Towards establishing the matrix language in Russian-Estonian code-switching

A corpus-based approach

Anastassia Zabrodskaja
Tallinn University

This paper explores different options for establishing the matrix language in Russian-Estonian code-switching. First, the notions of matrix and embedded language are introduced, followed by a short overview of the Russian-Estonian bilingual situation in Estonia. The Matrix Language Frame model (Myers-Scotton 1993, 1997, 2002) is introduced in the third part of the paper, where the corporal data are described as well. A possible application of the Matrix Language Frame model to the Russian-Estonian code-switching data is presented in part four. It is argued that in most cases the matrix language is clearly definable. However, there are instances which can be best described as “congruent lexicalisation” (Muysken 2000).

1. Introduction*

By examining which language provides more (content and system) morphemes in an utterance and by looking at determiners, verbs and verb inflection, the first word in a sentence and word order, code-switching (henceforth CS) research (see Zabrodskaja 2007a:25–28) attempts to identify one of the languages involved in CS as the base or matrix language (henceforth ML) and the other as the embedded language (henceforth EL).

Clauses have been argued to be the most suitable units for a grammatical analysis of CS (see Myers-Scotton 1993; Clyne 2003). The ML can change from

*This paper is part of the work conducted on a Russian-Estonian code-switching corpus. This research was supported by the Estonian Science Foundation grant no 6151 “Child language, interlanguage and code-switching corpora creation and description”. The author is grateful to Tatjana Baškirova, Indrek Konnapere, Jekaterina Ozernova and Anna Kull for their help in data collection. I would also like to express my deepest gratitude to Prof Anna Verschik (Tallinn University) for her tremendous help in the data analysis, to Prof em. Carol Myers-Scotton (Michigan State University, University of South Carolina) and to the anonymous reviewers, whose comments helped to improve my paper considerably.
one clause to the next during a bilingual conversation. Myers-Scotton (1997) argues that the ML is the language whose grammar is dominant in a bilingual clause. A clause is in Language X if the grammar is that of Language X, and this irrespectively of the fact that content words may come from the other language. According to Myers-Scotton (1993, 1997), the term EL refers to the other language(s) into which CS may occur, but which are deployed to a smaller extent than the ML. Thus, the ML is the base language which sets the grammatical frame in mixed constituents, into which items from the other language, the EL, are inserted. This paper attempts to describe general trends in Russian-Estonian CS, and it aims to discuss some of the difficulties inherent in defining the ML in mixed Russian-Estonian clauses; it will be argued that sometimes a clearly definable ML simply does not exist. Constraints on CS proposed in the literature (e.g., Poplack 1980; McSwan 1999) remain out of the scope of the present article, especially as the universal applicability of such constraints is disputed (see Clyne 2003; Gardner-Chloros & Edwards 2004).

2. A short overview of the bilingual situation in Estonia

Russian is one of the ‘migrant’ languages spoken in Estonia and it has been in prolonged contact with Estonian, the majority language. Although Russian speakers in post-Soviet Estonia represent 31.7% of the population according to the 2000 census (see Statistical Office of Estonia), they cannot be considered typical minorities due to socio-historical factors. The migration of Russian speakers to the Baltic countries was encouraged during the Soviet occupation (1940–1941, 1944–1991) by the central authorities and the newcomers are reminiscent of colonisers more than of immigrant minorities because they never conceptualised Estonia as a separate country (on the non-typical language situation in Estonia see Kolstø 1995; Rannut 1995, 2004, 2008; Smith 1998; Verschik 2005, 2008:25–47).

While Estonian was officially formally taught in all Russian-medium schools as a subject during Soviet occupation, most Russians remained monolingual, because Russians were not expected to master Estonian. Despite the fact that Russian was a compulsory subject in all Estonian-medium schools, the proficiency in Russian among Estonians varied according to personal needs, occupation, work requirements and region of residence. In the predominantly Russian-speaking North-Eastern part of Estonia, proficiency in Russian was crucial for Estonians. In the other Estonian areas, functional bilingualism was characteristic of those Estonians who had to work in the public sphere, the civil service etc.

Radical changes came about in 1989, when the first Language Law decreed that Estonian, which had acquired a de facto minority status in its own territory, should attain full national status as the language of the state, administration and
Towards establishing the matrix language in Russian-Estonian code-switching

public discourse (Ozolins 1994:161). In 1991, the Republic of Estonia was restored de facto, with Estonian as the sole official language of the state. As a result, knowledge of the Estonian language among non-Estonian native speakers increased from 14% in 1989 to 44.5% in 2000, according to the census data (Statistical Office of Estonia). Although the census does not define proficiency and the data are anonymous and self-reported, such self-descriptions can nevertheless be seen as an act of identity. According to the 2000 census data younger Russian speakers have a better command of the Estonian language than their parents (see Statistical Office of Estonia) and use Russian with their parents, but are essentially moving towards using primarily Estonian, concurrently with leaving school and getting jobs (Zabrodskaja 2006a). Thus, bilingual communication has been emerging during the past decade (Verschik 2004, 2007, 2008; Zabrodskaja 2006a).

The following types of variation among local Russian speakers can be established (see also Rannut 2008:155–158; Verschik 2008:25–47):

1. Regional: Tallinn is bilingual, the North-East is predominantly Russian-speaking and the rest of the areas are predominantly Estonian-speaking.
2. Generational: younger people are more likely to know Estonian.
3. Individual: the predominance of Estonian or Russian may depend on identification with Estonia or Russia, proficiency in Estonian, conversational goals, social networks etc.
4. Internal diversity within the Russian-language community: the wish and the need to communicate in Estonian may vary between indigenous Russian-speaking groups (e.g., Old Believers\(^1\) etc.) vs. Soviet-era newcomers.

The degree of proficiency in Estonian and the practical needs determining its use vary greatly among the Russian-speaking population, but as their surrounding environment becomes more and more 'Estonianised', more and more Russian-Estonian CS can be detected.

3. Introducing the MLF model

Myers-Scotton (1993, 1997) proposes the Matrix Language Frame (MLF) model for describing 'classic' CS, which consists of switching between two languages in the same clause but with only one of the participating languages clearly supplying the

---

1. The Old Believers (starovery or staroobrjadcy) abandoned the Russian Orthodox Church after 1666–1667 in protest against church reforms introduced by Patriarch Nikon; in order to escape from religious persecution in Russia, they settled on the western coast of Lake Peipus in Estonia.
abstract morphosyntactic frame. The model is based on the distinction between the ML and the EL on the one hand, and the distinction between system and content morphemes on the other. Content morphemes (such as nouns and verbs) assign or receive thematic roles while system morphemes (such as determiners and clitics) perform neither of these functions. The distinction between system and content morphemes is however problematic. Firstly, as Muysken (2000:16–19) points out, there are at least four different criteria relevant to this kind of classification in different languages; also, the distinction does not operate in the same way across languages. Jake (1998:354) emphasises that “there is variation across languages in the assignment of particular lexical “concepts” to content or system morpheme status”. According to Myers-Scotton (2002), although the content and system morpheme division may not be entirely clear, it is quite clear-cut for most morphemes, and is therefore a useful analytical tool.

The grammatical frame of the MLF model is defined in terms of morpheme order and system morphemes (The System Morpheme Principle). Most function words are system morphemes and they have the feature (−thematic roles receiver/assigner) and often (+quantification). Most nouns, adjectives (+thematic role receivers), most verbs and some prepositions (+thematic role assigners) are content morphemes. Myers-Scotton (1997:77–78) distinguishes between three types of constituents:

| a. ML  | (= ML) islands: | all morphemes come from the ML |
| b. EL  | (= EL) islands: | all morphemes come from the EL |
| c. ML+EL | (= ML + EL) mixed constituents: | the system morphemes come from the ML, the content morphemes come only from the EL or from both the EL and the ML |

Islands only occur within a bilingual clause. They cannot be full sentences or full clauses. Islands are embedded in a bilingual clause. EL islands are occasionally what makes the clause bilingual; having single content words from the EL can also make the clause bilingual.

According to Myers-Scotton (1997), in a conversation with dense CS, the ML can be recognised on the basis of frequency. Roughly, the ML is the language which more grammatical morphemes come from. Yet, Myers-Scotton (1997:68) adds: “How large is large enough is an unresolved issue”.

Myers-Scotton and Jake (2001) later modified the MLF model. In the new version, there are two models which are closely linked: the 4-M model and the Abstract Level model (see Figure 1).

The notion of 4-M refers to the morphemes earlier defined as content and system morphemes. In the new model, there are four types of morphemes, one
Towards establishing the matrix language in Russian-Estonian code-switching

... type of content morphemes and three types of system morphemes (one of them “early” and two of them “late” system morphemes). According to Myers-Scotton and Jake (2000:1066), both content and early system morphemes are activated at the lemma level, but the latter do not assign or receive thematic roles, in contrast to the former. Early system morphemes are “indirectly elected” because content morphemes that are “directly elected” by a semantic/pragmatic feature bundle “point to them”. They may be in a different lemma from the content morpheme pointing to them (e.g., regular plural -s in English) or in the same lemma (e.g., irregular plural in English). Early system morphemes, for example the plural morpheme, can act similarly to content morphemes. Myers-Scotton and Jake (2000:1063) further suggest two types of late system morphemes – bridges and outsiders. Neither type is activated at the lemma level, and neither receives or assigns thematic roles.

In spite of increasingly more refined definitions of the ML, there is no clear agreement on the criteria for identifying the ML in CS utterances. While Doron (1983) and Joshi (1985) define it as the language of the first word in a sentence, for Klavans (1985) and Treffers-Daller (1991) it is the language of the verb or inflection phrase. Some researchers have suggested that the ML is determined by the language of the main verb (Klavans 1985; Treffers-Daller & van den Hauwe 1990).

It should be noted that the base language of a conversation and the ML of a clause are not the same entity. Dealing with Spanish-English data gathered in Gibraltar, Moyer (1998:222) suggests that bilingual conversations can be analysed on three levels. At the first (highest, to use her terminology) level of analysis, the main language is the language of the entire conversation. The criteria for its identifying can be social information about the community together with the language...
in which most words and morphemes are uttered (Moyer 1998:223). At the second level (intermediate, to use her terminology), the ML is determined by the frequency of use of both languages in conversational turns. At the third (lowest, to use her terminology) level the main language is recognisable by examining the frequency of switches between the two languages within a turn or “turn constructional unit”. However, the frequency is hard to define, and in the case of CS between closely related languages it is often unclear which item belongs to which language. So, the criterion of frequency of occurrence of each language on each level as proposed by Moyer (1998) are not applicable to all cases.

With regard to Russian-Estonian CS, Zabrodskaja (2007a, 2007b, forthc.) provides an exhaustive description of a Russian-Estonian CS corpus which consists of a wide range of oral and written text samples. The texts were collected from 2000 to 2008 by Anastassia Zabrodskaja, Anna Verschik and their students. The corpus contains CS samples from a variety of speakers with different degrees of proficiency in Estonian, interacting in different settings, belonging to different age groups etc. The corpus data come from two socially and demographically different areas, bilingual Tallinn and predominantly Russian-speaking Ida-Virumaa (the North-Eastern part of Estonia). The data include:

1. Excerpts from bilingual TV programs and Russian-language newspapers.
2. Language use in bilingual or Russian-language commercials, leaflets, written public information.
3. Rapid, on-the-spot transcriptions of bilingual conversations in shops, markets, (secondary) schools etc. in Tallinn, Kohtla-Järve and Narva.
4. Recorded interviews with Russian-speaking students of Tallinn University and Narva College of Tartu University as well as samples of CS in their written assignments and field notes on students’ everyday speech.

4. On the applicability of the MLF model to Russian-Estonian CS

In this study, CS is a term that covers the alternating use of two or more ‘codes’ within one conversational episode (Auer 1998:1). Following the approach in Romaine (2000) and Thomason (2001), it can be claimed that in addition to grammatical factors, sociolinguistic and conversational factors may play a role in CS and can override the rules of the two monolingual grammars.

In the bilingual conversations and interviews, Russian-Estonian switches are mostly intrasentential, i.e., they take place within a sentence or within a clause, or, more rarely, intersentential, i.e., they occur across two sentences or clauses. In Muysken’s (2000:60) terminology, these types of switching are referred to as insertion: a base language structure – Russian – is the matrix frame into which lexical items or entire constituents are inserted from another language, Estonian. Longer stretches of Estonian within Russian occur quite rarely.3

Thus, Russian-Estonian intra-sentential and constituent-internal or word-internal CS (or insertion) prevails in the corpora, but instances of alternational CS, where Estonian and Russian remain relatively separate in the clause (see Muysken 2000:96) and congruent lexicalisation, where Estonian and Russian jointly provide the grammatical structure of the clause, and the vocabulary comes from both languages (see Muysken 2000:122) can also be found. Such mixed production is probably caused by the fact that the social and demographic conditions favour different degrees of proficiency in the Estonian language and accordingly the production of differing types of CS by Russian-speakers.

The Russian-Estonian data suggest that in the case of frequent insertional CS, determining the ML is sometimes straightforward if we count morphemes (see Zabrodskaja 2006b, 2007c:133 on Russian-speaking students’ bilingual speech). In the corpus, there are plenty of examples where the ML can be unambiguously established on the basis of morpheme count, i.e., it is clear what language provides the grammatical frame. However, this section focuses on instances where a ML cannot be defined according to any criteria proposed in the literature. Russian is a fusional language: one morpheme may simultaneously encode more than one grammatical feature (e.g., case and number). Sometimes the determination of the ML is impossible according to quantitative criteria.

It is not entirely clear whether and how the MLF and 4-M models are applicable in cases where function markers are not added agglutinatively or are free function words. For example, in Estonian, quantity4 may distinguish case forms laine “wave” (2nd duration): “wave.nom” and laine (3rd duration): “wave.gen”. We cannot therefore talk of early or late system morphemes in such cases. Suprasegmental morphology is arguably realised together with the whole morpheme. Probably,

3. In the data from bilingual TV programs CS is mostly alternational but this phenomenon is affected by the genre; for instance, in the bilingual program Unetus/Bessonica “Insomnia” there were two hosts, one Russian-speaking and one Estonian-speaking. They used predominantly their mother tongue and switched codes due to pragmatic reasons; citation and repetition were the most frequent reasons for of CS.
4. Estonian has three quantities.
since suprasegmental morphological features are part of the word, they are early system morphemes (i.e., “indirectly elected”).

Both Estonian and Russian have a complex inflectional morphology. Estonian, a sociolinguistically dominant language today, has an even more developed inflectional morphology than Russian (see Zabrodskaja forthc. on the various possibilities of morphological integration of Estonian nouns into the Russian matrix frame).

4.1 ML and EL islands in Russian-Estonian CS

The following excerpt from a conversation between a teacher and a student demonstrates how the principles of the MLF model work. This is a typical example of the most straightforward, least challenging case.

The discussion begins in Estonian, but during the course of the interaction Russian phrases are inserted. The student alternates between two languages for pragmatic reasons. For her, it is more convenient to express herself through a Russian proverb (see Zabrodskaja 2006b, 2007c on pragmatic functions of CS in student speech). The Russian part of the sentence is in italics, Estonian in bold. In the glosses and translations, Estonian items are in upper-case:

(1) (i) S: Ma olen ju veel tei-le võlg-u.  
I AM AFTER ALL STILL YOUR-DAT DEBT-PART  
“I STILL OWE YOU SOMETHING.”

[T looks at her in wide-eyed astonishment. She does not understand that S has in mind an unsubmitted written assignment]

(ii) S: Mul on see ülesanne küll meelles. 
I-ADJS IS THIS TASK ENOUGH IN MIND  
“I REALLY HAVE THIS ASSIGNMENT ON MY MIND.”

(iii) T: Ah, jah-jah. Seda küll.
OH YES-YES. THIS-PART REALLY.

Mul endal enam pol-nud meeles. 
I-ADJS MYSELF-ADJS ANY MORE IS-PAST-PARTC IN MIND  
“OH, YES, YES. THAT’S RIGHT. I’D FORGOTTEN ABOUT IT MYSELF”

(iv) S: Aga ruk-i ne dohod-jat. 
BUT hand-PL no go-3PL  
“But there is no time.”

(v) T: Selge, saa-n aru. 
CLEAR, UNDERSTAND-1SG  
“That’s right, I understand.”

(vii) S: No do-j-dut objazateljno. 
But go-FUT-3PL certainly  
“But I will definitely make some time.”
In general, example (1) is in accordance with the MLF model. Turns (i)–(iii) and (v) are clear ML islands, where constituents consist entirely of ML morphemes (see Myers-Scotton 1997:78) that come from Estonian. Turn (vii) is a clear-cut EL island, where well-formed constituents originate from Russian and behave according to Russian grammar rules (see Myers-Scotton 1997:78). In (iv), CS between Estonian and Russian occurs (ML + EL constituents): this sentence contains morphemes from both the ML and the EL. According to Myers-Scotton (1997:77), the prototypical ML + EL constituent contains a singly-occurring EL lexeme (aga in our example) in a frame of any number of ML morphemes (ruk-i ne dohod-jat respectively). Here, the ML is clearly Russian because all the content and system morphemes (except for the Estonian early system morpheme aga “BUT”) as well as the syntactic structure are Russian.

4.2 The MLF model and double-marking

It is interesting to note that both the MLF model and the 4-M model are not intended to deal with CS phenomena such as double marking, where a grammatical function is marked by two functionally equivalent but structurally divergent strategies from the two languages (see Auer 1999:328).

In example (2a), double marking of the place adverb by functionally equivalent but structurally divergent strategies occurs: the same case relation is marked both by the Russian preposition v “in” and the Estonian inessive case marker -s (the whole stretch of talk is presented in Zabrodskaja 2007d).

(2) a. v kapi-s! in CLOSET-INESS “In the CLOSET!”

b. v škaf-u (Russian)
in closet-LOC “in the closet”

c. kapi-s (Estonian)
CLOSET-INESS “IN THE CLOSET”

This is an instance where a Russian preposition phrase corresponds to an Estonian noun in the inessive (internal locative) case. In Russian, the preposition v “in” denotes spatial relations, and the choice of an oblique case is secondary; that is, if the case marker is erroneously chosen or is not present altogether, this does not affect the intelligibility of the utterance (see example (2b)). In Estonian, locative cases belong to the group of so-called semantic cases (as in example (2c)), whose meaning is not entirely abstract, unlike that of the three grammatical cases (nominative, genitive and partitive). Thus, the same meaning indicated through
the internal locative case is rendered by the function markers of both languages (a Russian preposition and an Estonian case-marker).

According to the 4-M model, both a Russian preposition and an Estonian case-marker would be late outsider system morphemes. Estonian kapī “CLOSET. stem” is a content morpheme. Given the occurrence of outsider system morphemes from Russian and Estonian and an Estonian content morpheme, on what grounds could we determine the ML in this particular clause? This question still remains unanswered because the System Morpheme Principle only allows double marking in early system morphemes, not late system morphemes such as case. Myers-Scotton (1997:110) calls this phenomenon “simultaneous system morphemes (double morphology)”: 

The formulator accesses not only ML system morphemes but also those EL system morphemes which are at the same lemma address as an EL noun or verb stem (or somehow automatically accessed with the stem). Here, however, the resemblance between the morphological doublets ends: these EL system morphemes have no relationships external to their heads; i.e., they show no interrelations with other items in the sentence, such as agreement.

Auer (1999) describes the evolution of bilingual speech as a movement along a continuum that starts with CS and ends up over a period of time in a conventionalised fused lect via code-mixing. In his model of bilingual speech development, Auer (1999) proposes different stages of conventionalisation (“grammaticalisation” being the authors’ term) of CS. When a pattern of CS is grammaticalised, it becomes obligatory and regular, and the contrast between new and old items fades away. The example of double marking (v kapī-s) represents the third stage in Auer’s continuum of CS, code-mixing and fused lects. According to Auer (1999:323–329), who also discusses examples of double marking as in example (2a), this strategy is characteristic of a more advanced stage of proficiency in two languages (the stage of “code-mixing” as opposed to the earlier stage of “CS” and the later stage of “fused lects”, to use his terminology). Note, however, that the strategy employed in example (2a) occurs in the speech of a Russian-speaking child who has limited command of Estonian.

4.3 The ML in Russian-Estonian CS across sentence or clause boundaries

As will be shown in example (3a), in some cases it is not possible to decide unambiguously what the basic language is in the Russian-Estonian data.

(3) a. See že on legko!
    THIS but IS easy
    “But THIS IS easy!”
Towards establishing the matrix language in Russian-Estonian code-switching

b. See on ju lihtne (Estonian)
   THIS IS BUT EASY
   “BUT THIS IS EASY!”

c. Éto že legko (Russian)
   This but easy
   “But this is easy!”

Code-switched utterances like (3a) compared with non-code-switched examples (3b) and (3c) prove that the grammatical rules used to construct CS utterances may be drawn from both languages. CS is dense because every second word is code-switched. The sentence was pronounced quickly and without hesitations; I would call the speaker a proficient code-switcher, with an excellent command of both languages. Russian word order is predominant in the utterance. Following Klavans (1985) and Treffers-Daller & van den Hauwe (1990), Estonian would be the ML because the subject and predicate are in Estonian. However if we count the number of morphemes, the ML would be Russian: Russian že, leg-, -k-, -o (four morphemes) and Estonian see on (two morphemes).

In recent formulations, the ML is thought to provide the majority of system morphemes (Myers-Scotton & Jake 2001). Here, the content morphemes are an Estonian verb on and a Russian adverb legko, while early system morphemes are Estonian see and Russian že. The Estonian pronoun see “THIS” opens the sentence. Russian že “but” is a focus particle. The verb on “BE.3sg” is clearly Estonian and follows Estonian inflection. Russian legko “easy” is an adverb. The syntactic position of each item does not require grammatical markers: see is a nominative\(^5\), že is a discourse particle, on is a fusional form, legko is an adverb. In such cases, where there are no markers and every other word is code-switched, the determination of the ML appears to be an impossible task. Alternatively, the definitions of content morpheme and system morpheme might be reconsidered. In their 4-M model, Myers-Scotton and Jake (2000) have on the one hand content morphemes and on the other early and late system morphemes (e.g., see from example (3a) and za- from example (6) respectively). However, the question arises as how one should treat cases such as a Russian adverb leg-k-o, which clearly consists of three morphemes (stem-adjectival/adverb suffix-adverb suffix), but which does not comply with the proposal in Myers-Scotton and Jake (2000:1059), according to which analogous cases are considered as a single content morpheme: “[...] within the thematic grid, only those morphemes that directly receive or assign thematic roles are content morphemes”.

---

5. Nominative is not morphologically marked in Estonian.
The analysis of example (3a) shows that the base code sometimes just does not exist. In these cases congruent lexicalisation (see Muysken 2000:122–153 and Section 4 above) occurs. Generally it is believed that because of the higher degree of equivalence, these types of switching occur more often with closely related languages (in fact, it is probably the dominant type of CS in these cases), but since typological distance is a gradient phenomenon, one would expect it to affect CS between unrelated languages as well.

In example (4), the two first words are in Estonian, while the rest of the sentence is in Russian. Here we have a combination of two clauses in different languages. The switch occurs before the subordinate clause; thus we have inter-clausal CS or alternation in Muysken’s (2000) view.

(4) Nad küsi-si-d, est’ li u nas v prodaže šerst’.
THEY ASK-pst-pl is Q at us on sale yarn
“THEY ASKED whether we sell yarn.”

Example (5) shows that the sentence can begin and end in Estonian, but even the combination of a compound noun (järelek-sam) at the beginning of the sentence and a verb (sobi-b) at the end does not determine the ML.

(5) Järeleksam v sledujuščij četverg budet, Kat’ka skaza-l-a čto ej tože sobi-b.
AFTER-EXAM on next Thursday will be, Kat’ka say-pst-3sg that her also SUIT-3sg

“The ADDITIONAL EXAM will be on next Thursday, Kat’ka [colloquial form of Katja] said it SUITS her too.”

The ML of the conversation is Russian. The use of the Estonian noun järelek-sam “ADDITIONAL EXAM” may be conditioned pragmatically and semantically. Russian-speaking students employ in their casual speech Estonian nouns referring to the university domain. The motive for such insertion is to express concepts whose appropriate Russian equivalents are not as readily available to them (like järelek-sam “ADDITIONAL EXAM”, ainepunkt “CREDIT POINT”, matrikkel “STUDENT’S RECORD-BOOK”). The inflection of the Estonian verb sobi-b is Estonian. That is a potential problem for the MLF model, unless the 3sg inflection is “directly elected” with the stem. At the same time it can be claimed that the inflectional morpheme comes with the stem. Note that the Estonian verb sobib has strong modal semantics: it has been grammaticalised into an independent discourse particle (see Keevallik 2003). Thus in example (5) sobib may be repeated as an integral whole because of grammaticalisation. The question arises whether forms such as sobib, which are both fully inflected verb forms (sobi-b) and grammaticalised discourse particles (sobib) can have a distinct lemma (-b in our case) or not. This is a psycholinguistic question that can not be answered here.
4.4 The ML in the Russian-Estonian constituent-internal or word-internal CS

In example (6), an Estonian content morpheme with two Russian late system morphemes (prefix and suffix) opens the sentence:

(6) Za-maks-aj, požalujsta, i za menja, ja zavtra
    pref-PAY-suf.imp2sg please and for I.gen I tomorrow
    tebe den'gi vernu.
    you-dat money return-fut-1sg
    “PAY for me too, please, I will pay you back tomorrow”

The first content morpheme is code-switched but this does not lead to the use of Estonian in the rest of the sentence. This is a clear example of insertion, because only the stem maks- is Estonian. The ML of the conversation is Russian because other morphemes, combinatorial rules and the word order are Russian.

Example (7) is a Russian clause with two inserted Estonian content morphemes, the subject noun and the direct object noun; and both are inserted into the Russian matrix grammar.

(7) Skol'ko nam keeleteadus ainepunkt-ov da-jot?
    how many us LINGUISTICS CREDIT-POINT-gen pl give-3sg
    “How many CREDITS does the LINGUISTICS (course) give us?”

Here we have Russian grammatical morphemes and a verb. The code-switched subject keeleteadus “LINGUISTICS.nom.sg” is an Estonian word in the nominative. The code-switched noun ainepunkt-ov “CREDIT-POINT-gen pl” is an Estonian noun in the nominative ainepunkt “CREDIT-POINT” with the Russian plural genitive ending -ov. The Estonian item ainepunkt can be interpreted in terms of Russian morphology as a masculine second declension noun with zero ending, because punkt is a common “internationalism” in the two languages. However, the sentence is not entirely intelligible either to a monolingual Russian or to monolingual Estonian speakers.

Still, example (7) shows that the main verb cannot be decisive for the determination of the ML. Here, the main verb dajot “gives” is Russian, so the ML should be Russian, but if we count the morphemes, the ML appears to be Estonian. On the other hand, the morphology and word order is Russian. On what basis should we then define the ML?

5. Conclusions

The ML as defined in Myers-Scotton (2002:66) is an abstract theoretical concept, not a descriptive tool. As was shown above, such a definition of a ML is problematic
in light of the Russian-Estonian data. In the MLF model, the ML is claimed to provide the morphosyntactic frame of the utterance (Myers-Scotton 1993).

The MLF model does not deal with convergence either in its old or in its new version; yet as Backus (2004, 2005) corroborates, convergence is connected to CS and via that to language change. This is why instances which are problematic for the MLF model are particularly interesting.

The data discussed in this paper show that Russian-Estonian CS is mostly inser
tional or alternational. Russian speakers basically use isolated items from Estonian in a Russian ML or introduce Estonian lexical morphemes combined with Russian grammatical morphemes. It should be stressed that whatever strategy is employed, it is the innovative character of its use that attests to the fact that Estonian Russians are creative communicators (e.g., example (3a)).

*Congruent lexicalisation* as proposed by Muysken (2000) is a more appropriate concept for describing dense CS, because it takes into account both morphemes and combinatorial principles (morphosyntax, word order etc). Congruent lexicalisation is a valid alternative to proposals for a composite ML comprising grammatical morphemes from both languages. At the same time, instances where no morphemes from the second language occur but where the morphosyntactic pattern is clearly ‘foreign’ require an explanation, and here the concept of a composite ML cannot help.

As the determination of the ML according to the MLF model turns out to be difficult for different reasons, more research is required to indicate what kind of model would be more appropriate for determining the ML on the basis of the Russian-Estonian CS corpus. I suggest that if the ML is an empirically-driven concept and not an abstract theoretical construct with universal application, then it is to be expected that sometimes the ML is simply not definable.

References


Toward establishing the matrix language in Russian-Estonian code-switching


