections, the analysis of utterances with insertions and congruent lexicalization follows. First, I focus on discourse particles (section 7.2.1), then on nominal constructions (section 7.2.2), and finally, on verbal constructions (section 7.2.3). As mentioned earlier, flagged switches as a special type of insertions and alternations are discussed separately (in section 7.3). I catalogue the main types of flagged switches attested in my data and provide a possible explanation for their emergence.

In Chapter 8, I discuss the attested CS patterns and focus on the categorization of Erzya–Russian bilinguals on the basis of their language use. I place the speakers on a continuum (in section 8.1) which facilitates the description of the transition from a monolingual variety to a mixed variety, and of variation in CS in general. Section 8.2 discusses the main characteristics of Erzya–Russian bilingual discourse, whereas section 8.3 outlines future research perspectives. As my study provides only qualitative data on Erzya–Russian CS, I argue that further studies could focus on the quantitative analysis of the CS types. By applying an interdisciplinary method described in Backus's 2015 usage-based model, an answer could be provided for the question what kind of sociolinguistic factors are behind the attested variation in bilingual Erzya–Russian speech.

Chapter 9 sums up the results of the paper chapter by chapter.

Finally, the appendix contains a listing of the interviews with the number of the recording and the categories to which I assigned the speakers on the basis of their switching types.

Chapter 2. Sociohistorical background of the Erzya–Russian contact situation

In this chapter, I discuss the relationship between Erzya(-Mordvin) and Moksha(-Mordvin), the main characteristics of the Erzya speech community, and the historical and present aspects of the Erzya–Russian contact situation.

Section 2.1 elaborates on the dialect versus language debate concerning the Mordvin varieties. Section 2.2 analyzes the present situation, while 2.3 section dwells upon the history of the contact situation.

2.1. Mordvin or Erzya and Moksha?

Erzya is a Finno-Ugric language spoken in the Russian Federation. Along with Moksha, it belongs to the Mordvin subgroup of Finno-Ugric languages. The dialect versus language statuses of Erzya and Moksha have been long debated. According to Bartens (1999: 10), it is mainly foreign linguists who argue for Erzya's dialect status, whereas native speakers predominantly prefer to call their variety a language. (The picture is even more complex if we take into account Shoksha, considered by some of the linguists a dialect of Erzya, whereas, by some of the speakers, a language in its own right.)

In this paper, I discuss Erzya and Moksha as separate languages, taking into consideration both the preference of the speech communities and the diverging development of these varieties since their separation, which occurred in the 13th century, or even earlier, according to Kreindler (1985a: 237) and Lallukka (1990: 53), in the 6th century at the latest. Furthermore, two standard varieties (standard Moksha and standard Erzya) were established for the two in the 1920s, partly due to the nature of Soviet language policy, but mainly due to the separatist intentions of the Moksha and Erzya intelligentsias. This separation had both positive and negative consequences for the subsequent development of the languages, and it has been judged in a variety of ways by the community members. (I discuss this further in section 2.3.)

In the 1920s, there were also efforts to create a unified standard which would have been based on a mixed dialect spoken in the Samara region (Agafonova 1995). Moreover, this new

standard would have been written using the Latin alphabet. Finally, both the idea of the unified standard and that of the Latin script were rejected, partly due to political reasons. Ostler (2006: 442) mentions that "all the different nationalities (except for the Baltics, Georgian, Armenian and Yiddish) came to declare in favour of switching their orthographies to some variant of the Cyrillic alphabet used for Russian."

Rueter (2010: 6) refers to another attempt for the creation of a unified Erzya-Moksha standard or at least a common vocabulary in the late 1980s when the Bible was being translated into the Mordvin languages. However, the unification failed again, as the distance between the two languages was considered insurmountable, and translators feared that the new vocabulary would scare away readers. Rueter also argues that projects aiming to create a common standard might jeopardize the official language statuses of Erzya and Moksha and could lead to the loss of funding for their research and development.

The split does not only concern the languages but, to some extent, also the cultures, traditions, and the institutions of the two ethnic groups. According to Salo (1991: 158), this separation has a long history, as Erzyas and Mokshas rarely intermarried before Christianization (in the 18th century), and the tradition continues to this day. In some organizations, there is tangible ethnic tension between the two groups, usually more radicals being on the Erzya side. Taagepera (1999) refers to anti-Moksha statements in Erzya newspapers and also mentions that the nationalist organization Erzyan' Mastor wanted to create an Erzya national district within Mordovia and "[i]ts leaders strongly opposed the term 'Mordovia' and even 'Moksherzia' because they wanted nothing to do with the Mokshans" (1999: 190–191).

The separation of the two communities is also institutionalized in the Mordvin Republic. This policy is clearly manifested by the establishment of two departments, one for Erzya and one for Moksha, for the study of these languages at the university in Saransk, the capital of the republic. The media is also divided between Erzya and Moksha newspapers, magazines, and broadcasting. The degree of cooperation between them varies from just sharing the same building to sharing information and running joint projects.

Nonetheless, within the community, there are still supporters of a unified language who are afraid that the division of the Mordvin people will facilitate Russification (Prina 2016: 143). As

seen above, there is a debate concerning the name of this unified language, Mordvin, "Erzya-Moksho", and "Moksherzya" being the most popular variants. This naming problem arises because there is no common word in these languages which would refer to the two groups together. Mordvin is an exonym with a long history, mentioned first as *Mordens* in the 6th century Gothic historian Jordanes's codex. This exonym has been used mainly in Russian discourse by Russian authorities and, to some extent, also in the Western tradition (Rueter 2013: 2). Erzyas and Mokshas, however, avoid this term in native discourse in designating themselves, due to its derogatory status.

Supporters of the unified language coined the term *Moksherzya* and use it as an equivalent of the Russian word *Mordvin* for both the two ethnic groups and the two languages together. Usually, Mokserzya activists intend to unify the two communities and facilitate language maintenance by creating a homogenous linguistic community and one standard variety. According to Salo (1991: 158), support for the unified language is a characteristic of urban intellectuals. These efforts are met with strong opposition by other community members.

Interestingly, many foreign (especially Hungarian) linguists also support the idea of unification and insist on the dialect status of the two varieties. Maticsák (2012: 3) suggests that the division between the varieties comes handy for the Russian authorities, who encourage this split by applying "the divide and conquer" principle. Keresztes (2011: 129–132) even gives practical guidelines along which the unified standard Mordvin variety should be created. The parallel use of the exonym by outsiders and the endonym by members of the community has been going on for centuries. In the first written account of Erzya, Strahlenberg's word list from 1730, the language and the people are designated as Mordvins, although from the word forms it is clear that the material was collected only in an Erzya-speaking area (Feoktistov and Saarinen 2005: 13–14).

Different levels of mutual intelligibility also contribute to the failure of unifying intentions. (Opinions vary on the extent of this intelligibility, cf., for instance, Keresztes 2011: 132 or Rueter 2010: 7–8.) It is usually only Erzya and Moksha speakers who live in mixed areas or have been otherwise exposed to the other variety that can communicate with each other without difficulties. Although individual coping strategies (e.g. the use of constructions and vocabulary

the two languages share) exist, the typical method for extra-group communication is the use of Russian. The same is the case in Erzya–Moksha mixed groups and even in marriages. During Soviet times, this tendency seems to have been welcomed by the Soviet ideologues. As Isaev (1977: 133) writes, "a unifying influence on both languages is exerted by Russian which is becoming a second mother tongue to an increasing number of both Erzya and Moksha". Haarmann (1998) argues that another motivating factor for learning and using Russian has been social mobility. Although Soviet ideologues tried to keep this a secret, people learnt Russian to gain access to social goods and power, because "social advancement meant a higher and specialised education, access to higher and better-paid posts, access to influential positions in the party hierarchy (i.e. access to the domain of the nomenclature), and access to social and economic privileges relating to such positions" (Haarmann 1998: 249).

This lack of clarity as regards the use of the different designations causes additional problems. It is impossible, on the basis of statistics, to define the number of ethnic Erzyas and of Erzya speakers because census data do not differentiate between the Erzya and Moksha (speakers). It did only in 1926 and in a 1994 mini census, but even in the latter, instructions to differentiate the two ethnic groups were only given in the Mordvin Republic, not in the diaspora settlements (Grenoble 2003: 80). According to the latest, 2010 Russian census (Russian census 2010), there are 744,237 ethnically Mordvin people (Moksha and Erzya) in the Russian Federation. It is also difficult to estimate the number of Erzya speakers, as the census data provide only a common number (392,941) for the speakers of the two Mordvin languages. Some of the speakers explicitly declared which speech community they belong to (36,726 Erzyas and 2,025 Mokshas). According to the 1926 census data and informal estimates, the situation remained the same since then, Erzya speakers are in a majority, constituting about two-thirds of all Mordvin speakers (Bartens 1999, Keresztes 2011, Rueter 2013).

2.2. Present situation of the Erzya

In discussing both the present situation and the history of the contact situation, my focus is on the Erzya language and speech community. I also mention Moksha in some of the places where

commenting on both of the languages is relevant or they show similar patterns of development, but my main aim is to describe the sociohistorical background of the Erzya–Russian contact situation.

Along with Russian, Moksha and Erzya have both been *de jure* official languages of the Mordvin Republic since 1995. However, this does not mean that they have a *de facto* equal status with Russian. In principle, Mordovia is a multilingual republic and part of a plurilingual state, but as Prina (2016: i) argues, "the Putin leadership's overwhelming statism and promotion of Russian patriotism are inexorably leading to a reduction of Russia's diversity". Prina cites data from the 2010 Russian census and draws attention to the fact that 80.9% of the respondents identified themselves as Russians, and 99.49% claimed to speak the Russian language, even when they belong to the other 193 ethnic groups, or have one of the other 169 languages as their mother tongue. Although Russia and the Soviet Union have had a history of diversity, and Russian policy and laws often refer to the country's multinational character, assimilation is still a strong ongoing process, partly due to globalization, partly due to the homogenizing tendencies mentioned above, e.g. the requirement of the Cyrillic alphabet as a prerequisite for the status of an official language. According to Prina (2016), one of the main problems is that "policies for the enhancement of linguistic and cultural diversity are generally dependent on the discretion and goodwill of public officials, and may be easily neglected" (2016: 4).

Another problem is that Erzya and Moksha people constitute a minority within the republic, both ethnically and linguistically (Keresztes 2011: 11): Mordvins make up only 33% of the population. There is a fluctuation in the census data, depending on the regional policy and the extent of centralization attempts, the number of inhabitants identifying themselves as Erzyas and Mokshas varying in different census years. The way the question of ethnic origin is formulated in the census can also influence the results. For instance, census makers using *Mordvin*, a term widely considered derogatory, might also have caused speakers associating themselves with the Russian speech community instead (Bartens 1999: 11).

As regards the Mordvin Republic, there was a decrease in the number of the Mordvin population in 2002 as compared to the 1989 census data, followed by an increase in 2010. Using Zamyatin's 2012 results, Prina claims that "this change is to be attributed to an increase in the

number of persons self-identifying as Mordovians in the censuses, possibly due to shifts in individual approaches to ethnic selfidentification as a result of regional policies" (2016: 37). This increase in the titular republic can be explained by the policy of the census takers aimed to distinguish Moksha and Erzya (Grenoble 2003: 80).

The other obstacle for language maintenance is the fact that only 27% of the Mordvin speakers live in the Mordvin Republic, with the rest of the speech community being in a diaspora situation (Bartens 1999: 10). Most of these Erzyas live in other parts of the Russian Federation (mainly in the Ulyanovsk, Penza, Samara, and Orenburg Oblasts) and in the now independent states of the former Soviet Union such as Kazakhstan, Armenia, etc.

Positive changes, such as the occasional increase in the percentage of the minority population within the Mordvin Republic, does not change the fact that Erzya and Moksha are endangered languages, with the number of speakers constantly decreasing. They belong to the 131 languages spoken in the Russian Federation that are in danger, both of them classified as 'definitely endangered' (UNESCO 2016). If we compare data from the 2002 census (Russian census 2002) and the 2010 census (Russian census 2010), we can see a radical decline in the number of Mordvin speakers: 614,260 in 2002 and 392,941 in 2010. According to Bartens (1999: 11), the number of Mordvin speakers was still 1,023,841 in 1897, and it gradually increased until the 1939 census. After World War II, however, the number of speakers decreased despite the fact that fertility rates were relatively high among Erzyas and Mokshas as well. Consequently, the decline can only be attributed to language shift to Russian.

Still, as Comrie (1981) points out, in the 1980s, Mordvins were the most numerous Uralic people in the Soviet Union, slightly outnumbering even Estonians. Census data, however, can be misleading, "despite the large absolute numbers, the Mordva live scattered across a huge territory, where they are almost everywhere a minority among a Russian majority" (Comrie 1981: 102), which explains the lower and decreasing language maintenance rate among the Erzyas and Mokshas (78.1% in 1959 and 77.8% in 1970). This tendency has continued after the Soviet times to the present.

There is a complex set of factors behind the language shift process. The main problem concerns the transmission of the language to the next generation. Especially in urban settings,

parents are inclined to choose Russian as the language of the home in order to provide children better opportunities in a mostly monolingual Russian environment.

Rueter (2013) gives an overview of the present situation concerning the settings in which the Erzya language is spoken. He differentiates between three types of settings: the village, the settlement (*posyolok*) and the city. In the village, the proportion of monolingual speakers (elderly women and pre-school children) is highest, non-official gatherings are conducted in Erzya. At school, Erzya can be the medium of instruction in the first four grades if an Erzya-speaking teacher is available. The problems in connection with the educational system are discussed in detail in this section below.

The *posyolok* is a bigger settlement with a mixed population. Although Erzyas might be in the majority, they come from different dialect backgrounds and live among various non-Erzya people. As a result, the default language of communication is Russian in these places. Erzya is spoken only at home, and children usually acquire a passive knowledge of Erzya in mixed marriages.

Although the number of Erzyas reaches thousands in the cities, their proportion usually stays under 10% (Rueter 2013: 5). If parents do not make an extra effort to create an Erzya language environment for their children, the next generation will only have a passive knowledge of the language. Rueter (2013) mentions that cities have a tradition of intolerance towards minorities. This is especially true in the Mordvin Republic and can be explained by its proximity to Moscow. Rueter also argues that "[t]he lower toleration of minority languages in Mordovia might readily be associated with the exodus which began with the fall of Kazan' in 1552; those who wished to retain their way of life crossed the Volga" (2013: 6).

Speakers of the Mordvin languages who moved from rural areas to Saransk, the capital of the republic, or to Moscow, consider their languages useless and remnants of an outdated lifestyle. The low prestige of the Mordvin languages can also be explained by historical factors, especially by the language and educational policy of the Soviet Union (cf. section 2.3 below). Purist attitudes also have a controversial effect. On the one hand, intellectuals provide a model for "pure", i.e. monolingual language use, by avoiding Russian loanwords and constructions and by propagating neologisms. On the other hand, disparaging heavy code-switchers and speakers

with a limited competence in Erzya and Moksha accelerates language shift to Russian and inhibit the transmission of the minority languages to the next generations.

The variation in the language use of Erzyas can be characterized as a type of diglossia with a semi-monolingual High variety used predominantly by intellectuals and a Low variety involving moderate to heavy switching. These issues and their relation to language endangerment will be discussed in Chapter 8 in detail.

Besides these factors, mixed marriages (even between Erzyas and Mokshas) also contribute to language shift. In such cases, parents in Russia typically choose Russian as their language of communication and also as the first language for their children (Haarmann 1998: 247).

The media is another domain where the Erzya language is used. A limited number of radio and television programs is available in Erzya, and the printed media products (including daily newspapers, literary journals, and children's magazines) also have an important role in creating possibilities for Erzya language use. The Internet provides new opportunities for the maintenance of minority languages in general. Speakers can create new materials and involve in the discussions Erzyas living at great distances from each other. The Internet can improve the accessibility of Erzya materials, and even the Erzya speakers living in the diaspora have a more direct access to the standard. One of these positive changes brought about by the Internet is the availability of radio programs, not only in the Mordvin Republic, but anywhere in the world. (This is the case with Radio Vaygel, although there are occasionally problems with the regular upload of the programs to the website.)

Education plays a vital role in the language maintenance process of the Erzya speech community. Prina (2016) cites 2010 Russian government data, according to which, in that year in Russia there were 200 schools with Moksha or Erzya as the language of instruction and 275 with Erzya or Moksha as a subject. The educational reform of the beginning of the 2000s brought about negative consequences for the minority languages of Russia. Zamyatin (2012) compares these reforms to the changes implemented in the years 1958–1959 (see section 2.3 for details) when the study of minority languages became non-compulsory. The minority languages have a disadvantage over Russian because of global economic tendencies favoring languages of

international use. Prina (2016: 180) also emphasizes that high level proficiency in Russian is required for the school-leaving exam (EGE), and parents want to provide their children with a good education and save them from the extra burden of studying yet another language.

Finally, the degree of language maintenance depends on one more factor. Haarmann argues that there is a significant difference between male and female speakers in this respect:

"language maintenance is the strongest among women living in rural areas and the weakest among men in urban surroundings – e.g. 90.9% for rural women versus 53.2% for urban men in the Mordvin community. […] It can be conjectured that the stronger tendency toward language maintenance among the female population [in Russia] relates to the fact that non-Russian women are more susceptible to and show a greater readiness for preserving traditional cultural patterns (language for one) than do men. Their attitude is apparently characterised by a marked protective and preservative spirit (1998: 247).

This is an intriguing question and it needs further investigation, whether this is still the case in the Erzya speech community. Earlier accounts of Erzya/Moksha–Russian CS partly contradict these statements. On the basis of his findings during fieldtrips to Erzya and Moksha territories in the 1950s, Hallap (1960) writes that (especially Moksha) women use a lot of Russian constructions, even though they are not fluent in Russian. However, Hallap also mentions that men's speech contains even more Russian elements due to their better Russian proficiency, which is in accordance with Haarmann's data.

2.3. History of the Erzya–Russian contacts

The location of the Mordvin settlements has significantly changed over the centuries, especially what concerns the Erzyas. The ancestors of the Mordvin people lived to the south and southwest of the areas where they do today. Erzyas lived in the Nizhniy Novgorod area and had contacts with the Finnic languages, whereas Mokshas, who occupied lands more to the south, in the area of the contemporary Mordvin Republic and to the west, were influenced by Iranian and Turkic

tribes (Taagepera 1999). The cities of Arzamas and Ryazan' are both supposed to have been named after the Erzya population which settled in the area (Taagepera 1999: 152).

Murom and Merya, now extinct Finn-Volgaic languages, were among the first Finno-Ugric languages that came into contact with Russian. There are very little data on these languages, and their speakers are believed to have assimilated to the Russians by the 18th century at the latest.

On the basis of toponymic evidence (Kreindler 1985b: 44), first contacts between the Mordvin languages and the Russian varieties are estimated to have taken place as early as the 9th century. After founding the city of Murom in the 11th century, Russian settlers started to move northwards and eastwards into the Mordvin areas, and contacts and trade connections intensified in the 11th century (Keresztes 2011: 15). First contacts are claimed to have been peaceful (Kreindler 1985a: 238), followed by fierce battles in the 12th century. A Russian invasion or slow infiltration was for a long time prevented by the resistance of the Mordvin peoples and the strong Volga Bulgarian Empire (Taagepera 1999: 143). The first bloody encounter of Mordvins and Russians was recorded in the Russian Chronicles as well, dating back to 1103. (In 2012, these series of invasions were celebrated in the Mordvin Republic as the unification of the Mordvin people with other peoples of Russia.) Starting from this period, the contact situation was characterized by an assimilation of the Mordvin groups to the Russian majority. As a result, there were no significant Russian-induced changes in the lexical or structural characteristics of either of the Mordvin languages at this time.

There were no natural barriers and powerful neighbors who could prevent the Russians from expanding their territory eastwards (Kreindler 1985b: 45) and from founding new settlements in the conquered lands, e.g. the city of Ryazan in the end of the 12th century and the stronghold of Nizhny Novgorod in 1221. The latter became the starting point for further conquests (Kreindler 1985a: 238). In the 13th century, however, Russian influence was replaced by Tatar dominance, which lasted until 1552, the fall of Kazan.

The 16–17th centuries brought along the restructuring of the Mordvin-inhabited areas. While Tatars were mainly interested in collecting taxes, Russians moved into traditionally Erzya areas and forced the indigenous people to migrate eastwards, thereby depopulating entire villages (Kreindler 1985a: 239). Migration was also triggered by army service, high taxes and the

spread of the Orthodox Christian faith (Keresztes 2011: 15). Russians used colonial strategies in purchasing the lands "legally" from Mordvins who could neither speak Russian, nor knew the laws and regulations (Taagepera 1999: 157). The lands were given to the tsar's boyars and the church (Nagy 2004: 93).

The state encouraged and backed the church's missionary plans. Due to these Russian administrative and religious policies of the late Middle Ages, Mordvin groups moved eastwards to the Transvolga region and beyond. By the 17th century, Erzya and Moksha people already constituted a minority in their own lands. "According to the census of 1762–66 the number of those living in the Mordvin lands was about 334,000 with 25% Mordvins, 70% Russians and 5% Tatars" (Nagy 2004: 93). The migration also had its linguistic consequences, as it resulted in the mixing of the dialects, and in some places an Erzya–Moksha mixed variety emerged (Salo 1991: 162).

The Mordvin communities could not avoid Christianization. With its community scattered, the Mordvin population was more vulnerable to assimilation. In the 1580s, several monasteries were built in their areas, to serve as the bases for the missionary work carried out among Mordvins and Udmurts (Salo 1991: 162). In the 17th and 18th centuries, Mordvins participated in many peasant uprisings, along with Maris and Udmurts, and these events left their traces in Mordvin epic songs, mainly about Pugachev, leader of the 1773–1775 insurrection (Salo 1991: 163). These uprisings resulted in retaliations. After the sacred woods were destroyed and indigenous leaders and elders killed, both the Moksha and the Erzya were officially converted to the Orthodox faith by the 18th century. This process is usually considered a tragic event in the history of the Mordvins, but there are scholars (e.g. Maticsák 2012) who also emphasize the benefits of Christianization. Maticsák (2012) claims that the translation of liturgical texts and the establishment of the Kazan school where Mordvin people could study came about because the Orthodox Church wanted to gain more popularity. Elements of the old religion faded so quickly that Kuzma Aleksejev, a Mordvin prophet who wanted to revive the animist religion in 1808, could not reconstruct it any more, as he found that the Virgin Mary, the angels, and the saints had become integral parts of people's belief system (Kreindler 1985a: 239). It also made the resistance to Christianization difficult that the Mordvins did not develop their own animist church with any

clergy or writing system, and the attempts of Kuzma Aleksejev were discouraged by the state (Kreindler 1985b: 46).

Christianization had two severe consequences for the Mordvin people. First, being Orthodox, they were not regarded as culturally different from Russians any more, as opposed to the Udmurts or the Maris, who maintained their animist faith. What is more, Mordvins, especially Erzyas, had similar physical traits as Russians, therefore they could blend in more quickly (Kreindler 1985b: 47). As a result, Mordvins were not entitled to the same special offers (e.g. mother tongue education) as other originally animist peoples were (Taagepera 1999: 165). Second, Christianization resulted in the increase of literacy among Mordvins who acquired reading and writing skills in Russian. (Mordvins did not have a writing system before.) In this way, the number of Mordvins with knowledge of Russian also increased, and the Russian influence on the Mordvin languages intensified. The Russian language spread so quickly that some of the folk songs and even animist prayers had been preserved only in Russian (Kreindler 1985a: 240). As evidenced by tax records, Mordvin names were replaced by Russian ones as early as the end of the 17th century. Kreindler (1985a: 241) mentions that already in the first accounts of the Moksha (1692) and the Erzya language (1787) numerous Russian loanwords are represent (cf. Maticsák 2012, Chapter 5), with Erzya showing stronger Russian influence and a higher degree of phonological integration: e.g. Russian *sťena* 'wall' was borrowed as Moksha *osťena* 'wall', whereas Russian *sčasťje* 'luck, happiness' as Erzya *ućaska* 'luck, happiness'.

Besides the loanwords, consonants were borrowed into Erzya from Russian, and syntactic changes also occurred. For instance, Russian conjunctions were borrowed, and complex Erzya verbal constructions disappeared (Salo 1991: 164).

According to Kreindler (1985a), Mordvins were reported speaking two mixed varieties in the 19th century: the Erzyas used a mixture of Erzya and Russian, while Mokshas a mixture of Moksha and Russian. The knowledge of Russian was high in the whole speech community, especially among men who worked outside the villages. As a result, religious books were not translated into the Mordvin languages, and mother-tongue education had also been deemed unnecessary. While the number of speakers was constantly decreasing, more than one million people regarded one of the Mordvin languages as their mother tongue in the 1897 census.

The end of the 19th century was characterized by Russian nationalism and a policy of Russification, but as Comrie (1981: 21) points out, "there was no explicit plan envisaged or executed to enable non-Russians to learn Russian and assimilate, and of course no possibility was provided for them to develop within their own culture, in their own language."

This was the period of national awakening, led by the missionary Ilminsky, a linguist and the head of the Orthodox Church's Translating Commission. His goal was to bring Orthodoxy closer to non-Russian speakers by translating religious texts into their mother tongues. He believed that the high level of Russian proficiency among Erzyas and Mokshas was exaggerated, and that there was an actual need for these texts to be translated (Salo 1991: 165). As there were no standards for Erzya and Moksha, they had to be thoroughly studied before translation could begin. Some of the texts were translated into varieties now considered to be (Erzya or Moksha) dialects. The Cyrillic alphabet was used for all the translations, for Moksha with additional characters. The alphabet for Erzya was the same as the one used for Russian, because the sound systems of the two languages are rather similar (as described in section 3.1 below). Ilminsky also published folklore texts and primers (Kreindler 1985b: 48), created schools (although with Russian as the language of instruction), and sought help among the Erzya and Moksha to work on the Mordvin languages. While earlier educated Erzyas and Mokshas concealed their origin and tried to make a career as Russians (Kreindler 1985b: 48), Ilminsky's students and colleagues did not try to assimilate into the Russian community, instead organizing a basis for the Mordvin intelligentsia in the beginning of the 20th century.

Despite the problems (such as a scattered population, small elite, illiteracy, and Erzya and Moksha conflicts), the new elite managed to create two standard varieties to be used in the written media and to establish an education system up to 10th grade with Erzya and Moksha as the medium of instruction by the mid-1930s. (Earlier the medium of instruction was Russian, with the Mordvin languages taught as school subjects.)

Status and corpus planning were carried out simultaneously, which resulted in a chaotic situation, with new schools having the mother tongue as the medium of instruction but facing also the scarcity of teachers and textbooks. The quality of textbooks improved as the standardization of the languages went on. Another problem was that the scattered Mordvin

population did not have an administrative center until the 1934 formation of the Mordvin Republic, with the Mordvin Autonomous District, formed in 1930, as its predecessor (Nagy 2004: 94). Kreindler (1985b) also points out that in the 1920s, the main meeting places of the Erzyas and Mokshas were Moscow, Simbirsk, Samara, and Saratov, but not Saransk. By the time the autonomous district was created, these cultural centers had lost their power due to the centralizing policy in the 1930s. Mordvins were in a minority in all the provinces and districts, and Stalin especially disliked the idea of Mordvinian autonomy (Kreindler 1985b: 49).

Positive changes in the status of the Mordvin languages can be attributed to Lenin's nationality policy and the activity of a small elite. Schwarz (1953: 42) describes how many new minority-language domains were created during these years, from cultural and educational institutions to government agencies. However, he also mentions that Russian remained the language of many administrative agencies because there was not enough native staff with the required specialization. The same problem occurred with education: certified teachers with adequate Russian knowledge were scarce, but still, the broadening of the educational system brought along the improvement of the literacy rate.

For the successful Sovietization of minority peoples, party workers and local cadres were needed who were native speakers of minority tongues. Sovietization created a similar tendency to Christianization: both involved spreading ideology using the language of the locals. In the 1920s Stalin's general policy allowed for publications to be national in form (i.e. to be written in the minority language), as long as the content was proletarian (Weinreich 1953).

Only a dozen or so people formed the Erzya-Moksha intelligentsia, and they were also divided, for instance, along the lines of unity vs. division of the two peoples and languages. In the 1920s the question of the creation of a unified Mordvin language arose again. The new language was supposed to be based mostly on Erzya, as it had a richer literary tradition and more speakers. According to some Western researchers, the creation of a unified standard failed because of the Soviet policy. As opposed to this, Kreindler (1985a: 248) suggests that "[t]he practical difficulties and the general populist approach to language prevailing in the 1920s rather than a deliberate Soviet policy of 'divide and rule' had led to the formation of two separate literary Mordvinian languages."

One of the main questions concerning the creation of Standard Erzya and Standard Moksha was which dialects to choose as their basis. Kreindler (1985a: 248–249) and Rueter (2010: 6) describe the process in detail. In creating the new standard, Mordvin linguists faced the problem of scarcity of materials on the varieties. For Erzya, the first suggestion was to base the standard on the Ardatovsko-Alatyrskii dialect, best known at that time, but after the first linguistic fieldwork expedition linguists realized that the Ardatov and the Alatyr dialects were quite different. At the 1928 linguistic conference in Moscow the Kozlovka dialect (in the Atyashevo District of the Republic of Mordovia) was chosen for Erzya, and the Krasnoslobodsk–Temnikov dialect for Moksha. While the standard created in the 1920s in Moscow by Erzya teachers has been used to some extent, in other Erzya cultural centers journalists and other members of the intelligentsia wrote in their own dialect. "In the 1930s, however, the Kozlovka-Mokshalei (Central-dialect) base of the language was broadened to include more features from the Insar or Western dialect, which meant development away from the Alatyr' or Northwestern dialect, familiar in the Erzya literature of the nineteenth century" (Rueter 2010: 6). As a result, Erzya linguists became aware of the fact that the Kozlovka dialect was the basis for the standard, but they still do not have a thorough actual knowledge of this Erzya variety (Rueter 2010: 9).

Another idea was to introduce the Latin alphabet (the same reform was planned for Russian), but due to more pressing matters, this was postponed. Erzya linguists generally agreed to switch to the Latin alphabet, but they still debated when to carry out the change, before or after Russians abandon the Cyrillic alphabet, which was considered inadequate for socialist thoughts and regarded as a reminder of tsarist imperialism. Some linguists argued that through the use of the Latin alphabet, money could be saved on printing, as less paper and lead would be needed, not to mention the advantage of making everything that had been written before the alphabet reform eventually unreadable (Weinreich 1953: 50). These changes were in accordance with Lenin's ideas: "On the one hand Lenin believed that no one language, and especially not Russian, should be given the status of 'state language'. On the other he promoted the concepts of national equality and national self-determination" (Kirkwood 1991: 61).

This policy also included an educational component, according to which all the languages were to be used as the medium of instruction so that none of the ethnic groups had an advantage

over the others. (This relative equality of the languages, however, was replaced with a policy of Russification in the 1930s.)

> "Language planning in the 1920s, therefore, was devoted in the main to obtaining linguistic data in the field, describing that data in a form which would be adaptable for pedagogic purposes, and devising alphabets and orthographic systems which would underpin writing systems for languages which did not have them before" (Kirkwood 1991: 62).

Despite the difficulties encountered in the beginning (high illiteracy rates, a lack of teachers, no written standards), in the mid-1920s the new intelligentsia achieved important results, such as creating new orthographies, enhancing the vocabulary and printing textbooks (for 25 languages in 1924, and for 104 languages in 1934). Kreindler (1985b: 51) describes the development in education using the following numbers: in 1923 there were 950 Mordvin schools, partially with Russian as the language of instruction, while by 1925–1926 50% of the subjects in the first grade were taught in the Mordvin languages, by 1931 the entire primary education used Erzya and Moksha as the medium of instruction, and by 1936, this extended up to the tenth grade.

Achievements of the new intelligentsia included also publishing school textbooks and developing the Erzya written media. The first Erzya primer *Tundon' chi* ('Spring day') was published in Cyrillic script in 1922 in Moscow, while the first Moksha primer came only three years later. Unfortunately, the book was withdrawn because of both orthographic mistakes and political issues (only Trotsky was depicted at Lenin's side, while Stalin was missing from the book altogether) (Salo 1991: 166).

The first newspaper, *Chin' stamo* ('Sunrise') was published in 1920/1921 in Simbirsk (today's Ulyanovsk), followed by several others in Moscow (1921 *Jakstere teshte* 'Red star'), Penza, Samara, and even in Novosibirsk, Siberia. First, literature was published in newspapers, and literary journals started to emerge in the late 1920s. The standard varieties had not yet been created, and, as a result, the authors used their own dialects in their writing (Salo 1991: 166), which led to misunderstandings. The Erzya–Moksha bilingualism of the Penza paper, *Od vele/Od*

Vel'ä ('New Village') also caused conflicts. This periodical started off as an Erzya and Moksha newspaper, but later as the readers complained, the Erzya parts were given up and the newspaper became the main Moksha media institution.

Kreindler (1985b) describes another problem which hindered language revitalization, namely, that Erzya and Moksha were considered backward languages with an underdeveloped culture, which caused low self-esteem for members of its communities: "After centuries of Russian domination, many Mordvinians felt that only Russians could be leaders and only Russian could be the language of culture and power" (1985b: 50). For example, new textbooks were not used, because a school could not be imagined without the Russian language. This attitude did not work in the favor of language maintenance, especially when the 1938's educational reform made Russian a compulsory subject from the age of seven.

The 1930s brought about both positive and negative changes for the Mordvins. When the Mordvin State was created (1934), more cultural institutions (the State Theater, the Pedagogical Institute, which later became the Mordvin State University, a publishing house, the Association of Writers, etc.) were established. The 1930s also brought about centralization, and a decrease in the development of national cultures and languages. The autonomy also involved the closing down of cultural institutions outside the republic, for example, the Erzya newspaper published in Siberia was discontinued (Kreindler 1985b: 52).

The 1930s are also regarded as the era of the rebirth of Russian nationalism. Although the Soviet Union did not have Russian as its *de jure* official language, Russian *de facto* became the most important promoted language, as a more and more centralized state needed a common language (Comrie 1981: 22). From 1917 onwards the minority languages had been used in new domains, therefore the missing terminology for e.g. educational and administrative purposes had to be created. In the 1920s, neologisms in the standard language were borrowed from the vocabulary of different dialects, and, in order to collect the necessary material, extensive fieldwork was conducted among the rural population. From the 1930s, however, the Russian terminology had to be employed, neologisms of the 1920s were rejected, which was a "colossal waste of scholarly and educational effort" (Weinreich 1953: 54). The status of the Russian language improved as the new terminology, although originating from Western European

languages, was adopted in the minority languages via Russian. However, as the minority languages gained more and more function, their status also rose, and this "gave the speakers of those languages an enhanced sense of their own national worth and pride in their own cultural traditions" (Kirkwood 1991: 63).

By 1939 the Cyrillic alphabet was established for almost all the languages in the Soviet Union (the exceptions are mentioned above in section 2.1), as another tool of "unification". In the 1940s the orthography of Russian loanwords in the minority languages was also standardized, the form had to follow Russian orthographic rules even in case of phonologically and morphologically adapted established borrowings (Weinreich 1953: 51). "The Russian spelling often violated the phonological system of the indigenous language and in many instances resulted in rather large differences between the written form of a word and its actual pronunciation" (Grenoble 2003: 53). According to Comrie (1981: 34), even the pronunciation of words had to follow the Russian originals. This made acquiring literacy in the minority languages even harder, because children had to learn Russian first in order to understand the orthographic rules of their mother tongue.

Some of these strict rules were revoked in the 1950s, and, thus, loanwords were to be written as they are pronounced in the recipient language. Grenoble (2003: 53) points out that the situation actually changed only partially in practice, because there was a loophole in the policy: new Russian borrowings (even international sociopolitical terms entering Erzya via Russian) constituted exceptions and had to follow Russian orthographical rules.

It was not only the language policy that changed in the 1930s. The development of the Erzya and Moksha cultural life was harshly stalled when the intelligentsia was deported or executed during the Stalinist era of terror in the 1930s. "The national and cultural renaissance among the non-Russian peoples had received a blow from which it never recovered" (Schwarz 1953: 44). Stalin's original national form vs. Socialist content distinction became yet another tool of Russification, as it was used to eliminate even more national elements from culture. For instance, Mordvin topics disappeared from the literature and were replaced by typical Soviet ones. The new aim was "a single, uniform, russianized culture, translated into languages which are themselves made ever more uniform" (Schwarz 1953: 45). Research institutes established

earlier to carry out studies specifically on the local minority cultures shifted their focus as well, involving also Russian topics in their research (Kreindler 1985b: 54).

All of this brought about changes in the educational policy, favoring Russian over other languages of the Soviet Union. In 1938 Russian became a compulsory subject in all schools. Although native languages still preserved their status as the medium of education (Kreindler 1985b: 54), "[d]e facto if not de jure Russian was thereby acknowledged as being of more importance than other Soviet languages" (Kirkwood 1991: 63). The academic year of 1938–1939 was the last one when the Mordvin languages were mentioned as means of instruction in 10-grade schools, while in 1958 subjects were taught in Erzya and Moksha only through the 7th grade. In 1936 there existed 18 newspapers and 4 periodicals in the Mordvin languages, but only two newspapers and two magazines remained by 1954.

The late 1930s and the 1940s were characterized by a kind of Erzya–Russian mixed language, and nobody continued the corpus planning for the Mordvin languages (Salo 1991: 167). Russian loans entered the minority languages without adaptation, formerly successful neologisms were replaced: a good example is the use of the Russian word *slovar'* 'dictionary' from 1941 instead of the coined Erzya term *valks* 'dictionary' (Kreindler 1985b: 53).

In some respects, language policy in the late 1930s was similar to that of the end of the 19th century. Weinreich (1953) acknowledges this similarity but identifies two major differences between the language policies of the two eras. First, during the Tsarist regime mother tongue education was not available for the minorities. Second, the Soviet Union's policy of Russification applied a different approach: while the use of minority languages was encouraged in some domains, the structure of these languages was under heavy Russian influence due to the loanwords and constructions borrowed from the dominant language and to the Russian writing system and spelling rules imposed on them.

Although in the first years of Soviet Russia and the Soviet Union, in principle, equal rights were provided to each nation to counteract the policy of Russification in the Tsarist times (Weinreich 1953: 46), Lenin's loathing of nationalism was in conflict with the separatist intentions of some minorities. Stalin's formula (culture had to be proletarian in content, national in form) was never defined in detail, which had serious consequences for the minorities.

"This ambiguity constituted, in effect, a Damocles' sword over the heads of the national minorities, and it proved to be the undoing of countless minority intellectuals who one day found their work branded as 'bourgeois nationalism', that is, a preoccupation with the national content, instead of the mere form, of their cultures" (Weinreich 1953: 46).

Post-war events, including increased urbanization, migration, and the centralization of administration reinforced the importance of the Russian language (Kirkwood 1991: 64). Comrie (1981) also discusses the effects of urbanization, how it influences language maintenance, and how it leads to matrix language turnover (a term explained in section 5.3.1 below). He suggests that the influx of new Russian loanwords into the minority languages needed for new technology and typical urban phenomena resulted in a hybrid language, in which the vocabulary came predominantly from Russian. In the urban context, the maintenance of this emerging hybrid variety was futile, because the increased number of Russian loanwords eased the way to shift to Russian. However, Comrie (1981) also emphasizes the importance of social factors in this language shift, e.g. the lingua franca status of Russian for the communication purposes of different nationalities, for instance, in the military.

As mentioned above, the policy to use the unmodified orthography and pronunciation of Russian loanwords changed in the late 1950s. Neologisms created in the 1920s and early 1930s could be used again, especially constructions connected to everyday life. Among the sources of new vocabulary, dialects were included anew, so expeditions to collect dialect material were, again, on their way.

In the educational reform of 1958–59, parents were given the choice of medium of instruction for their children, which might seem as a change to a more democratic system, but it also involved a decrease in the importance of the minority languages. The use of Erzya declined even in the years when teaching in the mother tongue was compulsory, and the reform accelerated this process, as parents chose Russian as the medium of instruction for their children to secure a better future for them. From 1958 Erzya was taught only until the 7th grade, in the 1960s until the 5th grade, while some of the schools, especially those outside the republic,

altogether switched to Russian from the first year. By 1970 Russian as the medium of instruction was introduced in the fourth grade. This presented problems in areas where the children's level of Russian was not sufficient, and the teachers explained the subjects in Erzya/Moksha even in the upper grades (5th through 8th). According to Kreindler (1985b: 55), in the academic year 1973/74 there were 391 Mordvin schools (using the Mordvin languages as a medium of instruction up to third grade) which 41,000 Moksha and 36,000 Erzya students attended, and 34.2% of the Mordvin students were taught in Russian in the first grade.

Kirkwood (1991: 64) points out that while Khrushchev also promoted Russian as the second mother tongue, the status of Russian as the common language of the Soviet people improved in the Brezhnev era, especially between 1962 and 1984, when the concept of the Soviet nation as one historical unity emerged in which minority national identity lost its meaning and prestige, while Russian claimed a leading role.

The educational reform of 1979 was another blow to the Erzyas, as it introduced a standardized way of teaching Russian, which meant that the contrastive approach was lost, and consequently, the prestige of minority languages decreased (Kirkwood 1991: 66). Moreover, the influence of Russian increased also via the Pioneer and Komsomol organizations.

Lewis (1972) sums up the reasons explaining the low language maintenance rates of the Erzyas and Mokshas. Factors leading to language shift and assimilation (low national growth, scattered population, emigration, and urbanization) are all present in the case of the Mordvin community. While all the other peoples living in the Volga region display growth after World War II, the birth rate was low among Mordvins. According to the 1979 census, 71.6% of the Mordvin population lived in the diaspora outside Mordovia, especially in urban settings where they formed a minority and mixed marriages were very common. The rate of assimilation was faster in the diaspora where the educational and cultural support system of the Mordvin Republic was missing, Erzya was taught only as a subject at schools – if a teacher was available. According to Kreindler (1985b: 55), the language maintenance rate in the republic was 72.7%, while in the diaspora only 64%.

In the Mordvin Republic, the urbanization rate was also lower in the 1970s than in the western parts of the Soviet Union, while among the Udmurts the rate of rural vs. urban population

rate was 12:1, among the Chuvash and Mari 16:1, while it was 26:1 among the Mordvins (Lewis 1972: 145). As compared to the Erzyas, Turkic peoples in general managed to maintain their language and identity more successfully due to their different religion and the fact that until 1920 the prestige language for them was Arabic, not Russian. Haarmann (1985: 337–338) describes these differences as follows:

> "Bilingual Buriats, Kalmyks or Tatars, for instance, are known to possess a strong mobility to use Russian and mostly speak good Russian. At the same time their rate of language maintenance is comparatively high. Other ethnic groups like Latvians, Mordvinians and Kazakhs show the opposite correlation. This means that higher rates for Russian in the status as a second language correspond to lower rates for language maintenance (and vice versa)."

According to Haarmann (1998), in 1989 the language maintenance rate among Tatars was 83.2%, among Bashkirs 72.3% (many of them switched to Tatar, not Russian), while among Mordvins only 67.1%.

In the 1980s the shift to Russian continued, even though there was no official policy of Russification. The language shift was accompanied by the development of Russian-influenced varieties of the minority languages. The general phenomenon was asymmetrical bilingualism, only members of the minority communities spoke Russian, while Russians did not speak minority languages. (There are some exceptions, individual cases of Russians learning Erzya in villages mentioned by Rueter 2013.) Comrie (1981: 31) argues that other languages have hardly had any influence on Russian, except for the pronunciation, and to a lesser extent, also for borrowings of mainly cultural phenomena. Bátori (1980: 156) predicted that this contradictory tendency would continue: Finno-Ugric languages would be more and more influenced by Russian, especially as far as their syntax was concerned, while Russian would be less and less influenced by Finno-Ugric languages.[1] Comrie (1981) described this assimilation as a natural process, and emphasized that

---

[1] "Insoweit eine zeitliche Einordnung der eizelnen Interferenzarten möglich ist, kann man sagen, daß der Einfluß des Russischen auf die finnougrischen Sprachen stärker wird, während umgekehrt, der Einfluß der finnougrischen Sprachen auf das Russische zurückgeht. Dies ist besonders deutlich im syntaktischen Bereich" (Bátori 1980: 156).

despite the intensive language mixing, it was not likely that a stable mixed language would develop: "[l]inguistically, the fact that Russian influences other languages of the U.S.S.R., but is not influenced by them, would seem to preclude the possibility of an amalgam language distinct from Russian finally developing" (1981: 37).

The educational reforms in 1984 and 1989 did not work in the favor of the Erzyas either. In 1984 the 10-year school system was changed to a 11-year one, and Russian became a compulsory subject beginning with first grade, and children were required to become fluent in Russian by the time they left the school (Kirkwood 1991: 67).

Despite these setbacks, there were some positive changes in the language policy of the late 1980s and early 1990s, similar to the reforms of the 1920s, with language use expanding to new domains. Haarmann (1998) also compares these two eras arguing that language policy was more permissive in both periods in order to counterbalance the Russification in the former (Tsarist and Soviet) times.

"Then, the extension of social functions to include the use of the mother tongue in school education, in the mass media (i.e. press, radio) and in local administration was directed against the dominance of Russian as a tool of czarist assimilation policies. Today, the revitalisation of local languages in Russia is a function of the drive toward democratisation which crystallizes in the rejection of Soviet-type Russian dominance in the public sectors" (Haarmann 1998: 227).

To sum up, the history of the Erzya–Russian contact situation can be characterized as gradual assimilation. Erzyas form "a speech community where the situational pressure of Russian has already caused strong assimilation processes and where Russian as a second language dominates in all age-groups" (Haarmann 1985: 334). Although earlier phases in the contact situation involving taxation and occupation of the ancient lands also took their toll, the most important steps of this assimilation process were Christianization in the 18th century, and the introduction of Russian-medium school education in the Tsarist era in the 19th century and also in the Soviet times. This prolonged process of acculturation caused changes in all subsystems of Erzya: Russian conjunctions and particles were borrowed, and body part and kinship terminology

was also replaced with Russian equivalents. In contemporary Erzya, native terminology is used in literature and formal written texts, while Russian terms dominate the spoken discourse. The 20th century also brought about important changes, as an Erzya standard was created and status and corpus planning began, followed by a period of Russification from the late 1930s. The fall of the Soviet Union created new hope for the Erzya community, as Erzya along with Moksha became the official languages of the Mordvin Republic, their use expanded to new domains: new media was established, and new educational opportunities were created. The Internet has also provided the Erzya communities with new possibilities to share the benefits of the cultural support system in the republic with people living in the diaspora.

Chapter 3. Typological overview of the Erzya and the Russian language

In this chapter, I discuss the main characteristics of the Erzya and the Russian languages using a typological perspective. The overview does not consider all structural phenomena attested in these languages: I focus on the features that are relevant for my study, because these particular constructions occur in code-switched utterances in my data. As I discuss in Chapter 5, I argue that there is a possible connection between (in)congruence of the constructions in the two languages and the type of CS attested in bilingual mixed utterances.

In the cases of both Russian and Erzya, the standard variety of each language is taken into consideration. This can be regarded as a questionable decision, as the linguistic repertoire of a speaker does not necessarily contain the standard varieties (depending on various sociolinguistic factors involving their level of education, occupation, etc.). The nature of my data and the number of my subjects, however, do not make it possible to consider all of the local varieties the speakers might have knowledge of. It is especially true in the case of the radio data (cf. Chapter 6), in which the sociolinguistic and, thus, the linguistic background and language proficiency of the interviewees remain unknown in the majority of the cases. In further research, this problem can be eliminated through the study of a smaller speech community.

In this chapter, I predominantly discuss morphosyntactic characteristics of the two languages, as the analysis of intrasentential CS constructions (in Chapter 7) also focuses on the morphosyntax of these CS patterns. I also give a brief overview of the phonological system of the two languages below (in section 3.1), partly to support the claim that the similarities between the phonological systems of the two languages make it, most of the time, impossible to differentiate between code-switches and borrowings. There are other reasons for my rejection of the distinction between these two contact phenomena (cf. section 5.1.2 below for details).

In discussing morphosyntax (section 3.2), I focus on both nominal (3.2.1) and verbal (3.2.2) constructions. I discuss the case system of the two languages and the characteristics of constructions with a possessive or a numeral modifier in detail. I also refer to differences concerning the tense and aspect system of the two languages. Next, I discuss two types of constructions, one of them expressing possession (section 3.2.3) and the other necessity and

obligation (section 3.2.4). Finally, I elaborate on issues of word order in the two languages (section 3.2.5).

3.1. Phonology

The phonological systems of the two languages display significant similarities (Saarinen 2014: 541). Both Erzya and Russian have five vowel phonemes: front vowels /i/ and /e/, back vowels /u/ and /o/, and central vowel /a/ (Raun 1988: 99 for Erzya, Rusgram 2016 for Russian). (In dialects, these numbers can differ – however, as mentioned above, the overview focuses on the standard varieties.) The number of consonant phonemes is different in the two languages: there are thirty-seven consonants in Russian as opposed to the twenty-eight consonants in Erzya. Both languages have palatalized (or soft, see below) consonants, but palatalization is more widespread in Russian, because in Erzya only alveolar consonants have palatal counterparts (for a detailed description of the consonant systems, cf. Raun 1988: 99 for Erzya, and Rusgram 2016 for Russian).

Russian and Erzya consonants are commonly divided into hard and soft types. Timberlake (2004: 29) also discusses this distinction used in Russian linguistics: "It is convenient, following the Russian tradition, to refer informally to non-palatalized consonants as HARD and palatalized consonants as SOFT. […] Palatalization, though a property of consonants, affects how vowels are pronounced." This is also relevant concerning the debate about the sixth vowel (the [i̯]) in Russian. Following the academic grammar of the Russian language (Rusgram 2016), I do not consider it a phoneme, as it occurs after hard consonants, while its allophone [i] occurs after soft consonants and in word-initial position ([i̯] can also occur in word-initial position, but only rarely, in foreign place names and onomatopoetic words). In Erzya, both /i/ and /e/ have a mid back allophone occurring after non-palatalized consonants, the [i̯] and the [ẹ], respectively (Bartens 1999: 27). These questions are relevant as regards the transcription of the examples, which I discuss in detail in Chapter 6.

Apart from the similar phonological systems, Erzya and Russian display different (morpho)phonological characteristics as well. Erzya has vowel and palatal harmony: the choice between allomorphs in both derivation and inflection (e.g. in the case of the allomorphs of case

suffixes) depends on the phonological make-up of the stem, whereas in the case of vowel harmony (Zaicz 1998: 190), on the last stem vowel. It is a complex phenomenon (for a more detailed description of both vowel and palatal harmony and their relevance in morphology, cf. Rueter 2010). However, as it is not the focus of my paper, I consider here only one example. For instance, the inessive case ending has two variants (*-so/-se*), the allomorph *-so* is attached to stems with back vocalism, e.g. *kudo* 'house' – *kudoso* 'in the house', while *veľe* 'village' has front vowels, the suffix variant *-se* is attached to it: *veľese* 'in the village'. This rule applies also to Russian origin elements, cf. *miṅisťerstvoso* 'in the ministry' (*miṅisťerstvo* 'ministry') and *MVD-se* 'in the interior ministry' (*MVD* [emvede]: *Ministerstvo vnutrennih del* 'Ministry of Internal Affairs of the Russian Federation').

Finally, the question of stress has to be discussed. In Erzya, word stress is virtually free (Rueter 2010: 14), but in most of the cases it falls on the first syllable of the word (for a diachronic explanation, cf. Estill 2004; for the relation of prosody and stress in Erzya, cf. Lehiste et al. 2003). In Russian, the place of the word stress can change even within the paradigm of a word (Timberlake 2004: 29). In Standard Russian, lack of stress causes vowel reduction. As opposed to this, there is no such reduction in unstressed syllables in Erzya. In some dialects of Russian, reduction does not occur in unstressed syllables (the phenomenon is called *okan'e*). It is also the case in Russian dialects spoken on the territory of the Mordvin Republic (Vasil'ev and Rusinov 1992: 41), which can be connected to the lack of reduction in the Erzya language. The phenomenon of *okan'e* is also prevailing in the monolingual Russian speech of Erzya speakers. This has relevance to the differentiation of CS and borrowing: Russian origin forms which show no reduction in the speech of Erzyas cannot be necessarily categorized as adapted loanwords because of the lack of reduction. The CS vs. borrowing question is otherwise also very complex, I elaborate on it in Chapter 5.

3.2. Morphosyntax

In this section, I study the main morphosyntactic characteristics of the two languages from a typological point of view. I focus on five aspects of morphosyntax in Erzya and Russian which are involved in the bilingual CS discourse.

In section 3.2.1, I discuss nominal constructions, followed by an overview of verbal constructions in section 3.2.2. Both possessive constructions and the ways in which necessity is expressed in the two languages display incongruent and congruent structures – I investigate them in sections 3.2.3 and 3.2.4, respectively. Finally, the difficulties presented by word order are elaborated on in section 3.2.5.

3.2.1. Nominal constructions

For the discussion of nominal phrases, especially constructions with a numeral modifier and time expressions, in Chapter 7, it is essential to give an overview of the nominal structures in both languages.

3.2.1.1. Declensions and cases

In Erzya (Raun 1988: 100), nouns have three declensions: indefinite (*kudo* 'house'), definite (*kudoś* 'the house'), and possessive (*kudom* 'my house', *kudot* 'your house', etc). The number of cases differs in the three declension types. In the indefinite declension, there are 12 cases in the singular; in plural, there is only nominative; e.g. *kudo* 'a house', *kudot* 'houses'. In the definite declension, there are 10 cases both in the singular and in the plural, e.g. *kudoś* 'the house', *kudotńe* 'the houses'. Finally, the possessive declension is limited, there is no clear plural vs. singular distinction, and there are only three cases (nominative, genitive, and dative), but apart from terms referring to family members and other relatives, words have only possessive nominative forms. In the first person and third person there are two different suffixes, one referring to one single possession, the other to two or more possessions (EKM 2000: 97): with a first person singular possessor, e.g. *kudom* 'my house' versus *kudon* 'my houses'; and with a third

person singular possessor, *kudozo* 'his/her house' and *kudonzo* 'his/her houses' (Aasmäe 2014: 26).

As mentioned above, the indefinite declension has the most number of cases (12): nominative, genitive, (allative-)dative, ablative, inessive, elative, illative, lative, prolative, translative, comparative, and abessive or caritive. Apart from these case suffixes, Erzya has many postpositions but no prepositions. Like all Finno-Ugric languages, Erzya has no gender distinction.

In Russian, case, number and gender can be marked on nominals. (Timberlake 1993: 836). There are two numbers (singular and plural) and six primary cases (nominative, accusative, genitive, locative, dative, and instrumental; the two secondary cases – secondary genitive, and secondary locative – which are used only with a small number of masculine nouns). As opposed to Erzya, Russian does not have postpositions but has a complex system of prepositions. This incongruence also causes the emergence of hybrid and double-marked constructions, which I analyze in Chapter 7.

Russian distinguishes between three grammatical genders: masculine, feminine, and neutral. (Syntactic) gender is present in the agreement of adjectives (both in the attributive and predicative positions), verbal predicates in the past tense, and pronouns (Timberlake 1993: 836). As the category of gender is missing from Erzya, Russian constructions with gender agreement pose a problem in CS because of the incongruence. In Chapter 7, I analyze how the mismatch of these constructions is "solved" in code-switched utterances. I predominantly focus on gender agreement in predicative adjectives and verbs in the past tense.

3.2.1.2. Constituent order

There are significant differences in the order of constituents in Erzya and Russian. In noun phrases, the modifier precedes the modified in both languages: Russian *starij dom* 'old house' (*starij* 'old' and *dom* 'house') and Erzya *tašto kudo* 'old house' (*tašto* 'old' and *kudo* 'house'). In Erzya, it is the general order of constituents (Keresztes 1990: 77), consequently, the same order is prevailing also in possessive structures: *babań kudozo* 'the grandmother's house' (*baba* 'grandmother' with the genitive singular suffix *-ń*, and the possessive suffix *-zo*). In Russian,

however, the possessor follows the possessed: *dom babuški* 'the house of the grandmother' (*dom* 'house', and *babuška* 'grandmother' in the genitive singular). In Erzya, the Russian-type (possession–possessor) reversed order is marked, occurring usually in poems, for reasons of rhythm. Although the construction occurs in descriptive grammars of Erzya, there is no explanation pertaining to its (stylistic) function. In *Grammatika* (1962: 109) there is one instance of a reversed-order possessive genitive construction. It is cited from a novel to illustrate that the possessor–possession relation can reflect a part–whole relation, and there is no comment on the unusual constituent order.

(1) *peľe-nze   veľe-ńť   soda-sa*
   half-POSS.3SG   village-GEN.DEF.SG   know-SG3<SG1
   'I know half of the village.'

The reversed constituent order, however, can lead to misunderstandings. Consider the following: in example 2, the possessor in the genitive case (*Sv'etań* 'Sveta's') is in a modifier position, while in example 3 it is the predicate.

(2) *Sv'eta-ń   kudo-zo*
   Sveta-GEN   house-POSS.3SG
   'Sveta's house.'

(3) *kudo-zo   Sv'eta-ń*
   house-POSS.3SG   Sveta-GEN
   'His/her house belongs to Sveta.'

Example 3 can be analyzed also as a Russian-type reversed order construction with the word *Sv'etań* 'Sveta's' as a modifier. While the context usually helps in distinguishing these two functions, the application of the Russian-like constituent order can cause ambiguities in the interpretation of the utterance.

3.2.1.3. Numeral phrases

Numeral phrases show different characteristics in Erzya and in Russian. In Russian, numeral phrases are constructed on the basis of the following rules: the nominative singular is used after

the number 'one' and compound numbers ending in one (with the exception of compound numbers ending in eleven, which require the genitive plural); the genitive singular is applied after the numbers 'two', 'three' and 'four', and their compound counterparts (with the exception of numbers ending in twelve, thirteen and fourteen requiring the genitive plural); and the genitive plural is the required form in all other cases. Timberlake (2004: 185) also argues that in oblique cases (i.e. not the nominative or the accusative) the numeral agrees with the noun in case.

The Erzya system, however, is less complex: the argument of the numeral phrase is in the nominative case; if the attribute is the number 'one', the singular form is used, whereas after numbers 'two' and higher, a plural ending is usually required (EKM 2000). As opposed to Russian, the form of the Erzya numeral does not change even if the phrase is in an oblique case, which can occur both with an indefinite and a definite head. Rueter (2010) argues that the difference between the following two constructions is that the second one is definite: *kavto kudoso* 'in two houses' and *kavto kudoťńese* 'in the two houses'. Rueter also emphasizes another important aspect of these constructions:

> "Since definiteness is an entailment of demonstrative pronouns, it will be noted that quantifiers indicating numbers larger than one can only co-occur with the plural demonstrative pronoun *ńe* 'these (anaphoric)' and not its singular counterpart *te* 'this'" (2010: 178).

There is another difference between Erzya and Russian numeral phrases, and it concerns the constituent order: generally the numeral precedes the modified in both languages. In Russian numeral phrases, however, approximation is expressed by a reversed order – the numeral follows the complement: *dva časa* 'two hours', but *časa dva* 'approximately two hours' (*dva* 'two' and *časa* 'hour', in genitive singular). In Erzya, approximation can be expressed by the comparative case suffix -*ška*: *kavto čast* 'two hours' and *kavtoška čast* 'approximately two hours' (*kavto* 'two' and *čas* 'hour').

3.2.1.4. Nominal predicates

There is a special construction in Erzya that does not occur in Russian, the nominal conjugation, which involves the inflection of nominals and locational predicates for person, number and tense (Turunen 2006). The present tense paradigm for 'I am beautiful, you are beautiful, etc.' as compared to verbal predication looks as follows:

nominal predication	verbal predication
mazij-an 'I am beautiful'	*važoď-an* 'I work'
mazij-at 'you are beautiful'	*važoď-at* 'you work'
mazij 'he/she/it is beautiful'	*važoď-i* 'he/she/it works'
mazij-ťano 'we are beautiful'	*važoď-ťano* 'we work'
mazij-ťado 'you [PL] are beautiful'	*važoď-ťado* 'you [PL] work'
mazij-ť 'they are beautiful'	*važoď-iť* 'they work'

Table 1. *The comparison of the nominal and verbal predication in the present tense*

In the present tense, the third person singular forms are unmarked, but otherwise the suffixes coincide with the suffixes of the verbal paradigm in the present tense. Past tense predicative suffixes are grammaticalized forms of the copula verb *uľems* 'to be', which are also used with verbs to express what is called a "second", habitual past, cf. *kolmoľ* 'there were three of them' (*kolmo* 'three') and *lovniľ* 'he used to read' (*lovnoms* 'to read').

In the Erzya–Russian CS corpus, Russian elements can also occur in these constructions. This is, however, not a recent development: Turunen (2006: 184) cites an example from Paasonen and Ravila's (1938: 284) folklore collection in which a Russian origin adjective *plohoj* 'bad' occurs as a nominal predicate with a first person singular verbal suffix. It is even more puzzling that the Erzya equivalent of the construction is also used in the utterance, and that both versions display a marked constituent order:[2]

---

[2] I retained Turunen's transcription and glosses.

(4) *mon lomań-eś b'eŕań-an, mon lomań-eś plohojń-an*

    I   man-DEF   bad-1SG   I   man-DEF wretched-1SG

    'I am a bad man, I am a wretched man.'

These types of mixed constructions are attested in my data as well, cf. Chapter 7.

3.2.2. Verbal constructions

Both languages, but especially Erzya, have a complex verbal morphology system. In this overview, I focus on the main characteristics of the Erzya and Russian conjugation paradigms and incongruences of the two systems that can explain the switch types occurring in the data.

Erzya verbs are divided into three main classes: the *a*-, *o*- and *e*-stem verbs. The categorization is based on the infinite form of the verbs. In the present tense, the verbal endings in the first and second person singular are the same in all three classes: *-an* (1SG) and *-at* (2SG). *A*-type verbs include *jarsams* 'to eat' – *jarsan* 'I eat', *jarsat* 'you eat'; *o*-type verbs *lovnoms* 'to read' – *lovnan* 'I read', *lovnat* 'you read'; *e*-type verbs *śimems* 'to drink' – *śiman* 'I drink', *śimat* 'you drink'. Bartens (1999: 123) offers an explanation for the spread of the *a*-paradigm to the other classes in these two persons. According to her, the change occurred to avoid homonymy. The verbal suffixes in the first past tense are *-ń* and *-ť* (as a result of palatalization after the past tense marker **j > i*). The spread of the back-vowel verbal suffixes characteristic originally of the *a*-stem (*-an*, *-at*) prevented the emergence of palatalized verbal endings in the *e*-stem verbs in the present tense. The prevalence of the *a*-type verbs seems also significant in the adaptation of Russian verbs (cf. section 7.2.3).

There are two conjugations in Erzya: the indefinite and the definite. The indefinite conjugation is used with intransitive verbs and transitive verbs with an indirect object expressing a continuous action. The definite conjugation is used if the verb is transitive, the direct object is definite, and the aspect of the action is perfective. Erzya distinguishes between four tenses: one present, two past tenses, and a periphrastic future. (As mentioned above, the second past tense refers to habitually recurring or long-lasting events in the past.)

There are different categorizations for the conjugation types in Russian. Timberlake (2004) differentiates between two types of stems (the past-infinitive and the present-stem) and between two conjugation types: the *i*-conjugation and the *e*-conjugation which both can be divided into further subtypes. (Since these types are not directly relevant in the case of my data, I do not focus on them further. For a detailed categorization, cf. Timberlake 2004: 99).

The category of aspect, on the other hand, plays an important role in Erzya–Russian CS. Russian differentiates between imperfective and perfective aspects. According to Timberlake (1993: 849), "[i]mperfectives distinguish past, present and future […] Perfectives distinguish past and a morphological present, which reports true future or singularized habitual situations". The present is inflected for person and number, while the past is inflected for gender and number.

Erzya does not have the grammatical category of gender, whereas gender, with its ternary distinction of masculine, feminine and neutral, presents a substantial segment of the grammatical system of Russian. In the case of verbs, gender agreement between the predicate and the subject is relevant in the past tense. In the past tense forms of the Russian main verbs, there are four markers used: *-l* (masculine), *-la* (feminine), *-lo* (neutral), and *-li* (plural). To be more precise, the *-l* element is the past tense suffix, the masculine form is unmarked, *-a* is the feminine, *-o* is the neutral, and *-i* is the plural marker. In bilingual Erzya–Russian speech, instances of gender agreement occur in cases where the predicate is a code-switch to Russian. In section 7.2.3.3, I investigate how and to what extent utterances with code-switched past tense forms display gender agreement in mixed sentences. My aim is to identify the factors triggering gender agreement in these bilingual utterances.

As Russian past tense forms are not inflected for person, the Erzya pronominal subject becomes overt in these mixed sentences. This involves the investigation of pro-drop properties in both languages. While Erzya is a clearly pro-drop language (Rueter 2007), the status of Russian in this respect is debated (Franks 1995). I discuss this in detail in connection with CS constructions in Chapter 7.

3.2.3. Possessive constructions

In Erzya there is one construction used for the expression of possession, whereas in Standard Russian possessive sentences can be formed two ways:

1. The first type of construction involves the use of the preposition *u* 'at, by' + the possessor in genitive + (the copula) + the possession in nominative in assertive sentences:

(5) *u meńa   jesť   dom*

    at I.GEN   there.is house

    'I have a house.'

In some cases, the copula used in assertive sentences in the present tense (*jesť* 'there is') can be omitted, for example, if the possession is a body part or some kind of clothing (Timberlake 2004: 313–314).

In negative sentences, there are two differences from the affirmative construction: the copula is negated and the possession is in the genitive. "The subjects of existential intransitive predicates, which would otherwise be in the nominative, appear in the genitive when the predicate is negated" (Timberlake 2004: 460).

(6) *u meńa   ńet   dom-a*

    at I.GEN   there.is.not house-GEN.SG

    'I do not have a house.'

2. The second type of possessive construction is formed through the use of the transitive verb *iḿeť* 'to have'. This is a marked construction used in special contexts, for example, if the possession is an abstract noun or is considered a property of the subject (cf. Timberlake 2004: 312).

Erzya possessive construction resembles the first type of the Russian constructions, and it does not have a transitive *habeo*-type verb. In Erzya, the structure of the possessive construction is the following: the possessor in the genitive case + the copula + possession in nominative and marked for the possessive suffix (Keresztes 2011: 96).

(7) *moń   uľ-i   kudo-m*

    I.GEN   be-3SG   house-POSS.1SG

    'I have a house.'

Negation does not change the case of the possession. The only change as compared to the assertive sentence is the replacement of the copula with the negative predicative.

(8) *moń*   *araś*   *kudo-m*

I.GEN   there.is.not   house-POSS.1SG

'I do not have a house.'

In Erzya, the copula cannot be omitted from the possessive construction, because it would transfer the possessive construction into a noun phrase with a possessive modifier: *moń kudom* 'my house'. Keresztes (2011) argues that there is a historical connection between the two constructions (*moń kudom* 'my house' and *moń uľi kudom* 'I have a house'), the use of the possessive suffix spread to the possessive constructions analogically from the noun phrases with possessive modifiers.

In Chapter 7, I discuss in detail how possessive constructions are code-switched in my data, what types of mixed constructions are attested, and how speakers cope with the incongruences of the two systems.

3.2.4. Necessity

There is a variety of ways in both languages as to how constructions expressing necessity can be formed. Here I focus on only two types of constructions, one which the two standard varieties share, and another one present only in Russian.

In the first type of construction there is a non-verbal predicate (in Russian: *nado* 'it is needed', example 9; in Erzya: the *ma*-form of the verb, the passive necessive participle, cf. Zaicz 1998: 205; or the infinitive-participle, cf. Bartens 1999: 151, example 10), or a verb in the third person (only in Erzya, *eŕavi* 'has to', example 11). The argument of the predicate, to which the obligation is directed, is in the dative (*ťeb'e* 'to you' in Russian and *tońeť* 'to you' in Erzya).

(9) Russian:

zavtra   *ťeb'e*   *nado*   *id-ťi*   *v*   *ľes*

tomorrow   you.DAT.SG   have.to   go-INF into   forest

'Tomorrow you have to go to the forest.'

(10)   Erzya: necessive participle

Vandi   tońeť   viŕ-ev   moľema

tomorrow you.DAT.SG forest-LAT   going

'Tomorrow you have to go to the forest.' (Raun 1988: 107)

(11)   Erzya: finite verb

Vandi   tońeť   eŕav-i   moľ-ems viŕ-ev

tomorrow you.DAT.SG have.to-3SG go-INF1   forest-LAT

'Tomorrow you have to go to the forest.'

The other construction expressing necessity has no equivalent in Erzya. This construction has the following elements in the present tense: the subject in nominative + the predicative adjective *dolžen* 'must' which agree with the subject in gender and number, and accordingly has four forms, namely, the singular masculine (*dolžen*), the singular feminine (*dolžna*), the singular neutral (*dolžno*), and the plural form (*dolžni*) which is used with plural subjects of all genders + the infinitive (Wade 2011: 341–343):

(12)   *zavtra   ti dolž-en id-ťi   v   ľes*

tomorrow you must-M go-INF into forest

'Tomorrow you have to go to the forest.'

The main difference between the above mentioned constructions and the *dolžen*-type one concerns the subject of the construction, or the Experiencer, and the case assigned to it. Another issue is gender assignment, which is present only in the case of the second construction. My hypothesis is that if the second type of construction occurs in mixed utterances, there will be variation as regards the presence or lack of gender agreement between an Erzya subject and the Russian predicate.

3.2.5. Word order

It has been long debated among researchers of Erzya what the word order prevailing in the language is. Zaicz (1998: 206) claims that SVO is the "textually preponderant" word order in Erzya, which evolved under the influence of Russian from an earlier SOV word order. If that is the case,

the equivalence in the word order patterns in the two languages provides more smooth switching points, making CS easier. (Zaicz's statement is widely accepted, the description of Erzya word order in *The World Atlas of Language Structures*, WALS 2016 also relies on his account.) Zaicz formulates the following basic rule: "topic(s) to the head of the sentence, focalized element immediately before the finite verb; a focalized verb must therefore stand sentence-initial" (1998: 206). Rueter (2010), however, brings an additional perspective to this picture by arguing that in Erzya it is not the word order which determines the main functions of elements in the sentence, but rather case marking does: "Person cross-referential marking on the finite verb, or other points of predication, appear to supersede Subject and Object arguments" (2010: 26). As a result, word order shows variation in Erzya. I agree that the case is more complex than Zaicz's description suggests, and the question of word order in bilingual sentences would need further investigation.

Various word orders co-exist in Russian, but SVO is considered dominant (WALS 2016). The different word orders can have different stylistic functions, whereas syntactic role and semantics can also influence the choice of word order. According to Timberlake (1993: 858), "the word order of the predicate and its major noun phrases (subject and objects) is relatively free in Russian. [...] The naturalness and frequency of various orders depends on the role of the noun phrase and the semantics of the verb".

If we accept either the dominance of the SVO pattern or the relatively free order view of the word order in the two languages, both lead to a conclusion that presumably switching is not inhibited by the position of the constituents in the sentence.

To sum up, in this chapter, I have discussed the main incongruent morphosyntactic (and to some extent phonological) characteristics of the two languages. My hypothesis is that Erzya–Russian bilinguals apply different strategies when it comes to CS between these incongruent structures, which leads to variation in CS constructions. Typologically congruent Erzya and Russian structures are also relevant to my study, as I argue that typologically congruent constructions facilitate CS, and through CS also structural interference on the long run. I discuss Thomason's 2014 findings related to this question in section 5.

Chapter 4. Earlier research on contacts between
Finno-Ugric languages and Russian

CS research on the Finno-Ugric minority languages spoken in the Russian Federation has started only recently. Studies carried out in the 20th century mainly focused on the Russian lexical (and to some extent structural) borrowings of these languages, CS was or still is considered a chaotic phenomenon not worth studying (Saarinen 2014). During Soviet times, researchers from outside the Soviet Union could not (easily) get access to naturally occurring spoken data in these languages. What is more, the technical possibilities did not allow for the collection and storage of such large amounts of recordings as they do today.

The end of the 20th century and the beginning of the 21th century brought about the necessary political changes and technical advancements which accelerated the pace of data collection and empowered new fields of research. At the same time, CS research also evolved, the main models that have been used to this day were created, criticized, reformed, and tested against newly accumulated data (I discuss some of these CS models in Chapter 5). The Russian Federation has been rediscovered as a treasure chest for CS studies as it is so diverse and hosts so many (sociolinguistically) different contact scenarios involving a large variety of languages.

My aim of this chapter is twofold. In the first part (section 4.1), I discuss contact phenomena and CS patterns attested in contact situations between Russian and Finno-Ugric minority languages other than Erzya. My overview is far from exhaustive, partly because, since publications on this topic are not always available outside the Russian Federation, some have not been available to me, but I intend to emphasize the striking similarities of CS patterns visible in these cases and common problems arising during their analysis.

In the second half of this chapter (4.2), I discuss the main findings of earlier studies on Erzya–Russian contacts. I refer to sources these studies rely on and emphasize the importance of studying the (written) materials from the mid-20th century that are still available today in order to understand how extensive bilingualism prevailing now in the Erzya speech community has evolved.

Finally, there is another aspect of CS research I have to mention here. There are extensive studies on contact situations between Slavic and major Finno-Ugric languages (Hungarian, Finnish, and Estonian, i.e the state languages), with especially the Finnish–Russian, the Estonian–Russian, and the Hungarian–Slovak contacts (Lanstyák 2011) described widely. Although they can give important insights into CS between languages that are typologically similar to Erzya and Russian, respectively, they all represent a different type of contact situation sociolinguistically, given the existence of a monolingual (or in case of Finland, a bilingual) state where these languages have an official status. As a result, I do not discuss them here. However, these cases could be also involved in a future comparative typological study of CS between Finno-Ugric and Slavic languages.

4.1. The study of code-switching in other minor Finno-Ugric languages

In this section, I give an overview of studies on CS between Russian and minor Finno-Ugric languages (Finno-Ugric languages spoken only as minority languages). I do not discuss the methods or the data of these studies in detail, I only focus on parallel Russian origin mixed constructions that occur in these and in the Erzya–Russian CS corpus as well: e.g. necessive constructions and numeral phrases.

CS between the Finnic languages and Russian has been extensively studied, recently there has been even more research on the Permic languages and Mari, and also on other languages to some extent. Here, I discuss the results of research on Karelian (Pyöli 1996, Sarhimaa 1999) and Votic (Turunen 1997) from among the Finnic languages, Udmurt (Salánki 2007, Shirobokova 2011, Edygarova 2014) and Ižma Komi (Leinonen 2009) from among the Permic languages, as well as Gavrilova's 2012 and 2013 findings on Mari and Pineda's 2009 examples from Lovozero Saami. I include a paper on the other Mordvin language, Moksha in this part, although I discuss some earlier and recent studies in 4.2 which involve both Mordvin languages or do not differentiate between them at all.

These minor Finno-Ugric languages have been in contact with Russian for several hundred years. Sarhimaa (1999) argues on the basis of Karelian–Russian CS that during intensive contact,

languages become more similar typologically. For example, the results of this convergence in the Karelian–Russian contact situation are the rich case system and the free word order in both languages. As will be mentioned below, it seems plausible that in cases where there is a clash between the forms of the two languages, CS is not necessarily avoided, but as a result of the incongruence, hybrid constructions occur.

4.1.1. The differentiation between code-switching and borrowing

One of the main debates in CS research concerns the differentiation between CS and borrowing (cf. section 5.1.2). Such differentiation is especially problematic in Karelian–Russian and Moksha–Russian CS. Sarhimaa (1999: 194) argues that because of the convergence of the Karelian and the Russian phonological systems, it is not possible to claim that a Russian origin lexeme has been adapted phonologically to Karelian. The same problem arises in the case of CS between the Mordvin languages (Erzya or Moksha) and Russian, because the sound systems of both Erzya and Moksha are similar to Russian (this is true especially in the case of Erzya, cf. section 3.1). Saarinen (2014: 541) argues that this makes the differentiation between code-switches and borrowings difficult in her Moksha–Russian data. In the analysis of code-switching between such languages, phonological adaptation cannot be used as a single criterion to differentiate between these two contact phenomena. As opposed to this, it is easier to attest switches in the case of bilingual Mari–Russian discourse, as Mari phonotactics is significantly different from Russian.

Saarinen (2014) regards CS and borrowing as two sides of the same process, however, she differentiates between them using a variety of criteria. Although she lists the problems and exceptions where this method does not work, she considers phonological and morphological adaptation the main criteria that can be used to differentiate between these two phenomena. On the one hand, problems with the phonological adaptation criterion lie in the similarity of the phonological system of the two languages. On the other hand, the morphological adaptation criterion does not work, for example, with adverbs that are inserted into Moksha discourse without any adaptation. The same dilemma is created by longer Russian sequences inserted into

the Moksha structure using Moksha morphological markers (e.g. a possessive suffix). The same type of construction occurs in the Erzya–Russian CS data, analyzed in Chapter 7.

Saarinen (2014) discusses the lexical criterion as well, namely, she takes into consideration whether the word is represented in the dictionaries of the Moksha language or not. If the element is not included in them, it should be regarded as a code-switch. This, however, might be problematic in cases of words that are widely used by the speakers but are not considered as part of the Moksha vocabulary by compilers of dictionaries. The same happens when the words sound exactly the same in the two languages and, as a result, they are, again, not included in the dictionaries due to spatial considerations. Saarinen also mentions that although there is a complete and regularly updated dictionary for the Moksha standard variety, it is not the case with other varieties of Moksha. For instance, there are no complete dictionaries for Moksha regional dialects, which makes it impossible to decide whether a Russian word is a borrowing or a code-switch in a given variety of Moksha.

Leinonen (2009) relies on the evaluations of her two Komi participants who were asked to judge how entrenched a Russian element was, whether it can be considered a code-switch or a borrowing. According to Leinonen (2009), the native speakers of Komi based their decisions on inflection: "If a Russian lexeme, even such as they had never heard being used in Komi speech, was grammatically embedded with Komi suffixes in the sentence, it was deemed as an individual loan by the speaker" (Leinonen 2009: 319). (Evaluation by the members of the speech community is used as a distinguishing criterion in Backus's 2015 usage-based model as well, cf. section 5.3.3.2.)

Finally, Turunen (1997) raises a puzzling problem he stumbled upon in his Votic data, which is relevant in other CS discourses as well, also in the case of the Erzya–Russian CS patterns. Turunen claims that some of the constructions in his CS data do not seem to abide by the rules of either Votic or Russian. He cites the example *nojābre* 'in November', which contains the Russian stem and the instrumental case ending, but the preposition *v* 'in', required in Russian, is missing. Turunen suggests that the name of the month was borrowed together with the instrumental ending from Russian. There is, however, another explanation for this phenomenon. In my opinion,

this is not a borrowing, rather an instance of CS, and the preposition is missing as a result of preposition deletion typical in Russian spoken language (Shagal 2016).

In my data, there are also code-switched elements that are longer than one word and are transferred to Erzya as a chunk, as one unanalyzed segment. This is especially typical with possessive constructions, place names, and dates. The evidence that they are chunks can be found in examples in which the native case ending is added at the end of the phrase, whereas if it was analyzed, the case would be attached to the head.

4.1.2. Code-switching types in other minor Finno-Ugric languages

Studies of CS between Russian and minor Finno-Ugric languages often argue that in these bilingual discourses some topics are typically discussed in Russian, which triggers CS to Russian even more (Pyöli 1996, Salánki 2007, Shirobokova 2011, etc.). Despite the fact that Komi, Udmurt, Mari and both Mordvin languages have official status in their respective republics, it is typically Russian which is used as the language of communication in official domains, e.g. in administration, in the mass media, in education, and in the military, etc. This phenomenon is not only characteristic of the Finno-Ugric languages, consider what Comrie (1981: 34) reports on Chukchi texts: "in looking at a page of modern Chukchi, for instance, unless it deals with traditional life, the impression given to the reader is of basically Russian vocabulary with a few strange words and a number of strange inflectional affixes."

CS patterns in these different contact situations show both semantic and structural similarities, and the typical switched elements belong to the same parts of speech. The most frequent one-word switches are the following: nouns, verbal stems, adjectives, adverbs, discourse particles, and interjections. These one-word switches are the most problematic when it comes to the analysis, because they raise the question how to differentiate between code-switches and borrowings. In all the case studies, longer switches are also discussed, predominantly numeral phrases, prepositional phrases, verbal constructions, and idioms (inserted with an unchanged Russian construction, or as partly adapted). Finally, hybrid

constructions are also attested in all of these contact situations: these involve mainly double morphology and code switches with structural copies.

Russian discourse particles and conjunctions are established borrowings in many minor Finno-Ugric languages. Even if they are not loanwords, they easily undergo switching, as argued by Matras (1998). If they are shared by the two languages, they can function as bridge words triggering the insertion of further Russian elements. Saarinen (2014) discusses an example of this phenomenon in Moksha–Russian CS. The use of the borrowed conjunction *i* 'and' triggers the use of the Russian adverb *srazu* 'immediately', resulting in a longer switch *i srazu* 'and immediately'.

Russian numeral phrases are very commonly switched in all of the above mentioned contact situations, although their use also shows variation to some extent. One of the common features is the use of Russian numerals when expressing prices, dates, years of birth, etc. (cf. Laakso 1996: 226 about Votic). Pineda reports the same for Lovozero Saami: "The year of birth and age are almost always given in Russian"[3] (Pineda 2009: 33, my translation).

This can be explained by the fact that speakers of these Finno-Ugric minority languages communicate with authorities in Russian and usually learn school mathematics in Russian. Shirobokova (2011: 117) also suggests that, in the Udmurt community she examined, this can be attributed to the fact that in the kindergarten it is only the Russian numerals (and also only the Russian color terms) that are taught to the children.

Compound numbers are typically expressed in Russian, smaller numbers, however, can occur in their native forms. Sarhimaa (1999: 234) describes the Karelian situation as follows: "Russian numeral phrases carrying Russian system morphology are almost exclusively used in place of the inherited numerals in Karelian in general. Inherited Karelian compound numerals, especially, are seldom found, even in the speech of the most traditional speakers." According to Turunen (1997), Votic numerals are used from 1 to 7, and also dates which are important to the speaker are expressed in Votic, not in Russian (1997: 219).

Numeral phrases are often switched as one unit, as chunks which are "multimorphemic elements of a language that are accessed as unanalyzed (i.e.: lexical) units in speech" (Backus 1999: 93). The phenomenon is not a recent development in the Finno-Ugric languages. Tsvetkov

---

[3] "Год рождения и возраст почти всегда даются на русском".

(1925: 44) wrote about this already in the 1920s. In example 13, the numeral and its complement are in the forms required by the rules of the Russian language, the word *veṛo* 'tax' is in Votic and has the partitive ending that is the required form in Votic after attributives expressing quantities (Tsvetkov 1925: 44). (The examples are in italics, Russian elements are highlighted with boldface while the Votic elements are indicated with normal typeface.):

(13)　*ṗatnadcať*　**rubľ-ej**　　　　*veṛo-a*

　　　fifteen　　　ruble-GEN.PL　tax-PART

　　　'fifteen rubles tax'

Besides numeral phrases, verbal constructions are also commonly switched using Russian morphological markers. According to Pineda (2009), it is especially common in the case of motion verbs, resulting from the fact that the system of motion verbs is more complex in Russian than in Saami. Incongruence and the greater complexity of the Russian form explain the use of Russian code-switches in one other case as well. Saarinen (2014) discusses an example which shows that the use of the Russian *habeo*-type transitive verb *imeť* 'to have' is also attestable in Moksha instead of the native possessive construction with the copula. Both of these Russian verbal constructions are used in Erzya–Russian bilingual discourse as well (Chapter 7). None of the Finno-Ugric languages has gender as a grammatical category, nevertheless, gender agreement in case of a code-switched Russian predicate sometimes occurs. Gender agreement is studied to some extent by Sarhimaa (1999), but extensive research on the topic has not been carried out yet. The use of Russian verbal constructions can also be attested in earlier records of CS, for example in Juhász's 1929 and Hallap's 1960 Moksha examples. However, there is no indication in these studies of the application of gender agreement, the masculine form is used generally, even with feminine subjects. In the case of Moksha, there is no indication of gender agreement even in dialect texts collected between 1960 and 1990 (Saarinen 2014). In one of the examples, the speaker uses the masculine form with an Erzya pronominal subject, although the feminine form should be required by the Russian rules, as we can judge by the Russian translation of the sentence. Unfortunately, there is no information on the age and background of the speaker.

Finally, there is a type of switching in Russian–Finno-Ugric contact situations which involves not only lexical but also structural transfer, namely, the use of the Russian constituent

order (head–genitive modifier) in constructions expressing possession. The Russian order is preserved if the whole Russian construction is switched as a chunk, but the Russian type order occurs also in cases in which part of the Russian construction is replaced by the native counterpart. Leinonen (2009: 324) cites an example in which the Komi constituent order and morphological markers are combined with their Russian equivalents.

It is also typical in the contact situations between minor Finno-Ugric languages and Russian that speakers use double marked forms, partly for emphasis, partly to overcome structural incongruence. Gavrilova (2012: 58) analyzes examples of double morphology in Mari–Russian CS. For example, the Mari comparative suffix (-*rak* 'more') can be used on a Russian comparative form *boľše* 'more' in a double comparative construction *boľšịrak* 'even more, lit. morer'. There are also CS cases in which not only one function is expressed with two markers, but the whole construction is repeated. In Gavrilova's data, these can occur in both directions (Russian to Mari and Mari to Russian). According to Gavrilova (2012: 58), if the Russian element is followed by the Mari element, this happens because the speaker unconsciously intends to "get rid" of a foreign element that "does not belong" there. If the switch happens the other way round and the Russian construction follows the Mari one, it has an explanatory function, the speaker wants to make sure that the audience understands them. On the basis of his Votic data, Turunen (1997) claims that these repetitions are mainly for emphasis, or they are used to insert a pause during which the speaker can think.

Even though the Finno-Ugric languages are genetically related and display similar typological characteristics, types of code-switching attested in each contact situation exhibit variation. The sociolinguistic background of the given contact situation also plays a role in the CS behavior of the speakers. The extent of convergence between the two languages involved can limit the variety of CS, too. Moreover, the type of the data can also determine the extent of CS, according to Saarinen (2014). For instance, in her Moksha data there are more switches in interviews when subjects are talking about their lives than in folk tales and other narratives. My data are more homogenous in this respect, as I did not collect folkloristic material and the radio data also mainly consist of interviews. I elaborate on this question in the methodology section of my paper (Chapter 6).

The extent of CS in a speech community also depends on the varieties the speakers are familiar with. Edygarova (2014) studies vernacular and standard language varieties of Udmurt, and the CS styles occurring in them. How widespread and accepted CS is in a community can also influence its occurrence. As I mentioned above, purism has prevented native Mordvin linguists to study CS considered a kind of corrupt, "impure" form of the language. This of course is rooted in the fear that CS leads to language shift, through the attrition and simplification of morphology and to an increase in the number of borrowings in a language (Saarinen 2014).

4.2. Earlier studies on Erzya–Russian code-switching

As we have seen in Chapter 2, contacts between the ancestors of Russians and Erzyas date back to the 9th century, widespread bilingualism is, however, a fairly recent phenomenon. The history of Erzya–Russian CS has not been studied extensively, earlier research predominantly focused on borrowings (Mosin 2002, Agafonova 2002). In some cases, language mixing and other interference phenomena are still mentioned in the studies of Erzya/Moksha–Russian contacts, such as in Juhász's 1929 article, which analyzes CS (he uses the term language mixture for it) in Moksha written texts from the late 1920s. Hallap (1960) studies Moksha as well but includes also Erzya data in his analysis and does not only rely on written sources but uses his fieldwork materials collected in the 1950s. He claims that language mixing is not random, rather representing the accepted community norm. This indicates that the CS variety of Erzya has been developing for a longer time. However, more data would be needed to corroborate this view.

As I mentioned above, earlier research on Erzya–Russian contacts mainly focused on loanwords. In his monograph, Maticsák (2012) analyzes the earliest Mordvin written sources, making occasional remarks on the Russian influence as well. Maticsák writes about the beginnings of "Mordvin" literacy, the first Erzya and Moksha vocabulary lists, texts and translations. He also mentions the extent of Russian loanwords in these collections. In this section, I refer only to some of the publications which are relevant for the study of CS in contemporary Erzya–Russian bilingual discourse.

Apart from the loanwords, Maticsák (2012: 156) also discusses Russian structural borrowings in Damaskin's dictionary from the 18th century, in which Russian-type constituent order (head–genitive modifier) is present: e.g. *mikšńića jarsamoń* 'seller of food' in which *mikšńića* 'seller, salesperson' and *jarsamo* 'food' is used, instead of the Erzya order *jarsamoń mikšńića* 'seller of food, lit. food's seller'.

Another relevant early publication Maticsák (2012) mentions is the 1821 Erzya translation of the Gospels which contain many Russian loanwords, constructions and particles. The translator is criticized by Gabelentz (1839: 237–238 in Rueter 2010: 12) as follows:

"He has shown little concern for the purity of the language, and even where he has had several good Mordvinian phrases to choose from, he has carelessly mixed in Russian words, which may well have slipped into the text from everyday vernacular usage, or, perhaps, he has just found it difficult to replace a purely biblical word with the corresponding native words."[4]

The last remark is very important: the translator used Russian elements from spoken discourse, elements that are characteristic of the vernacular. This means that CS was accepted and used already in the first half of the 19th century, at least in spoken discourse.

Apart from a handful of studies, including Maticsák's, there are only few publications on even the Russian loanwords in Erzya, not to mention CS. However, in the first half of the 20th century, a new Erzya intelligentsia was formed and the corpus and status planning of the Erzya language became their priority (cf. section 2.3). It was at this time that the first discussions of the Russian influence on Erzya started, and the first papers were written in this topic. For example, Endjukowski (1930) analyzes an ethnographic collection published in 1922 in Saratov. In this 40-page book, he finds 428 Russian words. He categorizes them according to semantic classes and also analyzes their forms to some extent. His findings on Russian origin verbs are relevant to the

---

[4]"Dabei ist er um die Reinheit der Sprache wenig besorgt gewesen; auch wo ihm ein oder mehrere gute mordwinische Ausdrücke zu Gebote standen, hat er unbekümmert russische Wörter eingemischt, die ebenfalls theils wohl durch täglichen Verkehr sich in die Umgangssprache eingeschlichen haben mögen, theils vielleicht, als rein biblisch, sich nur mit einiger Mühe durch ein entsprechendes heimisches Wort ersetzen liessen." Rueter's translation from German (Rueter 2010: 12).

study of Erzya–Russian CS. Endjukowski (1930) lists only two verb forms with Russian imperative markers, all the other elements are adapted to Erzya morphology. All his examples are instances of the *a*-types verbs: e.g. *naŕažamo* 'to dress up' from Russian *naŕažať* 'to dress up'. His thorough analysis of the Russian borrowings in this article might have contributed to a complex account of Russian influence on Erzya, but he became a victim of Stalinist terror in 1938 and the research on this topic was halted for decades.

The next step in the Erzya–Russian CS research is Hallap's 1960 article, in which the main CS types attested in contemporary Erzya–Russian bilingual discourse are discussed, including finite verbs (without gender agreement, however) and numeral phrases. In the case of verbs, Hallap provides an explanation for the transfer of the whole paradigm of the Russian verb *moč* 'to be able to'. In Erzya, this modal meaning is expressed by the reflexive suffix *-v*:

(14)

a. *moľ-an*

go-1SG

'I go'

b. *moľe-v-an*

go-REFL-1SG

'I can go'

This syncretic construction is replaced by its Russian analytic equivalent, the finite form of the auxiliary *moč* 'to be able to' with the infinitive form of the main verb. In example 15, we can see this mixed construction in which the Russian auxiliary *mogu* 'I can' is in the first person singular and the main verb *moľems* 'to go' is in the infinitive:

(15)   **mog-u** *moľ-ems*

        can-1SG go-INF1

        'I can go'

This is also the main reason for the use of the whole paradigm of the verb *hoťeť* 'to want'. Hallap (1960) argues that these forms are widespread and represent the community norm. He judges negatively this type of language mixing which involves utterances with morphological

markers from two languages. He argues that it is not only the written media's "bad language", but the mixed variety is used by the writers of literary works as well.

He also discusses another type of CS, the use of Russian numeral phrases, cf. example 16. (In the examples, Russian origin elements are highlighted with boldface while Erzya elements are indicated with normal typeface and the discussed construction is underlined.)

(16) ***dv'esťi   rubľ-ej***   *jarmak   maks-iń*

two.hundred ruble-GEN.PL money   give-PST.1SG

'I gave two hundred rubles.'

In example 16, the construction is formed following the rules of Russian, the nominal head is in genitive plural as required in the case of the numeral modifier 'twenty'. As we have seen in section 4.1 in the case of other contact situations, it is also typical in Erzya/Moksha–Russian CS that "constructions meaning dates, time, quantity, length, weight, etc. are usually expressed only using Russian structures"[5] (Hallap 1960: 222, my translation).

As publications are available from the mid-20th century (even online), further study of Erzya–Russian CS could focus on the analysis of these written materials which reflect the language use of the era, the beginnings of the intensive language mixing period. On the basis of the fact that these constructions are attested in the written variety, in literary texts, we can imply that these structures were also present in the spoken variety of the intelligentsia. It remains a question, however, whether these forms were the only choice bilingual speakers had, and whether the use of these Russian structures had special pragmatic function in their language use (expressing, for instance, their elite status, emphasizing their language knowledge, dual identity, etc.). Another reason for the dominance of Russian elements in these texts can be the fact that many of these are translations: "Since much of the literature in languages of the U.S.S.R. is translated from Russian, there has also been Russian influence on the syntax of many languages, especially those whose basic syntactic structure differs most from Russian" (Comrie 1981: 34). In any case, there is a significant difference in the scope of CS as compared to the present situation. Nowadays, Russian borrowings and code-switched forms are eliminated from official documents,

---

[5] "Harilikult on puhtvenekeelsed ka konstruktsioonid, mis tähistavad daatumeid, kellaaega, rahahulka, pikkust, kaalu jne" (Hallap 1960: 222).

CS is present in the spoken varieties and in informal written genres such as text messages and emails.

In addition to the research from the Soviet times, there are also some recent studies on Erzya–Russian language contact, but their approach or the examples they cite raise many questions. For instance, Lomshina's 2011 article analyzes aspects of Mordvin–Russian language contact, more specifically, the phonological and morphological adaptation of Russian borrowings in the Mordvin language (sic!), the same topic her PhD dissertation discussed. The main problem with her paper is that she does not differentiate between the Erzya and the Moksha languages but uses the term Mordvin. As mentioned in Chapter 2, some of the local linguists consider the two Mordvin languages dialects. It would be an acceptable approach if Lomshina (2011) made it clear throughout the paper which dialect, Erzya or Moksha, she is currently discussing, as there are significant differences between the adaptation strategies of the two. There are only two cases in the article where she mentions the designations Moksha and Erzya, once in the phonology part of her paper, when discussing vowels in syllables preceding the stressed syllable. (It is important to make the differentiation here, as Moksha has reduced vowels in these syllables, while Erzya does not.) In the second case, the distinction is not relevant, she only emphasizes that all four dictionaries (Moksha–Russian, Russian–Moksha, Erzya–Russian, Russian–Erzya) contain a lot of Russian loanwords, which proves that the Russian language has exerted a considerable amount of influence on the development of the Mordvin language. The lack of differentiation between the two languages can result in serious misconceptions in readers. For example, Lomshina discusses the adaptation of Russian verbs and claims that while earlier borrowed verbs were adapted directly, without a thematizing suffix (Mordvin *mešams* 'to disturb' from the Russian *mešať* 'to disturb', with *-ms* and *-ať* being the respective infinitive endings), more recent borrowings use the thematizing suffix *-nda* (Mordvin *kriťikovandams* 'to criticize' from Russian *kriťikovať* 'to criticize'). On the basis of this, one might draw the conclusion that both Mordvin varieties use this thematizing suffix. However, it is attested only in Moksha, Erzya uses direct transfer instead. I discuss this issue in detail in 7.2.3. The part of Lomshina's 2011 article focusing on phonological adaptation is even more problematic: for example, she uses no transcription or any phonological notation to illustrate (de)palatalization but relies on the orthographical form of

56

the verb. To understand what she means by the example Mordvin *апельсин* 'orange (fruit)' < Russian *апельсин* 'orange (fruit)', the reader has to be aware of the fact that the pronunciation of these words are [ɐpelʲˈsʲin] and [ɐpʲɪlʲˈsʲin], respectively.

While Lomshina (2011) focuses on loanwords, there are other recent studies which also mention cases of CS: Agafonova (2002) discusses the diaspora speakers in the Transvolga region where Erzya speakers have little connection with the Mordvin Republic, no education is available in Erzya, and their language use shows traces of heavy Russian influence, while Mosin (2002) concentrates on the bilingualism of Erzyas living in the Mordvin Republic.

Agafonova (2002) catalogues contact phenomena typical of the diaspora varieties. Numeral phrases (especially those expressing numbers higher than eleven), time expressions, dates, and quantitative constructions are commonly used in Russian in these communities. Synthetic structures are replaced by analytic ones, and Erzya suffixes are substituted by Russian prepositions. Some of these phenomena are also widely applied in the spoken variety of Erzya in the Mordvin Republic (cf. Chapter 7). The extent to which Russian contact phenomena are applied differs from generation to generation; Erzya forms are characteristic of older speakers, while variation between Erzya and Russian forms is typical of the middle-aged and younger generation.

Mosin's article (2002) reflects the general view that the extent to which Russian forms are used varies according to the age, profession, and location of the speakers. Erzya speakers living in regional centers and villages closer to the capital, Saransk or other Russian majority cities apply more Russian forms in their speech. Extensive use of Russian elements is also typical of younger speakers, especially those who had no schooling in Erzya and could not learn the language as a subject at school. Blue-collar workers also tend to apply more Russian constructions than white-collar workers. Mosin does not carry out quantitative analysis of the contact phenomena, nor does he present the sociolinguistic background of his subjects; instead, he reports tendencies in Erzya–Russian CS discourse. His arguments are based on common knowledge shared by the community. In my opinion, however, further research should concentrate on the quantitative analysis of these constructions and also on the study of how the sociolinguistic background of the speakers correlates with the number and type of Russian constructions in their speech.

Mosin does not differentiate between the various contact phenomena; he lists earlier Russian borrowings and code-switches in the same group. The categorization he uses is also ambiguous. He differentiates between partial mixing and total mixing on the basis of the nominal and verbal stems applied in the clause. Example 17 is considered total mixing because all the word stems originate from Russian. The fact that the grammatical markers are all from the Erzya language is not taken into consideration (Mosin 2002: 162, my transcription, translation, and glossing; bold also by me in example 17 and 18).[6]

(17) **sobrańija**-*so-ńť*   **ŕeša**-*śť*   **śakoj**   **vopros**-*t*

     meeting-INE-DEF.SG decide-PST.3PL   every.kind   question-PL

     'They decided on every kind of questions at the meeting.'

Example 18 is considered to be an instance of partial mixing, due to the fact that one of the elements (*siń* 'they') is from the Erzya language. In this case, however, the predicate is a Russian verb used in its finite form in the sentence, so, in my opinion, it is a more mixed sentence than example 17, in which the grammatical frame is clearly set by Erzya.

(18) *siń*   **karta**-*so*   **najaŕiva**-*jut*

     they   card-INE   play-3PL

     'They play cards.'

Finally, Mosina's 2002 book has to be mentioned, which discusses a special part of the Erzya–Russian bilingual community, the bilingual children and their language acquisition process. In her examples, the use of Russian finite verbs is typical, and in one instance gender agreement is also present in an utterance with an animate nominal subject. In Mosina's study, children were asked to describe pictures, one of which was a picture of a dog that wanted to climb a tree. In example 19, the subject is the Erzya word *kiskaś* 'the dog' and the predicate is the past tense feminine form of the verb 'to want' *hoťela* 'she wanted' (my transcription, translation, and glossing; bold also by me).

(19) *ťese*   **_hoťe-l-a_**   *kiska-ś*   *čuvto* *pŕa-s*   *kuź-ems*

     here   want-PST-F   dog-DEF.SG tree   top-ILL   climb-INF1

     'Here the dog wanted to climb to the top of the tree.'

---

[6] We do not know anything about Mosin's sources in general, or where these examples come from in particular.

We can explain the use of gender agreement in two ways. On the one hand, it is possible that the child recognized that the dog in the picture was female. On the other, it is also plausible that the feminine gender of the Russian word *sobaka* 'dog' triggered the gender agreement. Unfortunately, there are no more past tense examples in Mosina's study, and she does not comment even on this one – all the other cases are present tense forms of finite verbs, which do not involve gender agreement in Russian. It would be worth investigating the language use of different generations, in order to see whether gender agreement is spreading to include not only Russian origin words but also Erzya stems and possibly also inanimate objects.

Finally, Luutonen's 2014 article also discusses issues related to CS and borrowing. He compares the language use of two Erzya generations using the Erzya translation of a Russian novel and corrections made during proofreading the text. Two native speakers participated in this project, the younger one (born in 1983) translated the text, while the elderly speaker (born in 1952) proofread it. Luutonen (2014) discusses different types of changes applied to the text (corrections of the grammar, vocabulary, and stlye, etc.). One of the significant modifications concerns Russian discourse particles and adverbs. While the translator considered these acceptable and natural in standard or even in literary Erzya language use, the proofreader replaced them with the Erzya equivalents. Luutonen (2014: 191) argues that, however conscious they might be as regards their language use, young intellectuals do not avoid Russian elements when it comes to the above mentioned two word classes (discourse particles and adverbs), which suggests that they consider these Russian origin words parts of the Erzya vocabulary.

On the basis of this data, a hypothesis can be formed concerning the diachronic development of a CS mode in Erzya–Russian bilingual discourse. While Russian influence has earlier also resulted in convergence, e.g. changes in word order, and lexical borrowing, widespread bilingualism and extensive CS is presumably a 20th century phenomenon. Hallap (1960) already considered CS a community norm in 1960. The Russification policy prevailing in the 1950s might have also contributed to the wide use of Russian in the Erzya community. In contrast to Hallap's findings, the present situation seems more varied, due to the changes after the dissolution of the Soviet Union (Chapter 2). It became possible to carry out both status and corpus planning for Erzya, by coining neologisms and by creating new domains for language use,

and to create a near-monolingual Erzya style used mainly by intellectuals through the conscious elimination of Russian code-switches, which was not against official policy any more. As a result, the present situation is characterized by a variety of CS styles, ranging from a nearly monolingual to a heavy mixing variety. I discuss this variation in detail in Chapter 7.

Chapter 5. Theoretical background

In this chapter, I discuss the main theoretical issues concerning the study of CS, typical research questions of the field, and possible answers offered by various models. As the scope of the CS research is vast, I only focus on issues relevant to my study. In section 5.1, I discuss definitions of CS (5.1.1), the possible connection of CS to various other contact phenomena, and especially the ways it is usually distinguished from borrowings (5.1.2). CS frameworks relate differently to the scope of CS (5.2.1). Structural models predominantly disagree on two aspects of CS: the existence and universality of grammatical constraints on CS (5.2.2), and the existence of a matrix language (5.2.3).

In section 5.3, I provide an overview of the various CS models I rely on in my study. First, I discuss an influential framework widely used for the analysis of CS, the Matrix Language Frame (MLF) model by Myers-Scotton (1993, 2002) in section 5.3.1. While the MLF model is applicable to instances of *classic CS*, it cannot be used for the explanation of more complex cases of *composite CS* involving mixed constructions. Using Muysken's 2000 typology (cf. section 5.3.2), not only clearcut CS types can be categorized but also hybrid cases, which are abundant in the Erzya–Russian data. Although I carry out a synchronic analysis, I agree with researchers who take into consideration both the synchronic and the diachronic aspect of CS, linking CS to contact induced language change. That is why I also discuss models connecting different contact phenomena and structural change, e.g. Johanson's 1999 code-copying framework and Backus's 2015 usage-based model (in section 5.3.3).

Typological distance and the level of congruence between the constructions of the two languages should also be taken into consideration when analyzing mixed CS forms. Congruence is discussed in several structural models of CS (including e.g. the MLF model), but it is a central concept in Sebba's 2009 typology (section 5.3.4), which operates with three main strategies, harmonization, neutralization, and compromise.

There are CS types in the Erzya–Russian data, the analysis of which is not possible if we apply a purely structural approach. There have been a variety of attempts to combine various models (Muysken 2000, Isurin et al. 2009, Backus 2015, Stell 2015, etc.), but a complex

interdisciplinary framework has not been created yet. In this dissertation, I intend to combine various approaches to explain one specific phenomenon in the Erzya–Russian data, flagged switches. In order to achieve a more complex view of this type of CS, I combine the grammatical approaches discussed in the previous sections with Gafaranga's 2000 model (section 5.3.5).

In section 5.4, my main focus is continuum models (mainly Auer's 1999 and Kovács's 2001 continuum) which can be used to illustrate how CS in the synchronic state can predict further changes in the CS scenario and which provide an insight into what an earlier stage of the contact situation looked like, thus connecting the synchronic and diachronic aspects of CS. They can also be used to describe the variation currently present within the speech community.

5.1. Code-switching

Code-switching has been one of the most widely studied language phenomena in contact linguistics since the second half of the 20th century. Although research into it looks back to a relatively short, albeit complex history, a consensus concerning the definition or the scope of the term has not been reached yet. Code-switching has been studied from a variety of perspectives (structural, pragmatic, sociolinguistic, and psycholinguistic, to name a few). Recently, claims have also been made for the need for an interdisciplinary approach (Isurin et al. 2009, Stell and Yakpo 2015). In the present dissertation, the main focus is on the grammatical analysis of the morphosyntactic structure of code-switched utterances. However, I also refer to other possible directions of further research which could apply the results of this structural study.

5.1.1. The definition of code-switching

Code-switching is a complex contact phenomenon which has been defined in many different ways depending on the focus of the particular study and the type of the data under scrutiny. The main difference between the definitions concerns the scope of the term. The situation is aptly described by Milroy and Muysken (1995) as follows: "the field of code-switching research is replete with a confusing range of terms descriptive of various aspects of the phenomenon.

Sometimes the referential scope of a set of these terms overlaps and sometimes particular terms are used in different ways by different writers" (1995: 12).

I will refrain from discussing the history of the term here, as there are detailed overviews available on the changes in the meaning and definition of CS one can refer to (cf. Milroy and Muysken 1995; Clyne 2003, or Mahootian 2006, etc.). However, there are some aspects of CS that have to be discussed here as well. As mentioned above, the definition of CS varies from study to study, ranging from broad to narrow worded descriptions. When it comes to the definition of CS, one of the most cited CS terms is Gumperz's (1982: 59), whose research revealed that CS is not random but rule-governed. He defines CS as follows: ''the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems''. Other definitions are less specific, e.g. Grosjean (1982: 145) does not even differentiate between intrasentential and intersentential CSs, defining CS as "the alternate use of two or more languages in the same utterance or conversation". Definitions differ not just as regards structural aspects, but also according to the view concerning the actors of CS, for instance, Gumperz's and Grosjean's definitions do not refer to the speakers, whereas Gardner-Chloros's term involves the bilingual language users as well. In Gardner-Chloros's (2009: 4) view, CS is "the use of several languages or dialects in the same conversation or sentence by bilingual people".

Besides its definition, the term CS itself has evoked heated debates. The term *code* was originally borrowed from information theory (Alvarez-Caccámo 1998), and in contemporary CS research it is applied widely "as a neutral umbrella term for languages, dialects, styles/registers, etc., and partly usurps the place of the more usual 'catch-all' term *variety* to cover the different sub-divisions of 'language'" (Gardner-Chloros 2009: 11). Alvarez-Caccámo (1998) reconceptualizes the meaning of *code* and argues that we have to differentiate between the code and the involved languages, as the basic code in a mixed discourse can be considered bilingual. Gafaranga (2000) also applies this approach in his model of CS, using the term *medium* to designate the scheme of interpretation and claiming that speakers have a preference for same-medium talk. If a switch occurs in the medium, it is either repaired or has its own function (cf. section 5.3.5).

The other part of the term, *switching*, however, is more ambiguous, referring to different phenomena in different models. While it is used as an umbrella term by Myers-Scotton (1993), Muysken (2000) applies it only to designate a specific type of contact phenomenon, while he uses *code-mixing* as a general term. Muysken reserves the term CS to designate only a subtype of mixing, namely, *alternation*, in which "there is a true switch from one language to the other, involving both grammar and lexicon" (2000: 5). As a result, in Kovács's terms (2001), a terminological jungle is created, in which authors focusing on the same contact phenomenon use different designation for it. For example, Clyne (2003: 75) also limits the scope of code-switching, but, in contrast to Muysken, he refers to a complete switch as transversion and labels other types of lexical transference as CS. Johanson (1999) coins the term *code-copying* and rejects both the terms *switching* and *mixing*, as in his view elements are not just switched but adjusted to the recipient language which they become a part of. According to Johanson (1999: 40), mixing is also not an adequate term, because code-copying "does not produce any fusion or amalgam of codes, nor any 'mixing' with two participating languages", there is always only one system, the original is just copied from the model code, but the new element becomes a part of the recipient base code's system.

In my paper, I use the term CS and apply a broad definition of it: the use of two languages in the same discourse. However, my analysis focuses only on intrasentential switches, as the structural analysis makes sense only in these cases. In intersentential switches, the switch occurs on clause boundaries and, thus, the languages do not combine, because the morphosyntactic frame of each clause is provided by one language.

5.1.2. Code-switching and borrowing

In addition to the debate concerning the definition and scope of the term CS, another major issue in the field is how to distinguish between borrowings and code-switches. Winford (2003: 107) gives an overview of different approaches and provides the two main criteria used to differentiate between the two phenomena: the "degree of use by monolingual speakers" and the "degree of morphophonemic integration". Then he discusses the problems with these criteria, e.g. the fact

that the difference between the two is not clear for the speakers themselves, and how phonological and morphological integration is possible but not always present both in the case of borrowings and switches. There are a number of studies in which empirical data also seem to corroborate this claim. For example, Zabrodskaja and Verschik (2015: 449) emphasize on the basis of their findings on the insertion of Estonian elements into Russian discourse that phonological adaptation and morphological integration do not depend on each other. In another study on Serbian–English contact, Savić (1996) argues that English discourse markers are syntactically integrated into Serbian but not adapted phonologically.

Resulting from these ambiguities, there are researchers who do not consider borrowing and code-switching to be distinct categories but rather as points on a (diachronic) continuum (e.g. Treffers-Daller 1991, and Thomason 2001). Others make a clear distinction between the two phenomena. For instance, Poplack's 1988 model reflects this kind of a clear distinction between code-switches, nonce borrowings, and borrowings. While borrowings form a continuum between established and nonce borrowings, code-switching is considered to be an entirely different phenomenon in this model. Nonce borrowings are idiosyncratic, morphologically, syntactically, and optionally also phonologically integrated borrowings, while established borrowings are recurrent in the individual's speech, widespread within the community, and adapted to the recipient language. Poplack's ideas have been criticized (for example by Croft 2000: 211) for including the category nonce borrowing in the model only to explain why some insertions violate the constraints which were supposed to be universal (cf. section 5.2.2 for details of the constraints). Croft (2000) also emphasizes that distinguishing between CS and borrowing is even more problematic in heavy-switching discourses: "when the so-called nonce borrowing occurs in the middle of a highly code-mixed utterance, it hardly differs in status from code-mixing itself" (Croft 2000: 211). Romaine (1995: 140) describes this tendency as follows: "There is an inverse relationship between the degree of similarity between any two languages in contact, which will have the effect of maximizing the potential sites for code-switches, and the extent to which code-switching and borrowing can be distinguished".

On the basis of the above mentioned difficulties in distinguishing between CS and borrowing, the continuum model of contact phenomena seems more plausible. I agree with

Backus's 2015 view in that the connection between borrowings and code-switches is usually diachronic, a code-switch in the synchronic state can become a borrowing later if frequently used by a number of speakers, and if it becomes entrenched. Croft (2000: 211) describes this process as follows:

> "borrowing begins with some bilingual speaker using a foreign word, which then is propagated through the speaker's community, including to other monolingual speakers. This is also a gradual process, with variation in the degrees of integration of the foreign word into the system of the language until it is fully conventionalized."

According to Backus's 2015 usage-based model (discussed in detail in section 5.3.3.2 below), there are two methods which can provide insights into the status and the entrenchment of a given lexeme even in synchrony: frequency analysis of big data and judgement tests of native speakers.

5.2. Main issues of structural studies on code-switching

In the first part of this section, I discuss the main questions grammatical studies of CS has focused on. The overview is far from all inclusive, and I highlight only some aspects this field of research touches upon, as they are also relevant for the analysis of the Erzya–Russian corpus.

Section 5.2.1 describes changes in the scope of CS research in general. In section 5.2.2 and 5.2.3, I discuss two central issues which have been debated in structural studies of CS: the existence of constraints and the presupposition of a matrix language.

5.2.1. The scope of code-switching research

A substantial proportion of researchers of CS focus on only one aspect of this language contact phenomenon. Following Gardner-Chloros (2009: 10), we can differentiate between three main types of descriptions, the sociolinguistic/ethnographic, the pragmatic/conversation analytic, and

the grammatical approaches. While sociolinguistic studies focus on the speech communities' sociolinguistic situation, pragmatic studies analyze the conversation organizing functions of CS and new meanings brought about by the switches, and grammatical ones investigate the structural characteristics of code-switched utterances and possible constraints governing them.

Studies of CS generally give an overview of the history of CS research (cf. Milroy and Muysken 1995 and Gardner-Chloros 2009, etc.). In this paper, I only mention certain milestones in this field of research and discuss in detail only the models I based my analysis on. Early studies of language contact (Haugen 1950 and Weinreich 1953) focused on borrowing and interference and dismissed CS as arbitrary mixing. Milroy and Muysken (1995: 8–9) discuss the possible reasons why these otherwise seminal works on language contact regarded CS as a sign of speakers' deficient language proficiency. CS became the center of attention only later, in the 1960s and early 1970s, when Gumperz's studies revealed its social and discourse organizing functions. This established the socio-functional approach to CS, which has two main branches: studies offering an organizational explanation of CS using a conversation analytic model, and studies focusing on the identity-related functions and metaphorical uses of CS strategies.

Gafaranga (2007: 279) emphasizes that the focus of code-switching research has been characterized by the debate with this early monolingual linguistic ideology involving puristic and negative attitudes towards CS. Authors have intended to prove that switching in bilingual discourse is not random. He claims that the structural and the socio-functional studies both tried to demonstrate the orderliness of this phenomenon, but generally independent of each other. However, there are recently more and more studies which intend to apply an interdisciplinary approach, e.g. unite the structural and the sociolinguistic aspects of the study of CS. In section 5.3, I discuss two such approaches: Muysken's 2000 typology and Backus's 2015 usage-based model.

In the next two sections, I focus on two problems structural models have faced in searching for universal rules concerning the mechanisms of CS and for an explanation where a switch can occur.

5.2.2. Constraints

CS is often considered negatively by the general public and also by the speakers themselves (Bullock and Toribio 2009: 1). The early researchers on contact linguistics also regarded this bilingual phenomenon a sign of interference, and a proof of the speakers' limited proficiency. In his typology, Muysken (2000) discusses these early studies and points out that they acknowledge only CS cases in which the interlocutor or other aspect of the communicative setting has changed.

From the 1980s, studies focused on formulating constraints indicating where switching is possible between two languages. There is a debate regarding the existence and scope of constraints within CS. Following Chan (2009), we can differentiate between three approaches. Studies which focus on the pragmatic aspect of CS do not operate with constraints, as they believe that CS is governed by communicative functions, not structural constraints (e.g. Gardner-Chloros 2009). As opposed to this, researchers of the structural aspect of switching propose constraints, and the debate concerns their scope (Poplack 1980, Myers-Scotton 2002). While Poplack (1980) argues that principles (as the Free Morpheme and the Equivalence Constraints) are universal, other researchers, including Muysken (2000), refine this claim and talk about tendencies rather than rules which are followed in all types of CS discourses. Finally, supporters of the Null Theory (e.g. Mahootian 1993) deny the existence of specific constraints which only operate in CS discourse, and take into consideration rules that are present in monolingual speech as well.

This brings us to another major question in CS studies concerning the existence of a specific code-switching grammar. As we have seen above, Mahootian (1993) and other supporters of the Null Theory argue that the analysis of code-switched constructions does not require additional rules, they can be explained on the basis of the rules found in the two languages in contact. Poplack (1980) also suggests that the structure of the two languages determines which switches are grammatical, and that there are no additional norms or a third grammar.

This is a rather monolingual view of CS, as it analyzes the structure of mixed utterances based on the (standard, monolingual) grammar of the two participating languages. I agree with Gafaranga (2007) that CS speakers themselves do not separate the two languages as linguists do, but for practical reasons of grammatical analysis, it is necessary to take into consideration the typological and structural characteristics of the two languages. It is crucial, however, that we do

not propose constraints as prescriptive rules for CS but rather try to describe the tendencies of CS in a particular contact situation. Interdisciplinary frameworks, e.g. the usage-based model (section 5.3.3.2), take into consideration the pragmatic aspects of CS, i.e. how members of the speech community think about certain CS elements, and how entrenched they are in their opinion. In order to carry out such an analysis, however, basic studies are needed on the occurring CS types.

5.2.3. The matrix language

The other major debate concerns the existence of a matrix language, i.e. whether there is a dominant language which determines the morphosyntactic structure of the utterance, or whether the two languages provide the frame jointly. The claim of one dominant language was already present in the early studies on language contact: Weinreich (1953: 7) believed that each utterance belongs to a particular language, and Haugen (1950: 211) argued that there is no convergence between the grammatical structure of the two languages: "The introduction of elements from one language into the other means merely an alteration of the second language, not a mixture of the two."

Some more recent models also reject the idea of grammatical mixing and presuppose the dominance of one language, called e.g. the matrix language (ML) in Myers-Scotton's 1993 Matrix Language Frame (MLF) model, or the basic or base language in Johanson's (1999) code-copying framework. Gardner-Chloros (2009) emphasizes that the presupposition of a ML makes the grammatical analysis of switched utterances easier but entails a relatively recent assumption that languages are "discrete, identifiable and internally consistent wholes" (2009: 9). Gafaranga (2007: 304) also argues that the matrix language is a scheme of interpretation which has to be vital for the analysis, because "if the base language is not clearly identified, it becomes difficult to tell which elements, in a bilingual conversation, need interpreting".

While the matrix language can be identified in cases involving the insertion of lexemes from the other language, in more mixed utterances, however, the basic language might be very difficult to determine. Chan (2009: 198) acknowledges that there are cases in which there is one

dominant language, but he also suggests that "there may well be other strategies activating grammatical rules from both languages, when these choices are functionally motivated by sociolinguistic, pragmatic, or processing factors". If mixing is extensive, the language mixture might result in a new code which is different from both participating codes, but it is ambiguous how stable and conventionalized this variety is (Romaine 1995: 160).

I discuss the role of the ML, possible ambiguous cases, and the emergence of a new code in sections 5.3.1 and 5.3.5 below.

5.3. The applied models

In this section, I discuss a selection of structural models used in the analysis of the grammatical structure and the pragmatic functions of CS. I believe that by combining different frameworks, we can obtain a more versatile picture of the CS patterns present in a contact situation. Recent CS models apply an interdisciplinary approach (cf. Backus's usage-based model in section 5.3.3.2), and take into account the results of different studies (Isurin et al. 2009, Stell and Yakpo 2015, etc.).

In sections 5.3.1 and 5.3.2, I give an overview of two structural models: the MLF model and Muysken's 2000 typology, which both have connections to other disciplines (sociolinguistics, pragmatics, etc.) as well. My analysis of the Erzya–Russian data adopts a predominantly structural approach and is based on the framework of CS types formulated by Muysken. I am also interested in the connection of CS and contact induced language change, and discuss two models related to this question: Johanson's 1999 code-copying framework (section 5.3.3.1), and Backus's 2015 usage-based model (section 5.3.3.2).

In section 5.3.4, a discussion follows about the role of equivalence and congruence in CS. I discuss Sebba's 2009 model, which provides a typological explanation for the occurrence of different CS types.

In case of flagged switches, structural analysis does not always yield sufficient results, which is why I involve in the analysis Gafaranga's 2000 pragmatic model (section 5.3.5) that provides an explanation for structurally unmotivated flagged switches and tackles how speakers

negotiate language use norms and deviations from it, and which type of (a monolingual or a bilingual) medium they prefer.

In section 5.4, I discuss two continuum models which both can be adapted to the Erzya–Russian situation: Auer's 1999 continuum model demonstrates how the CS behavior of a speech community can change in time, but it is also applicable for the description of the present variation in CS patterns. Kovács's 2001 distance-based continuum model demonstrates CS patterns and their changes in immigrant communities. Although the Erzya and Russian languages have been in contact for centuries, which can involve different CS types, Kovács's model provides important insights into CS in general, and the Erzya–Russian contact situation in particular.

5.3.1. The Matrix Language Frame model

Myers-Scotton's 1993 and 2002 MLF model is one of the most widely used and widely criticized frameworks in CS studies. It differentiates between a matrix language (ML) and an embedded language (EL), which are both active during CS, but ML is more activated. While the ML provides the morphosyntactic grammatical frame of the CPs (complementizer phrase, the maximal projection of a complementizer) and the system morphemes (which do not assign or receive thematic roles), the EL is the source of the content morphemes (which assign and receive thematic roles) inserted into the ML. Muysken (2000) argues that the MLF model is therefore insertional as opposed to alternational models based on constraints.

The MLF model works as a tool for analysis in classic CS scenarios. Myers-Scotton defines these cases as follows:

"Classic CS is switching between two (or more) participating languages/varieties when speakers have strong enough proficiency in one of the languages to make it the sole source of the morphosyntactic frame that structures the unit of analysis, the bilingual CP (a clause that is the projection of Complementiser)" (2003: 189).

As opposed to this, Myers-Scotton talks about composite CS if mixing is rather intense and the EL also provides part of the utterance's grammatical frame. Classic CS is not flexible, EL elements can only be inserted in the ML if they are congruent with the ML's structures and do not violate the ML's frame. Myers-Scotton (2003) argues that as a result of change in the sociolinguistic situation, EL structural elements can also become part of the grammatical frame, these cases involve a great degree of variation and flexibility, but they are also usually short-lived and end in a switch of dominance relations, with the former ML becoming the EL and vice versa. This change is called the ML turnover in the MLF model and is considered a typical outcome after the phase of intensive mixing and a composite ML. There are also a small number of cases in which ML turnover does not occur, the mixed variety fossilizes, and a mixed language becomes the main language of a community.

The ML differentiates between three constituent types: the ML constituents, the EL islands and the mixed constituents. In ML and mixed constituents the morphosyntactic structure is set by the ML and the system morphemes also come from ML, while the structure of EL islands abides by the rules of the EL: "morphemes come from the Embedded Language and follow other well-formedness requirements of the Embedded Language (e.g. word order)" (Myers-Scotton 2006: 261).

The MLF is based on two principles which both refer to mixed constituents: the morpheme order principle claims that the surface morpheme order follows the rules of the ML, while the system morpheme principle argues that systems morphemes that are responsible for the construction of the thematic grid in the utterance are all elements of the ML (Myers-Scotton 2006: 244).

When mixed constructions do not follow these two principles, we talk about composite CS. In these instances, code-switching is accompanied by convergence, i.e. bilingual speech in which the structure of the utterance (e.g. constituent order) is set by one of the languages, while the other provides the lexical structure (Myers-Scotton 2006: 271). The MLF model does not go further in the analysis of composite CS, which is why we need another framework for the analysis of more complex CS types.

5.3.2. Muysken's typology

Muysken (2000) provides a straightforward framework for the categorization of different CS types. He uses *code-mixing* as a general term, reserving the term *code-switching* for one type of mixing, *alternation* (discussed below). While I consistently refer to the phenomenon as CS throughout the paper, in discussing Muysken's categories I use the term CS and code-mixing as synonyms.

In contrast to the long tradition of separating the different approaches to CS, Muysken links results of the grammatical analysis with sociolinguistic factors present in a given contact situation, suggesting that the sociocultural environment predicts the type of CS occurring in the given situation. Muysken manages to reconcile sociolinguistic models concentrating on variation and syntactic theories focusing on regularity. He suggests that constraints do apply, but they are not universal.

In his typology, Muysken (2007: 320) distinguishes between three categories: insertions, alternations, and congruent lexicalizations. In case of insertional switches, an element of language B is inserted into "a frame constituted by the rules of language A". There has to be congruence between the characteristics of "the inserted element and the properties of the slot into which it is inserted" (Muysken 2007: 320). Insertions are usually one-word switches or fixed phrases, resembling borrowings. In the MLF model, insertions are either considered EL islands or mixed constituents depending on the language the system morphemes come from, the EL in EL islands and the ML in mixed constituents.

The second type of CS is labelled as alternation and involves a total switch, which concerns both the grammatical frame and the lexicon. Alternation typically occurs at clause boundaries but can also happen clause internally. The MLF model does not provide an explanation for alternation, as it focuses on the internal structure of CPs (i.e. clauses), and alternation on clause boundaries involves interclausal switching. Alternations can occur if the constructions in the two languages are not congruent, elements from these languages are thus juxtaposed rather than combined. This CS type can also have an emphasizing or highlighting function, especially in case of clause-peripheral adverbial cases and discourse markers. Discourse marker switches can also be considered a different type of CS or a subtype of alternation.

The third type in Muysken's model is congruent lexicalization which involves repeated switching in a clause with a morphosyntactic structure in alignment with the rules of both languages (Muysken 2007: 322). Congruent lexicalization is very common in case of codes having gone through significant convergence (e.g. standard and dialect varieties, or closely related languages). According to Aikhenvald (2006: 47), "[i]n the situation of one language dominating the others, convergence may involve gradual adoption of the other language's structures at the expense of its own". Convergence in the Erzya–Russian data corroborates this view, as the Erzya language adopted many Russian constructions (e.g. word order) during the several hundred years of contact.

In Muysken's typology, the language contact situation and prevailing CS types are in correlation. Insertions characterize newer contact cases, while alternation and congruent lexicalization are typical in long-established contact situations. If the languages in contact are not related to each other or are very different typologically, alternation prevails, while if they are closely related or typologically similar languages, congruent lexicalization is the typical CS form. In Table 2, connection between the formal and the sociolinguistic aspect of CS structures can be observed as presented by Muysken (2000). This is, however, a simplified version of Musyken's 2000 complex model compiled to show the main occurrences of these CS patterns.

alternation	balanced bilinguals, stable bilingualism
insertion	former colonial settings, L1 is dominant
congruent lexicalization	typologically similar languages with equal prestige

Table 2. *Connection between CS types and the sociolinguistic situation* (Muysken 2000)

If the sociolinguistic situation changes, it is possible that another CS type gains dominance in the bilingual discourse. Auer's 1999 continuum model (discussed in detail in section 5.4.1) represents this change and provides an explanation for the emergence of mixed codes.

5.3.3. Code-switching and contact induced change

The connection between CS and contact induced change, i.e. the diachronic aspect of CS, has been long discussed in various CS models. Typological congruence is also related to the topic of CS. There has been a debate about the connection between the possibility and extent of CS and the typological distance between the languages. It is also worth analyzing how these two aspects are related to contact induced change. This section (5.3.3) will discuss two models which connect CS to contact induced change as well as a framework displaying the different strategies bilingual speakers apply when their two languages are typologically similar and when they are different.

I also agree with Thomason (2014), who claims that the typological congruence between languages can make contact induced change easier, but it is not a prerequisite of the change. She disagrees with the claims that structural borrowing is impossible without typological similarity and that an element cannot be borrowed if it is unattested in the recipient language.

5.3.3.1. Johanson's Code-Copying Model

Another widely used and influential framework is Johanson's 1998 and 1999 Code-Copying Model (CCM). This framework is similar to the MLF model in that it presupposes hierarchical relations between the two languages: it differentiates between a base code and a model code, the former setting the morphosyntactic frame for insertions from the model code. As opposed to the majority of CS frameworks, the CCM does not draw a line between synchronic and diachronic aspects of CS, i.e. it connects contact induced change with copying. This model does not differentiate between code-switches and loanwords, and links lexical and structural borrowing together, allowing also for mixed constructions. Johanson argues, however, that a case described as composite CS is not possible, it is always the basic language that determines the structural frame of the utterances, and elements inserted into the basic code are always adjusted to it, even 'momentary copies', they do not belong to the original system any more. "Code-copying does not produce any fusion or amalgam of codes, nor any 'mixing' with two participating languages" (Johanson 1999: 40).

In my opinion, the CCM is inconsistent to some extent, as it dismisses the existence of a mixed code but still allows for what Johanson calls mixed copies in which "[f]oreign blocks are copied *into the frame* of the basic code, whereas selected foreign properties are copied *onto units* of the basic code" (Johanson 1998: 327). Adopting a historical perspective, he argues that continuous copying can lead to increasing structural equivalence between the languages, thus creating new congruent points for copies, and the basic code is permanently reformulated. A dynamic concept of grammar is needed to describe such tendencies in prolonged contact situations, in which "it is mostly impossible to decide what is inserted and what constitutes the native code at a given point in time" (1998: 335). Where the synchronic analysis fails, however, the diachronic perspective can help see the different layers of copying.

Johanson (1999) aims at the creation of a highly accurate descriptive model with which all types of copying can be analyzed, and a coherent model of different language contact phenomena can be provided, linking individual copying to contact induced change. According to Johanson (1999), in addition to structural features, social, and psycholinguistic factors also have to be taken into account when studying copying. Johanson distinguishes between two types of code interactions: code alternation (intra- or extrasentential) and code-copying. In case of code-alternation, elements of the other language are like citations, not incorporated into the basic language. Johanson considers the widely used term code switching inadequate, as, in his view, in code-copying no actual switching occurs. While heavy mixing on the synchronic level is excluded from the CCM, Johanson (1999: 53) argues that it is possible as a result of extensive copying and frame changes:

"The emergence of new equivalence positions leads to further morphosyntactic convergence. Increasing similarity between the two codes may favour and accelerate the convergence. Successive copying processes may thus have veritable 'snowball effects'. It becomes increasingly difficult to distinguish between nativized copies and native, non-copied elements. Long series of changes produce copies which are established as part of the 'inherited' inventory and cannot easily be traced back to the original copying processes."

76

Johanson (1999) differentiates between global and selective copies, global copies involving copies of the material, semantic, combinational and frequential properties of the model language element. Global copies are usually lexical copies which cover both types of borrowings (borrowings proper and nonce borrowings, or momentary copies in his terms) and code-switches. They can also involve idioms copied as one unit, as chunks which can later result in the copying of a bound morpheme originally an unanalyzed part of the phraseological unit. Selective copies involve only one or some of the characteristics of the model language element. For example, if only the semantic feature is copied, a loan translation occurs. If only combinational patterns are copied, this can change the constituent order. It is also possible that a construction is present in the recipient language, but it is a marked combinational pattern. Due to frequent copying of the model language construction, the similar but originally marked native construction becomes less foreign. A similar tendency can be observed in the genitive construction in the Erzya–Russian bilingual varieties (cf. section 7.2.2.2 for details).

Johanson's 1999 model is applicable for the description of contact situations in which a variety of forms coexist, the dominant language exerts both lexical and structural influence on the minority languages, and the differentiation between borrowings and code-switches is not possible.

5.3.3.2. Backus's usage-based model

Backus (2015) creates an integrated model of different contact phenomena (borrowing, CS, loan translations, and structural transfer) often occurring simultaneously, and takes into consideration both the synchronic and the diachronic aspects of language change. His description of language contact situations in general aptly characterize Erzya–Russian bilingual discourse as well:

"Bilingual speech, especially of the informal in-group type in communities exhibiting intense language contact and high levels of bilingualism, tends to include instances of

insertional and alternational codeswitching, loan translation and grammatical interference side-by-side, and sometimes interlocking" (2015: 39).

Backus argues that the problems of categorization faced by CS researchers could be solved if studies applied an interdisciplinary approach. According to him, CS research has reached a point at which there are enough empirical data on CS patterns from different contact situations. The next step would require interdisciplinary studies in which results and methods of the various approaches would be combined. He claims that there are three main dichotomies of CS research that set back the development of this field.

One of the main problems is the separation of the diachronic and the synchronic aspects of CS, which caused the failure of the attempts of earlier studies to differentiate between CS and borrowing. As a second point, he emphasizes that the study of lexical and structural transfer should be integrated, because these phenomena often co-occur. The third distinction, characteristic of earlier CS research is between form and function, the structural studies of CS focusing on form versus the sociolinguistic and pragmatic studies focusing on function. Backus (2015) claims that with the unification of these approaches in his usage-based model, a complex framework of CS patterns can be created. His model relies on Croft's 2000 ideas and adopts a holistic view of CS involving also its sociolinguistic and cognitive aspects of CS (Stell 2015: 18).

Backus (2015) argues that researchers should take into consideration the frequency of given code-switched forms and the acceptability judgments made by the speakers, because these are better diagnostic tools for CS than the analysis of purely structural features of the languages. Backus differentiates between normal replication and altered replication in connection with CS (and also other contact phenomena). Normal replication involves the use of a linguistic pattern as applied before, while altered replication introduces a new pattern. In this way, many patterns co-exist at the same time and their entrenchment varies. If a new pattern is used many times, it will become even more entrenched. Whether the variation of patterns is lost through the maintenance of the "normal", older pattern or through the establishment of the new pattern as the new norm depends on extralinguistic factors. If the altered replication, the once new CS constructions, becomes the norm, it can lead to the fossilization of a mixed language.

Backus's usage-based model does not differentiate between borrowings and code-switches, all one-word insertions can be considered code-switches, and even established borrowings can function as code-switches in his model. Elements differ only as regards their entrenchment. A higher degree of entrenchment entails that the sociopragmatic indexicality of a given element will diminish, and it loses its pragmatic function due to frequent use. Backus also argues that there might be a correlation between "lack of morphosyntactic integration and low degree of entrenchment and/or conventionalization" (2015: 38). Leinonen (2009) also found this correlation in her Komi data, whereby Komi speakers considered morphologically integrated switches to be loanwords, even in idiosyncratic cases, and insertions with Russian morphological markers to be code-switches (cf. section 4.1.2).

According to Backus (2015: 42), there are two mechanisms of change: conscious selection and interference, the former being typical in case of lexical borrowings (the more entrenched they are, the more likely they are to be selected automatically), and the latter being more important in cases of structural borrowing with a more general meaning.

As mentioned above, the entrenchment of an element can vary from speaker to speaker and change over time. The usage-based model uses two methods to establish the level of entrenchment of a given element or construction: "Usage-based linguistics tends to base assessments of entrenchment on two types of measurement: corpus frequency of the words or constructions in question, and response data from participants in linguistic experiments, such as lexical decision tasks or acceptability judgments" (Backus 2015: 38). I argue that this type of studies would be required as a next step in the Erzya–Russian CS research.

5.3.4. Sebba's model

Issues of congruence and equivalence that are addressed also in the models presented above are central concepts in Sebba's 2009 categorization. (These concepts play an important role in the structural analysis of Erzya–Russian bilingual discourse, discussed in detail in Chapter 7.) Sebba's model (2009) provides a clearcut categorization of the structural reasons behind the different types of CS. I rely on Deuchar's (2005: 256) definition, according to which congruence "refers to

a notion of equivalence between the grammatical categories or word classes of different languages".

There are many predecessors to Sebba's model. Poplack (1988) suggest, for instance, that the typological distance or closeness of the participating languages can determine whether borrowing or CS is preferred in a given contact situation. Romaine (1995: 144) claims that "[s]imilar typologies will be conducive to equivalence-based code-switching, while conflicting typologies are more likely to result in nonce borrowing and/or constituent insertion".

Sebba (2009) distinguishes between three strategies: harmonization, neutralization, and compromise. If there is congruence between the constructions in the two languages, different strategies of harmonization are applied, and switching can occur within the construction. For example, if both languages place the attribute before the head, the switch can take place, since the head-attribute order does not violate the rules of either language. This latter idea is not new to CS research: as mentioned above, it was also part of Poplack's 1988 equivalence model.

If there is no congruence, there are two possible strategies: neutralization and compromise. If speakers insert EL elements into the ML as EL islands, thus avoiding the necessity to adapt the EL element to the ML, the strategy is labeled neutralization. The other option is the use of a compromise strategy which violates the rules of both languages. A typical compromise strategy is the use of bare forms or a hybrid construction, e.g. using the former example, when speakers keep the head-attribute order of one of the languages but fill the construction with lexical elements from the other language. In the case of compromise strategies, the switch is often flagged or preceded by pause, hesitation, fillers or an insertion of dummy words (Muysken 2000: 106), etc.

Sebba (2009) poses an intriguing question: why do speakers consider structures between two languages equivalent? Is it because universal categories do exist?; is the equivalence constructed by stable bilingual intellectuals?; or is the harmonized system conventional and acquired during the process of language acquisition? In the present dissertation, I consider congruence connected to the equivalence of structural features discussed in Chapter 3. This typological overview focuses on incongruent constructions of the two languages, for instance, the possessive construction and adpositional phrases. We need, however, further cognitive research

to analyze this aspect of CS, in the Erzya–Russian contact situation, the attested smooth switches could be used as a basis for planning this study (cf. Chapter 7).

5.3.5. Gafaranga's model

In this section, I discuss Gafaranga's (2000) model, which is profoundly different from the structural frameworks presented above. As Gardner-Chloros (2009: 112) argues, "[t]he difficulty with most models of CS grammar […] is that they seek to describe the data in terms of the interaction of discrete systems". The pragmatic approach applied by Gafaranga departs from the traditional monolingual view of bilingual language use. As opposed to the binary thinking common in structural models (e.g. in Myers-Scotton's MLF model or Johanson's CCM), Gafaranga does not consider CS as an interplay between two languages. In his framework, the medium of communication can be also bilingual, and CS occurs only if the speakers negotiate about this medium and consciously mark the switch points.

If we consider these switches from a structural point of view, they can be regarded flagged switches as opposed to smooth ones. In these cases, the switch is flagged with a non-verbal marker (laughter, pause, etc.) or a verbal one (a false start, hesitation marker, discourse particle, etc.). These switches are pragmatically salient. In Gafaranga's models, only these instances are defined as CS, "whereby speakers show awareness of alternating between two structurally distinguishable codes" (Stell 2015: 1). In Gafaranga's model (Gafaranga et al. 2001) language choice is a social action, and thus, the analysis focuses on the way speakers react to their own and their partner's language choice acts.

The norm of communication is defined by observing the deviations from the medium. According to Gafaranga et al. (2001: 210), "once a medium of an interaction has been adopted, any departure from it is repaired unless it is meant to be functional". Gafaranga's model (2000) focuses on these "medium repairs". Repairs can be self-repairs or other-repairs, the former being more common. Medium repairs are always initiated by the speaker, and they usually occur turn-internally. If the other speaker provides the missing expression, the initiator of the medium repair has to acknowledge it before the conversation can resume. "In medium repair, the speaker draws

on other languages in his/her repertoire and signals the other-languageness of the element used"
(Gafaranga 2000: 344). In these cases, speakers name the language they switch to, or they
indicate that this is the expression people usually use. Medium repair can involve translation, too.

Flagged code-switches occur in the Erzya–Russian bilingual data as well. In my opinion, it
is thus justified to combine different approaches in order to be able to analyze the whole
spectrum of CS patterns (cf. section 7.3).

5.4. Continuum models

The models presented above offer different ways to characterize CS patterns in a bilingual
discourse. It is rather typical in contact situations that many different CS types co-occur. In order
to be able to analyze them, a complex model is needed. Continuum models are especially suitable
for the depiction of this type of variation in the synchronic state, but they also allow for the
representation of diachronic processes.

In the present paper, I refer to two continuum models, Auer's (1999) and Kovács's (2001).
Both models take into account the CS patterns occurring in the bilingual discourse. Auer
differentiates between three phases: code-switching (CS) → language mixing (LM) → fused lects
(FL), which he regards as a type of grammaticalization process. In section 5.4.1, I analyze
characteristics of these phases in detail.

In Kovács's 2001 model, the most important factor is agreement. The terms used in the
model follow the terminology of the MLF model (Myers-Scotton 2002). The main emphasis is on
the fact whether the code-switched construction is formed obeying the rules of the ML or of the
EL. The place of each type on the continuum is determined by the distance of the code-switched
construction from the structure required by the ML.

5.4.1. Auer's continuum

Auer (1999) describes three phases, which differ according to the extent and types of CS. All of
the phases can involve insertions and alternations as well, but in the first one, in the code-

switching (CS) phase, the switches have a pragmatic function, which is lost in the second language mixing (LM) phase due to the increase in the number of switches. This second period is characterized by variation: many forms co-exist. Finally, in the phase of the fused lects (FL), the variation is lost, and the mixed structures fossilize and become compulsory for the expression of a given grammatical function.

If we analyze Auer's continuum model using the MLF model's terminology, the first phase is characterized by classic CS: there is a dominant language which can be identified in all cases, switches have a pragmatic function, and they are used as contextualization cues, as rhetorical and stylistic devices. In the second phase, it can be unambiguous which of the languages is the ML, here classic CS is replaced by composite CS. In the last phase, we cannot talk about CS any more. Although elements can belong to two different languages etymologically, they form one code, one fused lect. In Figure 1, we can see the most important characteristics of each phase on Auer's continuum model:

code-switching	language mixing	fused lects

→

| switches have a pragmatic function | variation | no variation switched forms are obligatory |
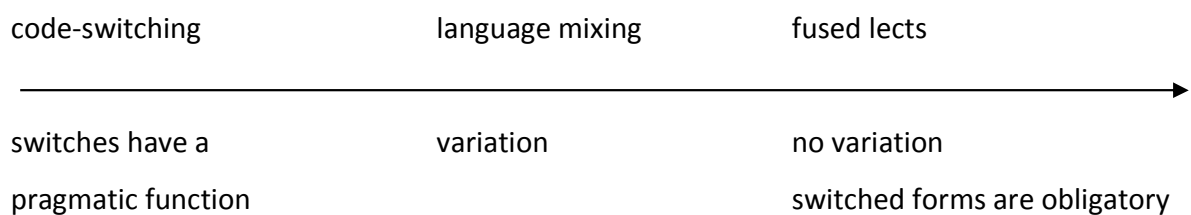
Figure 1. Auer's continuum (1999)

Auer's model is also compatible with Backus's (2015) usage-based model. The continuum depicts a diachronic change, a grammaticalization process during which the frequency and entrenchment of the elements (a key concept in Backus's model) changes: occasional switches will be fixed elements of the fused lect. Structures which are optional in the code-switching phase become compulsory, the unique option in fused lects. The LM phase is an in-between phase with variation, as the level of entrenchment varies from element to element.

The language mixing phase is more typical of relatively stable bilingual communities where the contact situation has existed for several centuries. Due to frequent switching, switches do not have a pragmatic function any more: "The more frequently codeswitching occurs, the less salient

it becomes; as a consequence, the potential for using it in locally meaningful ways is diminished" (Auer 1999: 320). This phase in a contact situation is also described by Myers-Scotton, labeled as the unmarked choice: "Each switch need have no special significance; rather it is the overall pattern of using two varieties which carries social meaning" (1998: 162).

On Auer's 1999 continuum, the last phase is the fused lect phase. A prolonged contact situation can have different outcomes: if it is not leading to a fused lect, speakers can become dominant in the majority language, and the ML changes from the original ML to the former EL. This process is labeled as Matrix Language Turnover in the MLF model (Myers-Scotton 1998). Aikhenvald also describes this process in connection with convergence:

"In the situation of one language dominating the others, convergence may involve gradual adoption of the other language's structures at the expense of its own. This kind of convergence – a typical result of displacive contact – is often the precursor of language attrition and obsolescence" (2006: 47).

Nevertheless, there are cases described in the language contact literature where an arrested turnover led to mixed languages, e.g. Gurindji Kriol (McConvell and Meakins 2005) or Ma'a (Mbugu) (Myers-Scotton 1998).

5.4.2. Kovács's distance-based model

Kovács (2001) has studied Finnish–English and Hungarian–English code-switching in Australia, and employs a model based on the Matrix Language Turnover hypothesis. The ML provides the morphosyntactic structure of the utterance, while the EL items are inserted into this grammatical frame. These insertions can be one-word switches or longer stretches of discourse.

Kovács's 2001 distance-based continuum model (Figure 2) focuses on instances of case marking in intrasentential switches.
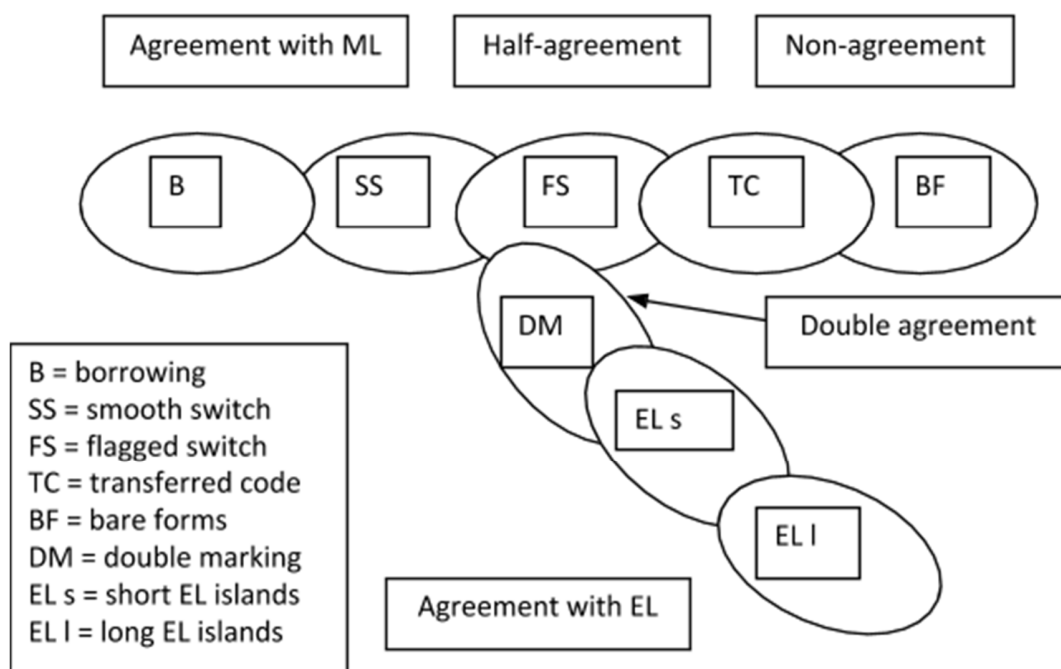
Figure 2. Kovács's (2001: 210) continuum model

At one end of the continuum, we find borrowings and smooth switches, which are both in agreement with the ML. In these cases, the structures of the two languages are congruent. Smooth switches might occur in cases where the congruence is missing but the grammatical frame is formed according to the rules of the ML (Kovács 2001: 189).

The continuum has two branches to the right, the first leading through a half-agreement phase involving flagged switches and a transferred-code phase to bare forms which show no agreement. The other branch, however, leads to agreement with the EL (first in short and then long EL islands) through a double-agreement phase characterized by double marking.

The model also adopts a diachronic approach and explains how a composite CS phase can result in the development of a fused lect or the turnover of the ML if the EL elements start to prevail in the utterances.

Kovács's 2001 model has been created to analyze CS between typologically different languages only one of which is morphologically complex. The model can be adapted to the Erzya–Russian CS discourse as well, although in this case both languages have a rich morphology.

Chapter 6. Data and methodology

In my dissertation, I analyze the CS patterns in two data sets: fieldwork interviews on the one hand, and radio interviews on the other. In section 6.1, I describe the two data sets and discuss the advantages and disadvantages of using this type of data. In section 6.2, I provide an overview of the methods used during the collection and analysis of the data. I also discuss problems connected to the transcription of bilingual utterances that I have encountered.

6.1. The two data sets

The data I analyze in my dissertation can be divided into two main data sets. The first part comprises fieldwork data, while the second part involves interviews recorded from Radio Vaygel. First, I present the characteristics of these two data sets, then I focus on the problems arising from the nature of the data.

The fieldwork corpus is versatile, it consists of data collected during field trips in 2005, 2008, 2010, and 2011. The first subset, from 2005, comprises 14 interviews (5 hours total) recorded for Rigina Turunen's project on nominal conjugation. She kindly let me use her data for my analysis of CS. The interviews were collected mostly in Saransk from speakers of different Erzya regions, including the diaspora. In this respect, these data are closer to the radio interviews in nature, in that the subjects are not from the same area.

The second subset of the fieldwork data contains 19 interviews (8 hours total) recorded at four different locations (Baevo, Atyashevo, Atrat', and Altyshevo) in two regions of the Mordvin Republic and in one diaspora area in the Chuvash Republic. The locations in the Mordvin Republic differ in respect of the status of the Erzya community: its majority or minority status in the settlement in question. Baevo (Ardatovsky District, Mordovia, Russia) is an Erzya-majority village (population: 1,628[7]) with a kindergarten and school providing Erzya-language education (which is only available in villages with an Erzya majority), and Atyashevo (Atyashevsky District, Mordovia, Russia) is a small Russian-majority administrative center (population: 6,095). Two villages in

---

[7] The source of the data is the Russian Federal State Statistics Service (FSRS 2016).

Chuvashia (Atrat', Alatyrsky District, Chuvashia, Russia, with a population of 1,265, and Altyshevo, Alatyrsky District, Chuvashia, Russia, with a population of 815) represent a unique type of Erzya speech community. These villages are outside the Mordvin Republic, and hence belong to the diaspora, but the proximity of Mordovia enables the speakers to be in contact with the Erzya communities there. As a result, they are in a better sociolinguistic situation than diaspora speakers in more remote regions like the Ulyanovsk or the Orenburg Oblast.

The corpus contains recordings of semi-structured interviews and group sessions. (Interviews coded as 201102 and 201103 are group sessions (1 hours 36 minutes total) in which more than one speaker takes part in the conversation with the interviewer.) However, in most of the cases (except for interviews 200802, 200806, and 201015) there is a friend of mine (fieldworker 1 in Chuvashia, fieldworker 2 in Baevo, and fieldworker 3 in Atyashevo, abbreviated as F1, F2, and F3), a native speaker of Erzya who introduced me to the local community, who is also present at the interviews and occasionally comments on the interviewee's narration or asks a question.

The other data set comprises 160 interviews (of approximately 26 hours) from Radio Vaygel broadcast between April 2013 and January 2016. The length of the individual interviews ranges between 2 and 31 minutes. They were conducted by four different reporters, 118 interviews by R1, 19 by R2, 14 by R3, 9 by R4. The trilingual (Erzya, Moksha, and Russian) radio station established in 2007 is based in Saransk, but for a limited time period (approximately for a month after airing) its broadcasts are available also online on the webpage of the radio (Radio Vaygel 2016). The radio station broadcasts three hours every weekday from 16:00 to 19:00 local time. In the first hour, the station plays mostly music, the second hour is dedicated to the Erzya, and the third is to the Moksha program. The programs include not only interviews but reports, news and music as well. I only analyze the interviews as they provide more diverse data on the language use of speakers from all walks of life than any other parts of the broadcasts. However, the (semi)-monolingual language use of the reporters and their possible puristic attitude would also be worth studying as part of a future research project. (The issue is discussed to some extent in Janurik 2016.)

The CS patterns of the two data types show similarities in many respects: for instance, all the interviews are conducted in Erzya by a native speaker, or, in the case of the fieldwork data, most of them at least in the presence of another native speaker of Erzya. The radio interviews, however, provide a more diverse material as compared to fieldtrip interviews, because in the former case speakers represent both genders, various ages and social groups, while fieldwork data are almost exclusively recorded from female speakers living in four locations and typically working in agriculture, education, or the service sector. At the same time, information on the sociolinguistic background of the speakers in the radio interviews is rather limited. Usually only the name and the occupation of the interviewee are provided when reporters introduce the guests. What also presents a problem is that some of the interviews appear edited. It is possible that hesitations and self-repair have been omitted from the recordings. These phenomena would also need further investigation, but that clearly goes beyond the scope of my paper.

6.2. Methods of data collection and analysis

In this section, I briefly discuss the methods of data collection focusing mainly on the methods used during the analysis of the CS patterns.

The sample of subjects used in my investigation was not representative: I used the snowball or friend-of-a-friend method. The radio data are also not representative, interviewees were chosen on the basis of their expertise in a given topic. As a result, the data cannot be analyzed quantitatively. Another obstacle to a quantitative analysis is the fact that almost all of the subjects were female speakers. In the 2005 data, all interviewees are women between the ages of 17 and 55 years at the time. In the second subset, there are two male speakers and the ages of the speakers range between 35 and 81 years. In the case of the radio interviews the lack of available data on the sociolinguistic background of the speakers prevents a representative quantitative analysis (of the 160 interviews 61 were conducted with male speakers).

Although I agree with Backus (2015) that specific data collection methods (frequency analysis of big data and judgement tests with native speakers) are necessary for a complex assessment of CS, the amount and nature of my data precludes the use of these methods. I argue,

however, that on the basis of my results, it will be possible to design a quantitative study and test the level of acceptance of given lexemes in the future.

For the analysis of the Erzya–Russian CS patterns, I use Muysken's 2000 categories, and I differentiate between insertions, alternations, and congruent lexicalization. In case of one-word Russian elements, the differentiation of CS and borrowing proves problematic, because the criteria used in order to distinguish borrowing from CS (phonological and morphological adaptation, frequency of usage, acceptance by the community, etc.) are inadequate for the analysis of the Erzya–Russian data. Phonological adaptation is not a reliable criterion, as the phonological systems of the two languages are similar (cf. discussion in section 3.1 above), partly due to a prolonged convergence process between the two languages. This is not a unique case in contact situations, e.g. Goss and Salmons (1990), studying the language use of English–German bilinguals in Texas, also argue that, as a result of convergence, the criterion of phonological adaptation cannot be used for differentiation between borrowings and code-switches. According to the census data (Russian census 2010), almost all of the Erzya speakers know Russian as well, consequently apart from some possible marginal exceptions, there are no monolingual Erzya speakers any more. As a result, we cannot decide on the borrowing–CS status of a given lexeme on the basis of their language use (otherwise if a lexeme occurs in monolinguals' utterances, it can be considered an established loanword). Finally, the frequency of an element cannot be studied on the basis of a small corpus, as the occurrence of a lexeme can depend on various factors, e.g. the topic of the conversation. The entrenchment of a given lexeme (i.e. how integrated it is) cannot be determined on the basis of spontaneous recorded data. The study of the dictionaries should be also involved as a further step in the analysis of the entrenchment of Russian origin elements in Erzya.

Consequently, I do not differentiate between these two types of contact phenomena, and in the examples I mark all Russian origin elements with bold face, irrespective of their supposed status as an established borrowing or an idiosyncratic switch. My main focus is on longer Russian sequences that are either morphologically integrated into the Erzya frame or inserted into it using Russian morphological markers.

My procedure for carrying out the analysis has been as follows. As the first step, I collected all Russian origin elements from the discourse. Next, I carried out a structural analysis of these constructions and categorized them as insertions, alternations, or instances of congruent lexicalization. I also referred to other models describing CS (mainly the MLF model), but the analysis of the data was carried out from a descriptive point of view. I argue that the avoidance of any theoretical frameworks in the analysis provides more possibilities for a future typological study of CS patterns in the Finno-Ugric languages.

In the second phase of my analysis, I focused on the incongruent constructions in the two languages. These are typically instances of insertions and congruent lexicalization in Muysken's 2000 terminology, and mixed islands and EL islands in the MLF model.

In the third phase of the analysis, I assigned speakers to categories on the basis of the CS patterns occurring in their speech. Then as the next step, I placed them on the Erzya-Russian continuum (cf. a preliminary form of the model in Figure 3), which combines Auer's (1999) and Kovács's (2001) models discussed in section 5.4.

In this continuum model, the first phase involves smooth switches (flagging only occurs with congruent constructions), usually one-word insertions embedded into the ML morphosyntactic frame with Erzya morphological markers. EL islands also occur occasionally, longer insertions and alternations have a pragmatic function. Speakers producing these types of switches are labelled as Category 1 (C1) speakers. The next two types of speakers belong to the LM phase on Auer's (1999) continuum. Their speech involves an increased number of EL islands. The utterances of the Category 2 (C2) speakers still have Erzya as their ML, while in Category 3 (C3), utterances can have a composite ML, and the CS type congruent lexicalization is present in their speech as well. Neither the FL phase of Auer's continuum, nor ML turnover are marked on the continuum, as my data do not contain mixed constructions that are compulsory throughout the speech community and no interviews have Russian as the ML in general. Figure 3 represents the Erzya–Russian continuum model (discussed in detail in Chapter 8):

| | | |
| Phase 1 | Phase 2 | Phase 3 |

Erzya is the ML | Erzya is the ML | composite ML
short EL islands | longer EL islands | mixed constructions
one-word insertions | | the ML cannot be determined
| | unambiguously
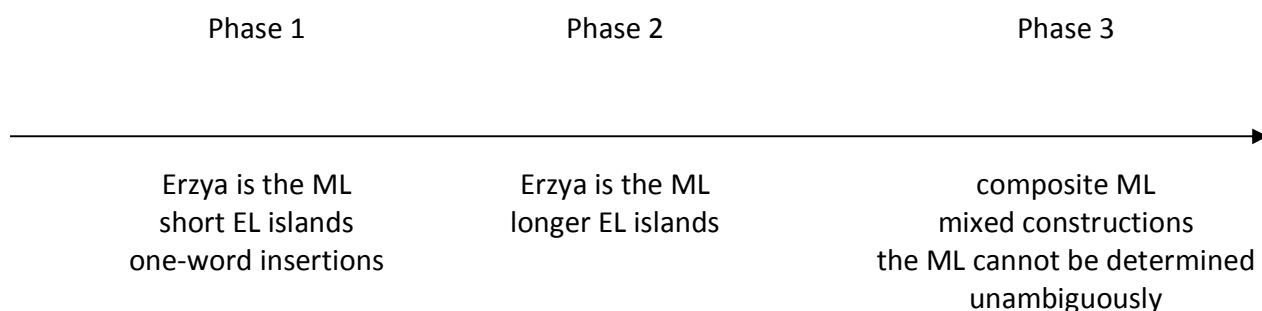
Figure 3. Continuum model for the Erzya–Russian bilingual discourse

Finally, I need to discuss some problems of transcription that I have encountered. Working with a bilingual corpus, one faces the dilemma of deciding between transcriptions. The main question is whether one should use more transcriptions or just one for both languages. As I hope for an international audience of readers of my dissertation, a logical choice would have been the International Phonetic Alphabet (IPA). However, as I carry out a morphosyntactic analysis of the data, the phonetic transcription would be too detailed. If I had opted for transliteration of the Russian data with the ISO9 standard and the use the Finno-Ugric Transcription (FUT) for Erzya, I would have had to decide on the (CS or borrowings) status of each word. Moreover, transcribing Russian words differently would be misleading in the case of bridge words in mixed constructions, as they can belong to both languages. Another problem this binary transcription would present is that there are words the forms of which do not differ in the ISO9 transliteration and FUT which would not show that the given element is of Russian origin, e.g. *kak* 'as', or *ladno* 'all right', to mention a few. As a result, I use the FUT for both languages and mark the Russian origin elements with boldface, without referring to their degree of possible entrenchment.

Chapter 7. The Erzya–Russian code-switching discourse

In this chapter, I discuss the CS types occurring in my data. I use Muysken's (2000) typology to categorize the examples, and consequently I differentiate between instances of alternation, insertion, and congruent lexicalization (for the definition of these phenomena cf. section 5.3.2 above). However, I take into consideration other models as well (mainly the MLF framework discussed in 5.3.1), to see how binary models insisting on the dominance of one of the participating languages analyze these mixed constructions.

First, I focus on alternation. Alternation involves different mechanisms of switching as compared to insertion and congruent lexicalization, which is why I discuss it in a separate subsection (7.1). Insertions and congruent lexicalizations are more related categories, partly involving the same mechanisms, but while in the case of insertions the morphosyntactic frame of the utterance is set by one language, in congruent lexicalization both languages are involved in the construction of the grammatical structure of the utterance. Alternation usually arises for two, occasionally also interconnected reasons: the complete switch can have a pragmatic function (e.g. it is a citation) (cf. section 7.1.1), or it can be used to avoid a point where the structures of the two languages clash (7.1.2). There are instances in which the alternational type of switching can be explained by both reasons. In the first case, we typically find that a longer stretch or a whole clause is switched, while the second type of alternation can have a scope of only one construction. Alternation at incongruent switch points is also usually accompanied by flagging, discussed partly separately in section 7.3.2.

There are borderline cases in which the differentiation between alternation and insertion might prove problematic, especially in the case of discourse particles. However, the position, i.e. the embeddedness, of the switched element can help the investigator to categorize the switches. In section 7.1.3, I focus on some of these ambiguous cases occurring in my data.

Section 7.2 is the most substantial part of my dissertation. It analyzes instances of insertions and congruent lexicalization together, as we can find a continuum ranging from occasional insertions to hybrid realizations of the given grammatical constructions in different speakers' language use. The main difference between these patterns is the fact that in case of

insertion the foreign elements are inserted into the ML, while in the case of congruent lexicalization, words are inserted into a single syntactic structure formed by the two languages together (Muysken 2000: 31).

When combining Muysken's (2000: 32) typology with the MLF model, insertions can be divided into two subtypes: mixed constituents and EL islands. My hypothesis is that if the constructions in the two languages are congruent, harmonization occurs (cf. Sebba's 2009 model in section 5.3.4), a switch can happen within the construction, and, as a result, a mixed constituent is formed. Consider example 20 (speaker 20130408, the discussed part is underlined):

(20) *son* **_očeń_** **_paŕśťe_**   *tonavtńe-ś*

s/he very   well   study-PST.3SG

'S/he studied well.'

In example 20, the two structures are congruent, both languages have the same constituent modifier–head order, so 'very well' in Erzya would be *pek paŕśťe*, while in Russian *očeń horošo*. These two elements are combined: the modifier is in Russian and the head in Erzya, yielding *očeń paŕśťe* 'very well'.

In some of the cases, both adverbs occur in the same utterance. In example 21 (speaker 20130930), both the Russian adverb *očeń* 'very' and its Erzya equivalent *pek* 'very' are applied as modifiers of the adjective *vadŕa* 'good'. Double forms and repetitions can be used, for the sake of emphasis, if the structures are congruent but also when they are incongruent, and the speaker avoids the lack of equivalence by juxtaposing the two structures.

(21)   **_očeń_**   *pek*   *vadŕa*   *ťe*   *uľńe-ś*   **_lager_**-*eś*

very   very   good   that   be-PST.3SG   camp-DEF.SG

'The camp was very very good.'

If the structures are not equivalent, the incongruence is neutralized by the insertion of an EL island. In example 22 (speaker 200811), there is a Russian prepositional phrase *na russkom* 'in Russian' inserted into the Erzya ML to avoid the incongruence between the constituent order of the Erzya form *ruzoks* 'in Russian' with the translative suffix *-oks* and the Russian prepositional phrase.

(22)   *veśe*   **_na_ _russk-om_**   *uľńe-ś*

everything   on   Russian-PREP.SG   be-PST.3SG

'Everything was in Russian.'

The same tendency (insertion of a Russian EL island) can be observed also when a postpositional construction is the Erzya equivalent of a Russian prepositional phrase. In example 23 (speaker 201001), the Russian *posľe vojni* construction is inserted as an EL island because the constituent order of this prepositional phrase and the Erzya equivalent *vojnado mejľe* 'after the war' with the postposition *mejľe* are incongruent.

(23) *miń s-ińek*   ***posľe vojn-i***   *nu*

     we   arrive-PST.1PL   after   war-GEN.SG   well

     i   *vot*   *miń*   *eŕ-ińek*   *toso*   ***šestnadcať ľet***

     and   look   we   live-PST.1PL   there   sixteen   year.GEN.PL

     'We arrived after the war, well, look, and we lived there sixteen years.'

There are cases in which insertion of an EL island is flagged, the speakers start to say the construction in one of the languages, but when they cannot combine the two, they switch in the middle of the construction. If they repeat the construction in the other language, it can be analyzed as an instance of EL island type insertion or alternation (cf. section 7.1.3).

(24) *s*   *vosm-i*   *ľet*   *škola-v*   *jak-iľť*

     from   eight-INS.SG   year.GEN.PL   school-LAT   go-PST2.3PL

     mińek   ***v***   ***naš-ih***   ***god-ah***

     we.GEN   in   our-PREP.PL   year-PREP.PL

     'In our time, people went to school at the age of eight.'

In example 24 (speaker 201001), the Russian prepositional construction *v naših godah* 'in our time, lit. in our years' is incongruent with its Erzya equivalent *mińek škastoňť* 'in our time', the difference in the order of the morphological markers (the Russian preposition in the beginning of the construction, the Erzya suffix at the end) prevents the switching.

If the construction is a combination of the two languages, the CS type can be labelled as congruent lexicalization, and in Sebba's typology this strategy is categorized as a compromise. If the compromise strategy is used (Sebba 2009), the emerging construction violates the rules of both languages, e.g. in example 25 (speaker 200803):

(25) **no** mińek **na** mińek pora-s(o) ška-ś **końečno**
but we.GEN on we.GEN time-INE time-DEF.SG of.course

koda šačoma-do mejľe ška-ś uľń-eś **trudnoj**
when birth-ABL after time-DEF.SG be-PST.3SG hard

'But in our time, times were, of course, as after birth, times were hard.'

In the phrase na mińek poras(o) 'in our time' has both the Russian preposition and the Erzya inessive ending which have the same function. This construction does not abide by the rules of either language, and both the lexical material and elements of the structural frame come from both languages, thus congruent lexicalization arises.

As I mentioned above, insertions and congruent lexicalizations have common features, and both congruence and incongruence can result in the emergence of both insertion and congruent lexicalization. In section 7.2, I analyze different types of constructions that can occur both as simple insertions or in a hybrid construction, as congruent lexicalizations. The only exceptions are discourse particles (discussed in section 7.2.1) that are instances either of insertions or of alternations. The other two major sections 7.2.2 and 7.2.3 focus on nominal constructions and verbal constructions, respectively. Although both constructions have many subtypes relevant in CS research, I only analyze particular constructions: numeral phrases and nominal phrases with a genitive modifier in section 7.2.2, necessive constructions and the question of gender agreement in 7.2.3.

Flagging is relevant in both in congruent and incongruent constructions. In section 7.3.1, I focus on equivalent constructions in which flagging has a pragmatic function motivated by sociolinguistic factors, and in section 7.3.2 I discuss incongruent constructions.

7.1. Alternation

In this section, I discuss instances of alternation attested in my data. These are CS types in which a complete switch of languages occurs that concern both the content and the function words. Alternations typically involve longer Russian stretches, but occasionally also single phrases, and

sometimes even one-word switches can be categorized as alternations. In such cases, switches are usually in a peripheral position in the clause.

The other distinction within the alternational type of CS concerns the existence or lack of a pragmatic function the switched sequence serves. First, I discuss alternations with a pragmatic function in section 7.1.1. These cases typically involve a switch at clause boundaries, and thus are instances of interclausal switches. Second, I focus on alternations without a pragmatic function in section 7.1.2. In this case, alternation entails the switch of a longer sequence within the clause. Such switches can be triggered by a code-switched Russian adverb or a discourse particle. There are also borderline cases between insertion and alternation which make the categorization of these CS types difficult. I discuss such problematic instances in section 7.1.3.

7.1.1. Alternation with a pragmatic function

Alternational switching is generally associated with the expression of pragmatic functions. For instance, Backus (2015: 42) argues that in the case of alternations, "speakers are said to switch for pragmatic reasons or to index their message with the values associated with the language switched to (such as modernity or local solidarity)." The main common pragmatic functions expressed by alternational CS "include drawing or directing attention, emphasizing or mitigating a message, expressing anger or emotion, etc." (Backus 2015: 42).

In my data, alternation occurred in the utterances of all speaker types. Its most common function was quotation. In example 26, a C1 speaker (speaker 200507), a mother is talking about her son with whom she speaks Russian, and the language use of the child is also exclusively characterized by Russian. Quoting the question of the boy triggers alternation, the Russian utterance appears in its original form as direct speech in the otherwise nearly monolingual Erzya discourse:

(26) *śeďe   śejeďste   karm-i   ńerńe-me*
     more   frequently   start-3SG   whine-INF2
     kogda mama   pŕiď-ot
     when   mother   come-3SG

iľi ješčo *meźejak*

or as.well something

'He starts to whine more frequently: "When is mom going to arrive?" Or something

else as well.'

The quotation is embedded into an Erzya sequence, the preceding clause is clearly monolingual Erzya, while the following one contains two elements, the conjunction *iľi* 'or' and the adverb *ješčo* 'still' which can belong to either language. Thus, they make the switch back to Erzya easier. Apart from this Russian clause, the speaker's language use is characterized by only insertions of Russian discourse particles and nominal stems embedded into the Erzya ML with Erzya morphological markers.

Most alternations occur in the utterances of C3 speakers. These alternations usually do not have a pragmatic function (discussed in section 7.1.2), but there are examples of alternational switches which serve a conversational function. These can be instances of quotations, as in example 28, but they have other functions as well, for example, an alternation can mark a change in topic. In example 27, the pragmatic function of the switch is topic change. A heavy-mixing C3 speaker (speaker number 200803) is using the construction: *vot takije ṕirogi* 'that's how it is, lit. these kinds of pastry' to end the story he has been telling, to signal that this topic has been discussed enough.

(27) *ťe-ste*   ***dvadcať ṕať***   ***kiloḿetr-ov***   *avoľ*   *vasolo*

   this-ELA.SG   twenty five   kilometer-GEN.PL   not   far

   no   *mon*   ***ježeńeďeľno***   *sinst*   ***nav'ešča-ju***

   but   I   every.week   they.GEN   visit-1SG

   osob'enno   *ťena-m*   *jalaks-om*

   especially   whatsit-POSS.1SG   little.brother-POSS.1SG

   <u>***vot***</u>   <u>***tak-ije***</u>   <u>***ṕirog-i***</u>

   well   like.this-PL   pastry-PL

   'It is not far from here, 25 kilometers away, but I visit them every week,

   especially my brother. That's how it is.'

97

Before this complete switch, the utterances contain elements of both languages, which is typical of the speech of C3 speakers. In the first clause, the ML is clearly Erzya into which a Russian numeral construction *dvadcať pať kiloḿetrov* 'twenty-five kilometers' is inserted. In the second utterance, the finite verbal predicate *nav'eščaju* 'I visit' is in Russian but the pronominal subject (*mon* 'I') and the object *sinst* 'them' are expressed with Erzya elements. The pronominal subject and object presumably would not be overt if the predicate was an Erzya verb, because the definite conjugation form could refer to the person and the number of the subject and the object. As a result, the ML in this utterance is ambiguous, and the finite predicate is Russian, but other core elements (the object and the subject) are in Erzya. The next clause belongs to the second utterance, containing a Russian adverb *osob'enno* 'especially', but the other two elements, which form the object construction, have both an Erzya dummy-word (Muysken 2000: 106) *ťeńa* 'whatsit' and an Erzya noun *jalaks* 'little brother' with an Erzya first person possessive singular marker.

Citations occur also in the utterances of C3 speakers. In example 28 (speaker 200511), the quotation is preceded by the first person singular form of the verb *meŕems* 'to say'.

(28) *mon* *sońenze* *meŕ-iń* ***ńe živ-ut* *ńe raďi***

     I s/he.DAT say-PST.1SG not live-3PL not for.the.sake

     ľubv-i* *a raďi* *ďeť-ej

     love-GEN.SG but for.the.sake children-GEN.PL

     'I told him/her, we do not live for the sake of love but for the sake of children.'

Citations are, however, not always clearcut cases of alternation. Because of the constant switching back and forth, switches are not complete in some cases, and the quotes are bilingual. There can be at least two reasons for this: the original conversation was bilingual, or the speaker does not cite the other speaker's words exactly.

Speaker 200511 produces many other examples of quotations in her narrative. Except for example 28, all of them are Erzya–Russian bilingual or Erzya monolingual utterances, even the citations from her Azeri husband, presumably with no knowledge of Erzya. All the following three bilingual quotations (both examples 29 and 30) are recorded from the speaker who provided example 28, from conversations with her son (the quotation is underlined).

(29) *son* *mońeń* *meźejak* *eź* *jovta-k*
s/he I.DAT anything not.PST.3SG tell-CONNEG

ćora-m **toľko** *sa-ś* **i** *meŕ-i*
son-POSS.3SG only arrive-PST.3SG and say-3SG

mońeń *eŕav-i* **vjeh-ať** **v** **Kazań**
I.DAT have.to-3SG travel-INF to Kazan

mon meŕ-an *meźe* **sluči-l-o-ś**
I say-1SG what happen-PST-N-REFL

sa-n *mejľe* **ob'esń-u**
come-1SG then explain-1SG

'He, my son, did not say anything to me, he just came and told me that "I have to go to Kazan". I told him: "What happened?" "I will come back and explain."'

(30) i *meŕ-ś* **mama**-*j* *mon* *toso* **ńe** **viďerži-l**
and say-PST.3SG mother-VOC[8] I there not bear-PST.M

'And said, "mother, I could not bear that place".'

These quotations are all examples of alternation, but the switch does not occur at the beginning of the citation, but inside the cited utterance, typically the second part of the utterance (the verbal predicate or the argument of the predicate) is in Russian. More examples would be needed to study whether this is a general alternation type. The following example from the same speaker seems to suggest that there is not only one pattern for quotations, and the original language they were uttered in does not matter. These citations are all instances of direct speech, and even the presence of the Russian complementizer *čto* 'that' does not trigger the use of reported speech, cf. example 31.

(31) *vejke* **aźerbajžańec** *sońenze* *mer-ś* **čto**
one Azeri s/he.DAT say-PST.3SG that

tońť *ťeťa-t* *Kazan-so*
you.GEN father-POSS.2SG Kazan-INE

---

[8] The status of vocative is debated: it has a restricted paradigm in Erzya and is only used with terms designating family members or other relatives.

'One Azeri told him: "Your father is in Kazan".'

It is not only longer sequences that can be categorized as alternations. In my data, one-word switches with a pragmatic function also occur. In example 32, the discourse particle *vśo* 'that's all' indicates a change in the topic, the speaker (200510) introduces her family, her siblings, and when she has discussed the last one, she moves on and indicates that the turn has ended, similarly to *vot takije ṕirogi* 'that's how it is, lit. these kinds of pastry' in example 27.

(32) *paťa-m*   *eŕ-i*   *Kaluga-so*   **robot**-*i*   **učiťelńića**-*ks*   **<u>vśo</u>**

sister-POSS.1SG live-3SG   Kaluga-INE   work-3SG teacher.F-TRA that's.all

'My sister lives in Kaluga, she works as a teacher. That's all.'

In the present paper, I apply a structural approach in the analysis of the data, and focus on incongruent constructions. Therefore, I do not discuss further examples of alternation with a pragmatic function but analyze intrasentential alternation which is connected to structural incongruence between the two languages.

7.1.2. Alternation without a pragmatic function

Due to the frequent switching, alternations in the utterances of heavy-mixing C3 type speakers (discussed above in 7.1.1) do not necessarily have a pragmatic function in my Erzya-Russian CS data. Let's consider the following example from speaker 200813:

(33) *śeste ška-ś*   *veľť* **trudnoj** *uľńe-ś*   *miń lamo-ľińek*

then   time-DEF.SG very   hard   be-PST.3SG   we many-PST2.1PL

ṕať **čelovek**   *uľńe-ś*   *ejste-ďe-ńek*

five   person.GEN.PL be-PST.3SG   from-ABL-POSS.1PL

mon **budto ṕatoj**   **četiŕe**   **śestr-i**   **brat**

I   as.if   fifth.M four   sister-GEN.SG   brother

<u>oťec   uḿer   v   vojn-u   pogib</u>

father   die.PST.M   in war-SG.ACC die.PST.M

<u>sorok   vtor-om   god-u   pod   Moskv-oj</u>

forty   two-INSTR.SG   year-GEN.SG   under Moscow-INSTR.SG

nu	*i*	*ška-ś*	*uľńe-ś*	*veľť*	**_trudnoj_**
well	and	time-DEF.SG	be-PST.3SG	very	hard

'At that time, times were hard, there were a lot of us. There were five of us (children), I was the fifth, four sisters and a brother. Father died in the war, at 42, near Moscow. Well, the times were really hard.'

In example 33, the switch is presumably triggered by the Russian particle *budto* 'as if', but it is not a complete switch, as the female speaker uses the masculine form of the adjective (*ṕatoj* 'fifth') instead of the feminine form (*ṕataja* 'fifth') required by the rules of the Russian language. Thus, it is possible that the numeral phrase *četire śestri* 'four sisters' triggers the complete switch to Russian.

Alternational code-switching has a unique role in the transfer of foreign elements into the receiving language. Although it involves a complete switch which enables the speaker to avoid possible incongruences between the two languages, the frequent use of Russian elements can change their degree of entrenchment (cf. Backus's model in section 5.3.3.2). Backus (2015) argues that alternation is also responsible for the entrenchment of the discourse structure allowing for further alterantional switsching (Backus 2015: 39).

Alternation is thus also connected to congruence. If a construction is incongruent, i.e. the structure of a given phrase is different in the two languages, it can result in a double construction which is a typical case of alternation. For instance, both the preposition from Russian and the suffix required by Erzya are present in the construction, or the whole phrase is repeated, as in example 34. The Russian prepositional phrase *do dvuh ľet* and the Erzya construction with the illative suffix are both used in a double construction by the speaker (201101). Russian numeral phrases are typical switch types in spoken Erzya, resulting partly from the fact that mathematics is taught in Russian in schools.

(34)	*pek*	*višińe-ť*	**_do_**	**_dv-uh_**	**_ľet_**	*kavto*	*ije-s*
	very	small-PL	until	two-GEN.PL	year.GEN.PL	two	year-ILL

'Very small ones, under the age of two, under the age of two.'

We can, of course, explain these examples as having the pragmatic function of making sure that the listener understood what was being said, but the structural aspect, the incongruence

of the forms can also be taken into consideration. Especially in the cases involving only partial repetition or backtracking, the explanation of structural incongruence seems more plausible:

(35) **nu** pek vasolo   Nuja **nav'erno kilo** kavto-ška **iľi** kolmo-ška   **kiloḿetra**-t
       well very far.away Nuja perhaps   kilo two-COMP or three-COMP kilometer-PL
       'Well, very far, the Nuja river, perhaps kilo…, two or three kilometers.'

In example 35, the alternation is triggered by the Russian adverb *nav'erno* 'perhaps'. The Russian numeral construction expressing approximation *kiloḿetra dva iľi tŕi* 'two or three kilometers' is started, but partly because of the incongruence of the constituent order, the speaker (200507) switches in the middle of the construction.

There are cases, however, in which partial repetition is not a result of incongruence, but the speakers monitor their speech and finally opt for the Erzya equivalent. Example 36 is from a C1 speaker (20151118) whose utterances are Erzya-dominant with a limited number of switches. The Erzya and the Russian constructions are congruent, consequently, the repetition should be analyzed as a self-repair.

(36) ńiľe **učastńik**-ť   uľńe-śť   di vejke **kak koda**   kunsolića
       four participant-PL be-PST.3PL and one as as   listener
       jaka-ś   tov
       go-PST.3SG   to.there
       'Four participants were (there) and one went there as as a listener.'

The switch type in example 36 can also be related to purism, since the speaker avoids Russian elements in general; if the switch occurs, it has to be repaired. I discuss the question of purism in more detail below, in Chapter 8, where I argue that more studies should focus on the analysis of the puristic tendencies in the Erzya speech community in general, and among intellectuals in particular.

There is variation as regards the triggering of alternation, because the same Russian code-switched element triggers switching sometimes but not other times. In example 37 (speaker 201007), the verb *ajdaťe* 'let's go' is followed by a Russian argument in two instances, and by the Erzya second infinitive in one. When the Russian argument is used, it involves a complete switch

to Russian both in the first case *ajdaťe na kosťor* 'let's go to the campfire' and the last case *ajdaťe uže sorťirovať kartušku* 'let's go to sort potatoes'.

(37) **ajda-ťe**   **na**   **kosťor**   **ajda-ťe**   **uš**   **piď-eme**
let's.go-PL   on   campfire   let's.go-PL   already   cook-INF2
meźejak   **to**   **uha**   **to**   **kartuška**-*t*   **potom**
something   sometimes   fish.soup   sometimes   potato-PL   then
mejľe   **uže**   *seďe*   *kas-ińek*
then   already   more   grow-PST.1PL
nu   **vot**   **ajda-ťe**   **uže**   **sorťirov-ať**   **kartušk-u**
well   well   let's.go-PL   already   sort-INF   potato-ACC.SG
'Let's go to the campfire, let's go and cook something, sometimes fish soup,
sometimes potatoes then then we got older, well well let's go and sort
potatoes.'

The second case when the Russian verb has an Erzya argument *ajdaťe uš piďeme* 'let's go and cook', is an instance of congruent lexicalization. The construction combines the morphosyntactic characteristics of two languages, the argument of the Russian verb *ajdaťe* 'let's go' is the second infinitive *piďeme* 'to cook', as required by motion verbs in Erzya. When the Russian infinitive is the argument, this distinction is lost. The difference between the two constructions can be observed in example 38: the dummy word *ťeńams* 'to do something, lit. to whatsit' takes the form of the Erzya second infinitive, while the Russian argument is in the basic infinitive form.

(38) **moľ-ťanok**   *ťeńa-mo*   **sortirov-ať**   **kartoška**-*t*
go-1PL   whatsit-INF2   sort-INF   potato-PL
'We go to, well, to sort potatoes.'

In example 38, the switch is flagged by the dummy word. Alternation is typically flagged by pauses and hesitation when the speaker is looking for the right word. In other cases, flagging can ease the transition in incongruent constructions (I discuss flagging in detail in section 7.3).

As mentioned above, alternations in the speech of C3 speakers usually do not have a pragmatic function, partly because of the high frequency of switches in this type of discourse. In

the following example, the C3 speaker (200511) changes the base language twice. These are unambiguous cases of alternation because the switches occur between the clauses, i.e. they can be considered intersentential switches.

(39) *miń eŕ-ińek*   *paŕśťe **požeńi-l-i-ś***

we live-PST.1PL well   get.married-PST-PL-REFL

eŕ-ińek   ***ľub'i-l-i***   *drug   drug-a*

live-PST.1PL love-PST-PL friend friend-GEN.SG

'We lived well, we got married, we loved each other.'

7.1.3. Alternation or insertion?

In this section, I analyze CS patterns that form a borderline category between alternation and insertion. My analysis focuses on intrasentential CS, but I discuss cases of intersentential alternation as well, usually involving a switch between turns in a dialogue. In example 40 (speaker 20130801), the Erzya numeral phrase *komś śiśemeće ije* 'twenty-seventh year' uttered by the guest is repeated by the reporter in Russian.

(40) Guest:

*mińek   komś   śiśemeće ije   koda   **panśionat**-oś ťese   panžov-s'*

we.GEN   twenty seventh year   how   home-DEF.SG here   open-PST.3SG

*i   uľ-it'   mińek   kona-t   ťese   **robot**-iť **pŕamo** panžoma-ń   či-ste*

and   be-3PL we.GEN which-PL here   work-3PL directly opening-GEN day-ELA

Reporter:

dvadcať   śeḿ   ľet

twenty   seven   year.PL

Guest: 'Our retirement home opened 27 years ago, and there are people who

have worked here since the very day the home opened.'

Reporter: 'Twenty-seven years.'

There are instances in which it is difficult to decide whether a case of alternation is involved or the Russian sequence is a longer insertion. Muysken (2000: 100) suggests that the

104

structural position of a switch can help in distinguishing between insertions and alternations. If the switch occurs at a peripheral position, it is clearly an instance of alternation, not an insertion. On the basis of this criterion, the sequence *s śeḿi utra do śeḿi večera* 'from seven in the morning until seven in the evening' in example 41 (from speaker 200811) should be considered an instance of alternation (time expressions are also typically switched in Erzya, as I discuss in detail in section 7.2.2.1).

(41) *tuj-iń*   **Saratov**-*ov* **robota**-*mo* **robot**-*ińek*   *pek*
     come-PST.1SG Saratov-LAT work-INF2   work-PST.1PL   very

     s śeḿ-i   ***utr-a***   ***do śeḿ-i***   ***večer-a***
     from seven-GEN.SG morning-GEN.SG until seven-GEN.SG evening-GEN.SG
     'I came to work in Saratov, we worked hard from seven in the morning until seven in the evening.'

However, I analyze these instances rather as insertions or congruent lexicalization and discuss them in section 7.2, despite the fact that they are not nested switches. Consequently, I regard example 41 as an insertion, while example 42 (from speaker 201006) is an instance of congruent lexicalization.

(42) *a*   *mon* *ľeľa-ń*   *marto*   ***nakladiva-l-a na voz***
     but I   elder.brother-POSS.1SG.GEN together load-PST-F   onto   cart
     'And I was loading the cart together with my elder brother'

There is a clear switch between the first and second part of the utterance, but the fact that there is gender agreement between the Erzya pronominal subject *mon* 'I' and the Russian predicate *nakladivala* 'I loaded' indicates that the switch does not occur immediately before the predicate, but the structure of the utterance is constructed in two languages. Thus, the utterance is an example of congruent lexicalization.

The most challenging part of the analysis is that of sequences that are entirely Russian and involve a longer sequence of Russian elements. Names and titles are typically used as chunks, cited in Russian within the Erzya base language. In example 43 (speaker 20151118), there is a complete switch to Russian after the Erzya copula *uľi* 'is', but this sequence can also be considered a longer insertion of two phrases: the first one is the name of the network *sóciaľnaja śeť*

pedagogov ŕespubľiki Mordovija 'social network of teachers in the Republic of Mordovia' and the second is the web address of their home page 'teacher-thirteen-dot-ru'.

(43) *koda **vot** soćialn-oj šeť-ťńe   koda **odnoklassńik-i***

as well   social-N network-DEF.PL as   classmate-PL

*koda **vkontakťe** isťa žo uľ-i   **soćiaľn-aja šeť**   **pedagog-ov***

as in.contact   so too be-3SG social-F   network teacher-GEN.PL

ŕespubľik-i   mordov'ija pedagog   tŕinadcať   točka ru

Republic-GEN.SG Mordovia   teacher   thirteen   dot   ru

'as, well, the social networking sites like Odnoklassniki or like Vkontakte, there is a social networking site of teachers in the Republic of Mordovia, pedagog13.ru'

The categorization of code-switched discourse particles also presents a problem. Muysken (2000) argues that discourse particles and adverbs occur rather as alternations, and not insertions. In my opinion, switched discourse particles and adverbs can also be instances of insertions if they occur in a nested position, i.e. within the utterance, surrounded by Erzya words, and not on the periphery. If they are categorized as alternation, they are clearly regarded as elements belonging to the Russian language, because alternation, by definition, involves a complete switch to another language. However, many Russian discourse particles are established borrowings in Erzya (e.g. *hoť* 'though, even', *uš* 'already', etc. in EKM 2000: 262). As a result, I analyze these non-peripheral switches of discourse markers as insertions in section 7.2.1.

Finally, I discuss two examples that are not clear cases of alternations. In example 44 (speaker 201101), the switch occurs inside the phrase between the Erzya quantifier *veśe* 'all' and the head *roďiťeľi* 'parents':

(44) *pek vadŕasto jut-iť ńe   vara   **vipuskn-oj bal**-tne*

very well   go-3PL these well   leaving-M ball-DEF.PL

*mińek   **tože** veśe   **roďiťeľ-i   pŕiglaša-jem***

we.GEN   too   all   parent-PL   invite-1PL

'Our [kindergarten-]leaving balls go very well, too, we invite all the parents.'

As a result, we face the following dilemma: the *roďiťeľi pŕiglašajem* 'we invite the parents' part can be considered alternation involving a complete switch, but as the switch occurs in the

106

middle of the construction, *roďiťeľi* 'parents' can also be considered an insertion which triggered the insertion of the Russian verbal phrase as well. In my opinion, categorizing the switch as an instance of congruent lexicalization is the least ambiguous analysis. The construction abides by the rules of two languages, as the Erzya pronoun *veśe* 'every' requires the plural form of the noun it modifies.

Example 44 also contains a Russian phrase *vipusknoj bal* 'kindergarten prom' which is inserted into the utterance with Erzya morphological markers. The phrase is a clear instance of insertion, but it is ambiguous whether the whole phrase is inserted as a chunk or it is adapted to Erzya. Russian adjectives are generally borrowed with the *-oj* ending into Erzya, and in this case the original Russian phrase also contains the masculine form of the adjective with the *-oj* ending.

As a result of these problems, I modify Muyken's (2000) definition of alternation. For the differentiation between insertion and alternation, I apply the predicate criterion, according to which, if the sequence is uniquely in Russian and has a Russian predicate, I consider the switch an instance of alternation, otherwise I regard it as an insertion in Muysken's terms or an EL as in the MLF model. The only exceptions are discourse markers in the peripheral position, which I regard as instances of alternation.

This method is applicable in most cases, although there are some borderline cases. In example 45 (speaker 20130402), a larger sequence, a compound sentence with two finite clauses is inserted into the Erzya utterance.

(45) **janvar**-ste sajeź **b'ibľioťeka**-ń važodića-ťńe ušod-śť **projekt**-eń
January-ELA from library-GEN worker-DEF.PL begin-PST.3PL project-GEN

ḿeropŕijaťija-ťńe-ń topavt-omo javoľavto-ź **zajď-i-ťe**
action-DEF.PL-GEN fulfil-INF2 announce-GER enter-IMP-2PL

v naš-u b'ibľioťek-u poradu-j-ťe roďiťeľ-ej
into our-ACC library-ACC make.happy-IMP-2PL parent-GEN.PL

i b'ibľioťekaŕ-ej akćija-ś
and librarian-GEN.PL action-DEF.SG

'Beginning with January, the library personnel have started to carry out the undertakings of the project by announcing the action: "Come to our library and

make the parents and the librarians happy."'

In this case, we cannot talk about alternation despite the complete switch for two reasons. On the one hand, the gerund *javoľavtoź* 'announced' and its object *akćijaś* 'action, project' are separated by the Russian sequence. On the other hand, the Russian compound sentence is a title, the name of the project, and, as a result, we can interpret it is as an unanalyzed chunk, an insertion in Muysken's (2000) terms and an EL island in the MLF model's (Myers-Scotton 2002).

To sum up, if we focus on the typical cases and disregard the borderline cases discussed above, the distinction between insertions and alternations can rely on two criteria: the position of the switched element in the utterance and the presence of the Russian predicate.

7.2. Insertion and congruent lexicalization

In this section, I study the intrasentential switch types occurring in my data. In the first section (7.2.1), the analysis focuses on discourse particles, followed by the study of nominal constructions in 7.2.2 and verbal constructions in 7.2.3.

The overview of these CS types is not comprehensive. In every section, I choose one type of construction I focus on. However, I refer to other phenomena if they are related to the CS pattern under scrutiny. I give an overview of CS types showing variation to the greatest extent as they are possible indicators of an ongoing change. The code-switches studied in this section are Russian constructions occurring in the bilingual discourse as insertions and congruent lexicalizations, except for discourse particles studied in section 7.2.1, which are also in-between categories, but showing characteristics of both insertions and alternations, discussed in section 7.1. The fact that CS constructions are realized differently by different speakers or even by the same speaker is connected to (in)congruence of the structures in the two languages. The main patterns are the following: mixed constituent insertions occur with congruent constructions, EL island-type insertions with incongruent ones, and congruent lexicalization occurs by definition in congruent constructions. However, there are deviations from these patterns, for instance, in the case of nominal constructions with a genitive modifier, congruent lexicalization occurs despite the incongruence of the constituent order in the two languages (cf. section 7.2.2.2).

7.2.1. Discourse particles

In this section, I discuss discourse particles inserted into the Erzya ML. These one-word switches are present in the utterances of even the least frequently switching C1 speakers, thus, they characterize all types of Erzya–Russian bilingual discourse.

Muysken (2000) categorizes discourse markers as subtypes of alternation. In my opinion, they can be considered also insertions in some of the cases depending on their position in the utterance. The switching of discourse markers could be analyzed as instances of alternation only if the two languages could be unambiguously separated in the discourse. However, as a result of intense contact and conversion, the discourse particles cannot be categorized as uniquely Russian elements. They are established borrowings in Erzya, discussed also in conservative descriptive grammars as elements of the Erzya vocabulary (cf. Russian origin particles *li* 'whether, if' *nu* 'well', *davaj* 'come on', etc. in EKM 2000: 260–267, or Jerina 1997). As a result, I consider them to be an in-between category which can occur as instances of insertion and alternation as well.

I argue that discourse particles in the Erzya–Russian CS discourse can be regarded as elements belonging to both languages, and as bridge words they facilitate smoother switches. The criteria traditionally used for the differentiation between code-switches and borrowings do not work in the case of discourse markers. With one-word switches in general, and discourse particles in particular, the problem of distinction between code-switches and borrowings arises, as these elements do not require morphological adaption and can be inserted into the Erzya ML without any modifications. The criterion often used for the distinction of borrowings and code-switches which takes into consideration the existence of native equivalents of a given Russian lexeme cannot be applied in this case either. While the discourse marker *vot* 'there' has an Erzya equivalent (*vana* 'there; look'), the Erzya form is not necessarily known all over the speech community. Some of the speakers use only the Russian forms, while there are Erzyas who apply the Russian and the Erzya equivalents interchangeably.

Discourse markers are regarded as a part of speech easily switched even in newer contact situations. According to Matras (1998), discourse particles, or in his terms, utterance modifiers, are the most susceptible to language change. He argues that utterance modifiers from another language are used for a very specific reason which is not connected to prestige or to the existence

of an equivalent in the recipient language. Instead, he considers the cognitive and pragmatic pressure of the dominant language the causes behind synchronic variation and diachronic change (1998: 281).

Matras (2009: 140) also discusses the universal nature of discourse markers: as they have gesture-like communicative functions, they are not connected to one language in the bilingual discourse:

"discourse operators or utterance modifiers carry out highly automated routine tasks, for which routine schemas appear to exist. Their role in managing the interaction and in mentally reaching out to the hearer gives them a gesture-like function. As such, they are less identifiable with just a particular subset of the repertoire and its permissible usage contexts – i.e. less attributable to a particular 'language' – and are instead accepted as more universal. They are, in other words, 'pragmatically detachable' from their source language".

In this overview, I focus only on the discourse particle *vot* 'here is, this is, look' which occurs in all of the interviews, and through the analysis of its versatile use, aspects of the switching of discourse particles can be thoroughly studied.

I argued above (and also in section 7.1) that the discourse particles can be regarded as both insertions and alternations. In Muysken's (2000) framework, they are considered instances of alternations. In Erzya–Russian bilingual discourse, they cannot be automatically categorized as alternations, because it is ambiguous which language they belong to. I categorize them as either insertions or alternations on the basis of their position in the utterance. If they occur in a peripheral position, they are analyzed as instances of alternation, while if they are nested, I regard them as insertions. If they occur at the periphery, they can be regarded also as extrasentential switches, which have a discourse organizing function. In example 46 (speaker 20160115), the discourse particle *vot* 'here is, this is, look' is used in a peripheral position:

(46) *i   vans-iť   i   kevksń-iť ťe   meźe ťe   koda ťe   źardo isťamo*
    and look-3PL and   ask-3PL   this what this how this when   like.this

uľńe-ś   **_vot_**

be-PST.3SG   here.is

'And they look, and they ask what this is, how and when it was like this, that's it.'

In example 47 (speaker 20151118), the discourse particle *vot* 'here is, this is, look' occurs in a nested position, consequently it can be considered an insertion.

(47)   *jutavt-ińek*   *nalkśema isťa **ž**o **urok**-sto-ńť*   **_to-jesť_**

        organize-PST.1PL   game   so   too   class-ELA-DEF.SG   that.means

        ejkakš-tne   *avoľ* **_vot_**   **_ćela_**   **_urok_** *asťe-śť*   **_parta_** *ekše*

        child-DEF.PL not   look   whole   class   stay-PST.3PL   desk   behind

        'We organized a game also during class so the children would not stay behind

        their desks the whole time.'

Russian discourse particles are used throughout the Erzya–Russian bilingual discourse. I analyze examples of the occurrence of the particle *vot* 'here is, this is, look' from different speakers' utterances. This discourse particle is attested both in a nested and in a peripheral position in the utterances of all types of speakers. Examples 48–51 demonstrate the use of this particle by all speaker types. Although Erzya discourse particles are described to some extent (Jerina 1997), further research should focus on the study of discourse particles based on corpus data and carried out with a conversation analytic approach.

The discourse particle *vot* is attested even in the utterances of the most purist speakers who monitor their speech and generally avoid code-switches (C1 speakers, cf. more Chapter 8). Particles (*daže* 'even' and *vot* 'here is, this is, look') as used by a C1 speaker (20130401) can be observed in example 48. The other Russian elements in these utterances are established loanwords (nouns: *sťih* 'poem', *poet* 'poet'; or conjunctions: *i* 'and', *no* 'but', *što* 'that').

(48)   *kojkona-tne-ń*   *mon vaśeńće-ďe ńej-iń*   *vaśeńće-ďe mar-iń*

        some-DEF.PL-GEN   I   first-ABL   see-PST.1SG   first-ABL   hear-PST.1SG

        *mon sinst **sťih**-est*   **_daže_** *a*   *soda-siń-gak*

        I   their poem-POSS.3PL   even   not   know-3PL<1SG-EMF

        _no_ *isťa-t vaďŕa-t **sťih**-t*   *siń lovno-śť*   **_vot_**   *i*   *keme-v-i*

        but   so-PL good-PL poem-PL   3PL   read-PST.3PL   there   and hope-REFL-3SG

što	ńe-ťńe	od	ľisevks-tne	kas-iť	paro	**poet**-eks
that	these-DEF.PL	new	comer-DEF.PL	grow-3PL	good	poet-TRA

'Some of them I have seen and heard for the first time. I did not even know their poems. But they read such good poems, so one can hope that these newcomers will grow to become good poets.'

In the utterance of the C2 speaker (20141217, example 49), the discourse particle *vot* occurs twice, in both cases on the left periphery of the utterance as an instance of alternation.

(49)

eś-enze	melga-nzo	**d'inastija**	eŕav-i	tŕ-ams	kast-oms
own-POSS.3SG	after-POSS.3G	dynasty	have.to-3SG	bring.up-INF1	bring.up-INF1

kadra-t	**potomu**	**što**	lomań-eń	**žizńa**-ś	iśťamo
personnel-PL	for.that	that	person-GEN	life-DEF.SG	like.this

rana	**iľi**	**pozdno**	lomań	śeŕeďi	**a**	**vot**	**štobu**	**koľľekťiv**-eś
early	or	late	person	ache-3SG	but	look	in.oder.to	collective-DEF.SG

eŕa-voľ	pinge-ďe	pinge-s	ije-ďe	ije-s	sonze	meľga-nzo
live-CONJ.3SG	era-ABL	era-ILL	year-ABL	year-ILL	(s)he.GEN	after-POSS.3SG

eŕav-iť	moŕića-t	polavtića-t	ťev-eń	nučka-nzo
have.to-3PL	singer-PL	continuer-PL	work-GEN	grandchild-POSS.3SG

ejkakš-onzo	**i**	isťa	**i**	isťa	**vot**	ťe	**proćess**-eś	**dolž-no**
child-POSS.3SG	and	so	and	so	like.that	this	process-DEF.SG	have.to-N

uľ-ems	**ńepŕerivnoj**	**ńepŕerivnoj**	**śisťemnoj**	**i**	**ńepŕerivnoj**
be-INF1	uninterrupted	uninterrupted	systematic	and	uninterrupted

'They have to bring up a dynasty, a personnel to continue the activities after they are gone, because the life of a person is like that, one becomes ill sooner or later, and for the society to live on for generations, for years after them, singers are needed, continuers, their grandchildren and children, and like that, like that, so the process has to be uninterrupted, uninterrupted, systematic and uninterrupted.'

In the utterance of the C3 speaker (20130411, example 50), the discourse particle is an insertion, embedded in the clause, occurring between the subject and the argument of the predicate, both expressed by Erzya elements.

(50) *son eŕavi   kińeń-buťi   ťej-ems **I   poka   no   krome***
it have.to-3SG   somebody.DAT do-INF1 and so.far but besides

poiskov'ik-ov   *ťe ťev-eńť*   *kijak*   ***ńe mož-et*** *ťej-ems*
searcher-GEN.PL this job-GEN.DEF.SG nobody not can-3SG do-INF1

pita-l-i-ś   *miń-ś*   ***popitka*** *ťej-ems* ***poiskovoj bataljon soldat-t***
try-PST-PL-REFL we-DEF.SG   initiative do-INF1 search   battalion soldier-PL

kona-t **_vot_**   *teń*   *ejste* ***zańima-jut-śa   i   služba-ń-gak***
who-PL there   this-GEN.SG with   engage-3PL-REFL and service-GEN.SG-EMF

'Somebody has to do it, and for now besides the searchers, this work cannot be done

by anyone else. We tried an initiative to create a searcher battalion, soldiers who,

here it is, will also engage in this work besides their service.'

The parallel use of multiple discourse particles is also typical in CS discourse, Auer (1995) describes this phenomenon as a tendency of contextualization cues to bundle. The discourse particles in the utterance can be both from the pragmatically dominant language, but there are cases in which both Erzya and Russian discourse particles are used. In example 51 (speaker 201101), only Russian particles occur, the particles *nu* 'well' and *vot* 'here is, this is, look' are grouped together at the right periphery of the utterance:

(51) ***duma**-tanok **što   sadïk-**ś*   *karm-i*   *pek ťena   śeďe pokš*
think-1PL   that kindergarten-DEF.SG will.be-3SG very whatsit more big

karm-i   ***vozmožnosť*** *uľ-ems* ***hoťa bu*** *veśe   vejse*
will.be-3SG possibility   be-INF1   at   least everything together

muzikaln-oj sporťivn-oj zal   aktovoj *toso veśe   vejse*   ***nu _vot_***
musical-M   sport-M   hall assembly there everything together well there

'We believe that the kindergarten will be very whatsit bigger, it will be possible

to have at least everything together, musical, sports and assemble hall there,

everything together, and that is it.'

In example 52 (speaker 20130401), two discourse particles are used, the Erzya particle *vana* 'there; look' is followed by its Russian origin equivalent, *vot*.

(52) *i   iśťamo   meľsparo uľń-eś   moń uľń-eś   toso   i   siń   i*
     and like.this happiness be-PST.3SG my   be-PST.3SG there and they and

     mora-śť   i   siń   korta-śť   i   zal-so-st   kadik a   pek uš
     sing-PST.3PL and they talk-PST.3PL and hall-INE-POSS.3PL let   not very
     all.that

     iśťa lamo loma-ťńe-ďe   uľńe-śť   no vana vot   iśťamo uľńe-śť
     so many   person-DEF.PL-ABL   be-PST.3PL but there there like.this be-PST.3PL

     atmosfʼera-ś   iśťamo   ľembe   iśťamo   meľsparo uľńe-śť
     atmosphere-DEF.SG like.this warm   like.this happiness be-PST.3PL

     veśe-ńeń
     everybody-DAT

     'And I felt such happiness there, and they sang, and they talked, and though in the
     auditorium there were not all that many people, but there, there the atmosphere was
     so warm, everybody had such a good time.'

Matras (2009: 140) explains this phenomenon with "the vulnerability of discourse
operators to selection malfunctions" which "leads to a relatively high frequency of 'slips' or
fallbacks into the pragmatically dominant language". In my opinion, here the use of the Erzya
particle does not suffice, the speaker "slips back" to Russian and uses the Russian equivalent of
the Erzya discourse particle.

It is also typical that the use of a discourse particle is followed by other switched elements,
as it triggers the use of more Russian forms. In example 53 (speaker 20151112), the adverb
napŕiḿer 'for example' is used after the particle *vot.*

(53) *a śe   źardo sakšn-iť mińeńek uže   sodaviks pojet-t*
     but that when come-3PL to.us   already famous   poet-PL

     vot napŕiḿer   Mariź Kemal meŕ-ems Išutkin Ńikolaj Ivanovič
     look for.example Mariz Kemal say-INF1 Ishutkin Nikolay Ivanovich

     źardo sakšn-i   ťe mińek   veľť keńardovt-I   maks-i
     when   come-3SG   this we.GEN very make.happy-3SG   give-3SG

     śeďe lamo vij-ť

more   much   strength-PL

'But that when already famous poets visit us, look, for example, Mariz Kemal, or say, when Ishutkin, Nikolay Ivanovich comes, it makes us very happy and gives us more strength.'

Finally, the level of entrenchment of the discourse particle *vot* has to be discussed. It occurs in all interviews, which corroborates Verschik's (2014: 50) argument that "in contact situations the domain of discourse pragmatics is easily borrowable, which means that L2 discourse particles either replace or co-occur with L1 equivalents". It can occur in a nested and a peripheral position, and can trigger the insertion of further Russian elements.

The use of the Russian origin particles can be explained by the pragmatic dominance of the Russian language in the discourse. According to Matras (1998: 396), speakers consider the pragmatically dominant language the norm which they are trying to conform to:

"The 'donor' language is one that is pragmatically dominant in the particular instance in which transfer occurs, that is, it is a linguistic system to which speakers show a special situative commitment and toward which their efforts at norm-conforming linguistic behavior are currently directed."

This is the language which is used in communication outside the ethnic group, in this case, with the Russian community. Erzya speakers do not take part in the formation of the language norm in Russian, consequently, they allow for mixed constructions, the repeated insertion of Russian discourse markers into the Erzya discourse. This can lead to the conventionalization of these forms. Matras (1998: 396) argues that prestige is responsible not only for the switching of discourse particles, but also for the borrowing of them.

7.2.2. Nominal constructions

In this section, I analyze nominal constructions that are either insertions or instances of congruent lexicalization. I only focus on two types of structures, namely, numeral phrases (section 7.2.2.1) and nominal constructions with a genitive modifier (section 7.2.2.2).

There are other nominal structures I do not discuss in detail, because they do not display a similar type of variation the two constructions in focus do. Further research could investigate, for instance, nominal phrases with adjectival attributes which present a borderline case between borrowings and code-switches. Object marking should also be studied in detail, especially with Russian code-switched predicates.

In Erzya–Russian bilingual discourse, there are longer switches, whole phrases, used as chunks, which are inserted in Erzya, similarly to one-word borrowings. This is especially true of numeral phrases (cf. section 7.2.2.1), which cannot be considered borrowings in the classical sense but are not idiosyncratic switches either, as they are used throughout the data by many speakers. In my opinion, these (and other chunk-like code-switches) can be considered the in-between stage on the borrowing–CS continuum. It is also typical in the Erzya–Russian CS data that Russian names of institutions and collocations (cf. section 7.2.2.2) are inserted into the Erzya utterances without any reanalysis. This phenomenon is not uncommon in Finno-Ugric languages spoken in Russia, for instance, Leinonen (2009: 319) also attests these cases in her Komi data.

7.2.2.1. Numeral phrases

In my corpus, there are two main types of numeral phrases inserted into Erzya utterances. On the one hand, there are constructions in which both the quantifier and the head noun are Russian origin elements and the structure of the phrase abides by the rules of Standard Russian. On the other hand, there are hybrid bilingual constructions in which the quantifier is typically Russian and the head noun is Erzya, but other mixed forms are also attested in the corpus. In this section, I intend to find an answer to the following questions: how these bilingual constructions arise, and how switching functions in these cases.

Differences between the two languages result in variation which concerns the choice of case and number, the constituent order, and the presence or lack of a demonstrative pronoun in the Erzya–Russian mixed numeral phrases.

7.2.2.1.1. Numeral phrases as Embedded Language islands

As discussed in the typological overview in section 3.2.1.3, Russian and Erzya numeral phrases are incongruent. In Erzya, the numeral phrase requires a nominative argument which is in the singular after the number 'one', and in the plural after numerals higher than two. Russian has a more complicated system: depending on the numeral, the nominative singular, the genitive singular, or the genitive plural is the required form (cf. section 3.2.1.3).

Incongruence is most typically avoided by the insertion of an EL island, i.e. a constituent that contains only EL system morphemes. In example 54, the Russian prepositional phrase *na tŕi goda* 'on three years' is inserted as an EL island in the Erzya frame. The Erzya phrase *kolmo ijeďe* (three year-ABL.SG) which would occur in a monolingual Erzya utterance is substituted by its Russian equivalent along with the preposition required by the rules of the Russian language.

(54) *śe   paťa-ś   **na tŕi god-a**   moń-ďe   **starši-ľ***

this   elder.sister-DEF.SG   on three year-SG.GEN   I-ABL.SG   old-PST2.3SG

'That elder sister was three years older than me.'

There are typical topics that trigger the use of Russian numeral phrases. These include time expressions, school grades and classes, quantities concerning distance (i.e. kilometers) and amounts of money. These are also inserted typically as EL islands.

In example 55 (speaker 200803), two types of time expressions occur, the Russian numeral phrase *šesť meśaćev* 'sixth months' as an EL island and *kolmo godt* 'three years' which is either a mixed construction or an Erzya constituent, as discussed below:

(55) *pŕad-ińek   **šesť meśać-ev**   toso **učebnoj otŕad**-so*

finish-PST.1PL   six   month-GEN.PL   there training battalion-INE

*tonavt-imiź   **na   sṕećiaľnosť***

teach-PST.1PL<3PL   to   specialty

torpedist-a		***i***	*mejľe* ***značit*** *pong-iń*	
torpedo.operator-GEN.SG		and	then well get-PST.1SG	

Kamčatka*-v*	***na***	***podvodn-uju***	***lodk-u***	***na***	***raketn-uju***
Kamchatka-LAT	to	underwater-ACC	ship-ACC	to	rocket-ACC

i	***vot*** *toso*	***služ****-iń*	<u>*kolmo* ***god****-t*</u>		
and	so there	serve-PST.1SG	three year-PL		

'We finished the training camp in six months, we were trained as torpedo operators, and then, well, I was assigned to Kamchatka to a submarine with missiles and so I served three years there.'

The numeral construction *kolmo godt* 'three years' is an Erzya constituent if the word *god* 'year' is considered a loanword in Erzya. If the existence of an Erzya equivalent (*ije* 'year', also used by the speaker) is taken into account, it can also be considered a code-switch which is morphologically integrated, as it is in the plural as required by the rules of the Erzya language in structures having numerals 'two' and higher as attributes.

In example 56, the speaker 201011 is explaining the traditions related to weddings. She uses the Russian time expression *tŕi časa* 'at three' inserted as a chunk into the utterance (the preposition *v* 'in' is deleted as typical in colloquial Russian), the numeral *čotiŕe* 'four' adapted to Erzya phonologically (standard Russian *četiŕe*) is added to the construction, as the speaker hesitates.

(56)	<u>***tŕi*** ***čas****-a*</u>	*valcke*	***iľi***	***čotiŕe***	*toso*	***rodńa****-va*	*jak-iť*
	three hour-GEN.SG	morning	or	four	there	relative-PROL	go-3PL

'At three or four in the morning, they visited relatives.'

In examples 57 (speaker 200501) and 58 (speaker 200811), the numeral phrases are connected to the domain of school and education and are inserted as EL islands. In example 57, the number of years the speaker has studied at school is expressed with a Russian chunk: *vośem klassov* 'eight grades', the noun *klass* 'grade' is in the genitive plural as required with the numeral *vośem* 'eight'.

(57)	*tonavl-iń*	***škola****-so*	<u>***vośem klass-ov***</u>		*pŕad-iń*
	study-PST.1SG	school-INE	eight class-GEN.PL		finish-PST.1SG

'I studied at school, I finished eight grades.'

In example 58, the prepositional phrases *na četeri* 'on four' and *na ṕať* 'on five' contain numerals, the names of school marks, and are inserted as chunks into the utterance.

(58) **pravda što** *tonavtń-iń*   *avoľ* **ovśe** *beŕańste* **na četeri na ṕať**

true   that   study-PST.1SG   not   at.all badly   on   four   on   five

no daľše *tonavtń-eme eź-iń*   *tuk*

but further   study-INF2   not.PST-1SG   go-CONNEG

'It is true that I studied quite well, earning fours and fives, but I did not study further.'

Example 59 (speaker 200502) illustrates how entrenched numerals referring to certain topics are. For example, school marks are named by their Russian forms even in the utterances of speakers who otherwise use Erzya numerals. Although referring to the same number, the mark is expressed in Russian *tŕi* 'a three', its modifier is in Erzya *kolmo* 'three'. The possessive suffix is added to the Russian numeral.

(59) *peľ* *ije-ń*   *juta-ź*   *liśńe-śť*   *kolmo*   *tŕi-m*

half   year-GEN elapse-GER   come.out-PST.3PL three   three-POSS.1SG

'Half a year later, I had three 3s.'

It is also typical to express distance with a Russian construction. In example 60 (also cited earlier as part of a longer sequence in example 27), the use of the EL island *dvadcať ṕať kiloḿetrov* 'twenty-five kilometers' can be explained by the fact that schools generally teach measurement and mathematics exclusively in Russian. Moreover, only Russian is used as the *de facto* official language of administration, despite the official status of Erzya in the Mordvin Republic.

(60) *ťe-ste*   **dvadcať ṕať kiloḿetr-ov**   *avoľ* *vasolo*

this-ELA   twenty   five   kilometer-GEN.PL   not   far

'It is twenty-five kilometers from here, not far.'

Russian numeral phrases are also used with expressions referring to other quantities, such as an amount of money. In example 61 (speaker 200507), the numeral phrase *dvatcať koṕejek* 'twenty kopeks' is an EL island, inserted as a chunk. The speaker adds the Erzya word *jarmak* 'money' to the construction and attaches the possessive suffix to this Erzya word, and not the

Russian numeral. In Erzya CS discourse it is common to attach the Erzya endings directly to the Russian chunk (cf. examples 62 and 63).

(61) koda ***dvatcať koṕej-e***k   jarmak-ot   moľ-at   meźejak   ram-at

   as   twenty kopek-GEN.PL money-POSS.2SG go-2SG   something buy-2SG

   'When you have twenty kopeks money, you go and buy something.'

This phenomenon is present in other minority languages of the Russian Federation as well (cf. Sarhimaa 1999: 234). In Karelian, it is also very typical to use Russian numerals for the expression of time, especially age and date of birth, and when referring to school grades and classes (Pyöli 1996: 295). Quantities of money are also expressed in Russian in Karelian discourse. According to Pyöli (1996), the use of Russian numerals in Karelian was recorded already 20-30 years before her study, so it has been a common practice in Karelian–Russian CS since the late 1960s. Although the data on earlier Erzya–Russian CS are scarce, we can claim on the basis of Hallap's 1960 investigation that insertion of Russian numeral phrases into the Erzya frame has been a long-standing practice among Erzya-Russian bilingual speakers (cf. section 4.2).

If the Erzya suffixes are attached to the end of the construction, the Russian numeral phrases can be considered EL islands inserted as chunks into the Erzya frame. In example 62 (speaker 201001), the Erzya inessive suffix is added to the construction, to the genitive plural form of the word 'year'.

(62) ***oďinnadcať ľet***-se   ***uš***   ***škola***-so   naš   ***zaṕreťi-l-i***   moń

   eleven   years-INE already school-INE well ban-PST-PL   I.GEN

   'When I was eleven years old, I was banned from school.'

Besides case endings, possessive suffixes can also be added to the Russian EL island which results in a mixed constituent containing morphological markers from both languages. In example 63 (speaker 200808), the second person singular possessive ending is added to the genitive plural form of the word *koṕejka* 'kopek'.

(63) ***ṕať   koṕej-ek***-eť   pojav-i

   five   kopek-GEN.PL-POSS.2SG   occur-3SG

   'If you manage to gather five kopeks.'

These constructions also have parallels in the Moksha–Russian CS discourse. In example 64 from Hallap (1960: 223), the suffix of nominal conjugation is attached to the genitive singular form of the word 'year'.

(64) *ton* **kako-və**   **god-a**-*t*

你 you which-GEN.SG   year-GEN.SG-2SG

'Which year are you from?' (Hallap 1960: 223)

In Karelian, it is also typical to insert the Russian numeral phrase as a chunk and add the Karelian endings to the whole construction. Pyöli (1996) cites an example (example 65 here), which can be considered a parallel construction to the Erzya CS types discussed in examples 62 and 63. However, Pyöli (1996: 287) analyzes it differently: "Numeral constructions typically decay in a way that the first part is a code-switch from Russian, while the last part is either an idiosyncratic or integrated borrowing" (my translation)[9]:

(65) *sygyzyl ol'l toine pruazn'iekku, девятого d'ekabr'al*

'There was another celebration in the autumn, on 9th December.' (Pyöli 1996: 287)

In example 65, the part *d'ekabr'a* 'of December' is the genitive singular form of *d'ekabr'* 'December'. Pyöli (1996) considers it a loanword, but in my opinion it forms one unit with *девятого* (*d'ev'atogo*) 'ninth', and the Russian phrase is inserted as a chunk into the Karelian utterance, the Karelian adessive ending -*l* 'on' attached to this chunk as a whole. Pyöli does not elaborate on this example further, so it is possible that other (phonetic and morphological) factors are also relevant for her analysis. However, it seems to me that this CS type is a parallel of the Erzya constructions and a widespread pattern in contact situations.

In example 66 (speaker 201001), there are three numeral phrase insertions. All of them are EL islands, but the first phrase *tŕidcatovo godań* 'of 1930' is inserted into the frame with an Erzya genitive marker (-*ń*). As a result, the Russian chunk is part of a mixed constituent. The numeral phrase *sorok ṕervom* would be a prepositional phrase in Standard Russian with the preposition *v* 'in' (*v sorok ṕervom* 'in 1941'), but the preposition is deleted, which is a common practice in spoken Russian.

---

[9] "Hyvin yleistä on lukusanailmausten hajoaminen niin, että alkuosa on koodinvaihto venäjästä, loppuosa joko omaperäinen tai integroitunut lainasana" Pyöli (1996: 287).

(66) *mon* **_třidcatovo god-a-ń_**   *a*   **_vojna_**-ś   **_sorok ṕerv-om_**

I   thirty-GEN year-GEN.SG-GEN but war-DEF.SG forty   first-INSTR.SG

karma-ś   *ťeńa-sto* **_značit_** *moń*   *topoď-iľ-ť*   **_śenťabŕa_**-*sto*

begin-PST.3SG   er-ELA   so   I.GEN   fill-PST2-3PL   September-ELA

vśego   **_oďinnadcať_** *a*   *v*   *ijuń-e*   *vojna*   *karma*-ś

only   eleven   but   in   June-PREP.SG   war   begin-PST.3SG

'I was born in 1930, the war started in 1941, it started in…, well, so and I had just

my 11th birthday in September and in June, the war started.'

In example 66, the numeral phrase *třidcatovo godań* 'of 1930' contains two genitive markers, one from either language. It is, however, not an instance of double marking, as the two suffixes have different functions. The Russian genitive is required by the form of the numeral, while the Erzya ending transforms the numeral phrase into an adjective.

7.2.2.1.2. Mixed numeral phrases

In this section, I analyze double forms and mixed structures, in which the incongruence of the two languages results in a hybrid construction. There are utterances in the data in which the numeral phrases in one of the languages are partially or fully repeated in the other language. These double constructions are typically flagged. Both orders are possible: if an Erzya nominal phrase is the first element, it typically involves repair in mid-construction, and the speakers indicate that they cannot produce the given numeral phrase in Erzya. The reverse cases, in which the Russian numeral phrase is followed by the Erzya equivalent, cannot be explained by a lexical gap, but rather by the speakers' monitoring their speech and applying medium repair. If the repair is offered by another communicative partner, it might be either ignored or accepted by the speaker.

In example 67, the speaker (200803, here S1) tries to produce the construction in Erzya but fails to remember the Erzya numeral 'nine hundred', and when he finally utters it in Russian, the other native speaker of Erzya (S2) present offers help and provides the numeral in Erzya, which the speaker accepts in the next turn.

(67) S1: *mon šač-iń   vejke **tišča**   vejkśe-će vejke **tišča***

      I   being.born-PST.1SG one thousand nineth   one thousand

  ďev'atsot

   nine.hundred

  S2: _vejkśeśadt_

   nine.hundred

  S1: _vejkśeśadt_   *ńiľeńgemeń kavkso-će*   **god**-*sto*

   nine.hundred   forty   eighth   year-ELA

  S1: 'I was born in one thousand ninth, one thousand nine hundred'

  S2: 'nine hundred'

  S1: 'in nine hundred forty eighth year (in 1948).'

The numeral *tišča* 'thousand' is an established Russian loanword in Erzya from the Russian *tiśača* 'thousand'. The word *tišča* 'thousand' functions here as a bridge word between the two languages, as its colloquial pronunciation even by monolingual Russians is also *tišča* 'thousand'. Consequently, the Russian numeral *tišča* 'thousand' could be considered a triggering element. However, this switch is a clear example of a lexical gap: the speaker tries to produce the Erzya equivalent of the numeral but cannot remember the required form. This is not an individual phenomenon, Erzya speakers typically produce (complex) Russian numerals with great ease due to the dominance (or exclusiveness) of the Russian language in education.

In example 68, the repair occurs within the phrase and the Russian construction is replaced by its Erzya equivalent.

(68) **_nu_**   *pek vasolo Nuja*   **_nav'erno_** **_kilo_**   _kavtoška_

   well very far   Nuja   maybe   kilo(meter) eight-COMP

  iľi *kolmoška*   **_kiloḿetra_**-*t iśťa*

  or three-COMP   kilometer-PL so

  'Well, very far, the Nuja river, maybe kilo…, two or three kilometers, so.'

Apart from discourse particles and adverbs, the speaker (200507) typically avoids Russian elements in her utterances. She begins the Russian type of construction expressing approximation *kiloḿetra dva iľi tŕi* 'two or three kilometers', but then she backtracks and eventually uses the

Erzya equivalent. The Russian construction can be triggered by the adverb *nav'erno* 'maybe', or by the fact that quantities are often expressed in Russian.

There are many instances of self-repair with numeral phrases in the data, since speakers try to use the Erzya numbers even if they do not know the Erzya equivalent of a Russian numeral. In example 69 (speaker 200802), the Erzya numeral is preceded by part of the Russian numeral and a pause, but nonetheless, it is provided correctly.

(69) ***mladšij*** *ťejťeŕka-ńťeń*   ***dvadsať*** *… komśvejke ije-ť*

younger girl-DAT.DEF.SG   twenty   twenty   year-PL

'The younger girl is twenty years old.'

In example 70 (also from speaker 200802), the same pattern can be found: the Russian numeral is repaired with the Erzya equivalent. However, the Erzya numeral is provided incorrectly in this case, as the speaker suggests that the girl is thirty years old instead of thirteen.

(70) *staršij*   *ťejťeŕka-ńťeń*   *ńej*   ***trinad*** *… kolmońǵemeń ije-ť*

older   girl-DAT.DEF.SG   now   thirteen   thirty

year-PL

'The older girl is now thir[teen] thirty years old.'

In example 71, the construction type is similar to the examples above, the Russian phrase *posľe ďev'aťi ľet* 'after nine years' is only partially uttered, then the speaker (200803) opts for the Erzya equivalent.

(71) ***posľe ďev'ať-i*** *vejkśe* ***god***-*oń jutaź mekev s-iń*   ***Mordovija***-*v*

after   nine-GEN.SG nine   year-GEN after back   come-PST.1SG   Mordovia-LAT

'After nine nine years later, I came back to Mordovia.'

The constructions in the two languages are incongruent, the order of the constituents is different, and the Russian prepositional phrase is contrasted by a postposition construction in Erzya.

In example 72, incongruence is avoided by a mixed construction in which the repetition is only partial, the numeral is uttered only in Erzya, the word 'years' occurs in both languages, and the order of the constituents follows partly the Russian, partly the Erzya pattern. There is a complete switch which might suggest that it is an instance of alternation. However, CS occurs

with the construction, thus, the resulting mixed phrase is rather an example of congruent lexicalization.

(72) **uže**   **prakťikuj-et-śa**   **ľet**   **koto**   **śiśem**   **ije-ť**

already practice-Sg3-REFL   year.GEN.PL   six   seven   year-PL

'They have already been doing it for 6 or 7 years.'

In this example, the incongruence concerns the order of the constituents. While in Russian approximation is expressed by a reversed word order, this construction does not change the order of the elements in Erzya. The first part of the construction is Russian, the numeral is uttered in Erzya, and the Erzya noun *ije* 'year' is in the plural, as required by the numeral in Standard Erzya.

In example 73, the speaker (201009) uses only the Russian numeral but flags her switch by admitting that she does not know the equivalent in Erzya:

(73) **trídcať śem**   *erźa-ks*   *a*   *mašt-an*   *jovtamo-st*

thirty   seven   Erzya-TRA   not   can-1SG   saying-POSS.3PL

'Thirty seven, I cannot say it in Erzya.'

There are other examples in which the switch points for Russian numeral phrases are flagged. In example 74, the speaker (200802) not only indicates that she does not know the Erzya equivalent of this Russian numeral but also argues that it is understandable for everybody present. This is a general tendency that the Russian numerals are considered default, and it is evident that everybody understands them. Let's compare this flagged CS to the speaker's other numeral phrases in examples 69 and 70. She provides the Russian form first in all three cases. Twice an Erzya repair follows, although one of the forms is incorrect. The third time in the interview, only the Russian form is provided with the excuse *erźaks a jovtavi ťenk* 'I cannot tell you in Erzya':

(74) **stop'aťďeśat**   *erźa-ks*   *a*   *jovt-av-i*   *ťenk*   **no pońatno stop'aťďeśat**

hundred.fifty   Erzya-TRA   not   tell-REFL-3SG   you.DAT   but   clear   hundred.fifty

ejkakš-t   *kona-ťńe*   *toso*   *eŕ-iť*   *ťeťa-vtomo*   *ava-vtomo*

child-PL   who-DEF.PL   there   live-3PL   father-ABESS   mother-ABESS

'One hundred and fifty, in Erzya I cannot tell you, but it's clear, one hundred and fifty children live there without parents.'

The examples above are instances of parallel constructions, in which the forms are partially or completely repeated. In example 74, there are no repeated elements in the discourse. The numeral 'one hundred and fifty' is expressed in Russian and the modified element 'children' is in Erzya. This is a mixed structure in which the switch occurs within the construction. In example 75, the speaker (200501, S1 here) uses the Russian construction as an EL island. The other native speaker present repeats the numeral phrase but replaces the Russian noun by its Erzya equivalent. As a result, a mixed construction arises in which the switch occurs within the construction.

(75)

S1:   ***četiŕesta***   ***p'aťďeśat*** **dom-ov**

      four.hundred   fifty   house-GEN.PL

S2:   ***četiŕesta***   ***ṕaťďeśat*** *kudo-t*   *ejse-nze*   *da*

      four.hundred   fifty   house-PL in.it-POSS.3SG right

'There are four hundred fifty houses in it, right?'

In this case, the speakers, or at least S2 considers these two constructions equivalent, the Russian genitive plural form, required by the Russian numeral *četiŕesta ṕaťďeśat* 'four hundred fifty' is replaced by the Erzya nominative plural, because in Erzya a numeral higher than one usually requires the use of nominative plural form of the noun it modifies.

The following examples are also instances of hybrid numeral phrases in which the Russian numeral is followed by an Erzya noun. In these cases, however, there is no hesitation, pause, or excuse at the switch point, the Erzya and the Russian numeral systems are combined.

(76)   *śe*   ***plaťija***-*ś*   *kis mońe*   *maks-śť*

       this   dress-DEF.SG for me.DAT   give-PST.3PL

       četiŕe kilogramm-a   *tovźuro-t*

       four   kilogram-GEN.SG   grain-PL

'They gave me four kilograms of grain in return for the dress.'

In example 76 (speaker 201006), the 'four kilograms' part of the 'four kilograms of grain' structure obeys the rules of the Russian language, while the Erzya word 'grain' is in plural in concordance with the Erzya system.

Example 77 (speaker 200811), contains both Erzya and Russian numerals – the number 'four' *ńiľe* in Erzya, and 'twenty' in Russian *dvadsať*– as part of an inserted chunk *dvadsať štuk* 'twenty pieces':

(77) *skal   kirď-an   ńiľe   tuvo-n   vaz-om*

cow   keep-1SG four pig-POSS.1SG.PL calf-POSS.1SG

*saraz-on   **dvadsať štuk***

hen-POSS.1SG.PL twenty   piece.GEN.PL

'I keep a cow, I have four pigs, one calf, and hens, twenty of them.'

There is one instance (example 78) in my data in which it is impossible to say to which language an element of the numeral phrase belongs to. In example 78 (speaker 201015), the Russian origin noun *kiloḿetra* 'kilometer' serves as a bridge word between the two languages. It can be interpreted as a Russian genitive singular form, or the nominative singular in Erzya.

(78) *mińek   viŕ-eńek…   **naverno kiloḿetra**   kavto ejste-ďe-ńek*

we.GEN forest-POSS.1PL perhaps kilometer   two   from-ABL-POSS.1PL

'Our forest is perhaps two kilometers from us.'

The morphosyntactic structure of the phrase relies on a composite matrix. Approximation is expressed through inversion as in Russian (*kiloḿetra dva tŕi* 'around two or three kilometers'). In monolingual Erzya, it would require the comparative suffix -*ška* (*kavtoška kolmoška kiloḿetrat* 'around two or three kilometers'). The numerals are in Erzya, which would entail that the noun is in the plural, thus *kiloḿetrat* 'kilometers' would be the required form. Consequently, the arising construction violates the rules of both languages. The most likely explanation is that the Russian construction is started and the switch occurs within the phrase without any hesitation, without any pauses.

Finally, we can conclude that CS patterns of numeral phrases in Erzya–Russian bilingual discourse show a great degree of variation. The co-existence of a variety of forms can be attributed to incongruence between Erzya and Russian constructions, which concerns especially the order of constituents in structures expressing approximation. Certain topics also trigger the use of Russian numerals, and the fact that Russian is the language used predominantly, or, in the

majority of cases, even exclusively, in teaching mathematics, contributes to the dominance of Russian forms, especially as far as compound numerals are concerned.

Similar tendencies can be observed in other contact situations as well. Halmari (2005) describes the CS patterns and language use characteristics of young Finnish American speakers who use Finnish numerals when referring to small quantities, but English compound numerals. On the basis of her data, Halmari (2005: 430) claims that "L1 loss is always lurking 'beyond twenty'".

There are bilingual numeral phrase constructions in the data that cannot be attributed to a diminished competence of the speaker. If the Russian form is partially or fully repeated in Erzya, the reasons for the switch can be attributed to pragmatic and sociolinguistic factors, especially purism. This question requires more thorough investigation in the future.

There are other individual differences in the realization of numeral phrases which should be investigated further. I argue that many factors contribute to the choice between the Erzya or the Russian structure, and to the choice between embedding the Russian element into the utterance as an EL island or a mixed constituent. These aspects include the sociolinguistic background of the speakers, their language proficiency, puristic tendencies in the community, the formality of the communicative situation, and the topic of discussion should all be investigated in a complex study, preferably involving other (Finno-Ugric) minority languages in contact with Russian, and also written sources.

7.2.2.2. Nominal phrases with a genitive modifier

In this section, I investigate a nominal construction which is present both in Erzya and Russian but is realized differently in the two languages. In these nominal phrases with a genitive modifier, the incongruence concerns the constituent order, and to some extent the grammatical markers attached to the modified element of the phrase. Resulting from this discrepancy, there are different ways CS can be realized in these constructions. The incongruence is neutralized if the Russian construction is inserted as an EL island, i.e. as a chunk, into the utterance. These CS instances are studied in section 7.2.2.2.1, while section 7.2.2.2.2 focuses on mixed nominal

phrases with a genitive modifier in which part of the original Russian construction is retained, but part of it is replaced with an Erzya equivalents.

7.2.2.2.1. Embedded language island-type insertions

As mentioned above, possessive structures are formed differently in the two languages. Erzya requires a possessor–possessee order, while in Russian, the possessee–possessor order is the default. There is a discrepancy in the form as well. While possessors are in the genitive case in both languages, in Erzya the possessee can have a possessive suffix as well.

In this section, I study constructions in which speakers apply neutralization strategies to avoid incongruence. The Russian structure is inserted into the Erzya frame as a chunk to avoid the necessity of adapting the Russian element.

In example 79 (speaker 20130930), the name of the device *aromaťizator vozduha* 'air aromatizer' is inserted as a chunk into the utterance. (The discussed construction is underlined in the examples.)

(79) *uľ-iť*　　*i*　　__*aromaťizator vozduh-a*__　　*kak vot*　　*sanatoŕija-so*

be-3PL　and aromatizer　　air-GEN.SG as　there　sanatorium-INE

'There is also an air aromatizer just like at a sanatorium.'

Although the majority of the lexemes in the utterance is of Russian origin, the function words are Erzya. (The word *sanatoŕija* 'sanatorium' is an adapted loanword in Erzya from the Russian *sanatoŕij* 'sanatorium'.) As a result, the ML is Erzya and the inserted chunk can be analyzed as an EL island. There is incongruence between the predicate and the subject, as the insertion is in the singular but the predicate *uľiť* 'there are' is in the plural. It can be explained by pragmatic factors: the speaker starts to enumerate the devices in the room but finally names only one of them.

In example 80 (speaker 20140318), the Russian possessive structure *v'id podďeržki* 'type of support' is inserted into the Erzya frame also without any Erzya morphological markers. However, the structure has an Erzya attribute *isťamo* 'such'.

(80) *uľ-i   ńej   isťamo **v'id podďeržk-i**   ľem-eze*

be-3SG now such   type support-GEN.SG name-POSS.3SG

ńesv'aźann-aja   podďeržka

unconstrained-F support

'There is now such type of support which is called unconstrained support.'

As we could see in section 7.2.2.1, numeral phrases are also typically inserted as chunks into the Erzya morphosyntactic frame. In example 81 (speaker 201015), there are two inserted Russian nominal phrases with a genitive modifier: the first is a numeral phrase, while the second is a nominal construction, the name of an event.

(81) ***vot**   avoľ umok   uľń-eś   **dvadcať ṕat-ogo   ijuľ-a***

there not long.ago be-PST.3SG twenty   five-GEN.SG July-GEN.SG

*mińek   **stŕeča   vipuskńik-ov***

we.GEN   meeting alumnus-GEN.PL

'Well, not long ago we had the meeting of alumni on 25 July.'

Names of professions and institutions are typically inserted as unaltered chunks into the Erzya frame. These EL islands can be part of a larger mixed constituent in which Erzya morphological markers are attached to the whole construction, to its end, and not to the head of the phrase. These endings can involve case endings (example 82) or the plural marker (example 83).

In example 82 (speaker 20130409), the inessive ending is attached to a hybrid construction: the word *kultura* 'culture' is used as a bridge word in the nominative instead of the genitive (*kulturi*) required by Russian.

(82) *mon robot-iń   alamoška-s **dom kultura-so**   **ďirektor-om***

1SG work-PST.1SG a.little-ILL   house culture-INE   manager-INSTR.SG

'I worked for a short time as the manager of the cultural center.'

The plural ending can also be attached to the inserted chunks, as in example 83 (speaker 200509) the Erzya plural ending *-t* is added to the whole phrase. The Russian equivalent *doski počjota* 'honor boards' of this plural construction would involve marking the plural on the possessee *doska* 'board'.

(83) ikeľe mińek   uľńe-śť   **doska počjot-a**-t   toso **vot** sval

before we.GEN   be-PST.3PL board honor-GEN.SG-PL there look often

uľń-iń   pongavt-oź **no** pŕad-ija   **škola**-ńť   **ḿedal**-ťeme

be-PST.1SG hang-GER   but finish-3SG<1SG school-GEN.DEF.SG   medal-ABESS

'Earlier we had honor boards, my name was often up there, still I finished school

without a medal.'

In example 84 (speaker 20130808), there are three possessive structures with a genitive modifier: *učiťeľ maťemaťiki* 'teacher of mathematics', *ďiŕektorom školi* 'headmaster of the school', and *ḿińisťerstvo obrazovańija* 'ministry of education'.

(84) mon **učiťeľ   učiťeľ   maťemaťik-i**   uľń-iń   **ďiŕektor-om**

I   teacher teacher mathematics-GEN.SG be-PST.1SG headmaster-INSTR.SG

škol-i   i   **mog-l-a   bi   dopusť-iť** što moń škola-so-ńť

school-GEN.SG and can-PST-F COND permit-INF that my   school-INE-DEF.SG

korta-v-olť   erźa-ń   keľ-se   **no ḿińisťerstvo obrazovańi-ja**

speak-CONJ-3PL Erzya-GEN language-INE but ministry   education-GEN.SG

sinďest ťe   a   eŕav-i

they.ABL this not   has.to-3SG

'I am a teacher, a mathematics teacher. I was the headmaster of the school. I could

have arranged that in my school the Erzya language would be spoken, but the

ministry of education, according to them, it is not needed.'

All three are inserted as chunks. A Russian genitive ending is attached to all the possessors. The possessees are in the nominative in the first and third cases. In the second insertion (*ďiŕektorom školi* 'headmaster of the school') the Russian instrumental case is used (*ďiŕektorom*) to express the meaning 'as a headmaster'. If the Russian instrumental case ending is replaced by its functional Erzya equivalent, the translative case ending, it is typically attached to the whole construction and not the possessee, as we can see in example 85 (speaker 201001). In this utterance, the Russian possessive construction *glava goroda* 'mayor, lit. head of town' is inserted into the Erzya utterance as one unit, and the Erzya translative suffix is not attached to the head of the construction (*glava* 'head') but to the last word (*goroda* 'city-GEN.SG'):

(85) moń ćora-m   ńej robot-i   **v Alatir-e**   meŕ-ems

my son-POSS.1SG now work-3SG   in Alatyr-PREP.SG say-INF1

glava   gorod-a-ks

head   town-GEN.SG-TRA

'My son now works as a mayor in Alatyr.'

As mentioned in section 7.2.2.1, this type of construction is also typical with numeral phrases, where the endings (plural, genitive, inessive, etc.) are attached to the whole construction. In example 86 (speaker 200806), the Erzya inessive ending is attached to the Russian noun *ľet* 'year' in genitive plural.

(86) *śestra-m*   ***šestnadcať ľet*-se**   *kul-oś*

sister-POSS.1SG sixteen   year.GEN.PL-INE die-PST.3SG

'My sister died at the age of sixteen.'

There are instances in which the Russian construction is not inserted as a chunk but is analyzed, and the Erzya endings replace the Russian markers and occur in the same position as the Russian elements would. I analyze these hybrid constructions in section 7.2.2.2 below.

7.2.2.2.2. Mixed nominal phrases

In this section, I discuss mixed constructions in which the elements of possessive structures come from both Erzya and Russian. In example 87 (speaker 20130918), two types of possessive structure occur, one as a Russian EL island and the other as a hybrid construction. The construction *praviťeľstvo ŕespubľiki Mordovijań* 'government of the Republic of Mordovia' is an inserted Russian chunk in which the Russian genitive is used to mark the possessor, while the Erzya genitive ending is required by the postposition *turtov* 'for'. The other possessive structure (*praviťeľstvo Rossijskoj Feďeraćijań* 'government of the Russian Federation'), however, is a hybrid construction, an instance of congruent lexicalization.

(87) *ťeči*   *śukpŕa* *mińek*   ***ďeda*-nok**   ***baba*-nok**   *turtov*

today   thanks we.GEN grandfather-POSS.1PL grandmother-POSS.1PL towards

ťeťa-nok   *ava-nok*   *turtov* ***praviťeľstvo Rossijsk-oj***

father-POSS.1PL mother-POSS.1PL towards government Russian-GEN.SG

Feďeraćija**-ń* *turtov* ***prav'iťeľstvo ŕespubľik-i ***Mordovija**-ń* *turtov*

Federation-GEN towards government Republic-GEN.SG Mordovija-GEN towards

ťev-eńek *mińek* *avoľ* *išťa-t* *beŕa-ť* ***di*** *avoľ* *staka-t*

thing-POSS.1PL we.GEN not such-PL bad-PL and not difficult-PL

'Today, thanks to our grandparents, our parents, the government of the Russian Federation, and the government of the Republic of Mordovia, our life is not so hard and not difficult.'

The monolingual Russian equivalent *praviťeľstvo Rossijskoj Feďeraćii* 'government of the Russian Federation' is modified, and the Russian genitive form of possessor *Feďeraćii* is replaced by the nominative stem *Feďeraćija* 'federation' to which the Erzya genitive marker -*ń* is added. In this way, the construction does not abide by the rules of either language. This type of words that could belong to both languages function as bridges and facilitate the use of Russian origin constructions in the Erzya utterance.

Below I discuss more straightforward and frequent CS types in which the Russian genitive marker is replaced by its Erzya equivalent, as in example 88 (speaker 201101): instead of the Russian genitive ending -*a* (*saďika* 'kindegarten's'), the Erzya definite genitive ending -*eńť* is applied (*saďikeńť* 'kindegarten's').

(88) *i **aťťestacćija** **saďik**-eńť* *ťeďeďe* ***nav'erno*** *karm-i*

  and attestation kindergarten-GEN.DEF.SG this.year probably will.be-3SG

  'The accreditation process of the kindergarten will probably take place this year.'

Example 89 (speaker 20151120) illustrates how these mixed structures arise, how the switch occurs within the construction. The switch is flagged, and the speaker indicates that he is going to use a Russian phrase, marking the CS with the expression *ruztne meŕiť* 'the Russians say'. After the insertion, he repeats the possessor, using the Erzya genitive form of the word *kultura* 'culture'.

(89) *viďste* *eŕav-i* *meŕ-ems* ***što*** ***pedučiľišče*** *śe* *ška-ńť*

  correctly have.to-3SG say-INF1 that pedagogical.college that time-GEN.DEF.SG

  uľńe-ś ***ruz**-tne* *meŕ-iť* ***rassadńik rassadńik kultur-i***

be-PST.3SG   Russian-DEF.PL   say-INF1   hotbed   hotbed   culture-GEN.SG

***kultura*-ń**

culture-GEN

'We have to say correctly that the pedagogical college at that time was, as

Russians say, a hotbed, a hotbed of culture, of culture.'

There are also cases in which one of the constituents is an Erzya noun, but the Russian-type possessee–possessor order is applied. Typically, the first element is in Russian and the second in Erzya (cf. example 91), but there are counterexamples as well, cf. example 90 (speaker 20130402) in which the possessee is an Erzya word *śormadićat* 'writers', whereas the possessor *pjesań* 'of the play' is a Russian lexeme:

(90) ***i   vot*** *sval   mon eńald-an* **buťi** *isťa-t*   *śormadića-t* **pjesa**-ń

and   look   always   I   beg-1SG   if   such-PL   writer-PL   play-GEN

uľ-iť   tink   śormado-do

be-3PL   you.PL.GEN   write-IMP.2PL

'And I always beg that if you have such playwrights, write (to us).'

In example 91 (speaker 200511, S2 here), the nominative form of the possessive declension is used instead of the indefinite genitive. To the question of the interviewer, the speaker answers with a Russian-type genitive construction, possibly triggered by the Russian possessee and the model Russian expression *kopija otci* 'copy of the father'.

(91) S1:

ťeťa-nzo   jonov **značit** *son   moľ-i*

father-POSS.3SG.GEN   towards   so   s/he   go-3SG

S2:

kopija   ťeťa-zo

copy   father-POSS.3SG

'So he resembles his father? He is a copy of his father.'

In the next type of constructions, Erzya markers are also attached to the possessee. As the Russian genitive marker is replaced by the Erzya equivalent, and the construction is not inserted as a chunk, an Erzya ending can appear on the first element of the construction. This ending can

be the definite marker (cf. examples 92), a possessive suffix (cf. examples 93 and 94), or a case marker (e.g. the elative in example 95).

In example 92, the reporter (interview 20151113, R3) uses the Russian type genitive construction, and the order of the constituents follows the Russian pattern, even though the possessor is expressed by an Erzya lexeme (*Umaŕina* 'Umarina, the name of an ensemble, lit. apple tree').

(92) *ki* **rukovoďiťeľ**-*eś* *Umaŕina-ńť*

who manager-DEF.SG Umarina-GEN.DEF.SG

'Who is the manager of [the ensemble] Umarina?'

In examples 93 (speaker 20160121) and 94 (speaker 200504) possessive suffixes are attached to the possessee, the third person singular in example 93 (*redakcijazo* 'its editorial office') and the first person singular in example 94 (*učiťeľem* 'my teacher'):

(93) **redakćija**-*zo* **gaźeta**-*ńť* *ašťekšń-eś* *Saranskoj*

editorial.office-POSS.3SG journal-GEN.DEF.SG situate-PST.3SG Saransk

uľića-*ń* **Volodarskoj** *kuro-ń* *kolońgemeń* *ńiľe-će kudo-so*

street-GEN Volodarskoj street-GEN thirty four-th house-INE

'The editorial office of the journal was in Saransk, at 34 Volodarskij Street.'

(94) **śexte** *pek* *uľń-eś* *vadŕa* **učiťeľ**-*em* **maťemaťika**-*ń*

most very be-PST.3SG good teacher-POSS.1SG mathematics-GEN

di *angľa*-*ń* *keľ-eń*

and English-GEN language-GEN

'My best teachers were the mathematics teacher and the teacher of English.'

Example 95 (speaker 201101) is also a hybrid construction: the elative case suffix is attached to the Russian origin word *zdańija* 'building' which is an adapted form of Russian *zdańije* 'building', functioning as a bridge word in the construction. The pronoun *mińek* 'our' is also in a post-head position.

(95) **i** *značit* *tov* **uže** **plańir-ujet-śa** *nana* *ťeke*

and so to.there already plan-3SG-REFL er also

ťe-ste **zdańija**-*sto* **škola**-*ń* *mińek* **stolovoj** *ťese*

this-ELA   building-ELA   school-GEN   we.GEN   canteen here

'And so they are already planning to (make) a canteen from this building of our
school.'

Example 96 (speaker 20130821) is also a clear instance of congruent lexicalization. The Russian construction would require the genitive plural after the head *časť* 'part'. In the hybrid construction, the word *erźaťńe* 'Erzyas' is in plural, but in nominative plural. The construction does not obey the rules of either language, because in the Erzya equivalent (*peľeze* 'its part') of *časť* would require a reversed order and the genitive case (*erźaťńeń peľeze* 'half of the Erzyas').

(96)   ***časť erźa-ťńe*** kadov-śť   ***časť vostok*** jono-v   ajg-śť

part Erzya-PL remain-PST.3PL part East   direction-LAT move-PST.3PL

'Part of the Erzyas remained (there), part (of them) moved to the East.'

The inverse genitive constructions are typical with Russian stems, but there are indications for their use also with Erzya stems. The following occurrence is from the utterance of a C1 speaker. This is an unambiguous case of congruent lexicalization or composite CS, as described by Myers-Scotton (2003: 189). The mechanisms in composite CS are similar to convergence "(i.e. all the surface morphemes come from one variety, but part of the abstract structure comes from another variety)" Myers-Scotton (2003: 190).

(97)   *stuvt-iń*   *ľem-eze   moro-ńť*

forget-PST.1SG   name-3SG.POSS   song-GEN.DEF.SG

'The name of the song is "I forgot".'

Finally, I discuss more complex constructions which contain also determiners or additional attributes. In examples 98, 99 (both by speaker 20160121) and 100 (20130401), the possessor and the possessee both have attributes. These are instances of congruent lexicalization, because the stems are predominantly Russian (except for *kuľa* 'news' in example 99), and the order of the constituents follows the Russian pattern (the Erzya order would be *mińek Mordov'ijań kuľaťńeďe* 'news of our Mordovia'). However, the morphological markers are all Erzya.

(98)   *kavto   ťoža-t   kemgotovo-će   ije-ń   jakšamkov-sto topoď-ś*

two   thousand-PL sixteen-th   year-GEN January-ELA   fill-DEF.SG

kavksońgemeń veťe ije-ť   koda ľiśekšńe-ś   vaśenće

136

twenty   five   year-PL as   come.out-PST.3SG first

nomer*-*eś   ***Ľeńin*-*eń***   *kija-va*   *erźa-ń*   ***gazeta*-*ńť***

number-DEF.SG Lenin-GEN   road-PROL Erzya-GEN journal-GEN.DEF.SG

'In January 2016, 25 years have passed from the publication of the first issue of

the Erzya newspaper named "On Lenin's road".'

(99) ***gaźeta*-*so*-*ńť***   *jovtńe-śť*   <u>*veśe*   *kuľa-ťńe-ďe*</u>

newspaper-INE-DEF.SG tell-PST.3PL all   news-DEF.PL-ABL

<u>*mińek*   ***Mordov'ija*-*ń***</u>

we.GEN   Mordovia-GEN

'In the newspaper they told all the news of our Mordovia.'

(100) ***ŕedaktor*-*oś*** *ťe* ***kńiga*-*ńť***   *mińek*   *sodaviks*   ***poet*-*eś***

editor-DEF.SG this book-GEN.DEF.SG we.GEN   famous   poet-DEF.SG

Ńikolaj Išutkin

Nikolay Ishutkin

'The editor of this book is the famous poet Nikolay Ishutkin.'

The different types of constructions can be modelled as represented in Figure 4. There is a continuum from the Russian embedded language islands through the hybrid constructions to the fully Erzya genitive structures, which have all the lexical and morphological elements from Erzya but whose structure (primarily, the order of constituents) still follows the Russian rules.

Genitive constructions can be assigned into three categories according to the origin of the nominal stems. The other aspect of categorization is the language whose morphological markers are used. Figure 4 illustrates the types attested in my data.

	modifier	stem	suffix	modifier	stem	genitive	suffix	example #
type 1	isťamo	v'id			podďeržk	i		80
		dom			kultur	a	so	82
type 2		aťestaćija			saďik	eńť		88
	veśe	kuľa-ťńe	ďe	mińek	Mordov'ija	ń		99
		rukovoďiťeľ-	eś		Umaŕina	ńť		92
type 3		ľem	eze		moro	ńť		97

Figure 4. Genitive constructions attested in the Erzya–Russian discourse

Grey background in cells indicates a Russian element, white background an Erzya element, and striped background indicates that additional markers are not possible in that slot. In type 1, all the stems are Russian and the Russian genitive marker is used – these are EL island-type insertions. Modifiers and additional markers can be of Erzya or Russian origin. Type 2 represents the hybrid constructions in which either both the possessor and the possessee are Russian lexemes, or one of them is expressed through an Erzya lexeme, while the other is Russian, but the morphological markers are Erzya. In type 3, only the constituent order is Russian, all the elements (both stems, modifiers and endings) are Erzya.

The variation in the realization of the possessive structure indicates an ongoing change in Erzya, CS leading to possible structural change. Due to frequent switches of these nominal constructions with a genitive modifier, the Russian-type constituent order can become widespread even with Erzya stems. The possessee–possessor order has been present in standard monolingual Erzya as well, usually used in poems and songs as a marked order. Due to CS the construction can become less marked. This corroborates Heine's (2008: 56) view that new word order patterns are usually the reinterpretation of an existing but marginal word order pattern that speakers reuse to mirror the construction in the source language.

7.2.3. Verbal constructions

In this section I analyze code-switched verbal constructions occurring in Erzya–Russian bilingual discourse. Using Muysken's (2000) categories, the majority of the code-switched verbal constructions can be categorized as insertions, both EL islands and mixed constructions as defined in the MLF model. In the first case, the Russian element is inserted into the morphosyntactic frame set by the ML with Russian system morphemes, whereas in the second one, the system morphemes come from Erzya. Beside these clearcut cases, however, there are hybrid constructions in which identifying the ML proves impossible. If the structures are congruent in the two languages, the switch can occur at any point in the utterance. In these cases, both languages provide the grammatical frame, and the code-switch can be categorized as congruent lexicalization.

As I discussed in the theoretical section (Chapter 5) of my paper, I do not differentiate between code-switches and borrowings. In my opinion, the distinction cannot be made in any synchronic state of the language solely relying on naturally occurring recorded data. Backus's (2015) usage-based model makes the distinction on the basis of frequency, elicited data and judgment tests carried out among speakers. Further research on Erzya–Russian CS should also employ these types of investigation. In discussing the Erzya–Russian mixed verbal phrases, I suggest possible directions for further analysis of these constructions.

The wide variety of parallel forms makes the distinction between borrowings and CS even more difficult. As all CS types in my data, Russian verbs and verbal constructions in Erzya–Russian bilingual discourse also show variation, and morphologically adapted and nonadapted forms co-exist even in the utterances of the same speakers.

After the discussion of general questions relevant in the study of code-switched verbal constructions (section 7.2.3.1), I focus on necessity (section 7.2.3.2) and gender agreement (section 7.2.3.3).

7.2.3.1. General questions of code-switched verbal constructions

The CS of verbal constructions has not been studied specifically yet in any language. However, there are models concerning the borrowing of verbs, and one of the widely used categorizations is Wichmann and Wohlgemuth's 2005 typology of verbal loans. As mentioned above, the dividing line between CS and borrowing cannot be drawn in the Erzya–Russian data, in my opinion. The typology created for borrowings, therefore, can be used for the categorization of verbal CS types as well. Wichmann and Wohlgemuth (2005: 6) operate with four strategy types:

1. the *light verb strategy*
2. *indirect transfer*
3. *direct transfer*
4. *paradigm transfer*

If we compare these types with the categories of the MLF model, the first three are equivalents of various types of mixed constituents, while the last one is similar to EL islands.

In Erzya–Russian bilingual discourse the two prevailing types of strategies are direct transfer and paradigm transfer, while the light verb strategy is not attestable in Erzya (it is present in Udmurt and typical in the Turkic languages, cf. Kantele 2016). The light verb strategy involves the use of the code-switched verb in the infinitive, while the finite verb is the recipient language's verb with the meaning 'do' or 'make' which serves as an auxiliary or light verb in the construction.

The second type, indirect transfer, involves the use of a thematizing suffix attached to the stem of the foreign verb. Usually this suffix has a verbalizer function in the recipient language. This can be observed in Hungarian where verbs from the source language are adapted by attaching either -*l* or the -*z(ik)* suffix to the stem, e.g. English *to stream* → Hungarian *streamel* 'to stream'. In the Erzya–Russian CS data, some of the verbal constructions might arise as a result of indirect transfer, but these rare cases can also be explained by analogy and considered a subtype of direct transfer (cf. in detail below).

In case of direct transfer, there are no thematizing suffixes attached to the foreign stem, instead, the source language verb stem is applied as any native verb stem. In Erzya–Russian CS, this is one of the most widespread strategies: the Russian *iť* or -*ať* infinitive ending is replaced by

the Erzya infinitive ending *-ams*, e.g. Russian *pomńiť* → Erzya *pomńams* 'to remember' and Russian *rešať* → Erzya *rešams* 'to decide'.

The fourth type of verbal borrowing in Wichmann and Wohlgemuth's (2005: 6) typology is paradigm transfer. In this case the Russian verb is embedded in the utterance with Russian finite endings, involving, for instance, Russian personal or tense markers.

In the following sections, I focus on the direct transfer (section 7.2.3.1.1) and paradigm transfer (section 7.2.3.1.2) of Russian verbs in Erzya–Russian bilingual discourse.

7.2.3.1.1. Direct transfer: inserted mixed verbal constituents

As discussed in the typological overview of the two languages (in Chapter 3), the majority of Russian verbs have two forms: a perfective and an imperfective one. Compare, for example, the imperfective infinitive *rešať* and the perfective infinitive *rešiť* of the Russian verb meaning 'to decide'. In the Erzya–Russian CS discourse, it is typical that the imperfective form of the verb is borrowed or code-switched. The Russian verb stems irrespective of their original Russian stem vowel all fit into the paradigm of *a*-verbs in Erzya. For instance, Russian *otv'ečať* – Erzya *otvečams* 'to answer', Russian *polučať* – Erzya *polučams* 'to get', but also Russian *pomńiť* – Erzya *pomńams* 'to remember', Russian *služiť* – Erzya *služams* 'to serve'. All Russian verbs listed in Keresztes's 2011 overview of Russian borrowings in the Mordvin languages also entered the class of *a*-verbs in Erzya, and all are back-vowel stems (e.g. Erzya *provožams* 'to accompany' from Russian *provožať* 'to accompany'). However, the question needs further investigation involving also a diachronic aspect. (The *a*-verbs had an important role in the history of the Erzya verb paradigm as well, as in the present tense the first and second person singular endings are *-an*, *-at*, respectively for all verb types, irrespective of the stem vowel, cf. section 3.2.2).

These latter cases in which the Russian verb does not have an *-a* in the stem (*pomńiť* 'to remember'), and in which, consequently nor does it have an *-ať* infinitive ending, cannot be considered as pure cases of direct transfer, as here the Erzya *-a* functions as a kind of thematizing suffix, making these, instead, instances of indirect transfer. However, the general pattern of

transfer can also be that 'take the consonant stem of the Russian verb' and insert it to the *a*-paradigm, in which case all the transfers are direct.

It is typical in my data that Russian verbs are transferred with their imperfective forms if they are inserted in the Erzya utterance with Erzya morphological markers. If they are EL islands inserted in the Erzya morphosyntactic frame using Russian morphological markers, both the imperfective and the perfective forms appear. However, there are few exceptions to this rule. In the next example, the perfective form of the Russian verb 'to give' (imperfective (imp.) *daŕiť* – perfective (perf.) *podaŕiť*) is chosen and inserted into the utterance with an Erzya infinitive marker (-*ams*):

(101) ***źorna vedra mog-ut*** **podaŕ**-*ams*
     wheat bushel can-3PL   give-INF1
     'They can also give a bushel of wheat.'

There can be various explanations for this example. It might be triggered by the Russian auxiliary *mogut* 'they can', which is inserted as an EL island into the utterance. In this scenario, the switch occurs within the word *podaŕams*. In this same interview (201103), there are another four instances of this form. Example 101 was uttered by another speaker present at the recording, the following four cases (examples 102–105) by the main speaker with whom the interview was recorded. The speaker in 101 does not pause before uttering the perfective form and does not backtrack, while the speaker in examples 102–105 initiates self-repair in three cases, in two of which she uses the imperfective form of the Russian verb (examples 102 and 103), in the third case (example 104), the Erzya equivalent of the verb, and in the last utterance (example 105), no repair occurs.

(102) ***ńevesta***-*ń* ***rodńa***-*t*   *s*-*iť*   ***žeńih***-*ńe*   *ťe koda* **uš**
     bride-GEN relative-PL come-3PL bridegroom-DAT it how already
     meŕ-*iť* *ťe*   **podaŕa**-*mo*   ***daŕa***-*mo*
     say-3PL that give.as.present-INF2   give-INF2
     'The relatives of the bride come to the bridegroom's place, well, how do they say, well, to give presents, to give.'

(103) **da** surks **da** ťe-ń   vejke-ń   **podaŕa**-*sazo*

yes ring yes that-GEN   one-GEN give.as.present-SG3<SG3

daŕa-*sazo*   **kŕestńica**-ń[]   kak **bojar**avaksči

give-SG3<SG3 goddaughter-DAT as   sign.of.wealth

'Yes, the ring, yes, that one is given as a present, as a sign of wealth to the

goddaughter.'

In example 104, the self-repair status of the correction is ambiguous, as all the people present are talking at the same time, but it is clear that the speaker finally opts for the Erzya form.

(104) moń vara **napŕiḿer-to** źaro   uľńeś-eľť   panar-t ťeńa-t

my   well for.example how.many be-PST2.3PL   shirt-PL something-PL

soročka-tne-ń   **podaŕakšno**-*śť*   da kazekšńe-śť

chemise-DEF.PL-GEN give.as.present-PST.3PL yes give.as.present-PST.3PL

'I, for example, had so many blouses, er, chemises given to me as presents, yes,

given to me as presents.'

The word *da* 'yes' indicates also that the repair was initiated by others and the speaker acknowledged it. The verb form is different from the verb forms in the other examples, because here the frequentative suffix -*kšno*- is added to the stem, as in the Erzya equivalent (in which the allomorph -*kšńe* is used). The ambiguity of the borrowing vs. code-switching status of this verb form can be observed in this example. On the one hand, the form is morphologically adapted, which may indicate its acknowledged status. On the other hand, the switch is repaired, which can be attributed to the word's lower level of entrenchment.

Example 105 is a repetition of the utterance in example 103, but this time it lacks the repair, and the perfective stem is the only one used as a predicate.

(105) ťe **podaŕa**-*sazo*   ťe   kak **bojar**avaksči   ťe   **krestńica**-*ńste*

it give-SG3<SG3   it   as sign.of.wealth it   goddaughter-DAT

'It [=the ring] is given as a sign a wealth to the goddaughter.'

Unfortunately, these five forms are not sufficient to decide on the status and entrenchment of the perfective form of the verb, especially considering that they all occurred in the same interview. It is not possible to make such distinctions on the basis of corpus data. As a

next step in the analysis of the transfer of the imperfective versus the perfective forms of Russian verb, we would need elicited data and also judgement tests with speakers to establish how entrenched the perfective forms are. However, it can be a telling sign that the second speaker (examples 102–103), who is in her sixties, typically initiates repair, while the other speaker (example 101), who is in her twenties, does not. It is possible that the younger generation feels more comfortable with the transfer of perfective forms as well, but this requires further investigation, preferably based on elicited and quantitative data.

Apart from the above mentioned counterexamples, it is typically the imperfective stem of the Russian verbs that occurs in mixed Erzya–Russian constituents, i.e. in cases in which the Russian stem is inserted into the utterance using Erzya verbal markers. As a result, the verb form pair imp. *otv'ečať* – perf. *otv'eťiť* 'to answer' will occur in the from *otvečams,* and the form **otveťams* cannot be attested.

It also seems to be a tendency that the stem of the directly transferred Russian verbs all contain back vocalism, and thus they fit into the Erzya *a*-paradigm (cf. the Erzya vowel harmony rules described in section 3.1). Although Russian stems could also fit into the *o*-stem class (e.g. *lovnoms* 'to read') on the basis of their back vocalism, they all enter the *a*-type, which is also historically the type used for the incorporation of loan verbs. In my data, there is one counterexample, the verb *rešams* 'to decide', which due to the vowel harmony rules should have entered the *e*-type of Erzya verbs as **rešems*.

Finally, in my data, there is a striking tendency for back vowel stems to be transferred directly, while Russian verbs with front vowel stems are inserted as EL islands with Russian verbal markers (e.g. Russian *viďeť* 'to see'). Of course, this might be a coincidence, but in my opinion this should be studied further and on more extensive data in order to reveal whether vowel harmony truly plays a role in the "decision" whether a verb is code-switched as an EL island or as a mixed construction in the utterance.

If we study from the point of view of semantics which types of verbs are code-switched as stems and inserted in the utterance frame using Erzya endings, we can find cognitive verbs and verbs connected to specific domains in which Russian is more dominant. For instance, verbs with general (cognitive) meanings are *dumams* 'to think', *otvečams* 'to answer', *polučams* 'to get',

pomńams 'to remember', and *pomogams* 'to help'. Verbs belonging to specific domains (army, education) with Russian dominanance are e.g. *služams* ('to serve in the army') or *postupams* ('to apply (to the university)'). In example 106, the use of the verb *polučams* 'to get' is illustrated. This is the only Russian origin element in the utterance:

(106) *źaro*   *miń **poluča**-tanok*   *ťe*   *jarmak-oń*   *langs*   *kuvať*

      how.many   1PL get-1PL   that money-GEN   on   for.a.long.time

      a   *eŕa-t*

      not   live-2SG

      'The money we get, you cannot live on that for a long time.'

There are some unambiguous cases which can already be considered established borrowings rather than code-switches. They involve verbs to which derivative verbal suffixes are added. For instance, the Russian verb *zvońiť* 'to call (on the phone)' is transferred to Erzya as *zvońams* 'to call'. In example 107, it occurs with the frequentative suffix *-ś-*:

(107) *mon*   ***zvoń**-ś-an*   *ńe-ť*   *lomań-ťńe-ńeń*

      1SG   call-FREQ-1SG   that-PL   person-DEF.PL-DAT

      'I often call these people.'

Finally, parallel constructions have to be mentioned. These structures do not involve double marking, as attested in prepositional phrases and comparative adjectives in the Erzya–Russian corpus, but are, instead, parallel structures in which both the Russian(-origin) and the Erzya verb forms are used.

There are verbs which can occur as two different types of insertions in the same discourse, both as direct transfer and as paradigm transfer, i.e. as mixed constituents and also as embedded language islands. Verbs having a more general meaning typically exhibit this type of variation, such as Russian *pomńiť* and Erzya *pomńams* 'to remember'. In example 108 (speaker 200507), the verb is inserted as a direct transfer with Erzya verbal markers, whereas in example 109 (also speaker 200507) it occurs with Russian verbal markers.

(108) ***ďeda**-m-gak*   ***pomń**a-sa*

      grandfather-POSS.1SG-EMF   remember-SG3<SG1

      'I also remember my grandfather.'

(109) śe   ije-ste   **_pomń-u_**   mińek   kudo-ńťeń   sov-ińek

that   year-ELA   remember-1SG we.GEN house-ALL.DEF.SG enter-PST.1PL

'I remember we moved to our house the same year.'

As discussed in the theoretical part of my paper (in Chapter 5), there are code-switched constructions the emergence of which cannot be explained relying purely on structural grounds. In most of the cases, the analysis has to take into account a variety of factors at the same time. For the analysis of parallel forms, one also has to see beyond the morphosyntactic frame of the utterance. In example 110 (speaker 20151118), there are two predicates, two synonymous verbs, both inserted into the frame using Erzya verbal suffixes. The first one is the Russian origin element followed by its Erzya equivalent. In parallel constructions in my data, the Russian form is usually the first to which its Erzya counterpart is also added in most cases. The use of the two forms can be explained by the speakers' wish to emphasize what they are saying. The Russian–Erzya order in the construction also suggests that the speaker tried to go back on her words and use the Erzya equivalent instead of the Russian origin one, which can be considered marked in formal situations due to the purism of Erzya intellectuals. (I analyze these types of flagged switches in secion 7.3.)

(110) valske   marto **_i_** čokšńe   **_zvoń_**-_iť qajńevť-iť_ eŕva   ška-sto

morning   with   and evening   call-3PL call-3PL   every   time-ELA

'They call us in the morning, in the evening, all the time.'

A stylistic explanation seems also plausible in example 110, i.e. the speaker emphasizes the intensity of the calls by using the synonym. This stylistic function of bilingual double forms can be observed in coordinate compounds as well, which can contain only Erzya synonyms (*jalgat-ojat* 'friends', both *jalga* and *oja* meaning 'friend' in Erzya, and both parts have the plural suffix, lit. 'friends-friends'; cf. Grammatika 1962: 57) or counterparts from two languages juxtaposed in the compound (*ladt-kojť* 'ways, customs', from the Russian *lad* 'manner, way' and the Erzya *koj* 'custom, way'; both parts have the plural suffix, or *pťicat-narmuńť* 'every kinds of birds' from the Russian *pťica* 'bird' and the Erzya equivalent *narmuń* 'bird' both in the plural, the latter example from Kelin 1969: 119).

In example 111 (speaker 20151113), the parallel constructions involve a finite Russian verb form and its Erzya equivalent. This switch to the Russian predicate may be triggered by the Russian prepositional phrase (*do P'etrovska* 'to Petrovsk') preceding it.

(111) **do P'etrovsk-a** <u>*saď-im*</u>*oza-tano* *Sińinka-so ard-tano* **do P'etrovsk-a**

to Petrovsk-GEN.SG sit-1PL sit-1PL Sininka-INE travel-1PL to Petrovsk-GEN.SG

'To (travel to) Petrovsk, we get (on the bus) in Sininka and travel to Petrovsk.'

7.2.3.1.2. Paradigm transfer: inserted Embedded Language islands

In this section, I discuss EL island predicates inserted into the utterance with Russian morphological markers. If we apply a diachronic perspective, we can find a connection between CS and borrowing in case of paradigm transfer: frequent insertion of the source language verbs as EL islands can result in the borrowing of the verb's entire paradigm.

As I mentioned in Chapter 4, journals, textbooks, but also literary works published in the late 1940s and 1950s show instances of paradigm transfer, i.e. the use of Russian verbs in otherwise exclusively Erzya-only utterances, with the verb *moč'* 'to be able to' being most attested in these publications. Hallap (1960: 221–222) also refers to this phenomenon, although not with a descriptive intention but rather disparagingly (cf. more on this question in Chapter 4 above). For example, there are many examples of paradigm transfer in the art almanac *Iznjamo* 'Victory' published in 1949 in Saransk, which contains 79 instances of the predicative adjective *možna* 'it is possible' or its form marked for past tense *možnaľ* 'it was possible' (the ending *ľ* is the Erzya past tense suffix marking third person forms of the predicative conjugation, the form is a hybrid, not grammatical in Russian, cf. also example 121), and 13 different forms of the verb *moč'* 'to be able to'. In example 112, there is an entirely Erzya sentence into which the finite form of the Russian verb *možet* 'he/she/it can' is inserted with a Russian verbal suffix (1949: 87).

(112) *Di eŕvaś aŕśeś, meźe son* **možet** *maksomo ťe od ťeveńťeń.*

di   *eŕva-ś*   *aŕśe-ś*   *meźe son* **mož-et** *maks-omo*

and each-DET.SG think-PST.3SG what (s)he can-3SG give-INF2

ťe   *od*   *ťev-eńťeń*

that   new   work-DAT.DEF.SG

'And each of them were thinking what they could contribute to this new task.'

In the Erzya–Russian CS discourse the use of the verb *moč'* 'to be able to' is widespread, its forms are attested both in the fieldwork materials and in the interviews recorded from the Radio Vaygel. The verb is mainly used in the third person singular form *možet* 's/he/it can', which can be explained by the fact that this form can be used as an adverb as well, as a shorter form of the construction *možet biť* 'it is possible'. In example 113 (speaker 200502), the adverb-like function of *možet* can be observed.

(113) *seďe tov   a   sod-an   **mož-et** kostojak*

more to.there   not know-1SG can-3SG from.somewhere

*ľija   **robota** muj-an*

other   job   find-1SG

'I do not know further. It is possible that I'll find another job somewhere.'

In the following examples, the verb *moč'* 'to be able to' is the predicate of the utterance: it is used in third person singular (examples 114 and 115), in first person singular (example 116), and third person plural (example 117).

Examples 114 (speaker 201001) and 115 (speaker 20140305) differ as regards the argument of the verb, where, in the first case, the first infinitive (*ťejems* 'to do'), and in the second example the second type of infinitive (*kortamo* 'to speak') is used.

(114) ***učťeľńića**-jak   meźejak **ńe mož-et**   ťej-ems   **vojna** ťeťa   araś*

teacher.F-EMF   nothing not can-3SG do-INF1   war   father there.is.not

'Even the teacher cannot do anything, there is war, the father is away.'

(115) *son **mož-et** korta-mo   **hoť**   meźe **no**   ťe   **jasno***

s/he can-3SG speak-INF2 any   what but   that   clear

***ef'ir**-s   ťeń   **uže**   a   eŕav-i   maks-oms*

broadcast-ILL   that.GEN already not   have.to-3SG give-INF1

'He can say anything, but it is clear that we should not broadcast it.'

The use of the second infinitive in example 115 contradicts Bartens's (1999: 149) description of the cases the second infinitive is used in. According to her, the second infinitive is

typically used with verbs of motion or verbs expressing transformation which require a lative nominal argument.

In my data, the first infinitive was generally used if the infinitive was the argument of the verb *moč'* 'to be able to'. The use of the second infinitive could be considered an idiosyncratic phenomenon, but in the example from the 1949 publication (example 112) the second infinitive was used as the argument of the verb. On the basis of this, I suggest that future investigations should also focus on the distribution of the first and second infinitive with code-switched Russian verbs.

In example 116 (speaker 200803), the verb *moč'* 'to be able to' is used in the first person singular and the first infinitive is its argument.

(116) *mon* **_mog-u_** **_daže_** **_fotograf'ija_** _ńevť-ems_

I   can-1SG   even   photograph   show-INF1

'I can even show you a photograph.'

In example 117 (speaker 20130801), the negated form of the verb is used. In the Erzya–Russian CS discourse there are no cases in which the switch would occur between the negative particle and the verb, since they form a chunk.

(117) *siń* **_prosto ľež-at_** *siń*   *meźejak* **_ńe mog-ut_** *eś*   *melga-st ťej-ems*

they   just   lie-3PL   they   nothing   not can-3PL own   after-3PL do-INF1

'They are just lying around, they cannot do anything for themselves.'

It is a general tendency in the corpus that there is no switching between the negative particle and the negated nominal, or between the negative verb and the connegative form. There are no examples in the corpus which would violate this constraint, so it is probably safe to identify this as a strong tendency.

In example 118 (speaker 20140305), the first clause contains two negated Erzya predicates with the past tense negative verb and its connegative form (*eźiń juta* 'I did not leave' and *eźiń moľe* 'I did not go'), while the second clause contains a Russian negative particle with a Russian finite verbal predicate *ńe hoťela* 'I did not want to'.

(118) *mon eź-iń   juta   eź-iń   moľe   kovgak*

      I   not.PST-1SG   leave.CONNEG   not.PST-1SG go.CONNEG   anywhere

      potomu čto ńe   hoťe-l-a *sonze   marto jav-oms*

      for   that not want-PST-F (s)he-GEN with   part-INF1
      'I did not leave, I did not go, because I did not want to part with her.'

The only instance in which a mixed construction can be found in my corpus is an utterance with an Erzya predicate *araś* 'there is not', in example 119 (speaker 201006).

(119) *meźejak* **_inťeŕesn-ogo_**   *araś*

      nothing   interesting-GEN.SG   there.is.not

      'There is nothing interesting.'

This is an instance of congruent lexicalization, as the morphosyntactic frame is set by two languages. In Russian, negated existential sentences require the subject to be in the genitive case, which expresses a partitive function. Consequently, the construction 'there isn't anything interesting' would be formed in Standard Russian as *ńičego inťeŕesnogo ńet*. In contrast, the subject of existential sentences is in the nominative in Erzya, and the Standard Erzya form is *meźejak inťeŕesnoj araś.* In example 119, a hybrid construction is used, as the negative predicate (*araś* 'there is not') is in Erzya, the Russian subject (*inťeŕesnogo* 'interesting') is in the genitive, and the Erzya modifier of the subject (*meźejak* 'nothing') is in the nominative (although the ablative case marker could have been used to express the partitive function *meźďejak*).

This structural transfer must be seen as an idiosyncratic phenomenon, as there are no other examples for use of the partitive-genitive in negated sentences with an Erzya predicate, and negation overall shows a homogenous picture in Erzya–Russian CS discourse, with restrictions on phrase-internal switching.

In case of paradigm transfer, one could argue that the ML of the utterance changes if the main verb of the utterance is a finite Russian verb form. These are usually instances of truly mixed sentences and in most cases it is impossible to decide which language provides the morphosyntactic frame of the utterance, as the argument structures of the two verbs are identical. In example 120 (speaker 20130429), the verb is a finite Russian verb form *iḿel* 'had' which is a transitive verb and would require its object to be in the accusative case, so instead of

v'era 'faith' in the nominative, the accusative form *v'eru* would abide by the rules of Standard Russian.

(120) **každoj** *lomań-eś* **iḿe-l** **v'era** *paz-ońťeń*

every person-DEF.SG have-PST.M faith god-DAT.DEF.SG

jesľi *lomań-eś* **v'eŕ-it** *paz-oś* *sval* **posob**-*i*

if person-DET-SG believe-3SG god-DEF.SG always help-3SG

maks-i *vij*

give-3SG strength

'That every person had faith in god. If you are a believer, God always helps you and gives you strength.'

In this case, the Erzya and the Russian constructions are not congruent. In Russian, there are two ways the possessive construction can be formed: one being the use of the verb *iḿeť* 'to have' with an accusative object, or the more typical construction in which the possessor is expressed with the preposition *u* 'at, by' and the complement in the genitive case + (the copula) + the possessee in the nominative case. For instance, *u čelov'eka* + (*jesť*) + *v'era* 'the person has faith'. In Erzya, however, there are no 'have'-type verbs, similarly to most Finno-Ugric languages, but the following construction is used to express possession: the possessor in the genitive case + the copula + the possessee in the nominative case with a possessive suffix (cf. section 3.2.3). As we can see, the possessee is in the nominative case in both languages in the constructions which are congruent in Erzya and Russian.

In example 120, if the predicate determined the structure of the utterance, the finite verb would require the object to be in the accusative case. Here we can see a combination of the possessive construction with the copula and the other construction containing the verb *iḿeť* 'to have'. As a result, it is impossible to determine which language is the base language, and which language provides the morphosyntactic frame.

The use of the transitive verb *iḿeť* 'to have' is not typical in these types of constructions even in Russian, but in idioms with abstract nouns such usage is possible (Timberlake 2004: 311–312). However, it is used with the noun *v'era* 'faith' mostly in the Bible. As the interviewee was an Orthodox bishop, it is possible that he was citing the New Testament (Matthew 17:20: *если*

вы будете иметь веру с горчичное зерно, Russian Bible; 'if you have faith like a grain of mustard seed', Biblehub), and, as a result, it is possible that the whole construction is a loan translation from Greek.[10]

Another instance of hybrid constructions is a mixed constituent which contains an EL island *ńeľźa* 'forbidden', but it is inserted into the utterance using a ML marker, the Erzya second (habitual) past tense suffix *-ľ* in example 121. In Standard Russian, no morphological markers can be added to *ńeľźa* (its part of speech is debated, it is considered an adverb used as a predicate by some and a predicative adverb by others), and the past tense form would involve an analytic construction in which the copula would have the past tense marker *ńeľźa bilo* 'it was forbidden'.

(121) **ńeľźa-ľ**   *pečke-ms*   *tuvo*

forbidden-PST2.3SG   slaughter-INF1 pig

'It was forbidden to slaughter pigs.'

There is only one instance of the use of the habitual with a Russian predicate, and it occurs in an elderly speaker's language use (201006). Elicited data could reveal more about the scope of this phenomenon and also about the possible generational differences speakers show in the use of these forms. In the corpus there is one more instance of this construction, but it mirrors the Russian analytic construction. Instead of the Russian copula the Erzya equivalent is used in example 122 (S1: speaker 200503, S2 is the interviewer).

(122) S1: *mińek*   *veľe-ńťeń*   *rudazškańe* **ńeľźa**   *uľńe-ś*

we.GEN   village-DAT.DEF.SG   dirty.weather impossible   be-PST.3SG

S2: *sov-ams*

enter-INF1

'It was impossible to go into our village when we had mud due to the weather.'

Besides these generally attested cases, there are verbs which have a variety of code-switched forms. Russian reflexive verbs are usually code-switched with Russian morphological markers in Erzya–Russian bilingual discourse, which can be explained by the fact that the reflexive verbs are formed differently in the two languages. While both Erzya and Russian have reflexive suffixes, their position is different: in Russian the reflexive suffix *-śa/-ś* occurs at the end of the

---

[10] I have to thank Ágnes Felföldi and Szabolcs Janurik for this suggestion.

verb form as the final morpheme, whereas in Erzya the reflexive suffix *-v* is attached to the stem of the verb and can be followed by the personal or time markers or the infinitive's derivative suffix, etc. (e.g. Erzya *kadoms* 'to leave [somebody]' *kadovoms* 'to stay, to remain'). Due to this discrepancy, Russian reflexive verbs are inserted in the utterance as one unit with Russian grammatical markers, as in example 123 (guest speaker at the interview 201103), the verb *razľičajutśa* 'differ' or the verb *usvajivajetśa* 'to be acquired' in example 124 (speaker 20130813).

(123)   *tatar marto* **nav'erno**   *mari*   **očeń**   **razľiča-jut-śa**

Tatar with   probably   Mari   very   differ-3PL-REFL

'The Mari language is probably very different from Tatar.'

(124)   **veśe**   **usvajiva-jet-śa**   *ejkakšpinge-ste*   *śeďe*   *šoždińeste*

everything acquire-3SG-REFL childhood-ELA   more   easily

'Everything can be acquired more easily in childhood.'

Besides these typical cases, there is one reflexive verb which is code-switched in a variety of ways in my data. The Russian reflexive verb *staraťśa* 'to try, to strive' can occur as an EL island, just like all the other reflexive verbs, cf. example 125 (speaker 20130429):

(125)   *od*   *loma-ťńe*   *pek*   *lamo karma-śť*   *jaka-mo*   **čerkova**-*v*

young person-DEF.PL very lot   begin-PST.3PL go-INF2   church-LAT

kona-ťńe   **stara-jut-śa**   **sobľud-ať**   *paz-oń*   *val-ońť*

who-DEF.PL   try-PL3-REFL   follow-INF   god-GEN   word-GEN.DEF.SG

'The young people have started to go to church a lot, the ones who try to follow God's words.'

In contrast with the other reflexive verbs, the Russian reflexive verb *staraťśa* 'to try, to strive' can be inserted into the utterance with Erzya morphological markers. The most frequently attested form in the corpus is the first person plural *starajatano* 'we try'. This verb contradicts the general patterns of verb insertion discussed above. On the one hand, the reflexive suffix *-śa* is deleted. On the other hand, the form of the stem also raises important questions. Relying on the other examples presented above, the logical form of the verb stem would be *stara-* and the Erzya infinitive *starams*. In the present case, however, the stem is *staraja-*, which form does not

occur in the original Russian paradigm at all. Moreover, the Russian past tense forms do not even have the *-j-* in the stem (cf. Russian *staralśa* 'he tried').

In Erzya, the stem is always *staraja-*, and the verb occurred in many speakers' utterances, both in the present (example 126, speaker 20130813) and in the past tense (example 127, speaker 200802).

(126) *meźe   mińek   ejste   **zav'iś-it**   miń   ťeń   **staraja**-tano ťej-ems*
what   we.GEN from   depend-3SG   we   that.GEN try-1PL   do-INF1
'What depends on us, we try to do.'

(127) *ťeťa-m   eś-eńek   **staraja**-ś   **tože**   kirďe-me   **skromnoj**-ste*
father-1SG   self-POSS.1PL   try-PST.3SG   also   keep-INF2   modest-ELA
'My father tried to keep us modest.'

I do not focus on this verb further, but I want to mention two other uses (both from written Erzya) which might help define the scope of usage for this particular verb. Future research could focus on the entrenchment of this verb throughout the speech community.

The two written instances of the verb are very far from each other timewise. Example 128 is a sentence attested on the Internet, at the Russian social media network *Vkontakte* ('in contact'), in the group called *Mordovija–Mordva,* in which people discuss topics related to the Mordvin Republic and its peoples. The members of the group can upload pictures of things and places Mordovia is famous for. One of the photos is taken of a bottle of vodka produced locally, and among the Russian comments, there is also an Erzya-language one criticizing the excessive drinking habit of Erzyas (Vkontakte 2010):

(128) *moń veľe-se   źaro   eŕźa-t   kulo-śť   śimma-do*
my   village-INE how.many   Erzya-PL   die-PST.3PL   drinking-ABL
*meź-ďe   **veď**   **toľko**   a   śim-iť*
what-ABL   after.all   just   not   drink-3PL
*mońś   **staraj**-an koda   bu   a   śim-ems   zńardo   karmavt-iť*
I.myself   try-1SG   how   some   not drink-INF1   when   force-3PL
'In my village so many Erzyas have died of drinking, after all what do they not drink.

154

I myself try to somehow not drink when they force me to.'

It is also worth considering that the verb is an established borrowing as it can be found in earlier publications. For example, in the above mentioned art almanac *Iznjamo* 'Victory' there are six instances of the verb used in a variety of forms. Example 129 gives an idea about the book's topics and atmosphere, which are hardly surprising since the year the book was published in 1949 (p. 204):

(129) *miń **robota**-do-ńť   a   peľ-ďano eś-eńek   turtov*

        we work-ABL-DEF.SG not fear-1PL self-POSS.1PL for

        ***rodǐna**-ńť   turtov   **staraja**-tano*

        homeland-GEN.DEF.SG   for   strive-1PL

        'We are not afraid of work. We strive for ourselves and for our homeland.'

This section has focused on special types of Russian verbs inserted as EL island into the Erzya utterance, namely, *moč'* 'to be able to', possessive construction, and reflexive verbs. Further research into code-switched verbal constructions could focus on the insertion of transitive verbs and object marking in mixed utterances.

7.2.3.2. Necessity

The insertion of predicates as EL islands into the utterance can be considered a strategy of neutralization, as discussed by Sebba (2009) (cf. section 5.3.4). While the congruent necessive constructions can be harmonized (section 7.2.3.2.1), the necessive constructions with the adjective *dolžen* 'has to' and the infinitive are incongruent, as they do not have an equivalent in Erzya. Consequently, they are instances of neutralization.

In the Erzya–Russian data, constructions with the adjective *dolžen* 'has to' and the infinitive are typically mixed: the switch occurs between the adjective and the infinitive, therefore, they cannot be considered instances of alternation.

7.2.3.2.1. Congruent necessive constructions

Necessity can be expressed in different ways in the two languages (cf. section 3.2.4). In Erzya the expression of obligation and necessity always requires the Experiencer to be in the genitive case: the predicate is either the third person singular finite form *eŕavi* 'must, have to' or the *ma*-form of the verb which is designated in the different Erzya grammars with various terms such as the passive necessive participle (Zaicz 1998: 205) or the infinitive-participle (Bartens 1999: 151). The two constructions are represented in the same turn by one of the radio reporters in example 130 (interview 20130408, R1).

(130) *źardo **ṕijut**-so   tŕaź   ejkakš-ońťeń   topoď-iť*
       when orphanage-INE brought.up child-DAT.DEF.SG   fill-3PL

       kemkavkso   ije-ť   tińenk   eŕav-i   sonze   nolda-ms
       eighteen   year-PL2   you.PL.DAT   must-3SG   s/he.GEN   let.go-INF1

       kov   tujema   ejkakš-ońťeń
       where.to going   child-DAT.DEF.SG

       'When a child brought up in the orphanage turns 18 years old, you have to let
       them go. Where should a child go?'

In the first part of example 130, the Experiencer is *tińenk* 'to you', the pronoun *tiń* 'you (plural)' in dative, and the predicate is *eŕavi* 'must, have to', while in the second utterance, necessity is expressed by the non-finite form *tujema* 'going', and the Experiencer is the definite dative form of the noun *ejkakš* 'child'.

The Russian equivalent of the Erzya *eŕavi*-type 'must, have to' construction also involves an Experiencer in the dative form and the non-verbal predicate *nado* 'it is needed'. In example 131 (speaker 20130712), there is no Experiencer in the sentence, and a switch occurs between the predicate and its argument, the Erzya infinitive *śukońams* 'to bow'.

(131) ***vot nada**   śukoń-ams eŕźa   raśke-nsteń   **čto** isťamo*
       look it.is.needed bow-INF1   Erzya   people-POSS.3SG.DAT   that like.this

       mazij   vadŕa ṕŕevej pokšči   kad-ś   mińek   isťamo
       beautiful good   smart celebration leave-PST.3SG we.GEN like.this

       koj   kirda

156

custom custom

'The Erzya people have to be thanked for transmitting to us such a beautiful,

good and smart celebration, for leaving us with this kind of custom, custom.'

Although this construction is congruent with its Erzya counterpart involving the Erzya predicate *eŕavi* 'must, have to', and switching would be possible within the construction, example 131 is the only case in the corpus in which the predicate adjective *nado* is used. In Erzya–Russian CS, necessity is rather expressed with the Erzya verb *eŕavi* 'must, have to' or with the Russian adjective *dolžen* (section 7.2.3.2.2).

7.2.3.2.2. Constructions with the adjective *dolžen* and the first infinitive

In addition to the reflexive constructions analyzed in section 7.2.3.1.2, another instance of neutralization concerns the insertion of the Russian predicate *dolžen* 'must' into the Erzya utterance. The predicate adjective expresses necessity and has four forms: the singular masculine (*dolžen*), the singular feminine (*dolžna*), the singular neutral (*dolžno*), and the plural form (*dolžni*) which is used with all genders.

Sarhimaa (1999) studied this construction (the Duty and Obligation Construction, or DOC, in her terms) thoroughly in Karelian, and I follow her methods of description in this brief overview of how the construction is used in Erzya–Russian CS discourse. Sarhimaa (1999: 118) discusses the elements of the DOC construction: the predicate *dolžen/dolžna/dolžno/dolžni* + an Experiencer in the nominative case + an infinitive (intransitive or with a Target complement). In example 132 (speaker 20141111), the Experiencer is the *kvaľif'ićirovannij staž marto lomať* 'qualified and experienced people', the predicate *dolžni* 'must' is in plural as it is congruent with the Experiencer in plural, the infinitive is a transitive verb *tonavtoms* 'to teach' and it has an object in the accusative case, the Target *ťe ṕedḿeteńť* 'this subject'.

(132) *ťe*   **ṕedḿet**-*eńť*   *tonavt-oms*   **dolžn-i**   **kvaľif'ićirovannij**

that subject-GEN.DEF.SG teach-INF1   must-PL   qualified.M

staž   *marto loma-ť*

tenure with person-PL

'This subject has to be taught by qualified and experienced people.'

There can be also adjuncts in the utterance: these are non-compulsory elements, not required by the argument structure of the predicate. For instance, example 133 (speaker 20130419a) contains a Locative adjunct *každoj školaso* 'in every school':

(133) ***každoj škola**-so* ***dolž-en muźej*** *uľ-ems*

      every school-INE must-M   museum be-INF1

      'In every school there has to be a museum.'

In contrast with the hybrid form in example 121, in which the predicative adverb *ńeľźa* 'forbidden' was inserted as an EL island into the utterance but was accommodated to its grammatical frame by an Erzya second past tense marker, in example 134 (speaker 20140305), the past tense form of the copula (*bili* 'were') in the necessive construction is not a hybrid but follows the rules of Standard Russian:

(134) ***ŕećenźija**-t* *sav-ś*   *ťej-ems* ***napŕiḿer***

      review-PL   have.to-PST.3SG do-INF1   for.example

      šesť čas-ov   ***radio**-s(o)* ***načina-jut*** *korta-mo*

      six   hour-GEN.PL radio-INE begin-3PL talk-INF2

      miń ***po***   ***očeŕeď-i***   <u>***dolžn-i*** ***bi-l-i***</u>   *kunsol-oms*

      we   according.to row-DAT.SG   must-PL be-PST-PL listen-INF1

      i   *jovtń-ems* *eś-eť*   *meľ-eť*   ***za*** ***ńeďeľ-u***

      and   tell-INF1 own-POSS.2SG opinion-POSS.2SG for week-ACC

      meźe   *koda*   *toso*

      what   how   there

      'Reviews had to be written. For example, they started to talk on the radio at six,

      we had to listen to it taking turns and tell our opinion in one week

      what was (in the interview) and how.'

A hypothetical hybrid form would involve the use of the second (habitual) past tense suffix, which is also the past tense marker in the nominal conjugation (discussed in section 3.2.1.4) and would look like as follows: *dolžniľińek* 'we had to'. However, no such construction is attested in the corpus, not even in utterances by elderly speakers.

As mentioned above, no mixed constituents occur in the case of negation either. In my corpus, it is a general tendency that no switching is attested between the negative particle and the predicate or the auxiliary negative verb and the connegative form. In example 135 (speaker 20130924), the predicative adjective and the negative particle are inserted as a chunk into the utterance. The arguments (the Experiencer *miń* 'we' and the infinitive *kadoms* 'leave') are both in Erzya:

(135) *miń* **_ńe dolžn-i_** _kad-oms_   *sonze eŕav-i*   *soka-ms*   **_i_**   **_viď-ems_**
        we not must-PL leave-INF1 it.ACC must-3SG plough-INF1 and   sow-INF1
        'We should not neglect it, we must plough and sow it.'

As regards negation, code-switched forms are homogenous. There is, however, variation in the choice of the language in which the infinitive, the complement of the predicate adjective, is expressed in. As seen above, a switch is possible between the predicate and the infinitive.

In examples 136 (speaker 201103) and 137 (speaker 20141111), an Erzya infinitive (*ramams* 'to buy' and *tonavtńems* 'to learn, to study', respectively) is the complement of the Russian adjective *dolžen*.

(136) *śe kolko-ś*   **_dolž-en_** _rama-ms_ **_vśe_**   **_čemodan_**-*ś-gak*
        that godfather-DEF.SG must-M buy-INF1 whole suitcase-DEF.SG-EMF
        'The godfather has to buy the whole suitcase as well.'

(137) **_vrač_**-*oś*   **_dolž-en_**   *tonavtń-ems*
        doctor-DEF.SG   must-M   learn-INF1

        koda lomań-eń   *marto* **_pravilno_**   *kort-ams*
        how   person-GEN with   right   talk-INF1
        'A doctor must learn how to talk right to people.'

Even though the infinitive is also a Russian code-switch, it is not inserted into the utterance as one unit with the adjective, and other elements might also occur between them. In example 138 (speaker 20130801), the Erzya pronominal object *sinst* ('them') is inserted between the adjective (*dolžni* 'they must') and its complement (*ob'eśṕečivať* 'to supply').

(138) *miń*   *veśe* **_dolž-ni_**   *sinst*   **_ob'espeč-ivať_** **_vśem_**
        we   all   must-PL they.GEN   supply-INF   everything-INS.SG

159

'We must supply them with everything.'

In example 138, both the infinitive *ob'eśṕečivať* 'to supply' and its complement *vśem* 'with everything' follow the rules of Standard Russian. Example 139 (speaker 20130606b), however, presents a mixed construction, in which the word *vńimańija* 'attention' possibly triggers the switch to the Russian infinitive.

(139) ***jesli*** *ton* *saj-at* *eśteť* *rakša* *ton* **_dolž-en_** *ejste-nze*

if   you   get-2SG   for.yourself   animal   you   must-M for-POSS.3SG

pokš **_vńimańi-ja_**   **_uďeľ-ať_**

big   attention(-GEN)   devote-INF

'If you get yourself an animal, you have to devote a lot of attention to it.'

The form *vńimańija* 'attention' is a bridge word which can be analyzed as a (Russian origin) Erzya lexeme in the nominative case or as a Russian word in the genitive case, and it provides an equivalence point for CS. In Standard Russian, the nominative form *vńimańije* 'attention' would be required.

In this section, I discussed the variation attested in one type of necessive construction and the tendencies the analysis of these bilingual utterances revealed. In addition to the choice which language provides the complements and adjuncts in these constructions, the question of gender agreement would be worth investigating further. Nowadays official publications, especially literary works, avoid Russian constructions, whereas earlier documents, even textbooks and art journals published after the Stalinist terror in the late 1930s, as a result of Russification (see section 4.2 for details), swarmed with Russian elements, including the necessive construction. In the *Iznjamo* almanac mentioned above, there are eight instances of the predicative adjective, seven occurrences of *dolžen,* and one of *dolžni*. Hallap (1960) also finds the construction in written sources from 1953. These recorded instances of Erzya–Russian CS suggest that the entrenchment of the necessive construction with the adjective *dolžen* started in the mid-20th century at the latest.

Russian-type necessive constructions can involve gender agreement. These predicates, however, are not so frequent in the data that clear tendencies could be observed concerning their use. I study the attested examples in section 7.2.3.3.1, and supplement these findings with the

analysis of gender agreement in code-switched verbal predicates (section 7.2.3.3.2). In contrast with constructions with adjectives, gender agreement in the past tense forms of verbal predicates is more widespread.

7.2.3.3. Gender agreement

As is typical among Finno-Ugric languages, Erzya does not have gender as a grammatical category. However, in code-switched Russian elements, both with adjectives (inserted as attributives or as predicates) and past tense verb forms, gender agreement has been attested. In this section, I investigate possible patterns of its occurrence in Erzya–Russian CS discourse.

One can argue that gender agreement with predicative adjectives should be discussed in the section on nominal constructions. However, as they also involve gender agreement, which is a prevailing phenomenon with verbal predicates, I analyze them in this section. Moreover, the dividing line between nominals and verbs is not so rigid in Erzya (cf. the nominal conjugation in section 3.2.1.4).

7.2.3.3.1. Gender agreement with predicative adjectives

In this section, I analyze gender agreement attested with code-switched Russian predicative adjectives. Adjectives can also show gender agreement in the attributive position, but these examples typically occur in nominal phrases inserted as chunks, and there are no instances in which an Erzya noun would have a Russian attribute agreeing in gender with the Russian equivalent of the Erzya noun. Consider example 140 (speaker 200501), in which the Erzya and the Russian constructions are used in parallel in the utterance.

(140) ***žeľezn**oj ki*   *langso*   ***robot**-i*   ***žeľezn-aja** **doroga***

      iron   road   on   work-3SG   iron-F   road

      'She works at the railway (company), railway.'

In example 140, the first adjective *žeľeznoj* 'iron' is morphologically adapted to Erzya. In Erzya, Russian adjectives are typically transferred with the *-oj* suffix, which can be the masculine ending in Russian, but in this case, this nominative form does not even exist in Standard Russian, and the masculine ending is *-ij*: *žeľeznij*. The second Russian adjective *žeľeznaja* 'iron' agrees in

gender with the Russian nominal head *doroga* 'road', and the construction is inserted into the utterance as a chunk, as an EL island.

As a next step, I analyze the use and scope of gender agreement in the Russian-type necessive constructions. In Standard Russian, the predicate is the short form of the adjective, and it can have gender and number markers. If a Russian code-switched word is the Experiencer in the sentence, we would expect gender agreement with the predicate in Erzya–Russian CS discourse, as in example 141 (speaker 201103):

(141)   *mejľe*   ***ńev'esta   dolž-na   podarok***   *maks-oms*

      then   bride   must-F   present   give-INF1

      'Then the bride has to give a present.'

In example 141, the Experiencer is a Russian origin word *ńev'esta* 'bride', and the feminine form of the predicate adjective is used in the utterance, with gender agreement present. However, this does not seem to be a general tendency, as the next utterance (example 142, speaker 201015) is a counterexample to the Standard Russian rule:

(142)   ***učiťeľńića****-nok*   ***dolž-en*** *sa-ms*

      teacher.F-POSS.1PL   must-M   come-INF1

      'Our teacher has to come.'

In this example, the Experiencer is again a Russian origin word, still, there is no gender agreement, the speaker inserts the masculine form of the adjective into the utterance. It might be explained by the fact that there is an Erzya first person plural possessive suffix attached to the Experiencer which can inhibit the triggering of the gender agreement. Sarhimaa (1999) also found the lack of gender agreement and the overall use of masculine forms with elderly speakers. However, more examples would be needed to confirm this hypothesis, as most of the examples in my data involve the use of the masculine or plural forms of adjectives.

Gender agreement occurs in other utterances as well, involving an adjective in the predicate position. In example 143 (speaker 200505), the Russian adjective *strogaja* 'strict' agrees in gender with the subject *babam* 'my grandmother'.

(143)   *ťeťa-ń*   *jondo*   ***baba****-m*   *son*   *uľńe-ś*   *pek*   ***strog-aja***

      father-GEN   side   grandmother-POSS.1SG   she   be-PST.3SG   very   strict-F

'My grandmother on my father's side was very strict.'

If we analyze example 143 using the MLF model, the adjective *strogaja* 'strict' can be considered an inserted EL island. In example 144, the negated form of the predicative adjective is also inserted into the utterance as an EL island, but it is integrated into the utterance frame by an Erzya verbal marker.

In example 144, the speaker (200509) uses the feminine adjective *ńezaščiščonnaja* 'someone who has not defended their dissertation yet' (or *ABD*, in North American English usage, for "all but dissertation") and attaches the first person singular verbal suffix to it as required by the nominal predication.

(144) ***robot****-an* ***aśśisťent****-aks* *śeks* *meks* *poka* ***ńezaščiščonn-aja****-n*
work-1SG   assistent-TRA for.that why   yet   not.defended-F-1SG

śeks *meks* ***ďiśśertaćija****-m* *avoľ* *anok*
for.that why   dissertation-POSS.1SG   not   ready

'I work as an assistant because I am ABD, because my dissertation is not ready.'

7.2.3.3.2. Gender agreement with verbal predicates

In addition to the reflexive verbs and constructions with the predicative adjective *dolžen* 'must', the third most typical case of neutralization in the Erzya–Russian CS data is the emergence of gender agreement involving Russian finite predicates with a feminine marker. As discussed above, this tendency concerns the necessive constructions as well, but the greatest number of examples can be found with verbal predicates in the past tense.

One of the most widespread verb, which has gone through paradigm transfer is the Russian verb *hoťeť* 'to want'. It has two Erzya equivalents: the verb *bažams* 'to want' and the construction *uľi* meľ + possessive suffix + infinitive 'I have a disposition to do something'. While the three constructions vary in the speech of Erzya–Russian bilinguals, apart from the conscious and puristic language use of intellectuals, *hoťeť* 'to want' is dominant in spoken language, at least in some speech communities. For example, this verb prevails in the data, and it is the exclusive form in the interviews recorded in Atrať, in all contexts, involving the entire paradigm of the verb. As variation is lost in this case, we can talk about a variety exhibiting features of fused lects (the

last phase on Auer's continuum). Ideally, this issue would also require further investigation with elicited quantitative data from a variety of speech communities and different generation of speakers.

As already mentioned in section 4.2, the verb *hoťeť* 'to want' was already used in the 1950s almost exclusively. Hallap (1960) mentions some verb forms which show variation, in some contexts they are used with Erzya (or Moksha), whereas in other cases with Russian verbal markers. There are a number of verbs, however, which occur predominantly or even exclusively as EL islands, i.e. with Russian suffixes: "So it is absolutely common in the case of the verb *hoťeť* 'to want' to conjugate it the Russian way"[11] (Hallap 1960: 222, my translation). Hallap does not give an explanation in this case for the popularity of the Russian paradigm, but he discusses this verb together with the verb *moč'* 'to be able to', which also lacks an Erzya verbal equivalent (cf. for details section 3.2.2).

The issue and complex nature of the Erzya equivalent is mentioned by other authors as well, e.g. in Keresztes (2011), who, while giving guidelines for the creation of the unified Mordvin standard, suggests that the canonization of the Moksha verb *jorams* 'to want' instead of the Erzya *moń uľi meľem* 'I want (lit. I have the mood to do something)' construction would make the new standard more practical.

The paradigm transfer in this case also facilitates gender agreement. In the past tense forms of the Russian main verbs, there are four markers used: *-l* (masculine), *-la* (feminine), *-lo* (neutral), and *-li* (plural). The following examples are instances in which the past tense forms of the verb *hoťeť* 'to want' occur. I investigate the extent of gender agreement in these utterances, and intend to provide an explanation for the absence or presence of gender agreement.

The emergence of the gender agreement rule even in the case of Erzya subjects has presumably started quite recently. When analyzing the language use of elderly speakers (around the age of 80), the use of the masculine forms prevails, even in cases involving a feminine subject of not only grammatical but also natural gender (the same tendency is attestable in Karelian, as mentioned above). In example 145, the speaker (201102) is talking about her life, using the Erzya

---

[11] "Nii on täiesti üldine verbi *хотеть* 'tahtma' pööramine venepäraselt" (Hallap 1960: 222).

first person singular personal pronoun *mon* 'I', and despite the fact that she can be sure of the gender of the subject, she uses the masculine form of the verb *hot́el* 'wanted to'.

(145) | *mon* | ***ńe*** | ***hot́e-l*** | ***ńe*** | ***hot́e-l*** | ***venča-ms*** |
| --- | --- | --- | --- | --- | --- |
| I | not | want-PST.M | not | want- PST.M | marry-INF1 |

'I did not want to get married.'

This speaker is the only completely illiterate consultant I interviewed. This 86-year-old woman grew up monolingual and illiterate, and started to learn Russian when she was evacuated from Mordovia to Mongolia during World War II.

This correlates with the results of Shagal (2016), who studied the Russian language use of Erzya–Russian bilinguals and also found gender mismatch in the Russian monolingual speech of elderly community members. This kind of gender mismatch, however, does not mean random mixture. Feminine predicate forms with grammatically masculine or neutral subjects do no occur in the data. So when I talk about the emergence of gender agreement in code-switched sentences, it involves the appearance of feminine and neutral past tense predicate forms.

Shagal (2016) raises a question related to this phenomenon: is the use of the masculine forms a result of final vowel reduction which Pussinen (2010) describes as a common phenomenon in the Russian language use of Erzyas. Shagal rejects the hypothesis by citing examples in which the feminine and masculine past tense forms do not differ only in the presence of the final *-a* marker. In my data, there is also an example from the same elderly speaker (201014), which also refutes the reduction hypothesis:

(146) | *meźe-st* | ***ostanovi-lśa*** | *mon* |
| --- | --- | --- |
| what-ELA | stop-PST.M.REFL | I |

'Where was I?'

In this example, the feminine past tense form of the reflexive verb would be *ostanovilaś* 'I stopped', so we cannot attribute the use of the masculine form to final vowel deletion as suggested by Pussinen (2010).

As compared to elderly speakers, the younger generations tend to apply the gender agreement rule to a greater extent, – however, not with all types of subjects. In the following examples, I aim to find a pattern as to where and why gender agreement occurs in the data.

The first question is whether subjects expressed through a personal pronoun trigger gender agreement. As mentioned above, gender as a grammatical category is missing from Erzya, and personal pronouns do not encode the gender of their referents. When studying the code-switched Russian predicates in the data, it is only the singular forms that are relevant for the investigation, because the plural *-li* ending does not mark the gender of the subject. The personal pronouns *mon* 'I', *ton* 'you' and *son* 'he, she, it' refer to animate subjects whose natural gender is masculine or feminine. Theoretically, the use of neutral would be also possible if the personal pronoun referred back to a Russian noun in a previous sentence which had the neutral grammatical gender, but this type of agreement between a past tense neutral predicate and an anaphoric personal pronoun is not attested in my data. As a result, gender agreement in this case equals the emergence of the feminine past tense ending (*-la*) on the predicate.

In the following examples, I first study the occurrence of the feminine form of the verb *hoťeť* 'to want'. The form *hoťela* 'wanted' can be attested with different subject types, among them, with personal pronouns.

(147) *mon* **uže** *viška ping-ste* **hoťe-l-a** *uľ-ems* **učiťeľ-eks**

I   already small age-ELA   want-PST-F be-INF1 teacher-TRA

'I have wanted to be a teacher since I was small.'

This type of examples, involving gender agreement, also raise the question of the ML. It would be a logical assumption in this case that Russian is the ML of the utterance because of the gender agreement. The subject is an Erzya personal pronoun not marked for gender, and if gender agreement occurs, the Russian morphological frame has to be dominant. This would mean that apart from the adverb *uže* 'already' and the verb *hoťela* 'she wanted', all other elements in the utterance are inserted EL islands, because EL system morphemes cannot otherwise occur in the ML frame. The complement of the infinitive *učiťeľeks* 'to a teacher' is also problematic, as it contains a Russian origin stem *učiťeľ* 'teacher', but it is included in the frame using an Erzya translative ending, required by the rules of such construction (meaning 'to become something'). In Standard Russian, instrumental case would be used involving the form *učiťeľem*. As *učiťeľeks* is a mixed constituent (Russian stem + Erzya translative suffix), it should be inserted into the frame

166

with a Russian system morpheme. Consequently, we have to consider this utterance as having a composite ML (Myers-Scotton 2002).

There is another aspect of this example worth discussing: the presence of the overt subject which can also be considered the result of the dominance of Russian. Erzya is a pro-drop language (Rueter 2007), and the person and number of the subject (in particular cases also that of the object in the definite conjugation) is marked on the predicate, while the subject pronoun is overt only if it is emphasized. There is a debate about the existence of (partial-)pro-drop in Russian. Franks (1995) argues that a sentence with a past tense predicate and without a subject pronoun would be elliptical, as Russian retains the non-emphatic pronominal subjects in these sentences (1995: 292). Franks compares Russian to Polish, where the pronominal subject is overt if it is emphasized, similarly to Erzya. As a result, the fact that the Erzya subject pronoun is overt in these past sentences can be considered an influence of Russian, with lexical transfer in these utterances being accompanied by structural interference. Since, according to Muysken (2000), pro-drop has not been studied in detail in the CS literature, further research would be needed on this aspect of CS as well.

The construction can be found in written language use as well in the informal language use of the above mentioned social media network Vkontakte ('in contact'). In a group aiming at the unification of Erzyas, one new member introduces herself not as an Erzya girl (*erźanočka*) but with the name used by outsiders, mainly Russians, to refer to a female Erzya person as *mordovočka* 'a Mordvin girl'. When other members criticize her, her reaction is the following (Vkontakte 2008):

(148) *mon* **_hoťe-l-a_** *śormad-oms* ***erźanočka*** ***no*** ***vot***
I want-PST-F write-INF1 Erzya.girl but well

ki *jak* *a* *pek* *sod-i* *meźe* *isťamo* ***i*** *śormad-iń*
who EMF not very know-3SG what like and write-PST.1SG

mordovočka *śide* *pek* ***pońatno*** *ki* *mon*
Mordva.girl more very clear who I

'I wanted to write "Erzya girl", but well nobody quite knows what that is, so I wrote "Mordva girl", it is clearer that way who I am.'

Gender agreement occurs with other verbs as well. These utterances most typically contain pronominal subjects, as example 149 (speaker 200511). The only Russian element of the utterance is the past tense predicate. The morphosyntactic frame is set by two languages.

(149)   *mon*   *mejľe*   **_reši-l-a_**   *tosto*   *tu-mo*   *kudo-v*

          I   later   decide-PST-F from.there   come-INF2   house-LAT

          'I decided later to come back home from there.'

Gender agreement occurs with nominal subjects as well if the natural gender of the subject is feminine. In example 150 (speaker 201101), the subject is the definite form of the numeral 'one'.

(150)   *vejke-ś*   **na**   **ṕerv-uju**   **kaťegoŕij-u**   **_poda-l-a_**

          one-DEF.SG on   first-ACC.SG   category-ACC.SG   give-PST-F

          'One of them passed the exam needed for the first level in the qualification

          system of teachers.'

However, in this case, it is possible that the gender agreement is attested in example 150, because a switch occurred after the subject, and the second part of the utterance can be considered an instance of alternation.

In example 151, the subject is expressed by a name (speaker 20130419), and gender agreement occurs even though an Erzya discourse particle is inserted between the subject and the predicate.

(151)   **Tamara** **Śergejevna** *vana*   **_izuči-l-a_**   *moń*   **maťeŕial**-*om*

          Tamara   Sergejevna look   study-PST-F   I.GEN   material-POSS.1SG

          'Look, Tamara Sergejevna studied my material.'

Finally, example 152 (speaker 201007) is an instance of hypercorrection or overcompensation for the lack of gender agreement in Erzya. In Standard Russian, the masculine noun *reb'onok* 'child' would not trigger the use of the feminine form of the predicate, in the mixed Erzya–Russian utterance, however, gender agreement occurs.

(152)   *obmoće-ś*   *moń*   **vtoroj**   **reb'onok** **tože**   **_postuṕi-l-a_**

          second-DEF.SG I.GEN   second.M child   too   enter-PST-F

          Ardatovskoj   **meďićinskoj**   **učiľišče**-*s*

Ardatovskoj   medicine   college-ILL

‘My second child also got accepted in the medical college.’

While the predicate *postupila* ‘she entered (college)’ agrees with the subject in gender (not the grammatical gender of the word, but the natural gender of the referent), the adjective *vtoroj* ‘second’ in the attributive position is the masculine form, because the word *reb'onok* ‘child’ has masculine as grammatical gender. This examples seems to prove that, at least in case of the past tense predicates, the natural gender (that the child is a girl) is the main factor controlling agreement and not the grammatical gender of the word. This, however, needs further investigation, taking into consideration also the level of the speakers' Russian proficiency.

7.2.3.3.3. Animacy hierarchy and gender agreement

In this section, I intend to establish a hierarchy of factors triggering gender agreement. In my data, all but one speaker used the feminine form of the verb with animate pronominal subjects. This elderly speaker (201001) consistently applied the masculine form in all of the cases, with a dropped subject and a predicative adjective (the morphologically adapted form of the Russian adjective), but also with verbal predicates in the past tense, with an animate (*avam* ‘my mother’) and with an inanimate subject (*vojnaś* ‘war’):

(153) *avoľ* **sovśem**   *ťena*   ***ńegarmotnoj-an***
      not   at.all   whatsit   illiterate-1SG

      a uš   *ťeńa-ś*   **gramota**-*ś*   *araś*   *moń*
      but already whatsit-DEF.SG certificate-DEF.SG there.is.not I.GEN

      ťeke **dva klass-a**   *ťejń-iń*   **treťij**-*ś*   *karm-iń*
      only   two class-GEN.SG do-PST.1SG   third-DEF.SG   start-PST.1SG

      treťij-*ś*   *karm-iń*   **vojna**-*ś*   **nača**-*ľ-śa*   ***v ijuń-e***
      third-DET.SG   start-PST.1SG   war-DET.SG   start-PST.M-REFL in June-PREP.SG

      a *ťeťa-ń*   *saj-iź*   **okťabŕ-a**   ***nav'erno štoli***
      but father-POSS.1SG.GEN get-PST.3SG<3PL October-GEN.SG maybe   or.what

      konć-e   **okťabŕ-a**   *i*   **vśo** *moń*   *ava-m*
      end-PREP.SG October-GEN.SG and   all   I.GEN   mother-POSS.1SG

169

> ***zapŕeťi-l*** ***škola***-*v* *jaka-ms*
>
> forbid-PST.M school-LAT go-INF1
>
> 'I am not completely illiterate, but I do not have a, whatsit, a certificate, I finished only two classes, I started the third, I started the third, the war had started in June, and my dad was drafted in October, maybe, or what, in the end of October, and that's all, my mother forbade me to go to school.'

The first clause is also worth studying as far as negation is concerned, since we find two types of negation in the utterance: the Russian adjective displays a Russian negative prefix, while the whole construction is negated by an Erzya particle *avoľ* 'not'.

Gender agreement with an inanimate subject was not attested in the corpus. In Standard Russian the subject *nacionaľnosť* 'nationality' would require the past tense predicate to have the feminine marker *pomogla* 'helped'. However, in example 154 the predicate is in the masculine form *pomog* 'helped':

> (154) ***nacionaľnosť***-*eś*   *vadŕa* ***i***   *mońeń*   ***pomog***
>
> nationality-DEF.SG   good   and   I.DAT   help.PST.M
>
> 'My nationality is good and it helped me out.'

There is no agreement, because the subject is an inanimate proper noun, with no natural gender that can be assigned to it.

On the basis of these examples, we can claim that there are two factors determing the use of gender agreement. On the one hand, the grammatical form of the subject is of decisive force. The use of the overt pronominal subject is typical in cases involving Russian verbs in the past tense. Nouns exhibit variation as regards the use of feminine forms. This is also connected to the animacy of the subject. On the other hand, variation is attested as regards the age of the speakers. This is a similar tendency as the one described in Karelian by Sarhimaa (1999). While gender agreement is missing from the utterances of elderly speakers (the use of masculine as the default forms prevail in all cases), younger speakers apply feminine forms exclusively with pronominal [+ animate] [+ human] subjects. Variation can be observed with nominal subjects, whereas [+ animate] [+ human] subjects involve agreement. In the speech of children already [+ animate] Erzya subjects such as *kiska* 'dog' can trigger agreement (cf. example 19 in Mosina's

2002 study). Natural gender seems to be the decisive factor, and that is why inanimate entities as subjects do not agree with the predicate in gender.

However, it also has to be noted that gender assignment is consistent in the Erzya–Russian CS disourse, the masculine form is used as default, and gender agreement involves the use of feminine verb forms with feminine subjects. There are no instances of other combinations in the corpus, for example, in which the feminine or the neutral verb forms would be used with masculine subjects. As a result, the application of gender agreement in Erzya cannot be explained with limited Russian proficiency, only with the above mentioned two factors, the age of the speakers and the animacy of the subject.

7.3. Flagged switches

In the present section, I analyze instances of flagged code-switches occurring in the Erzya–Russian bilingual data. While the emergence of flagged switches can be partly explained on structural grounds, with the avoidance of incongruent switch points, other instances of flagging are, instead, medium repairs. I study medium repairs in section 7.3.1, while I focus on flagging of incongruent constructions in section 7.3.2. I think it is justified to combine structural and pragmatic approaches in order to be able to analyze the whole spectrum of CS patterns.

Flagging types can be categorized in a variety of ways. Muysken (2000: 105) differentiates between two main cases, pragmatically and structurally motivated flagged switches, and cites examples using data from different case studies. While in case of typologically similar languages, flagging is related to puristic views of CS and the speakers' intention to avoid intraclausal switching, in case of typologically different languages, flagging is used to enable switching even in incongruent constructions:

"While true cases of flagging can be seen as revealing the hesitation of speakers to mix intra-sententially, as in French/English or French/Dutch mixing (perhaps due to strong pressures to produce monolingual, pure sentences or, to use Grosjean's terms, stay in the monolingual mode), the other cases of dummy word insertion, like Finnish/English, can be

seen as showing that flagging helps overcome conflicts in linearization patterns: from left to right in English and right to left in Finnish" (Muysken 2000: 106).

Romaine (1995: 141) also discusses this latter type of switching, in which the speaker uses the pause to stop the smooth flow of speech in order to avoid the clash of constructions and to mark the switch: "These switches are marked by pauses, hesitation phenomena, repetition and metalinguistic commentary, which draw attention to the switch and interrupt the smooth production of the sentence at the switch point." In my data, both types occur, and I discuss their typical uses below.

Flagged code-switches are often parallel or double constructions in which the given phrase occurs in both languages. Gafaranga (2007: 296) argues that in these cases of repetitions we have to take into consideration the direction of the switch: if it goes from the more intimate language to the language of dominance, we can talk about emphasis; in the reversed case, the harshness of the utterance is being toned down by the juxtaposition of sequences from different languages. The Erzya–Russian CS data also exhibit differences according to the direction of the switches: the Erzya–Russian direction is often an instance of a lexical gap, while the reversed order can be attributed to purism (cf. also section 7.1.2).

Muysken (2000: 105) draws attention to the connection between double constructions and alternations, as he argues that "[d]oubling can be best seen as indicative of alternation, since it involves an adjustment in the planning of the sentence." Laakso (1996: 227) also emphasizes the consciousness of the speakers in the case of flagged switches, especially if they make metalinguistic comments about their language use: "The speaker consciously monitors his/her language use, when he/she, for example, says 'we say something like this', lists synonyms, parallel constructions, contexts of use, or explains the meaning"[12] (my translation).

---

[12] "Puhuja tarkkailee tietoisesti kieltään esimerkiksi kertoessaan, että jokin asia »sanotaan meillä näin», luetellessaan synonyymeja, rinnakkaismuotoja tai käyttökonteksteja tai selittäessään merkitystä".

7.3.1. Flagging in congruent constructions

In this section, I study flagging in congruent constructions. These instances of flagging often occur in case of a lexical gap when the speaker cannot produce a word or a construction in Erzya. I discussed this phenomenon as related to numeral phrases in section 7.2.2.1. The tendency is partly attributable to the fact that mathematics is taught in Russian in schools. The production of numeral phrases presents a problem even to speakers who otherwise use only a limited number of Russian switches. Even though they can produce a correct form of the numeral in Erzya, these structures are often accompanied by hesitation, embarrassed laughter, self-repair, or a request for the correct form from other participants.

Flagged switches can be related to purism as well. Medium repair type of flagging, e.g. when the choice of the language is repaired, can be initiated by the speaker or other communicative partners. If it is not a self-repair, however, it is not necessarily accepted by the speaker. In example 155, the interviewer (S2) provides the Erzya equivalent of the Russian predicate, but the speaker (200507, S1) ignores the comment and does not backtrack.

(155) S1: *a   ńej   **narod**-oś   alamo   a   veľe-ś   alamoń*

but now people-DEF.SG little   but   village-DEF.SG   slowly

*alamoń **zatuha-jet***

slowly   attenuate-3SG

S2: *mad'-i   mad'-i*

attenuate-3SG   attenuate-3SG

S1: 'And now there are few people, and the village slowly slowly dies out.'

S2: 'Dies out, dies out.'

The medium repair initiated by the interviewer goes unnoticed. On the basis of this, we can assume that the speaker considers the medium of interaction bilingual and that both the language switched from and the language switched to are part of this medium. This phenomenon can be related to Myers-Scotton's (1993: 213) term "unmarked choice", which refers to the unmarked status of CS in bilingual communities where speakers identify themselves with both languages and cultures.

In other cases, the medium repair initiated by the communicative partner is accepted and repeated. In example 156, the speaker (20151118) follows suit and repeats the Erzya equivalent provided by the reporter.

(156) S1: *nama   neť   veśe   kevkśťńema-ťńe   langs*
          of.course these all   question-DEF.PL   onto

          *miń śeśke   maks-tano **otv'et***
          we   right.away give-1PL   answer

          Reporter: *karšoval*

                   answer

          S1: *karšoval*

               answer

          'We of course give an answer to all of these questions right away.'

Repairs can be initiated by the speakers themselves as well. In example 157 (speaker 200501), flagging is a self-repair, in which the reporter consciously monitors her speech and repairs the use of Russian elements.

(157) ***bol[še]*** *śeďe   śeďe   lamo   jovtńe-k   sonze langa*
          more   more   more   much tell-IMP.2SG   its   on
          'Tell me more more about it.'

Purism is a widespread phenomenon in the Finno-Ugric communities in the Russian Federation. In the Erzya media, for example, the avoidance of Russian elements is prevailing. It is a widely accepted belief among the Erzyas that Russian elements should be avoided in formal situations both in oral and written genres. In newspapers, texts are reviewed by editors who replace Russian constructions and delete even loanwords from texts. The same approach is present in (university) education: instructors reprimand students for using elements of the dominant language in their speech (whereas in primary education Russian is the dominant language, e.g. in mathematics). While the creation of neologisms is part of corpus planning in the language, this puristic tendency often counteracts achievements of language maintenance. In Chapter 2, I referred to the coexistence of a semi-monolingual High Variety of Erzya which is in a

diglossic situation with an Erzya–Russian CS Low Variety. In Chapter 8, I elaborate on the main characteristics of these two varieties.

In this section, I discuss different types of flagged switches occurring in my data. I focus both on the form of the flagged switch and the possible reasons for flagging. I also examine which words and constructions are flagged.

In example 158, the speaker (200501) indicates that she does not know the Erzya word for *pśih* 'psycho'. She uses the phrase 'everybody says it like this' and apologizes for the gap in her Erzya lexicon, which can be explained by the fact that the interviewer corrected her earlier in the recording and reminded her that she should speak Erzya.

(158) *ńej*   *mon* **robot**-*an* **dom** **invaľid-ov**   **naźiva-jet-śa** *mińek*

now   I   work-1SG home disabled-GEN.PL call-3SG-REFL we.GEN

pśihońeurologičesk-ij   **inťernat** **pśih**-*ť*   *mińek*   *toso*   *koda*

psycho-neurological-M   center   psycho-PL we.GEN   there how

bolše *meŕ-an* <u>*veśe*</u>   **tak** *meŕ-iť* *veśe*   *i*   *eŕźa-ks*   *a*   *sod-an*

more say-1SG everybody so   say-3PL everybody and Erzya-TRA not know-1SG

'I am working now in the home for the disabled, it is called, we have a psycho-neurological center, we have psychos there, how should I say it differently, everybody says it like this, everybody, and I do not know it in Erzya.'

During my fieldwork, I experienced it as a general tendency that speakers apologized for not speaking "real" or "pure" Erzya. In the interviews, speakers word this problem by admitting that they are unaware of the "correct" Erzya form of a given construction. This is often the case with numeral phrases, cf. examples 73 and 74 in section 7.2.2.1.2. Speakers' attitudes to their own variety should be investigated in detail in further research.

Flagged switches often involve the use of constructions such as *ruzks meŕems* 'to say it in Russian' or *koda ruztne meŕiť* 'as the Russians say'. It is also common to combine different medium repair types; Auer (1995) also describes this bundling tendency of contextualization cues. In example 159 (speaker 20130821a), there are two indicators of medium repair, the above mentioned expression *ruzks meŕems* 'to say it in Russian' and the determiner *iśťat* 'these kinds of' (the latter type of flagging can also be considered a deictic marker, cf. Halmari 1997):

(159) *erźa-ťńe-ń*   *toso eŕamo-do   uľ-iť' iśťa-t*

Erzya-DEF.PL-GEN   there life-ABL   be-3PL these.kind-PL

ruz-ks   meŕ-ems   sv'iďiťelstva-t

Russian-TRA   say-INF   1   proof-PL

'There is this kind of, to say it in Russian, evidence (lit. evidences) that the Erzyas

lived there.'

According to Pyöli (1996: 296), switches flagged by demonstrative pronouns are signs of a lexical search in Karelian. In example 159, purism is also a relevant factor in flagging in addition to the fact that the speaker can produce the word only in Russian.

In case of a lexical gap in Erzya in general, speakers often flag their switches with the special stem *ťeńa* 'whatsit' which can be used instead of both verbal and nominal stems. In example 160 (speaker 201101), it is used as a finite verb form in the third person plural, while in example 161 (speaker 200803), it is congruent with the noun *sanatoŕija* 'sanatorium' in the illative case.

(160) *i   klub-so sćena   langso ejkakš-tne   ťeń-iť   vistuṕ-iť*

and   club-INE stage   on   child-DEF.PL whatsit-3PL perform-3PL

avoľ beŕańste meŕ-ems   vadŕasto

not   badly   say-INF1   well

'And in the club, on the stage, the children whatsit performed not badly, let's say,

well.'

(161) *posľe každ-ogo   pohod-a   nas   mińek   poḿesťi-ľ-i*

after   every-GEN.SG cruise-GEN.SG we.GEN   we.GEN   place-PST-PL

ťeńa-s   sanatoŕija-s   na otdih

whatsit-ILL   sanatorium-ILL   on rest

'After every cruise (with the submarine), we were placed in a whatsit, in a

sanatorium to have a rest.'

In example 161, there is another flagged switch, the doubling of the pronominal object, which can be attributed to the tendency that the subject and object pronouns are typically expressed in Erzya, and the speaker backtracks after producing the Russian equivalent.

176

In case of the reporters' turns in the Vaygel radio interviews, there are a number of switches which are flagged through the use of the above mentioned type of phrases (*ruzks meŕems* 'to say it in Russian') but cannot be explained by a lexical gap, as the radio journalist provides the Erzya form first. In example 162 (interview 20130813), the reporter (R1) uses four ways for flagging the switch: she produces two synonyms of the switched word *iďeja* 'idea', names the language, and applies the determiner *kodatkak* 'some kind of':

(162)   *paŕak*   *ťe*   *ška-ś*   *uľ-iť*   *koda-t-kak*   **_ruz-ks_**

       maybe this   time-DEF.SG   be-3PL   some.kind-PL-EMP   Russian-TRA

       meŕ-ems   **_iďeja-t_**   _meľ-ť'_   _aŕśema-t_

       say-INF1   idea-PL   opinion-PL   thought-PL

       'Maybe at this time there are some kind of, to say it in Russian, ideas, thoughts.'

In example 163 (interview 20130606, reporter R1), the switch is marked by the translation and the flagging construction *ruzks jovtasa* 'to say it in Russian'.

(163)   *eŕav-i*   **_štobu_**   *uľe-veľ*   **_finanśirovańija-zo_**   **_i_**

       need-3SG   that   be-COND.3SG   financing-POSS.3SG   and

       finanśirovańija-ś *sval*   *ška-ń*   *uľe-ze*   **_ruz-ks_**

       financing-DEF.SG   always   time-GEN   be-3SG.OPT   Russian-TRA

       jovta-sa   **_stabiln-oj_**

       tell-DEF.3SG<1SG   stable-M

       'There has to be financing, and the financing has to be all-time, to say it Russian stable.'

This type of bundling of flagging types occurs not only in the utterances of the reporters but also in the speech of an invited guest. In example 164, the speaker (20130821) uses the construction *koda ruzoks meŕiť* 'as they say it in Russian' and the Erzya equivalent *važoďemanzo* 'his work' of the inserted Russian element *trudonzo* 'his work':

(164)   *sonze*   *važoďema-nzo*   **_trud-onzo_**   *koda*   *ruz-oks*

       his/her   work-POSS.3SG.PL   work-POSS.3SG.PL   as   Russian-TRA

       meŕ-iť   *lovnokšn-iń*

       say-3PL   read-PST.1SG

'I have read his works, his works, as they say in Russian.'

Studies on flagging also discuss the question what is typically flagged in bilingual utterances. For example, Rosignoli (2011) studies flagging in Italian–English code-switching and argues that in his data higher frequency elements are less typically flagged: "A similar relation holds between flagging and different grammatical categories, with nouns being less flagged than adjectives or verbs" (Rosignoli 2011: III).

In my data, nouns – even morphologically adapted forms of nominal constructions – are frequently flagged. While flagged nouns and adjectives are adapted to Erzya to some degree, flagged verbal structures and prepositional phrases are typically inserted as EL islands into the Erzya frame.

In example 165, the speaker (20151118) flags the EL island infinitive *poučastvovať* 'participate'. As we have seen in section 7.2.3, the infinitive is not adapted to Erzya most likely because it involves the perfective form of the Russian verb, which is typically code-switched as an EL island.

(165) *mon pokš meľsparo-s-an   meks mońeń   sav-ś   ťe*
        I   big   happiness-INE-1SG why I.DAT   have.to-PST.3SG this

         ***vśerośśijsk-ij masťerklas*-*so-ńť*   ***poučastv-ovať ruz*-*oks*   *meŕ-ems*
         all.Russian-M masterclass-INE-DEF.SG participate-INF   Russian-TRA say-INF1
         'I am very glad because I could participate, to say it in Russian, in this all-Russian masterclass.'

There are also instances of flagging involving a finite verb form, as in example 166. The speaker (200504) inserts the verbal predicate as an EL island, while its argument, the dative pronoun, is in Erzya. The Russian verb is part of a widely used Russian construction *mńe* (the dative form of the pronoun) *povezlo* 'I got lucky, lit. to me it was delivered', which has no equivalent in Erzya.

(166) ***veśe predmet*-*tne-ń*   *kořaś* *mońeń*   ***ruz*-*ks*   *meŕ-an* **pov'ez-l-o***
         all   subject-DEF.PL-GEN with   I.DAT   Russian-TRA say-1SG get.lucky-PST-N
         'I say it in Russian, I got lucky with all my subjects.'

Common expressions are inserted into the Erzya utterance as chunks. In example 167, the flagged prepositional phrase (*dľa galočki* 'for a checkmark') is an EL island, the speaker (20160115) flags the switch with the construction *ruzoks jovtams* 'to tell it in Russian'.

(167) *ńej   lamo-t   ťej-iť   iśťa-t   **vistavka**-t*

now   many-PL   make-PL   such-PL   exhibition-PL

***dľa   galočk-i   ruz**-oks   jovt-ams*

for   checkmark-GEN.SG Russian-TRA   say-INF1

'Now many people create exhibitions just to get a tick in the box.'

So far I discussed constructions in which the speaker flags the switch point typically with expressions such as 'Russians say it like this' or 'to say it in Russian'. There are other types of flagging as well, repetitions and translations can also be used, and false starts are also instances of flagging. The direction in these constructions is typically Russian–Erzya, which can be explained by puristic tendencies.

The following examples are false starts: the Russian form is just partially uttered, then it is replaced by the Erzya equivalent. In example 168, the Russian adverb *počťi* 'almost' is started, but then the speaker (200509) opts for the Erzya adverb *malav* 'almost'.

(168) ***poč…*   *malav veśe kudo-ťńe   čuvto-ń**

al[most]   almost all   house-DEF.PL   wood-GEN

'Almost all the houses are made of wood.'

In example 169, the speaker (200510) replaces an established Russian loanword *robotan* 'I work' with its synonym *važoďan* 'I work', which can be considered an instance of purism typical in the language use of journalists and other intellectuals.

(169) *mon **ro…   mon važoď-an**   erźa-ń   **pravda gaźeta**-so*

I   wo[rk] I   work-1SG   Erzya-GEN   Pravda   newspaper-INE

'I work I work at the newspaper Erzyan' Pravda.'

In example 169, the constructions are equivalent: a synonym loanword with the same argument structure is replaced in the flagged switch. In the next section, I discuss more complex switch types, in which flagging can be explained by structural factors, namely, the incongruence of the forms.

7.3.2. Flagging in incongruent constructions

In section 7.3.1, the part of speech of the original and the flagged switched element is the same in Russian and Erzya. However, flagging can also involve parallel structures in which the two forms belong to different word classes. In example 170 (interview 20130726), the reporter provides the Erzya equivalent (a finite verb *meŕevi* 'it is possible to say') as an equivalent of the Russian predicative adjective *možno* 'it is possible'.

(170) *poladića-ks   kińgak   **možn-o***   *ľemd́-ems*

successor-TRA somebody.GEN   possible-N   name-INF1

meŕe-v-i   ľemd́-ems

say-REFL-3SG name-INF1

'Can anybody be named as her successor?'

As seen in example 169, speakers flag also adapted words, not only clearly Russian elements. In example 171 (20151112), the flagged construction *nastolnoj kńigat* 'handbooks' is a hybrid, as in standard Russian it would be *nastolńie kńigi* 'handbooks', and the form *nastolnoj* does not even exist in Russian as a nominative form.

(171) *mon večk-an lovno-mo   iśťa meŕ-ems **nastolnoj***

I   love-1SG read-INF2   so   say-INF1 on.the.table

***kńiga**-t **ruz**-oks   moń   lamo*

book-PL Russian-TRA I.GEN   many

'I love reading, I have a lot of handbooks, to say it in Russian.'

In Erzya, Russian adjectives typically receive the *-oj* ending, which is the masculine ending in case of some Russian adjective types, but in many cases, including the given example, it creates a pseudo Russian element which is an in-between category, having characteristics of both code-switches and borrowings. In case of C1 speakers, it is typical to flag such hybrid constructions and label them as Russian.

Translation and double constructions can also be applied at incongruent switch points. In example 172, the reporter (interview 20151118, R1) uses a translated form *tonavtomań koŕaś insťitutoś* 'an institute concerning learning' instead of the common Russian form *insťitut*

obrazovańija 'institute of education', but to make sure that the communicative partner understands what she means, she inserts the original Russian name as well.

(172) *tink*   *tonavtoma-ń*   *kořaś*   **insťitut**-*oś*

       you.PL.GEN   learning-GEN   related   institute-DET.SG

       insťitut   obrazovańi-ja   *iśťa*

       institute   education-GEN.SG   so

       'You have an institute for education, institute for education, right?'

So, here repetition is used to clarify the meaning of the neologism. Reporters at Radio Vaygel use this puristic strategy often, and it is also a common practice in the written media. When journalists coin a new word, they propagate it by writing it first, while including the Russian equivalent in brackets to ensure understanding. In example 172, the two constructions are incongruent morphologically, as the constituent order in possessive structures differs in the two languages (cf. also section 7.2.2.2).

The flagging of the Russian construction through the insertion of a dummy element can also circumvent the need for harmonization of incongruent constructions in the two languages. The most frequent dummy element in Erzya is the stem *ťeńa* 'whatsit', which can be used both as a nominal and a verbal stem. In example 173 (speaker 201001), the equivalent of the Russian prepositional phrase is the Erzya dummy stem *ťeńa* 'whatsit' with the translative case suffix (-*ks*).

(173) *moń*   *meŕ-ems* <u>*ťeńa-ks*</u>   **po**   **muž-u**

       I.GEN   say-INF1 whatsit-TRA according.to husband-DAT.SG

       famiľija-*m*   *M...*[13]

       surname-POSS.1SG M...

       'My, let's say, er, after my husband, my surname is M...'

There is another flagging element in the utterance, the infinitive *meŕems* 'to say', which also indicates that the speaker is looking for the right word, so the motivation for flagging can be a lexical gap as well, but the dummy element is used to avoid the incongruence between the prepositional phrase in Russian and the case suffix in Erzya.

---

[13] I do not disclose the full name to protect the identity of the speaker.

Example 174 (speaker 201001) also involves two incongruent flagged time expressions. In the first one, the name of the month is expressed in Erzya *śeńťabŕasto* 'in September, lit. from September'. In the second one, the Russian construction is a prepositional phrase *v ijuńe* 'in June' with the preposition *v* 'in' (the Erzya equivalent would be *ijunste* 'in June, lit. from June' with the elative case suffix (-*sto*/-*ste* 'from'). The analysis of this example is complicated by the fact that my transcription differs from the transcription made by the native speaker linguist whom I collected the data with. While she added the elative suffix to the dummy stem *ťeńa* 'whatsit' in both cases, I clearly hear the inessive case endings (-*so*/-*se* 'in') uttered. I include both transcriptions below:

(174) Version 1:

moń	*topoďe-v-eľť*	*bu*	*ťeńa-so*	*dvadcať*	*pať*	*ľet*
I.GEN	fill-CONJ-3PL	if	whatsit-INE	twenty	five	year.GEN.PL

śeńťabŕa-sto	*a*	*ťeńa-so-sť*	*v*	*ijuń-e*	*zamuž*
September-ELA	but	whatsit-INE-DEF.SG	in	June-PREP	married

Version 2:

moń	*topoďe-v-eľť*	*bu*	*ťeńa-sto*	*dvadcať*	*pať*	*ľet*
I.GEN	fill-CONJ-3PL	if	whatsit-ELA	twenty	five	year.GEN.PL

śeńťabŕa-sto	*a*	*ťeńa-sto-sť*	*v*	*ijuńe*	*zamuž*
September-ELA	but	whatsit-ELA-DEF.SG	in	June-PREP	married

'In, well, September, I would have been twenty-five years old, but in June I got married.'

The two versions can be interpreted differently. Version 1 indicates that the dummy word is formed according to the Russian rules, as the inessive case ending is the equivalent of the Russian preposition *v* 'in' in Erzya. The dummy word follows the Russian rules irrespective of the fact whether it is followed by the Erzya time expression *senťabŕasto* 'in September' or the Russian form *v ijuńe* 'in June'. In Version 2, the stem *ťeńa* 'whatsit' follows the Erzya rules and has the elative suffix in both cases. The second analysis would be more plausible in the face of the other examples in which the dummy word *ťeńa* 'whatsit' displays the form required by Erzya.

In example 175 (speaker 20130702), the incongruence is between the object marking of the two languages. Erzya requires the genitive suffix of the definite declension, while in Russian, the nominative case is used. The Russian predicate *isľedovali* 'researched' occurs in two clauses and has two objects: the Erzya object *kosťumońť* 'the costume' and the Russian object *žizń ďeda* 'the life of the grandfather' is preceded by the word *ťeńa* 'whatsit' with the definite genitive ending (-*ńť*), as required by the Erzya rules.

(175)   *ańśak*   **kosťum**-*ońť*   **isľedova-l-i**   *i*   *isľedova-l-i*

         only   costume-GEN.DEF.SG research-PST-PL   and research-PST-PL

         <u>*ťeńa-ńť*</u>   <u>***žizń ďed-a***</u>   *kotorij bil*   *ŕepŕeśirovan*

         whatsit-GEN.DEF.SG life grandfather-GEN.SG who   be.PST.M repressed.M

         'They studied the (folk) costume and the, well, the life of grandfather who was

         persecuted.'

The insertion of the dummy word prevents the speaker from facing the problem of different object marking in the two languages. Another possibility to adhere to the Erzya object marking rules could be the use of the Erzya definite genitive ending on the Russian possessive structure, but it would involve another problem, whether to add the suffix to the phrase *žizń ďeda* 'the life of the grandfather' as a chunk (*žizń ďedańť* 'the life of the grandfather'), which is the usual solution in Erzya–Russian CS discourse (cf. section 7.2.2.2), or to the head of the construction *žizńańť ďeda* 'the life of the grandfather'.

To sum up, flagging of CS in Erzya–Russian discourse can be attested in a variety of contexts with different functions: when the speaker is looking for the right expression, or highlighting cases in which the constructions in the two languages do not match. Another type of flagging occurs when the speakers draw attention to the switch consciously and apologize for the use of the Russian element in their speech. It can also involve non-verbal flagging, pause and hesitation, as well as explicit flagging with constructions 'as they say', 'to tell it in Russian', or translation.

Chapter 8. Discussion

This chapter of the present dissertation is divided into three parts. Section 8.1 describes how the Erzya–Russian continuum model presented in section 6.2 can be used for the depiction of the variation in the investigated bilingual discourse. In section 8.2, I discuss the findings of my study and its connections to language endangerment and language shift. In section 8.3, I refer to possible avenues of further research on Erzya in particular, but also on CS in general.

8.1. Variation in Erzya–Russian bilingual discourse

In this section, I provide an overview of the main CS patterns used in my data and discuss how I categorize speakers on the basis of these CS types they use. These categories represent tendencies, or phases on a spectrum. Using earlier continuum models (Auer 1999, Kovács 2001), I devise an Erzya–Russian CS continuum model in order to describe the variation in the corpus (cf. also section 6.2) and to present the diglossic situation characteristic of the language use of Erzya speakers. All of these above mentioned models also have a diachronic aspect, and the categories can represent different stages in the history of a language contact situation. A gradual increase in the frequency of switching can lead to both language shift (the turnover of the ML in the MLF model) and the emergence of a mixed language.

In addition to the above mentioned frameworks, I also rely on Treffers-Daller's (1998: 185) continuum model, which ranges from the monolingual mode to a bilingual mode with in-between modes involving increasingly more switching. Treffers-Daller's model is enhanced with a hierarchy of constituents: different switches are likely to occur at different points of the continuum. Speakers on the monolingual end tend to make peripheral switches involving only interjections and nouns. On the bilingual end, longer constructions can also be switched along with words belonging to basic vocabulary.

In the Erzya–Russian CS data, there is also a connection between types of switches and the amount of switching. I distinguish between three main categories of CS language use types: Category 1 (C1), Category 2 (C2), and Category 3 (C3). Insertions occur in all categories, however, the type of insertion varies in the utterances of C1, C2, and C3 speakers. C1 speakers

predominantly have shorter, typically one-word switches which are inserted into the utterance with Erzya morphological markers. These switches often have a pragmatic function. Longer Russian insertions occur as alternations, e.g. as example 26, in which the Russian clause is a citation. Switches (or even earlier borrowings) are also often flagged in this (semi-)monolingual end of the continuum. In parallel constructions, Russian is typically the first element, repaired by the Erzya equivalent. These are the characteristics of the High Variety in the diglossic situation.

Example 176 is an excerpt from an interview with a speaker (20130412) whom I categorized as a representative of the C1 type.

(176) ***i*** ***vot*** *arś-an* ***što*** ***vot*** *ťe* *ńaka-ś* *ejkakš-tne-ńeń*
and look think-1SG that look this doll-DEF.SG girl-DEF.PL-DAT

vśo ***ravno*** *pŕa-s* *alamoška* *meźejak*
all the.same head-ILL little something

put-i ***što*** ***raz*** *ťeťa-ťńe* *ava-ťńe* ***sam****oj* *pokš*
put-3SG that once father-DEF.PL mother-DEF.PL most big

kaźńe-ń *tońeť* *kaź-iž* *sinst* ***tože***
present-GEN.SG youSG.DAT give-DET.3SG<3PL they.GEN also

eŕav-i *siń* ***uže*** *uľ-i* ***dolg****-ost* *ťe* *ťeťa-ťńe-ń*
need-3SG they already be-3SG debt-3PL.POSS this father-DEF.PL-GEN

ava-ťńe-ń *ikeľe* ***što*** *eŕav-i* *sińenst* *ľezd-ams* *eŕav -i* *sińenst*
mother-DEF.PL-GEN before that need-3SG them help-INF need-3SG them

sval *sińďest* *aŕś-ems* *sinst* ***uže*** *a* *ki* *langs* *ńežeď-ems*
always about.them think-INF1 their already not who onto rely-INF1

ťejťer-est ***eľi*** *ćora-do-st* *baška* ***i*** ***vot*** *ťe-ť* ***veśe***
daughter-POSS.3PL or son-ABL-POSS.3PL besides and look this-PL all

ńaka-ťne ***vot*** *isťamo* *pokš* ***smisla*** *marto*
doll-DEF.PL well like.this big meaning with

'Well, I think that this doll will put something into the head of the children anyway.

185

If the parents have given you the biggest possible present, they also need, they also owe the parents, they have to help them, they have to think of them often, they don't have anybody else to rely on apart from their daughters and sons. And well, all of these dolls have such an important meaning.'

Most of this speaker's switches are one-word insertions and established borrowings: conjunctions (*što* 'that' from Russian *čto* 'that', and *eľi* 'or' from Russian *iľi* 'or'), adverbs (*samoj* 'most', *raz* 'once'), or discourse markers (*vot* 'here is, this is, look'). The Russian nouns are morphologically integrated into Erzya (e.g. Erzya *smisla* 'meaning' from Russian *smisl* 'meaning'). These switches can also be analyzed as borrowings. Even longer switches as *vśo ravno* 'anyway' are inserted into the Erzya sequence as one unit, a chunk.

A typical C2 speaker also uses insertion types (discourse markers, adverbs, and nouns) found in C1. However, there are patterns in C2 speaker's CS which are entirely missing or very rare in C1 speakers' language use. C1 speakers usually avoid Russian verbs, using only older loanwords (e.g. *robotiń* 'I worked' from the Russian verb stem *rabot-* 'to work') and inserting them in the utterance using Erzya morphological markers. In contrast, C2 speakers insert the finite form of the Russian verb into the utterance with Russian morphological markers, transferring the paradigm. The use of Russian elements is not repaired, switches are smooth. Numeral phrases are also inserted as chunks, and the reverse order of possessive constructions is also possible.

Example 177 (speaker 20141111) is categorized as a C2 speaker utterance due to the increased amount of switching in this speaker's speech (as compared to a C1 speaker), but also because the type of CS patterns she uses.

(177) **boľše   vśe-h**   **vśe-go**   mon   **robot**-iń   **karďiologija**-so
more   everybody-GEN.PL   all-GEN.SG   I   work-PST.1SG   cardiology-INE

ďiśśertaćija-ńť   mon   **zaščišča-l-a**   **gastroenťerologija**-so   ťe
dissertation-GEN.SG.DEF   I   defend-PST-F   gastroenterology-INE   this

ťev-eńť   ejse   **tože**   kuvať   robot-iń   alamo-ń
work- GEN.DEF.SG inside   also   long.time   work-PST.1SG   little-GEN

alamo-ń   v'eśe   **razďel**-ťńe-ń   mon   **fakťičeski**   jut-iń   kona-t
little-GEN   all   section-DEF.PL-GEN   I   practically   go-PST.1SG   which-PL

otnoś-it-śa *potmo* *orma-ťńe-ńeń* ***di*** *ška-ś* *kuvaťs*

relate-3SG-REFL internal illness-DEF.PL-DAT and time-DET.SG for.long

veďgemeń-ďe *lamo* ***i*** ***prihoďi-l-o-ś*** *eŕva* *kodamo* ***robota-so***

fifty-ABL many and have.to-PST-N-DEF all kinds work-INE.SG

robot-ams

work-INF1

'Most of all I worked in cardiology, I defended my dissertation in

gastroenterology, in this field I also worked for a long time, little by little I

practically covered all the sections which are related to internal illnesses, this is a

long time, more than fifty years, and I had to do all kinds of jobs.'

In example 177, Russian origin conjunctions (*i* 'and', *di* 'and'), adverbs (*fakťičeski* 'practically') and borrowed nominal stems (*robota* 'work') occur, similarily to the switch types in C1. However, here we can also find finite verbal forms, both in the present (*otnośitśa* 'relates') and in the past tense (*prihoďiloś* 'had to'). In the case of the predicate *zaščiščala* 'I defended', there is gender agreement between the past tense predicate and the pronominal subject. In this type of CS style, the grammatical markers are typically Erzya. In example 177 only the three Russian verbs are inserted as EL islands, arguments of the predicate are expressed with constituents having Erzya endings.

Example 178 (speaker 20140318) is a typical case of a C3 speaker's utterance.

(178) ***kažnoj učastńik-eńť*** *ťe* ***programa-ńť*** *marto*

ḿińisťerstvo

every participant-GEN.DEF.SG this program-GEN.DEF.SG with ministry

seľsk-ogo ***hoźajstv-a-ś*** ***upravľeńie*** ***seľsk-ogo***

rural-GEN.SG economy-GEN.SG-DEF.SG administration rural-GEN.SG

hoźajstv-a ***rajon- oń*** *di* ***ferḿer-eś*** ***zakľuča-jut***

economy-GEN.SG region-GEN and farmer-DET.SG conclude-3PL

soglašeńie *veťe* *ije-t'* ***tak*** ***kak*** *ťe* ***ferḿer-tne***

agreement five year-PL this.way how this farmer-DEF.PL

187

dolžn-i	*veť-ams*		*eśe-st*	***hoźajstva****-so-ńť*	*veťe*	*ije-ť*
have.to-PL	lead-INF1	own-POSS.3PL	farm-INE-DEF.SG		five	year-PL

ťe	***soglašeńija****-so-nť*	*siń*	*ńevť-iť*	*žaro*	***skoťina****-ń*	*pŕa-t*
this	agreement-INE-DEF.SG	they	show-3PL	how.many	cattle-GEN	head-PL

každij	*ije-ń*		*sinst*	***dolžn-o***	*uľ-ems*	***no i***	***sobstv'enno***
every	year-GEN		their	have.to-N	be-INF1	well and	properly

kodamo	***kakom***	***objom-e***		*siń*	***dolžn-ie***	***proizv'es-ťi***	***produkcija***
what	what	amount-PREP.SG		they	have.to-PL	produce-INF	product

'Every participant in this program with the Ministry of Agriculture or the administration of agriculture of the region and the farmer make an agreement for

five years, so the farmers have to fulfill this agreement at their farm for five

years, they show how many cattle they have to have every year and also how much they have to produce.'

In this category, we can also find the characteristics of the earlier phases, but switching is more frequent, Russian-type and mixed constructions prevail. Discourse markers (*tak kak* 'so') and adverbs (*sobstvenno* 'practically') are also typically Russian, but these general switch types (common in C1 and C2, too) are accompanied by hybrid necessive constructions with the Russian predicative adjective (*dolžno* 'has to') and an Erzya infinitive (*uľems* 'to be'). The majority of the predicates (four out of five in this example) are also Russian finite forms.

In incongruent constructions, C3 speakers choose the Russian-type structure. For instance, the Russian constituent order is present with in the otherwise hybrid possessive construction *upravľeńie śelskogo hoźajstva rajonoń* 'administration of agriculture of the region' in which the genitive markers are from both languages (-*a* from Russian in *hoźajstva* and -*oń* from Erzya in *rajonoń*). However, the order of the constituents follows the rules of the Russian language.

In addition to single verb switches, there are cases in C3 in which the predicate and its argument(s) are all Russian elements inserted as one unit: *zaključajut soglašenie* 'make an agreement', or *dolžnie proizv'esťi produkćija* 'they have to produce', which can also be considered alternation.

As mentioned above, the categories are designed on the basis of the CS types in the given interview. The utterances of C1 speaker contain alternation with a pragmatic function, discourse particles, mixed constituent type insertions in which Erzya morphological markers are attached to the Russian element. EL islands sporadically occur, but this type of switches and even Russian borrowings are typically flagged. Numeral phrases are predominantly Erzya, and congruent lexicalization does not occur.

C2 speakers switch more frequently, consistently, the Russian elements do not have a pragmatic function any more. Russian origin words are inserted as EL islands as well, which in the case of finite past tense verbal forms involve the emergence of gender agreement as well. However, it is only the predicate that is Russian, arguments are typically Erzya lexemes or Russian origin lexemes which are, however, inserted into the utterance with Erzya system morphemes (e.g. the infinitive is Erzya in the necessive construction). Russian numeral phrases can be inserted as chunks, and inverse possessive structures are also attested.

Finally, I assigned speakers to the C3 category if the ML of their utterances could not be determined unambiguously. Their speech usually involved Russian predicates with Russian type or hybrid argument structures. In addition to frequent insertions, congruent lexicalization also prevailed in their utterances. In many cases the differentiation between insertion and alternation was made impossible by the level of mixture the bilingual utterance displayed.

In order to see the main types in one model, I repeat the Erzya–Russian continuum model here as Figure 5 (cf. also section 6.2):

Phase 1	Phase 2	Phase 3

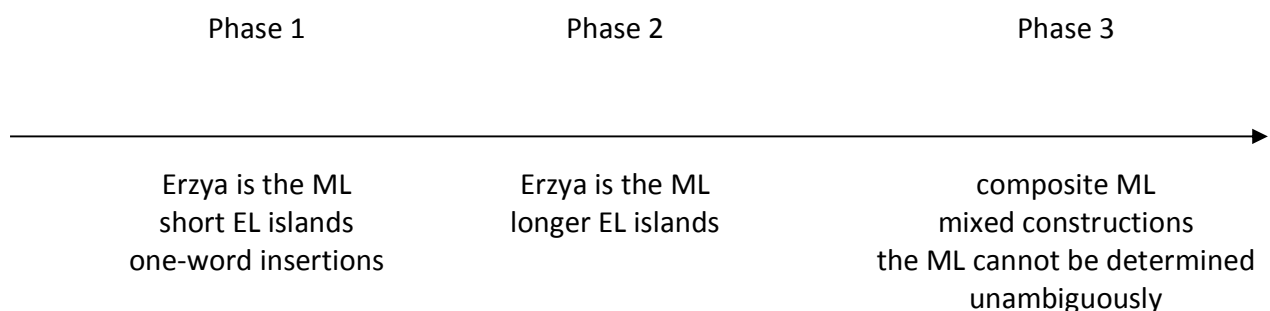| Erzya is the ML | Erzya is the ML | composite ML |
short EL islands	longer EL islands	mixed constructions
one-word insertions		the ML cannot be determined
		unambiguously

Figure 5. Continuum model for the Erzya–Russian bilingual discourse

While one-word switches, adverbs, and discourse particles occur in all categories, longer EL islands emerge in the second phase in C2, and congruent lexicalization is typically attested in the Phase3 in the mixed utterances of C3 speakers.

As mentioned above, categories are not rigid. It is possible that the same speaker uses a different type of CS in a more formal situation. For instance, the researcher (F1) who accompanied me in the fieldwork to Chuvashia switched more frequently when speaking to the locals than in the classes she taught at the university. I also observed this tendency during the spring semester of the academic year 2007/2008, when I studied at the university in Saransk. Instructors who applied a monolingual Erzya variety and reprimanded students for any kind of switching, used a rather mixed C2-type CS pattern during the tea breaks after classes. Obviously, I cannot draw conclusions on the basis of these chance instances of participant observation, so further studies are needed involving a scenario in which the formality of the situation and the language consciousness of the speakers can be controlled.

8.2. Findings of the investigation: the main characteristics of the Erzya–Russian bilingual discourse

In this section, I sum up the findings of the study on Erzya–Russian CS. I focus on the issues discussed in the theoretical part of the dissertation (Chapter 5), and try to provide answers to the general theoretical questions using the results of the Erzya case study.

The differentiation between CS and borrowing presented various problems throughout the analysis of Erzya–Russian bilingual utterances. One-word switches with no morphological markers (e.g. discourse particles, adverbs, nouns in the nominative case) cannot be labelled code-switches or loanwords based purely on morphological criteria. Phonological adaptation also breaks down as a single criterion for differentiation between these two contact phenomena due to the fact that Erzya and Russian have similar phonological systems (cf. section 3.1). Moreover, the nature and quantity of the data does not allow for a frequency analysis, and the repeated use of Russian origin elements does not necessarily mean that they are established loanwords.

For instance, numeral phrases and names of institutions are typically switched as chunks and they are inserted into the Erzya morphosyntactic frame in a variety of ways. Erzya endings can be attached to the head word of the construction or to the end of the unanalyzed chunk. The status of bridge words is also ambiguous. These are word forms that are grammatical in both languages although they have different functions. For example, it is impossible to decide if *kilometra* in example 35 is a Russian genitive form, and, thus, a code-switch, or an established Russian borrowing in the nominative case in Erzya. On the basis of these findings, I argue that code-switches and borrowings are not rigid categories, but points on a continuum with unambiguous cases of borrowings and code-switches located towards the ends of the continuum, while the more complicated cases are in between.

The differentiation between these two contact phenomena is problematic not only due to structural reasons, but also due to sociolinguistic factors. Elderly speakers utter even longer Russian sequences with Erzya phonology (placing the stress on the first syllable or pronuncing unstressed vowels as stressed, etc.), while younger speakers use Russian phonology even with established Russian borrowings. In my view, other sociolinguistic factors (e.g. profession of the speaker) might influence their pronunciation of Russian origin elements.

A possible way for telling apart borrowings and code-switches is described in Backus's 2015 usage based model. Studying frequency and the judgment of speakers, the degree of entrenchment of a certain Russian origin lexeme (e.g. the verb *hoťeť* 'to want') or even a construction (e.g. the necessive construction with the predicative adjective *dolžen*) can be defined. Moreover, the ease in the production and perception of these elements could be contrasted to the speakers' assessments concerning the question how well integrated these Russian origin stems and constructions are. Due to the nature of my data, the study of speakers' judgements was not part of my research aims. However, there are statements by the subjects in my study that reveal the speakers' opinion on the entrenchment of certain Russian elements. For instance, Russian numeral phrases (cf. example 74) are clearly understandable to all Erzya speakers, and it is rather the Erzya form that is difficult for them to produce, at least for a group of speakers, especially when compound numbers are concerned.

Another central question in the structural research on CS is the existence of constraints. In the Erzya–Russian bilingual discourse, CS is realized in a variety of ways, as we could see in section 7.2.2.2 in nominal constructions with a genitive modifier. Therefore, I agree with Muysken (2000) who considers the constraints tendencies and not rules. However, the lack of mixed examples in the case of negation suggests that some tendencies are stronger than the others. Another tendency which is possible in Erzya–Russian code-switching discourse involves a switch between the subject and the predicate. However, the pronominal subjects are predominantly Erzya, reversed cases (Russian pronominal subject – Erzya predicate) do not occur in the data. The tendency is so strong that it applies even to the utterances of C3 speakers (cf. example 178).

In the analysis of the data, I apply a combination of Muysken's 2000 typology and the MLF model (Myers-Scotton 2002). In the majority of the cases, the ML can be defined in the given utterance, so the MLF model works in these instances. However, constructions with the Erzya second infinitive and utterances with gender agreement present a problem. In these cases, the morphosyntactic frame of the utterance is formed by a composite ML. In addition to these examples, which are intances of congruent lexicalization, utterances with alternations (section 7.1) involving a total switch are also out of the scope of the MLF model.

I argue that these hybrid constructions are possible due to the high level of convergrnce between the two languages, which resulted in Erzya becoming more similar to Russian. Moreover, this process has accelarated in the last decades. On the basis of census data (Russian census 2010), we can claim that almost the entire Erzya community has become bilingual, the acceptedness of Russian constructions seems to have grown (cf. Luutonen 2014 on the use of Russian discourse particles by younger intellectuals). In order to assess these changes, the attitudes of the speech community towards Russian origin elements should be the subject of further studies.

The increasing Russian influence raises the question how endangered the Erzya language can be considered. As I mentioned in Chapter 1, Erzya is labelled as a definitely endangered language in the UNESCO framework, because intergenerational transmission is threatened (especially in urban communities) and domains of language use are shrinking. On Fishman's 1991 GIDS (Graded Intergenerational Disruption Scale), the Erzya speech community is in between level

6 (all generations use the language in an oral form and children acquire it as their first language) and level 7 (the parents' generation still speaks the language with the elderly but it is not transmitted to children). The categorization depends on which Erzya community (urban or rural) is under scrutiny. Language transmission is more effective in homogenous village communities as compared to the Russian dominant administrative centers and Saransk, the republic's capital. In EGIDS (Expanded Graded Intergenerational Disruption Scale) (Lewis and Simons 2010), the expanded version of GIDS, Erzya can be assigned to level 6b (Threatened) as it is used by all generations to communicate face to face, but it is continuously losing speakers. The urban communities show characteristics of EGIDS level 7 (Shifting) as the parents' generation can speak the language, but do not transmit it to the children. As my dissertation focuses on structural aspects of CS, and I carry out a qualitative study, the role of sociolinguistic factors cannot be determined unambiguously. However, it seems that there is a clear connection between the speakers' profession and the amount of CS they use in formal situations, but the knowledge of the High Variety does not seem to correlate with the language transmission intentions of the parents (leading cultural figures and even linguists choosing to transmit Russian to their children, while blue-collar workers are language activists). The validity of these observations should be investigated in future studies focusing on the attitudes of the speakers.

In order to describe the language shift process, data would be needed from semi-speakers and shifting diaspora speakers. Both my fieldwork data and the radio interviews were recorded from speakers who had a good command of Erzya, that is why they were invited into the radio or recommended to me by other community members. In my view, semi-speakers and shifting diaspora speakers could be assigned to C4 on the continuum which would show characteristics of the ML Turnover (Myers-Scotton 1998) with the majority of utterances having Russian as their ML into which Erzya elements would be embedded.

Variation in Erzya–Russian bilingual discourse can also be described as a type of diglossia. Although I mentioned in Chapter 4 that I do not elaborate on contact situations involving the major Finno-Ugric languages, one exception has to be made here. There are, namely, a number of parallel phenomena to the Erzya–Russian bilingual discourse in Lanstyák's 2011 description of the linguistic situation of Hungarians in Slovakia. On the one hand, the standard variety of the

minority language is restricted to the domain of education. In other formal situations, the standard variety of the majority language prevails. This results in a diglossic situation as concerns both the Hungarian community in Slovakia and Erzya speakers in Russia. Lanstyák (2011) argues that children acquire the standard variety only at the school and not in spontaneous situations. On the other hand, the High Variety and the Low Variety are becoming gradually more and more different, because the standard variety resists the influence of the majority language, at least the influx of loanwords and copied constructions is monitored, while the Low Variety is open to CS and borrowing. This diglossic situation is not an instance of classic diglossia (Ferguson 1959), as the division is not connected only to function and there are not any rigid norms where the semi-monolingual C1-type variety is allowed, although intellectuals typically apply the standard in formal situations.

8.3. Avenues of further research

In various subsections thoughout this dissertation, I have referred to further research possibilities. With my dissertation, I have hoped to create a basis for a later quantitative study relying on quantitative analysis of these CS types. Elicited data could also enhance the results obtained from naturally occurring data.

A correlation has been detected between the occupation and level of education of speakers, at least representatives of certain professions (e.g. journalists and teachers) tended to use fewer switches, especially in formal situations. Gender and generational differences in the use of CS would also need further research. This aspect of Erzya–Russian CS should be investigated further by planning a project in which speakers are selected in a representative fashion. Language attitudes and the extent of purism would also be worth studying, especially their connection to flagged switches.

In the paper, I have also pointed out two other further research aspects, a comparative and a diachronic one. On the one hand, a comparative-typological study of CS between Finno-Ugric and Slavic languages could reveal connections between CS patterns and shared mechanisms in these contact situations. On the other hand, studying the history of Erzya–Russian CS on the

texts available could provide evidence as to how the CS discourse evolved and what contact induced changes could be attributed to CS.

As far as CS research in general is concerned, the study of equivalence could be carried out further, to investigate what equivalence means from a cognitive point of view, and whether possible switch points can be predicted on the basis of this. Another aspect of this study could focus on entrenchment, the extent to which an element is integrated into the recipient language, by designing a methodology that relies on the ease of processing (reaction times of blinguals during both production and perception of code-switched forms).

Finally, the creation of a digital bilingual Erzya–Russian corpus could facilitate further investigations of the Erzya–Russian CS. As a next step, I intend to create this digital corpus on the basis of the Radio Vaygel interviews.

Chapter 9. Conclusion

In the present paper, I have reported on the investigation into the structural characteristics of intrasentential CS in Erzya–Russian bilingual discourse. In order to be able to describe the present scenario in the contact situation, the sociohistorical background of the Erzya–Russian relations has had to be taken into account as well as described (Chapter 2). The long duration of the contact explains how these two, typologically different languages have converged through time and have given rise to congruent constructions that facilitate CS further. The analysis of the present linguistic situation has revealed the fact that Russian has had an increasing influence on Erzya in the past decades, which has been especially visible when comparing my data to earlier accounts of Erzya–Russian CS (e.g. as far as the spread of gender agreement is concerned).

In Chapter 3, I have discussed the typological characteristics of the two languages. I have provided an overview of their phonological systems in order to explain why the similarity of the two cases makes the distinction between borrowings and code-switches nearly impossible. However, since my main focus is the examination of morphosyntactic features, I have intended to identify constructions that are (in)congruent to pinpoint possible switch places where certain CS types were predicted to occur. For example, I have described differences of the two case systems and of the constituent order in possessive structures. The discrepancy between the two languages in this respect predicted the emergence of mixed code-switched constructions. The differences in constituent order are relevant in the formation of numeral phrases expressing approximation. Finally, I have analyzed a unique feature of Erzya, the conjugation of nominals.

From section 3.2.2 to 3.2.4, I have described characteristics of verbal constructions, directing attention to the differences between past tense marking in the two languages, and related to it, the lack of the category of gender in Erzya. As a next step, I have discussed possessive constructions, both congruent structures involving the use of the copula, and the construction with the *habeo*-type verb which is present only in Russian. On the basis of this incongruence, I expected many hybrid code-switched utterances, but apart from sporadic examples, no such switching types have been attested in my data. In case of necessity, congruent structures do not

involve switching, the use of the Erzya verb *eŕavi* 'has to' and various forms of the Russian predicative adjective *dolžen* 'must' have been found to be used in parallel.

Finally, I have focused on word order in section 3.2.5, but I have not been able to find considerable differences in the two languages, as compared to the discrepancy between the constituent order in possessive nominal constructions. This, however, does not mean that no switches have been found to occur in utterances with a similar word order. On the contrary, congruence of the order of phrases has been found to prompt further switching.

Chapter 4 has presented an overview of earlier research into both contacts between minor Finno-Ugric languages and Russian (section 4.1), and the history of studies on the Erzya–Russian contact situation (section 4.2). The study of CS in other minor Finno-Ugric languages has revealed several parallel cases which characterize all the bilingual discourses. Researchers of these contact situations faced similar problems, even when methodology was concerned (e.g. the differentiation between instances of CS and borrowing). The comparison has uncovered parallel tendencies in the switching of discourse particles, adverbs and numeral phrases, to name a few. Sarhimaa's 1999 description of the Karelian–Russian mixed necessive constructions directed my attention to the characteristics of these constructions in the Erzya–Russian data. Apart from the similarities, the CS patterns in these minor Finno-Ugric languages have not always been found to coincide with the structures attested in Erzya. That is why, in section 8.3, I have argued for the need of further studies carried out from a comparative perspective.

Purism as an important factor in these contact situations was first mentioned in section 4.1, but I have analyzed its possible effects in section 4.2 in detail, because researchers studying the Russian influence in Erzya have always been biased and discussed the contact phenomena from a puristic point of view. These earlier studies revealed similarities between the present situation and the earlier periods of contact at different points of the 20th century, especially as far as verbal constructions and numeral phrases are concerned.

Chapter 5 has focused on the theoretical background of my study. First (in section 5.1), I have discussed problems concerning the definition of CS and the differentiation of CS from other contact phenomena. I have faced these challenges throughout my analysis, and in some borderline cases it proved impossible to make a distinction between CS and borrowing. This

decision would not have to be made if applying a diachronic approach, as these two contact phenomena could then be considered parts of the same continuum rather than distinct categories. Discrepancies in my data have corroborated the advantages of this approach.

In section 5.2, I have concentrated on the main questions posed when analyzing the structural aspects of CS. Following Muysken (2000), I have adopted a less strict approach that regarded constraints as tendencies rather than rules without exceptions. In the Erzya–Russian CS data, one of the most unambiguous constraints has concerned the inhibition of CS between the negative particle and the predicate, and between the negative verb and the connegative form. The final question in this subsection has been the existence (i.e. identifiability) (or not) of a matrix language. In my data, there are CS patterns that fit well into binary models such as the MLF model (Myers-Scotton 2002), whereas other, more mixed types are instances of a composite ML in which the two languages set the morphosyntactic frame together.

In section 5.3, I have introduced the models I have applied in the analysis of examples. The two main frameworks I have used are the MLF model (section 5.3.1) and Muysken's 2000 typology (section 5.3.2). I have based the analytical categories on the combination of these two models, differentiating between alternation, congruent lexicalization, and insertion; and dividing insertion into subtypes of ML constituents and EL islands (or mixed constituents with ML morphological markers).

I have also argued that we need to take into consideration Johanson's 1999 code-copying model and Backus's (2015) usage-based model which discuss the diachronic aspect of CS, i.e. its relation to contact induced change. However, the available diachronic (especially spoken) data on Erzya–Russian CS have not proven to be sufficient to draw conclusions on how this bilingual discourse has evolved, but the main tendencies could be described. Johanson's 1999 term "selective copies" could be applied to cases in which no lexical elements were switched, for example, as regards the use of Russian constituent order in possessive constructions with only Erzya lexemes and morphological markers (cf. section 7.2.2.2.2). I have also adopted the concept of entrenchment from Backus's 2015 model, because it could be used to explain the variation attested in Erzya–Russian CS discourse. For example, this model has accounted for the co-existence of different CS patterns in the verbal forms: Russian insertions into the utterance with

Russian or with Erzya morphological markers (section 7.2.3). In my paper, I have described these coexisting forms but also argued that the levels of entrenchment of these constructions could be determined in the course of further studies.

I have focused on Sebba's 2009 categorization next and adopted congruence as one of the main concepts in my analysis. Congruence facilitates switching, and if CS occurs, the constructions in the two languages are harmonized. For example, the fact that both languages have an attributive–head constituent order (except for possessive structures) makes the switch between attributes and heads possible. However, congruence does not mean that the switch actually happens. For instance, no switches have occurred in negated predicates, although the constructions are congruent in the two languages. If the switch concerns a non-congruent construction, the strategy of neutralization or compromise is used. In my data, neutralization has been found to be the prevailing strategy which involves the insertion of Russian EL islands (e.g. numeral phrases) into the Erzya utterance. In other cases, a compromise has been made between the two languages and a mixed construction has occurred which has not abided by the rules of either language. These examples could also be analyzed as instances of congruent lexicalization when using Muysken's 2000 terminology.

In addition to the structural frameworks of CS, I have also involved a pragmatic model (Gafaranga 2000 in section 5.3.5) in my study. I have applied his concept of medium repair in the analysis of flagged switches in my data.

Finally, I have discussed continuum models (Auer 1999 and Kovács 2001) on which I have based my Erzya–Russian CS continuum model (presented in section 8.1). I have used this model to display the variation attested in present Erzya–Russian bilingual discourse, but such a model could be also applied for the illustration of diachronic changes.

After discussing the sociohistorical, linguistic (typological) and theoretical background of the Erzya–Russian bilingual discourse, I have focused on my data in Chapter 6, characterizing the two data sets and pointing out the advantages and disadvantages of the types of data used, and, finally, discussing the methods of analysis.

Chapter 7 has presented the analysis of the data in three major sections. In section 7.1, I have focused on alternation which in part of the cases involved pragmatic function, mainly

citation and repetition. Section 7.2 contains the main body of my analysis, where I have focused on insertions and congruent lexicalization. The first subsection, section 7.2.1, has discussed discourse particles that have proved to be borderline categories between alternation and insertion. On the basis of my data, they are to be considered insertions, especially if they do not occur in a peripheral position.

From among nominal constructions (section 7.2.2), I have focused on numeral phrases that involved mainly Russian EL islands, but also hybrid structures used in expressing approximation. The incongruent constituent order also results in mixed constructions in possessive structures. I have argued that the Russian type order has been spreading as a result of CS, thus we are observing possible ongoing contact induced change.

Verbal constructions have been discussed in section 7.2.3. After I have described general categorization problems of verbal code-switches, I have focused on the different ways of expressing necessity, primarily on the predicative adjective *dolžen*. Although the majority of the examples involve the masculine and the plural form of the adjective, gender agreement has been found to be also relevant in some of the utterances with the adjective *dolžen*. Gender agreement, however, has been found more widespread with verbal predicates. On the basis of the attested utterances, two tendencies can be observed: the use of gender agreement seems to be spreading, and the point where the subject is located at the animacy hierarchy seems to influence the choice of the verbs form. While most elderly speakers have always applied the masculine past tense form of the Russian verbs, younger speakers already use the feminine forms with animate, especially [+ human] pronominal subjects.

Finally, I have discussed flagged switches which are also partly connected to alternation, but the majority of the flagged elements belong in the category of insertions and congruent lexicalizations.

In Chapter 8, the discussion part of my dissertation has followed in which I have summed up the main tendencies occurring in my data, and assigned the speakers to three categories on the basis of the CS patterns they apply. Finally, I have drawn a continuum of these categories to represent the variation in one model.

My hypothesis has been that there are structural reasons responsible for the emergence of mixed forms, and variation is caused by the speakers' different coping strategies when facing incongruence. This has proven only partly to be correct. In the analysis of flagged switches and alternations, the structural incongruences have explained CS only to some extent. Flagging, however, has been found to occur also in congruent constructions in which the flagging of the switch can be related to purism. There are examples of flagging that are instances of lexical gaps, where the speaker is looking for the right word. The ordering of the forms (Russian–Erzya vs. Erzya–Russian) is significant, since the former pattern can be explained by purism, while the latter with structural incongruence and lexical gaps. As a result, although structural (in)congruence has an explanatory force in the majority of cases involving mixed constructions, in a further complex analysis also pragmatic and sociolinguistic aspects would need to be taken into consideration.

Finally, the answers for the research questions posited in Chapter 1 are as follows:

"(1) What are the main types of insertional switches in the two Erzya–Russian data sets?"

The main types of insertional switches are discourse particles and adverbs, numeral phrases, prepositional phrases, possessive structures, and necessive constructions, accompanied by verbal switches inserted as EL islands or as parts of mixed constituents. Thus, CS patterns in Erzya–Russian bilingual discourse are partly similar to the types attested in other Finno-Ugric languages in contact with Russian. While Russian numeral phrases occur in many minority languages due to education in Russian, the structural transfer of possessive structures has not been described in other contact situations. Moreover, the emergence of gender agreement has proved to be a strong tendency in my data, which is significant as an indicator of the extent of Russian influence and the intensity of the contact.

"(2) Which CS constructions can be described with the binary model of the Matrix Language Frame model (MLF) (Myers-Scotton 2002)? In which cases does this binary model break down? Are these constructions instances of congruent lexicalization (Muysken 2000)? Can these cases be explained by the incongruence of the constructions in the two languages? What are the main strategies speakers use to realize CS in cases where the structures in the two languages are incongruent?"

The MLF model has been found to be applicable in the analysis of insertions but not concerning instances of congruent lexicalizations, in which the morphosyntactic frame of the utterance is provided by two languages, e.g. in bilingual numeral phrases expressing approximation, or mixed constructions showing gender agreement. Hybrid constructions have been found to occur as a result of incongruence. However, this compromise mixed form has not been the only strategy speakers have been found to apply in the case of incongruence, since neutralization, i.e. the avoiding of the "problematic" construction, is also a widespread strategy, and the speakers inserted Russian form as chunks, as EL islands in Erzya utterances.

"(3) How are CS and contact induced change connected? Are there any cases in which certain CS types might be indicators of ongoing change?"

There seems to be a clear connection between CS and contact induced change in that frequent CS contributed to the entrenchment of once foreign forms. In my opinion, the variation attested in the possessive nominal constructions and the spreading of gender agreement are indicators of ongoing change.

"(4) Can we define different CS styles? Is there a correlation between the amount and types of CS in speakers' language use?"

I have been able to identify different CS styles, ranging from a light-switcher speaker type I have designated as C1 speaker to a heavy-switcher C3 speaker with C2 speakers in-between, who switch more frequently than C2 speakers, but the ML of their utterances is still mainly Erzya. The CS styles have formed a continuum, the categories have not been defined as rigid entities, instead, as phases of a spectrum. While light-switchers have been found to apply mainly insertions, flagging, and switches with pragmatic functions, heavy-switchers seem to engage in frequent CS without a pragmatic function, and their utterances are often based on a composite ML, with mixed constructions characterizing their speech.

Bibliography

Aasmäe, Niina. 2014. *An introductory course of the Erzya language.* http://www.murre.ut.ee/arhiiv/naita_pilt.php?materjal=kasikiri&materjal_id=C0154&sari=C&formaat=A (Date of access: 30/11/2016)

Agafonova, Nina. 1995. Pronouns of mixed Mordovian dialects in the Samara region. In: Heikki Leskinen (ed.), *Congressus Octavus Internationalis Fenno-Ugristarum*. Jyväskylä: Kumerus, 17.

Agafonova, Nina. 2002. Роль языковых контактов в возникновении инноваций в смешанных мордовских говорах Заволжья [The role of language contact in the emergence of innovations in the mixed dialects of the Transvolga region]. In: Jorma Luutonen (ed.), *Volgan alueen kielikontaktit*. Turku: University of Turku, 65–69.

Aikhenvald, Alexandra Y. 2006. Grammars in contact: A cross-linguistic perspective. In: Alexandra Y. Aikhenvald and Robert M. W. Dixon (eds.), *Grammars in contact. A cross-linguistic typology*. New York: Oxford University Press, 1–66.

Alvarez-Caccámo, Celso. 1998. From 'switching code' to 'code-switching': Towards a reconceptualisation of communicative codes. In: Peter Auer (ed.), *Code-switching in conversation: Language, interaction and identity*. Routledge: London and New York, 29–50.

Auer, Peter. 1995. The pragmatics of codeswitching: A sequential approach. In: Leslie Milroy and Pieter Muysken (eds.), *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. New York: Cambridge University Press, 115–135.

Auer, Peter. 1999. From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *The International Journal of Bilingualism,* 3 (4): 309–332.

Backus, Ad. 1999. Evidence for lexical chunks in insertional codeswitching. In: Bernt Brendemoen, Elizabeth Lanza, and Else Ryen (eds.), *Language encounters across time and space*. Oslo: Novus Press, 93–109.

Backus, Ad. 2015. A usage-based approach to code-switching: The need for reconciling structure and function. In: Gerald Stell and Kofi Yakpo (eds.), *Code-switching between structural and sociolinguistic perspectives*. Berlin: Mouton de Gruyter, 19–38.

Bakker, Peter. 2003. Mixed languages as autonomous systems. In: Yaron Matras and Peter Bakker (eds.), *The mixed language debate: Theoretical and empirical advances*. Berlin/New York: Mouton de Gruyter, 107–150.

Bartens, Raija. 1999. *Mordvalaiskielten rakenne ja kehitys* [The structure and development of the Mordvin languages]. Helsinki: Suomalais-Ugrilainen Seura.

Bátori, István. 1980. *Russen und Finnougrier. Kontakt der Völker und Kontakt der Sprachen* [Russians and Finno-Ugrians. Contact between peoples and contact of the languages]. Wiesbaden: Veröffentlichungen der Societas Uralo-Altaica 13.

Biblehub 2016. http://biblehub.com/ (Date of access: 09/09/2016)

Bullock, Barbara E. and Almeida Jacqueline Toribio. 2009. Themes in the study of code-switching. In: Barbara E. Bullock and Almeida Jacqueline Toribio (eds.), *The Cambridge handbook of linguistic code-switching*. Cambridge: Cambridge University Press, 1–17.

Chan, Brian Hok-Shing. 2009. Code-switching between typologically distinct languages. In: Barbara E. Bullock and Almeida Jacqueline Toribio (eds.), *The Cambridge handbook of linguistic code-switching*. Cambridge: Cambridge University Press, 182–198.

Clyne, Michael. 2003. *Dynamics of language contact: English and immigrant languages*. Cambridge: Cambridge University Press.

Comrie, Bernard. 1981. *The languages of the Soviet Union*. Cambridge: Cambridge University Press.

Croft, William. 2000. *Explaining language change. An evolutionary approach*. Longman: Harlow.

Deuchar, Margaret. 2005. Congruence and Welsh–English code-switching. *Bilingualism: Language and Cognition*, 8(3): 255–269.

Edygarova, Svetlana. 2014. The varieties of the modern Udmurt language. *Finnisch-Ugrische Forschungen,* 62: 376–398.

EKM 2000 = Цыганкин, Дмитрий В. (отв. ред.) *Эрзянь кель, морфемика, валонь теевема ды морфология. Вузонь эрзянь ды финнэнь отделениянь тонавтницятнень туртов*

[The Erzya language, morphemics, word formation, and morphology. For the students of Erzya and Finnish majors at the university]. Саранск: Красный октябрь.

Endjukowski, Alexander Gawrilowitsch. 1930. К вопросу о русских заимствованиях в мордовском языке [On the issue of Russian borrowings in the Mordvin language]. *Учёные записки Саратовского университета*, 8:3. 55–112.

Estill, Dennis. 2004. *Diachronic change in Erzya word stress*. Helsinki: Société Finno-Ougrienne.

Feoktistov, Aleksandr and Sirkka Saarinen. 2005. *Mokšamordvan murteet* [Moksha-Mordvin dialects]. Helsinki: Suomalais-Ugrilainen Seura.

Fishman, Joshua A. 1991. *Reversing language shift: Theory and practice of assistance to threatened languages*. Clevedon, UK: Multilingual Matters.

Franks, Steven. 1995. *Parameters of Slavic morphosyntax*. New York and Oxford: Oxford University Press.

FSRS 2016 = Russian Federal State Statistics Service.
http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/afc8ea004d56a39ab251f2bafc3a6fce (Date of access 06/09/2016)

Gabelentz, Herr Conon von der. 1839. Versuch einer Mordwinischen Grammatik. *Zeitschrift für die Kunde des Morgenlandes,* II: 2–3, 235–284, 383–419.

Gafaranga, Joseph. 2000. Medium repair vs. other-language repair: Telling the medium of a bilingual conversation. *The International Journal of Bilingualism,* 4 (3): 327–350.

Gafaranga, Joseph. 2007. Code-switching as a conversational strategy. In: Peter Auer and Li Wei (eds.), *Handbook of multilingualism and multilingual communication*. Berlin: Mouton de Gruyter, 279–313.

Gafaranga, Joseph, Torras i Calvo, and Maria Carme. 2001. Language versus medium in the study of bilingual conversation. *International Journal of Bilingualism*, 5 (2): 195–219.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Gavrilova, Valentina Grigor'evna. 2012. *Дублирование как одно из проявлений переключения* [Doubling as one of the manifestations of code-switching]. *Финно-угорский мир*, 3 (4): 56–59.

Gavrilova, Valentina Grigor'evna. 2013. *Марийско-русский билингвизм: Переключение и смешение кодов* [Mari–Russian bilingualism: Code-switching and code-mixing]. Yoshkar Ola: Mari State University PhD dissertation.

Goss, Emily L. and Joseph C. Salmons 1990. The evolution of a bilingual discourse marking system. *International Journal of Bilingualism,* 4: 469–484.

Grammatika 1962 = Колядёнков, Михаил Никитич (ред.) 1962. *Грамматика мордовских языков. Часть I. Фонетика и морфология* [A grammar of the Mordvin languages. Part I. Phonetics and morphology]. Саранск: Мордовское книжное издательство.

Grenoble, Lenore. A. 2003. *Language policy in the former Soviet Union*. Dordrecht: Kluwer Academic Press.

Grosjean, François. 1982. *Life with two languages. An introduction to bilingualism*. Cambridge Mass.: Harvard University Press.

Gumperz, John J. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.

Haarmann, Harald. 1985. The impact of group bilingualism in the Soviet Union. In: Isabelle T. Kreindler (ed.), *Sociolinguistic perspectives on Soviet national languages: Their past, present and future*. Berlin: Mouton de Gruyter, 313–344.

Haarmann, Harald. 1998. Multilingual Russia and its Soviet heritage. In: Christina Bratt Paulston and Donald Peckham (eds.), *Linguistic minorities in Central and Eastern Europe*. Clevedon: Multilingual Matters, 224–254.

Hallap, Valmen. 1960. Mordva-vene segakeelsusest [On Mordvin–Russian mixed language]. *Keel ja kirjandus* 111: 220–223.

Halmari, Helena. 1997. *Government and codeswitching: Explaining American Finnish*. Amsterdam and Philadelphia: John Benjamins.

Halmari, Helena. 2005. "I'm forgetting both". L1 maintenance and codeswitching in Finnish–English language contact. *International Journal of Bilingualism*, 9: 397–433.

Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language*, 26: 210–232.

Heine, Bernd. 2008. Contact-induced word order change without word order change. In: Peter Siemund and Noemi Kintana (eds.), *Language contact and contact languages*. Amsterdam and Philadelphia: John Benjamins, 33–60.

Isaev, Magomet Izmailovich. 1977. *National languages in the USSR: Problems and solutions.* Moscow: Progress.

Isurin, Ludmila, Donald Winford, and Kees de Bot. 2009. *Multidisciplinary approaches to code-switching.* Amsterdam: John Benjamins.

Iznjamo 1949. https://fennougrica.kansalliskirjasto.fi/handle/10024/68002 (Date of access: 08/10/2016)

Janurik, Boglárka. 2011. Mixed structures in numeral phrases in the speech of Erzya–Russian bilinguals. In: Klára Agyagási (ed.), *Tractata Slavica* III. Debrecen: University of Debrecen, 127–141.

Janurik, Boglárka. 2013. Code-switching in an Erzya–Russian bilingual variety: An "endangered" transitory phase in a contact situation. In: Elena Mihas, Bernard Perley, Gabriel Rei-Doval, and Kathleen Wheatley (eds.), *Responses to Language Endangerment: In honor of Mickey Noonan.* Amsterdam: John Benjamins, 175–195.

Janurik, Boglárka. 2016. Varieties of Erzya–Russian code-switching in Radio Vaygel broadcasts. In: Reetta Toivanen and Janne Saarikivi (eds.), *Linguistic Genocide or Superdiversity.* Bristol and Buffalo: Multilingual Matters, 91–107.

Jerina, Olga. 1997. *Частицы в мордваских языках* [Particles in the Mordvin languages]. Tartu: Tartu Ülikool PhD dissertation.

Johanson, Lars. 1998. Code-copying in Irano-Turkic. *Language Sciences* 20: 325–337.

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

Johanson, Lars. 2002. Contact-induced change in a code-copying framework. In: Mari C. Jones and Edith Esch (eds.), *Language change. The internal, external and extra-linguistic factors.* Berlin and New York: Mouton de Gruyter, 285–313.

Juhász, Jenő. 1929. A moksamordvin–orosz nyelvkeveredés [Moksha Mordvin – Russian language mixture]. *Nyelvtudományi Közlemények* 47 (2): 296–300.

Kantele, Simo. 2016. *Venäläislähtöiset verbit udmurtissa – lainautumista ja koodinvaihtoa* [Russian origin verbs in Udmurt: Borrowing and code-switching]. Helsinki: Helnsingin yliopisto MA Thesis.

Kelin, Mihail Aleksandrovič. 1969. *Сложные слова сочинительного типа в мордовских языках* [Complex compounds in the Mordvin languages]. Саранск: Мордовское книжное издательство.

Keresztes, László. 1990. *Chrestomathia Morduinica*. Budapest: Tankönyvkiadó.

Keresztes, László. 2011. *Bevezetés a mordvin nyelvészetbe* [Introduction to Mordvin linguistics]. Debrecen: Debrecen University Press.

Kirkwood, Michael. 1991. Glasnost', 'The National Question' and Soviet Language Policy. *Soviet Studies* 43 (1): 61–81.

Kovács, Magdolna. 2001. *Code-switching and language shift in Australian Finnish in comparison with Australian Hungarian*. Åbo: Åbo Akademi University Press.

Kreindler, Isabelle T. 1985a. The Mordvinian languages: a survival saga. In: Isabelle T. Kreindler (ed.), *Sociolinguistic perspectives on Soviet national languages: Their past, present and future*. Berlin: Mouton de Gruyter, 237–264.

Kreindler, Isabelle T. 1985b. The Mordvinians. A doomed Soviet nationality? *Cahiers du monde russe et soviétique* 26 (1): 43–62.

Laakso, Johanna. 1996. Haisee provosoidulta: aineiston arvioinnin ongelmia Dmitri Tsvetkovin vatjan sanaston valossa [Suspicions of provocation. Problems of data assessment]. *Virittäjä* 100: 219–238.

Lallukka, Seppo. 1990. *The East Finnic minorities in the Soviet Union: An appraisal of erosive trends*. (Annales Academiae Scientiarum Fennicae, B-252.) Helsinki: Suomalainen tiedeakatemia.

Lanstyák, István. 2011. A határon túli magyar szókészlet eredet szerinti rétegei [Etymological layers in the vocabulary of Hungarians outside Hungary]. In: Szabómihály Gizella and Lanstyák István (eds.), *Magyarok Szlovákiában*, *VII. Nyelv*. Somorja: Fórum Kisebbségkutató Intézet, 83–111.

Lehiste, Ilse, Niina Aasmäe, Einar Meister, Karl Pajusalu, Pire Teras, and Tiit-Rein Viitso. 2003. *Erzya prosody*. Helsinki: Société Finno-Ougrienne.

Leinonen, Marja. 2009. Russian influence on the Ižma Komi dialect. *International Journal on Bilingualism*, 13: 309–329.

Lewis, E. Glyn. 1972. *Multilingualism in the Soviet Union*. The Hague: Mouton.

Lewis, M. Paul and Gary F. Simons. 2010. Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique*, 2: 103–120.

Lomshina, Svetlana Olegovna. 2011. К вопросу о мордовско-русских языковых контактах [On Mordvin–Russian language contact]. *Учёные записки Российского государственного социального университета*, 10: 181–188.

Luutonen, Jorma. 2014. Kahden sukupolven ersää – kielenhuoltoa ja muutoksen merkkejä [Erzya of two generations: Language planning and signs of changes]. In: Nobufumi Inaba, Jorma Luutonen, Arja Hamari, and Elina Ahola (eds.), *Juuret marin murteissa, latvus yltää Uraliin: Juhlakirja Sirkka Saarisen 60-vuotispäiväksi* 21.12.2014. Helsinki: Suomalais-Ugrilainen Seura.

Mahootian, Shahrzad. 1993. *A null theory of code switching*. Evanston: Northwestern University PhD dissertation.

Mahootian, Shahrzad. 2006. Code switching and mixing. In: Keith Brown (ed.), *Encyclopedia of language and linguistics*. 2nd edition. Amsterdam: Elsevier, 511–527.

Maticsák, Sándor. 2012. *A mordvin írásbeliség kezdetei* [The beginnings of Mordvin literacy]. Debrecen: Debreceni Egyetemi Kiadó.

Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics,* 36: 281–331.

Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

McConvell, Patrick and Felicity Meakins. 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics,* 25 (1): 9–30.

Milroy, Leslie and Pieter Muysken. 1995. Introduction: Code-switching and bilingualism research. In: Leslie Milroy and Pieter Muysken (eds.), *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. New York: Cambridge University Press.

Mosin, Mihail. 2002. Об особенностях мордовско-русского двухязычия у мордвы-эрзи в Республике Мордовия [On the characteristics of the Mordvin–Russian bilingualism of Erzya-Mordvins in the Republic of Mordovia]. In: Jorma Luutonen (ed.), *Volgan alueen kielikontaktit*. Turku: University of Turku, 160–166.

Mosina, Natal'ja Mihajlovna. 2002. *Эрзянская речь от детей 3 до 7 лет* [Erzya speech of 3- to 7-year-old children]. Саранск: Красный октябрь.

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

Muysken, Pieter. 2007. Mixed codes. In: Peter Auer and Li Wei (eds.), *Handbook of multilingualism and multilingual communication*. Berlin: Mouton de Gruyter, 315–339.

Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.

Myers-Scotton, Carol. 1998. A theoretical introduction to the markedness model. In: Carol Myers-Scotton (ed.), *Codes and consequences: Choosing linguistic varieties*. New York: Oxford University Press, 18–38.

Myers-Scotton, Carol. 2002. *Contact linguistics*. Oxford: Oxford University Press.

Myers-Scotton, Carol. 2003. Code-switching: Evidence of both flexibility and rigidity in language. In: Jean-Marc Dewaele, Alex Housen, and Li Wei (eds.), *Bilingualism: Beyond basic principles*. Bristol: Multilingual Matters, 189–203.

Myers-Scotton, Carol. 2006. *Multiple voices: An introduction to bilingualism*. Oxford: Blackwell.

Nagy, József. 2004. The Mordvins. A historical overview. In: György, Nanovfszky (ed. in chief), *The Finno-Ugric world*. Budapest: Teleki László Foundation, 91–94.

Ostler, Nicholas. 2006. *Empires of the word: A language history of the world*. London: Harper Perennial.

Paasonen, Heikki and Paavo Ravila. 1938 [1981]. *Mordwinische Volksdichtung* [Mordvin folklore]. Helsinki: Suomalais-Ugrilainen Seura.

Pineda, David. 2009. Переключение кодов или «смешанный лект»? [Code-switching or fused lects?] *Полярный вестник,* 12: 26–42.

Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: Towards a typology of code-switching. *Linguistics,* 18 (7–8): 581–618.

Poplack, Shana. 1988. Contrasting patterns of code-switching in two communities. In: Monica Heller (ed.), *Codeswitching. Anthropological and sociolinguistic perspectives*. Berlin and New York: Mouton de Gruyter, 215–244.

Prina, Federica. 2016. *National minorities in Putin's Russia: Diversity and assimilation*. London: Routledge.

Pussinen, Olga. 2010. Особенности языковой ситуации и русского языка в Мордовии [Characteristics of the language situation and of the Russian language in Mordovia]. In: Arto Mustajoki, Ekaterina Protassova, and Nikolai Vakhtin (eds.), *Instrumentarium of Linguistics: Sociolinguistic approaches to non-standard Russian*. (Slavica Helsingiensia 40.) Helsinki: Yliopistopaino, 106–126.

Pyöli, Raija. 1996. *Venäläistyvä aunuksenkarjala: kielenulkoiset ja -sisäiset indikaattorit kielenvaihtotilanteessa* [Olonets Karelian under pressure from Russian: Extralinguistic and linguistic indicators in language shift]. Joensuu: Joensuun yliopisto PhD dissertation.

Radio Vaygel. http://vaigel.ru/efir_er (Date of access: 10/11/2016)

Raun, Alo. 1988. The Mordvin language. In: Denis Sinor (ed.), *The Uralic languages: Description, history and foreign Influences*. Leiden: Brill, 96–110.

Romaine, Suzanne. 1995. *Bilingualism*, Oxford: Blackwell.

Rosignoli, Alberto. 2011. *Flagging in English–Italian code-switching*. Bangor: Prifysgol Bangor University PhD dissertation. http://e.bangor.ac.uk/5073/ (Date of access: 11/09/2016)

Rueter, Jack. 2007. *Subject–object versus topic agreement in Erzya*. http://www.ling.helsinki.fi/~rueter/MYV/Person/subjectObjectOrTopicAgreement.shtml (Date of access: 08/10/2016)

Rueter, Jack. 2010. *Adnominal person in the morphological system of Erzya*. (MSFOu 261.) Helsinki: Société Finno-Ougrienne.

Rueter, Jack. 2013. The Erzya language. Where is it spoken? *Études finno-ougriennes*, 45. https://efo.revues.org/1829?lang=fr (Date of access: 08/10/2016)

Rusgram 2016. http://rusgram.narod.ru (Date of access: 08/10/2016)

Russian Bible. http://allbible.info/bible/sinodal/mt/17/#!prettyPhoto[iframes]/19/ (Date of access: 08/10/2016)

Russian census. 2002. http://www.perepis2002.ru/index.html?id=17 (Date of access: 20/10/2016)

Russian census. 2010. http://www.perepis-2010.ru/results_of_the_census/results-inform.php (Date of access: 20/10/2016)

Saarinen, Sirkka. 2014. Переключение кодов в мокшанских диалектных текстах [Code-switching in Moksha dialect texts]. In: Нина Григорьевна Зайцева Ирма Ивановна Муллонен (ред.), *Всероссийская конференция финно-угроведов «Финно-угорские языки и культуры в социокультурном ландшафте России»*, Петрозаводск, 25–28 июня 2014 г., 540–543.

Salánki, Zsuzsanna. 2007. *Az udmurt nyelv mai helyzete* [The present situation of the Udmurt language]. Budapest: Eötvös Loránd Tudományegyetem PhD dissertation.

Salo, Merja. 1991. Mordvalaiset [The Mordvins]. In: Johanna Laakso (ed.), *Uralilaiset kansat: Tietoa suomen sukukielistä ja niiden puhujista* [The Uralic peoples: Information on the languages related to Finnish and their speakers]. Porvoo–Helsinki–Juva: WSOY, 156–184.

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

Savić, Jelena M. 1996. *Code-switching: Theoretical and methodological issues*. Belgrade: Belgrade University.

Schwarz, Solomon M. 1953. The Soviet concept and conquest of national cultures. *Problems of Communism, 2*: 41–46.

Sebba, Mark. 2009. On the notions of congruence and convergence in code-switching. In: Barbara E. Bullock and Almeida Jacqueline Toribio (eds.), *The Cambridge handbook of linguistic code-switching*. Cambridge: Cambridge University Press, 40–57.

Shagal, Ksenia. 2016. Contact-induced grammatical phenomena in the Russian of Erzya speakers. In: Arjava, Heini, Jouni Harjumäki, and Ksenia Shagal (eds.), *Mordvin languages in the field*. Helsinki: Suomalais-Ugrilainen Seura, 363–377.

Shirobokova, Larisa. 2011. *Az udmurt–orosz kétnyelvűség* [Udmurt–Russian bilingualism]. Budapest: Eötvös Loránd University PhD dissertation.

Stell, Gerald. 2015. Elusive or self-evident? Looking for common ground in approaches to code-switching. In: Gerald Stell and Kofi Yakpo (eds.), *Code-switching between structural and sociolinguistic perspectives*. Berlin: Mouton de Gruyter, 1–16.

Stell, Gerald and Kofi Yakpo. (eds.) 2015. *Code-switching between structural and sociolinguistic perspectives*. Berlin: Mouton de Gruyter.

Taagepera, Rein. 1999. *The Finno-Ugric republics and the Russian state*. London: Hurst and Co.

Thomason, Sarah G. 2001. *Language contact: An introduction*. Edinburgh: EUP.

Thomason, Sarah G. 2014. Contact-induced language change and typological congruence. In: Juliane Besters-Dilger, Cynthia Dermarkar, Stefan Pfänder, and Achim Rabus (eds.), *Congruence in contact-induced language change*: *Language families, typological resemblance, and perceived similarity.* Berlin and New York: Mouton de Gruyter, 201–218.

Timberlake, Alan. 1993. Russian. In: Bernard Comrie and Greville Corbett (eds.), *The Slavonic Languages*. London and New York: Routledge, 827–886.

Timberlake, Alan. 2004. *A reference grammar of Russian*. Cambridge: Cambridge University Press.

Treffers-Daller, Jeanine. 1991. Towards a uniform approach to code-switching and borrowing. In: ESF Network on codeswitching and language contact, *Papers for the workshop on constraints, conditions and models. European Science Foundation*, Strasbourg: ESF, 259–279.

Treffers-Daller, Jeanine. 1998. Variability in code-switching styles: Turkish–German code-switching patterns. In: Rodolfo Jacobson (ed.), *Codeswitching worldwide*. Berlin and New York: Mouton de Gruyter, 177–200.

Tsvetkov, Dmitri. 1925. Vadjalased [The Votic people]. *Eesti Keel,* 4: 39–44.

Turunen, Mikko. 1997. Nykyvatjan koodinvaihdosta [On code-switching in today's Votic]. *Virittäjä,* 2: 208–232.

Turunen, Rigina. 2006. Complex morphosyntactic features of nominal predicates in Erzya. *SKY Journal of Linguistics,* 19: 173–187.

UNESCO 2016. *UNESCO atlas of the world's languages in danger.* http://www.unesco.org/languages-atlas/index.php?hl=en&page=atlasmap&cc2=RU (Date of access: 08/10/2016)

Vasil'ev, Nikolaj Leonidovič and Nikolaj Dmitrievič Rusinov. 1992. Лексика, словообразование и фонетика среднерусских говоров Поволжья [The lexicon, word formation, and phonetics of varieties of Central Russian dialects in the Volga Region]. Саранск: Издательство Мордовского университета.

Verschik, Anna. 2014. Conjunctions in early Yiddish–Lithuanian bilingualism: Heritage language and contact linguistic perspectives. In: Heli Paulasto, Lea Meriläinen, Helka Riionheimo, and Maria Kok (eds.), *Language contacts at the crossroads of disciplines*. Cambridge: Cambridge Scholars Publishing, 33–58.

Vkontakte 2008. https://vk.com/topic-83623_1060400?offset=60 (Date of access: 08/10/2016)

Vkontakte 2010. https://vk.com/photo-1178752_117777774 (Date of access: 08/10/2016)

Wade, Terence. 2011. *A comprehensive Russian grammar*. Hoboken: Wiley-Blackwell.

WALS 2016. *The world atlas of language structures online* http://wals.info/ (Date of access: 09/01/2016)

Weinreich, Uriel. 1953. *Languages in contact*. The Hague: Mouton.

Wichmann, Søren and Jan Wohlgemuth. 2005. *Loan verbs in a typological perspective*. http://loanverb.linguist.de/Wichmann_Wohlgemuth_verb-borrowing-typology.pdf (Date of access: 23/09/2016)

Winford, Donald. 2003. *An introduction to contact linguistics*. Oxford: Blackwell.

Zabrodskaja, Anastassia. 2009. *Russian–Estonian language contacts: Grammatical aspects of language use and change*. Tallinn: Tallinna Ülikool PhD dissertation.

Zabrodskaja, Anastassia and Anna Verschik. 2015. Morphology of Estonian items at the interface of Russian–Estonian language contact data. *Sociolinguistic Studies,* 8 (3): 449–474.

Zaicz, Gábor 1998. Mordva. In: Daniel Abondolo (ed.), *The Uralic languages*. London: Routledge. 184–218.

Zamyatin, Konstantin. 2012. From language revival to language removal? The teaching of titular languages in the national republics of Post-Soviet Russia. *Journal on Ethnopolitics and Minority Issues in Europe*, 11 (2): 75–102.

Appendix

Fieldwork data

number	category	gender
2005		
1. 200501	2	F
2. 200502	1	F
3. 200503	2	F
4. 200504	1	F
5. 200505	2	F
6. 200506	3	F
7. 200507	1	F
8. 200508	1	F
9. 200509	1	F
10. 200510	1	F
11. 200511	3	F
12. 200512	1	F
13. 200513	2	F
14. 200515	2	F
2008		
1. 200802	2	F
2. 200803	3	M
3. 200806	2	F
4. 200808	2	F
5. 200811	2	F
6. 200812	2	F
7. 200813	2	F
2010		
1. 201001	2	F
2. 201006	2	F
3. 201007	2	F
4. 201009	2	F
5. 201011	2	F
6. 201013b	1	F
7. 201013c	2	M
8. 201014	2	F
9. 201015	2	F
2011		
1. 201101	1	F
2. 201102	2	F
3. 201103	2	F

Radio data

	number	gender	category
1.	20130401	F	1
2.	20130402a	F	1
3.	20130402b	M	1
4.	20130403	F	1
5.	20130408	M	2
6.	20130409	F	2
7.	20130411	M	2
8.	20130412	F	2
9.	20130416	F	1
10.	20130419a	M	2
11.	20130419b	F	1
12.	20130420	F	3
13.	20130422	M	1
14.	20130423	M	3
15.	20130429	M	3
16.	20130430	F	1
17.	20130514	F	1
18.	20130515	F	1
19.	20130516	F	2
20.	20130522	M	1
21.	20130523	F	1
22.	20130529	M	1
23.	20130530	F	2
24.	20130603	F	2
25.	20130604	M	2
26.	20130606a	F	1
27.	20130606b	M	2
28.	20130624a	F	2
29.	20130624b	M	2
30.	20130626	M	2
31.	20130628	M	3
32.	20130702	F	3
33.	20130703	F	2
34.	20130705	M	1
35.	20130708	M	2
36.	20130712	M	2
37.	20130716	F	2
38.	20130718	F	1

39.	20130722	F	1
40.	20130724	F	1
41.	20130726	F	1
42.	20130801	F	3
43.	20130805a	F	1
44.	20130805b	F	1
45.	20130808	F	2
46.	20130813	F	2
47.	20130815	F	3
48.	20130819	M	3
49.	20130820	F	1
50.	20130821a	M	1
51.	20130821b	M	1
52.	20130918	M	1
53.	20130924	M	2
54.	20130930	M	2
55.	20131004a	M	3
56.	20131004b	F	2
57.	20131016	F	1
58.	20131023	M	2
59.	20131202	F	1
60.	20131203	M	1
61.	20131223	F	1
62.	20140221	F	1
63.	20140221	M	1
64.	20140225	M	1
65.	20140303	F	3
66.	20140305	F	2
67.	20140318	M	3
68.	20140320	M	2
69.	20140321	F	1
70.	20140324	F	1
71.	20140326	F	1
72.	20140328	F	1
73.	20140407	F	1
74.	20140430	F	1
75.	20140505	F	1
76.	20140512	M	2
77.	20140514	F	1
78.	20140521	M	1
79.	20140523	F	1

80.	20140529	F	1
81.	20140602	F	1
82.	20140610	F	3
83.	20140616	M	1
84.	20140618	F	1
85.	20140624	M	1
86.	20140626	M	3
87.	20140708	F	3
88.	20140710	M	2
89.	20140714	F	2
90.	20140716	F	1
91.	20140718	F	1
92.	20140722	M	3
93.	20140813	F	1
94.	20140815	M	1
95.	20140916	M	1
96.	20141023	M	2
97.	20141024	M	3
98.	20141027	M	3
99.	20141028	F	2
100.	20141030	F	2
101.	20141107	F	1
102.	20141111	F	2
103.	20141215	M	1
104.	20141217	F	2
105.	20141223	F	1
106.	20150113	F	1
107.	20150119	M	1
108.	20150121	F	1
109.	20150123	F	1
110.	20150129	M	3
111.	20150202	F	2
112.	20150206	F	1
113.	20150311	F	1
114.	20150313	F	2
115.	20150317	F	1
116.	20150327	M	2
117.	20150423a	M	1
118.	20150423b	F	2
119.	20150520	M	2
120.	20150601	F	2

121.	20150602	M	2
122.	20150605	M	1
123.	20150611	F	1
124.	20150623	F	2
125.	20150702	M	1
126.	20150714	M	2
127.	20150716	F	2
128.	20150803	F	1
129.	20150818	F	3
130.	20150907	F	2
131.	20150908	F	2
132.	20150910	M	1
133.	20151002	M	1
134.	20151005	F	1
135.	20151006	M	1
136.	20151007	M	1
137.	20151008	F	3
138.	20151009	F	4
139.	20151012	F	1
140.	20151013	F	1
141.	20151014	F	1
142.	20151015	M	1
143.	20151016	F	1
144.	20151019	M	1
145.	20151020	F	3
146.	20151021	M	1
147.	20151102	F	1
148.	20151103	F	1
149.	20151105	F	1
150.	20151106	F	1
151.	20151109	M	3
152.	20151110	F	2
153.	20151112	F	1
154.	20151113	F	2
155.	20151118	F	1
156.	20151120	F	1
157.	20160115a	M	1
158.	20160115b	F	1
159.	20160119	F	1
160.	20160121	F	1