Morphology of Estonian items at the interface of Russian-Estonian language contact data

Anastassia Zabrodskaja and Anna Verschik

Abstract
This paper argues that integration of code-switched items into the Matrix Language is not a straightforward process and does not entirely depend on the structural characteristics of the languages involved. Estonian-Russian code-switching is instructive because both languages are rich in inflectional morphology. Full integration of Estonian nouns into the Russian matrix means gender assignment and adding of inflection (case, number, or case and number). Russian-Estonian code-switching data demonstrate a wide range of possibilities for integration. Zabrodskaja (2009) specifies six gradations of morphological integration, whereby morphological integration and phonological adaptation do not depend on each other: an Estonian item may show high phonological adaptation and no morphological integration into Russian, and vice versa. Empirical observations show that there are instances where Russian inflectional morphology is absent, although the Estonian noun fits structurally into Russian declension classes. This paper compares the integration of Estonian nouns into Russian morphosyntax in speech and in blogs. Both sets of data show similarities as far as types of integration are concerned. Morphological integration is often a matter of individual preference and variation rather than structural compatibility with the Russian matrix.

Keywords: bilingual speech, code-switching, computer-mediated-communication, morphology, Estonian, Russian
1 Introduction

The aim of this paper is to establish how principles of morphological integration of Estonian nouns into Russian (formulated by Zabrodskaja, 2009) work in two sets of data – oral communication and blogs. We compare the data collected among young bilingual Russian and Estonian speakers in the context of Estonian secondary schools in 2010–2013 with the data from Russian-Estonian code-switching (henceforth CS) in Russian-language Live Journal blogs (data collected in 2009–2011).

Estonian is the sociolinguistically dominant language, as the official language of the country. Estonian is an agglutinating inflectional SVO Finno-Ugric language, known for its rich morphology. Belonging to the family of Indo-European languages, Russian is a fusional East-Slavic language that also has a rich inflectional morphology but to a lesser extent than Estonian. The contact situation between two languages with highly developed inflectional morphology provides new empirical data relevant to test different approaches suggested in language contact theory. We analyse Russian-Estonian insertional CS: thus, the base language structure is Russian, into which lexical items or entire constituents are inserted from Estonian.

According to most Russian grammars published in Russia, there are three declension classes (see Table 1); this classification is also used in the present paper:

Table 1. Russian declension classes.

<table>
<thead>
<tr>
<th>1st declension class</th>
<th>2nd declension class</th>
<th>3rd declension class</th>
</tr>
</thead>
<tbody>
<tr>
<td>feminine and masculine nouns ending in -a and -ja</td>
<td>masculine nouns with zero ending</td>
<td>feminine nouns ending in a palatalized consonant</td>
</tr>
<tr>
<td>neuter nouns ending in -o</td>
<td>neuter nouns ending in -e</td>
<td>neuter nouns ending in -m’a</td>
</tr>
</tbody>
</table>

We investigate morphosyntactic integration of Estonian lexical material in Russian with regard to gender assignment and case marking. Similar research was undertaken on gender / case assignment by Leisiö (2001a, 2001b) and Chirsheva (2009), who made their case studies on Russian-Finnish and Russian-English data respectively, both looking for integration strategies of Finnish and English items in Russian matrix applied by bilingual speakers with Russian as L1. Leisiö (2001a, 2001b) found that bilinguals do not need the same mechanisms of adaptation as monolinguals.

We explore cases of morphosyntactic integration, or lack thereof, of Estonian insertional code-switches into the Russian matrix and compare data from oral speech and Russian-language blogs. The paper is organised as follows: It opens...
with a short overview of the previous research on the Estonian-Russian sociolinguis
tic situation. Second, certain theoretical considerations are introduced. Then the methodology of the study is described in detail. After, types of integration of Estonian nouns in oral and blog communication are compared. Such a comparison is necessary because, as Gardner-Chloros (2009:110) states, ‘CS between the same language-pairs in different sociolinguistic settings’ might exhibit ‘radically different’ CS patterns. Thus, it would be instructive to compare integration types and degree, looking at how different sociolinguistic factors and context interact with each other. A separate section deals with script choice and CS (applicable to blogs only). Finally, conclusions are drawn.

2 A brief overview of previous research on Estonian-Russian language contacts

Modern Russian–Estonian language contacts can be characterized as ‘multiple’ because, on the one hand, Estonians have some competence in Russian and, on the other hand, Russians have started using Estonian extensively (see Verschik, 2007). The setting can be viewed as a language laboratory since one can explore contacts between the same languages and Russian in a changed sociolinguistic situation. To reformulate it in a more figurative way, we quote Jacobs (2005:271): ‘The list of languages present in the [sociolinguistic] “soup” remained the same, while the “recipe” changed’.

Within the last decades, a number of related quantitative and qualitative studies have been carried out to describe the outcomes of this language contact situation, such as Ehala (1994, 2000, 2009, 2011a, 2011b) on Russian influence on Estonian, and Verschik (2008) and Zabrodskaja (2009a, 2009b, 2009c, 2010, 2013) on Estonian influence on Russian. The studies imply the emergence of a new variety called Estonian Russian. The study by Zabrodskaja (2013) shows that whole morphosyntactic patterns that often contradict monolingual Russian norms have been borrowed; for instance, the word order in the NP with the attribute in the genitive. Empirical data from Estonian Russian (Verschik, 2004, 2006, 2008; Zabrodskaja, 2009a, 2009b, 2009c, 2010) show that there are indeed constructions conventionalized in local Russian that are at odds with monolingual grammar rules.

3 Theoretical background

Almost all language contact studies on the topics of CS begin by trying to define such elusive terms like host language and guest language. For example, Myers-Scotton (1997 [1993]:82) states that the matrix language (henceforth ML) is the
base language which sets the grammatical frame in mixed constituents, that is, the frame into which items from the other language, the embedded language (henceforth EL), are inserted, retaining the word order and general structure of the ML. Her model is severely criticized for monolingual bias. Indeed, conceptualization of language contact as the confrontation of two monolingual systems is way too simplistic (see references in Zabrodskaja, 2009). This is a synchronic model, and it does not take into account conventionalization of innovations and emergence of a new non-monolingual variety.

Usually, CS models focus too exclusively on lexical aspects and thereby tend to ignore the structural effects CS can bring about. With that in mind, Clyne (2003) suggests looking at compromise forms that are absent from two monolingual grammars.

While considering the constraints and principles, we have seen that there are different positions with respect to CS. If constraints state what is possible, then general principles of the Matrix Language Frame model suggest what is probable. Oftentimes, the grammar of CS is viewed in terms of constraints. For instance, both Poplack (1980) and Myers-Scotton (1993 and later versions) in fact focus on two monolingual grammars. While the former lists restrictions on possible switches, the latter suggests principles that explain morphosyntactic structures on the basis of monolingual equivalents. Russian-Estonian CS data support Clyne’s (1987) claim, who points out that although the grammatical constraints may describe some general tendencies, other factors may play a role in CS. A researcher never knows how objective the constraints proposed by language contact theorists are. Empirical data show that constraints are often violated (Clyne, 1987, 2003; Zabrodskaja, 2009).

Monolingual bias has governed the field of contact linguistics since the times when its acknowledged pioneer Weinreich (1966 [1953]:7) stated that based on Lotz’s (1950:712) assumption that ‘every speech event belongs to a definite language’ it was possible to determine in an utterance some elements which belong to another language.

We argue against the claim that ‘every speech event belongs to a definite language’. One should bear in mind that sometimes it can be hard to determine what element belongs to what language and why; for example, this happens in Finnish-Estonian data where the languages in contact are closely related (Praakli, 2009; Frick, 2013). Verschik (2012) shows very convincingly that there are compromise forms and new creations that are not unambiguously assignable to a particular language. But such ambiguity may also be present in contacts between non-related languages. Thus, an analysis of CS as something that combines two monolingual varieties is wrong per se. It is not always possible to establish firmly
what item belongs to what language in Russian-Estonian bilingual utterances, even though the languages belong to different language families.

In the mid-1990s, Milroy and Muysken (1995:2–3) pointed out that in linguistics a monolingual speaker in a homogenous speech community is regarded as the norm. In contact linguistics, researchers are not concerned with ‘deviation’ from the respective monolingual norms. However, a monolingual bias exists in the sense that bilingual speech is frequently analyzed in terms of two monolingual varieties, e.g. when ‘linear’ CS models postulate constraints on switches that create ‘morphosyntactic conflict’; constituents in CS that are ‘well-formed’ in the terms of respective monolingual grammars, etc. (Backus, 1999; Auer, 2007).

The results of bilingualism for an individual speaker cannot be considered a simple ‘overlay’ of one language over another (Auer, 2007) because the resulting amalgamated constructions and items cannot always be labelled as belonging to a particular variety (see Zabrodskaja, 2013).

Researchers such as Myers-Scotton (1993) assume oftentimes (unconsciously, we presume) that the ML and the EL are indeed monolingual varieties. Accordingly, two monolingual grammars of the monolingual standard varieties that are in contact allegedly operate in the code-switched clause. But the phenomenon is more complex: grammar can mean different things. As Gardner-Chloros and Edwards (2004:105) demonstrate, the term ‘grammar’ can have a variety of meanings (dictated by particular linguistic rules or theories of principles and constraints that underlie the syntax and morphology, etc.) and it is not clear whether and how common assumptions about (monolingual) grammars are applicable to non-monolingual speech.

In this study, the continuously changing sociolinguistic situation and growing knowledge of Estonian among Russian-speakers has to be taken into account. For a microsociolinguist, patterns of Russian-to-Estonian communication are more subtle and diverse than just the opposition between two monolingual varieties. This is especially true of Tallinn, where people with differential degrees of command in Estonian and in Russian interact on a daily basis in institutions, over the service counter, in schools, blogs, etc.

Another question that a microsociolinguist might ask is exactly what kind of Russian (or Estonian) it is that speakers use? The utterance could belong to a new cluster of contact varieties, tentatively called Estonian Russian. We would also emphasize that Estonian Russian is not a clear-cut, fully crystallized variety; it is cover label because this new phenomenon has not yet become conventionalized and definitely varies across speakers. How internally consistent it is, and how different from standard Russian, is an empirical matter, and will be decided on
whether we can indeed legitimately talk about Estonian Russian as a new variety, or as a cluster of varieties.

4 Methodology: Data, informants and contexts

As argued by Jarvis and Pavlenko (2008:28), the methodology one adopts for the linguistic analysis depends on the scope of the investigation. The scope of our paper, on the one hand, is an exploration of Estonian-Russian contact phenomena in terms of morphosyntactic contact-induced changes whose result can be seen in the morphological integration of Estonian-language items inserted into seemingly monolingual Russian utterances.

The main challenge for the contact linguistic study of morphological integration lies in the fact that naturally occurring morphologically integrated items have a low type and token frequency. It is not uncommon that analyses often use only a handful of examples to draw far-reaching conclusions about the nature and constraints on morphological integration. For this reason, it is not possible to rely exclusively on naturally occurring oral data in real contact interaction. Instead, a triangulated methodological approach involving two main data sources is used here. This paper draws on a corpus of naturally occurring instances of morphologically integrated items in oral data and in written blogs. Such collected data provide the empirical basis for the integration phenomena that do occur in Estonian-Russian contact settings.

Our research is qualitative and we rather seek to answer the question ‘what?’ rather than to provide a thorough quantitative analysis. Nevertheless, we believe that it would be useful to suggest some figures in Table 3 (see Table 4 for written blog data in Section 4.2). We include not only nouns but noun phrases (NPs) N GEN + N NOM because the border between such NPs and compound nouns in Estonian is fuzzy. Potentially, a NP with attribute in the genitive can become conventionalized as a compound noun if it is used frequently in a contact situation. For instance, there are in-between cases: seltskonнатäht ‘celebrity’ (literally, ‘society star’, seltskonna ‘society.GEN’ + täht ‘star’) can be also rendered in writing as seltskonna täht (for more on Estonian compound nouns in Russian see Verschik, 2004). It is possible that these are perceived as one content morpheme (or even a content word) by an L2 speaker of Estonian who does not form these words from two components but memorizes and uses such items as a whole, without analyzing them into components.

4.1 Collection of the spoken data

Pupils form micro speech communities that interact throughout most of the day. Different schools located in different linguistic environments have different
intensities of language contact. For this reason, schools in three different linguistic environments were targeted in the study: 1) the dominantly Russian-speaking eastern Estonian towns; 2) the balanced bilingual area, the capital city Tallinn; and 3) the dominantly Estonian-speaking southern Estonian towns. The oral data presented in the current article were selected with the intention of getting examples from both male and female students with different language profiles and from different sociolinguistic regions, provided they all used Estonian items in the Russian matrix in their speech with a different degree of frequency (see Table 2). The latter aspect should provide more insights into the nature of morphological integration and phonological adaptations of Estonian nouns in Russian grammar and provide an opportunity to look at the continuum of how speakers of different language knowledge interpret and use possibilities of Russian declension classes and gender determination.

Table 2. Subset of data analysed.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Age</th>
<th>Regional background</th>
<th>Social network</th>
<th>Home language</th>
<th>School language</th>
<th>Exposure to Russ. (age)</th>
<th>Exposure to Est. (age)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria</td>
<td>16</td>
<td>NE Est.</td>
<td>monolingual Russ.</td>
<td>Russ.</td>
<td>Russ.</td>
<td>0+</td>
<td>7+</td>
</tr>
<tr>
<td>Andrei</td>
<td>16</td>
<td>NE Est.</td>
<td>bilingual but more Russ. than Est.</td>
<td>Russ./Est.</td>
<td>Est.</td>
<td>0+</td>
<td>0+</td>
</tr>
<tr>
<td>Helen</td>
<td>16</td>
<td>Tallinn</td>
<td>monolingual Est.</td>
<td>Est.</td>
<td>Est.</td>
<td>13+</td>
<td>0+</td>
</tr>
<tr>
<td>Martin</td>
<td>16</td>
<td>S Est.</td>
<td>bilingual but more Est. than Russ.</td>
<td>Est./Russ.</td>
<td>Est.</td>
<td>0+</td>
<td>0+</td>
</tr>
</tbody>
</table>

At the time of recording, Maria (for reasons of confidentiality, all participants have been given a pseudonym) was a secondary school pupil who had lived in the same area of north-eastern Estonia since birth. Andrei was a pupil aged 16 who had also lived in north-eastern Estonia his entire life. For Maria, Russian was the home and school language; she started to learn Estonian only in primary school. Andrei reported that Estonian was his dominant language at school; he acquired both Russian and Estonian simultaneously. Helen was born in Tallinn. Both her parents were monolingual speakers of Estonian. Her schooling was conducted through the medium of Estonian but she started to learn Russian at the age of 13. Martin had lived permanently in southern Estonia. Estonian was his dominant language at school, though he had acquired both Russian and Estonian simultaneously as home languages.
The analyzed data included recorded interviews with Russian- and Estonian-speaking students of Russian- and Estonian-medium secondary schools (see Table 3). Recorded unstructured interviews covered topics such as school, everyday life, family life, hobbies etc.

**Table 3. Number of CS.**

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Time of participation in conversations</th>
<th>Dominant language of conversation</th>
<th>CS instances</th>
<th>CS nouns and NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria</td>
<td>10 hours</td>
<td>Russian</td>
<td>416</td>
<td>368</td>
</tr>
<tr>
<td>Andrei</td>
<td>10 hours</td>
<td>Russian</td>
<td>289</td>
<td>271</td>
</tr>
<tr>
<td>Helen</td>
<td>10 hours</td>
<td>Estonian</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Martin</td>
<td>10 hours</td>
<td>Estonian</td>
<td>76</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>40 hours</td>
<td></td>
<td>814</td>
<td>639</td>
</tr>
</tbody>
</table>

The dominantly Russian character of Maria’s everyday linguistic practices was reflected in her choice of language: she communicated in Russian with a great number of Russian-Estonian CS instances when talking about school activities, plans for the future or other themes on which Estonian has an impact as a sociolinguistically dominant variety. Andrei also chose Russian as a main language during the conversations and sometimes, listening to his responses, it seems that as a fluent bilingual he tried to separate languages, not switching to Estonian too often and, if switching, then mainly for language play, pragmatically expressive purposes.

In the conversations with Maria and Andrei, Russian-Estonian switches are mostly intrasentential, i.e., within a sentence or within a clause, or, more rarely, intersentential, i.e., across two sentences or clauses. In Muysken’s (2000:60–61) terminology, therefore, most switches are of the ‘insertion’ type, because Russian is the base language into which lexical items or NPs are inserted from Estonian.

In Helen’s speech Russian plays a role of seldom insertional adjective or verbal switches (with the aim to quote somebody or add some colouring to an Estonian subject). It is also evident that although being bilingual from early childhood, at the time of the study Martin had departed from Russian as the main medium of his communication strategies: he preferred to hold a conversation in Estonian with some switches into Russian and in some sense the pragmatic character of his switches is reminiscent of Helen’s: switching to Russian he made reference to some activities or labels some Estonian nouns with, as he believes, a more suitable Russian adjective.

Although the four informants may be clearly divided into two groups – one consisted of Maria and Andrei, who clearly used passages (that is Russian-Estonian CS, where an Estonian noun is inserted into Russian grammatical frame)
needed for the current analysis, and the other comprised of Helen and Martin, who insert Russian items into their dominantly Estonian speech – nobody is excluded from the following analysis as the second group exemplifies strategies of lack of agreement (or partial agreement) of inserted Russian adjectives, nouns or all kinds of modifiers with Estonian nouns.

4.2 Collection of blog data
Multilingualism on the Internet (or in CMC, computer mediated communication) has received much scholarly attention during the recent decade (Androutsopoulos, 2006; Crystal, 2011). However, only a few scholars have so far employed multilingual Internet data for research in contact linguistics (Dorleijn and Notrier, 2009; Verschik, 2010; Frick, 2013). Mostly the focus has been on diasporic communities, immigrant identity, discourse analysis and the pragmatics of CS.

Investigation of CS in written texts and, in this connection, the sociolinguistics of orthography in general, has been expanding during the recent decade (Angermeyer, 2005; Koutsogiannis and Miydikopoulou, 2003; Sebba, 2007) but the field is still rather small (Sebba, 2011, personal communication). In the view of the current authors, the use of Internet data for contact linguistic research in general and research on CS in particular would contribute to our understanding of contact phenomena for several reasons: (1) generalizations, models and theories developed on the bases of spoken data can be tested; (2) it would be useful to look at applicability of modes and develop them further; (3) as it will be shown, issues of orthography are relevant because choice of or between orthographies can have some impact on non-monolingual items; (4) data is relatively easily obtainable and one can collect an ample empirical base in a short time.

In the particular case of Russian and Estonian, two factors make such an investigation especially instructive: first, the highly developed inflectional morphology of both languages; second, the languages in question use different orthographies: Roman orthography for Estonian and Cyrillic for Russian.

Thus, the questions to be addressed in the analysis of the blog data are: (1) what cases of morphosyntactic integration or non-integration (gender assignment and/or case markers addition) of Estonian nouns occur in blogs, and (2) whether rendition of Estonian nouns in Cyrillic (transliteration) affects morphosyntactic integration. Based on what is known about the integration of Estonian nouns into the Russian matrix (Zabrodskaja, 2009), we assume that also in blogs no strict predictions can be made and that there is a good deal of individual variation.

Needless to say that there is no ‘Internet language’ or ‘language of CMC’ as such because different genres call for different patterns of communication, language use, etc. In some genres language use may be regulated or restricted in
some way (for instance, use of obscenities in some Internet communities, etc.). Also, the very character of communication in a forum differs from that in a blog. As opposed to forums, the latter is a personal space, even if a blogger adopts a fictional persona and hides aspects of his/her offline self, personality and life.

Live Journal (henceforth LJ) is a blog platform favoured by Russian-speaking bloggers. It must be noted that Estonian-language bloggers prefer environments other than LJ. Blogs on platforms such as Blogspot or Wordpress are isolated in the sense that users cannot form groups and communities; neither is there an option to befriend a fellow blogger. This is different from LJ, where one can create communities, mark other bloggers as friends and so on. While the main features of blogs, that is, being a diary and at least to some extent a monologue, are present, features of social networks are added.

The data come from three LJ blogs. All three users are males who are not secretive about their background. They all come from Tallinn, are ethnic Russians and have acquired Russian as their first language, have higher education, use Estonian on a regular basis and are highly qualified specialists. All three code-switch (see Table 4 for figures) and their readers do the same in their comments.

**Table 4. Number of CS.**

<table>
<thead>
<tr>
<th>Bloggers</th>
<th>Tokens/Entries</th>
<th>CS instances</th>
<th>CS nouns and NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogger 1</td>
<td>18,251/51</td>
<td>116</td>
<td>99</td>
</tr>
<tr>
<td>Blogger 2</td>
<td>14,438/90</td>
<td>141</td>
<td>105</td>
</tr>
<tr>
<td>Blogger 3</td>
<td>26,297/44</td>
<td>182</td>
<td>123</td>
</tr>
<tr>
<td>TOTAL</td>
<td>58,986/185</td>
<td>439</td>
<td>327</td>
</tr>
</tbody>
</table>

Thus, in both sets of data we focus on insertional CS (Muysken, 1995; 2000:60), which is, in a sense, prototypical CS: both the base language and the language of insertions are easily definable (in the terms of Myers-Scotton, 1993, 1997, Matrix Language and Embedded Language respectively). We did not analyze the cases of alternational CS (Muysken, 1995; 2000:96; Johanson, 2002), that is, when both languages remain relatively separate in the clause, although at times it is difficult to say exactly what the characteristics of the shortest alternation are.

5 **Types of integration of Estonian nouns**

According to Zabrodskaja (2009:38), there are six gradations (types) on the scale of integration/non-integration. They are rather points on a continuum ranging from complete morphosyntactic integration (case assignment, when possible, also gender assignment) through instances where the syntactic position does not require case ending (nominative singular and, in some declension classes, also
accusative singular) through ambivalent cases to lack of case/gender assignment. As we shall see, not all types are present in both sets of data. In addition to that, there are some non-numerous instances where a single Estonian noun retains Estonian case markers. Of course, in blogs the choice of alphabet may play a role in the choice of integration strategies. It will be shown that choice of Cyrillic does not automatically imply full morphosyntactic integration, neither does it mean it is an ‘established borrowing’. We believe that the distinction between one-word code-switches and lexical borrowings cannot be drawn based on structural characteristics, neither do we find this distinction meaningful in a (micro-)community of multilingual speakers (on a similar position, see Johanson, 2002 on conventionalization).

In this section, we will consider the types of integration based on Zabrodskaja’s (2009) study and some additional types. Where applicable, we provide instances from both the oral and blog CS under each type. Examples are labelled accordingly. In the oral examples, Russian-language items are transliterated, and in the blog examples all items are presented as they appear in the original with a transliteration added below.

**Type 1**: Total morphological integration of an Estonian single noun into the Russian matrix (the Estonian noun receives Russian gender and case marker).

A noun behaves syntactically as it were a Russian noun (see Example 1; henceforth in the examples, the Russian part is in *italics*, Estonian in *bold*). This is expectable when an Estonian noun fits into one of the Russian declension classes (or can be easily reanalyzed; for instance, the final vowel of the Estonian stem is omitted and the remaining part is reinterpreted as a Russian stem).

(1) Oral

<table>
<thead>
<tr>
<th>Brat-u</th>
<th>kupi-l-i</th>
<th>krasivuj-u</th>
<th>jop-u</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brother-DAT</td>
<td>buy-PAST-PL</td>
<td>nice-ACC</td>
<td>jacket-ACC</td>
</tr>
</tbody>
</table>

‘they bought a nice jacket for my brother’

Cf. Estonian *jope* ‘jacket’

Compare to monolingual Russian *krasivuj-u kurtk-u* (nice-ACC jacket-ACC). As we see, the Estonian noun is reanalyzed (*jope > new stem *jop-*). The agreeing adjective shows that there is no ambiguity in gender and case assignment.

In Example (2) the Estonian noun *linnavalitsus* ‘municipal government’ fits into the second declension and an appropriate case marker (genitive) can be easily added:

(2) Blog

<table>
<thead>
<tr>
<th>linnavalitsus</th>
<th>krasivuj-u</th>
<th>jop-u</th>
<th>kurtk-u</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘municipal government’</td>
<td>nice-ACC</td>
<td>jacket-ACC</td>
<td>jacket-ACC</td>
</tr>
</tbody>
</table>
very clearly demonstrates the work style of the municipal government’
(Blogger 2)

**Type 2:** Non-integration of the Estonian noun even if the noun fits structurally into one of the Russian declension classes. Such non-integrated Estonian nouns appear in the Estonian nominative, although the syntactic position does not call for the nominative.

(3) Oral

*Kakoj toit predlagajetsja v nov-om kohvik-?*

‘what food is offered in the new cafeteria?’

Cf. Estonian *kohvik* ‘cafeteria’

The syntactic position requires the locative and the noun *kohvik* ‘cafeteria’ fits into Russian 2nd declension (masculine zero-ending nouns). However, no ending is added to the stem. The agreeing adjective *nov-om* (new-LOC MASC) confirms that the noun is treated as a masculine 2nd declension noun; a hypothetic locative singular form would be *v kohvik-e* (in cafeteria-LOC).

Consider Example (4a) from a blog. From the point of view of compatibility with Russian declensions, the Estonian compound noun *ehitus-planeerimisseadus* ‘planning and building act’ is not different from the Estonian noun in the Examples (2) and (3). However, it does not receive any case markers, although the syntactic position calls for the Dative case:

(4a) Blog

*Pričem po exitus-planeerimisseadus-? nel’zja perekryvat’ dorogi veduščije k morju!*

‘at that, according to the planning and building act one is not allowed to block the roads to the sea’

Cf. Estonian *ehitusplaneerimisseadus* ‘planning and building act’
(Blogger 3)

Theoretically speaking, the noun could be integrated like in (4b):

(4b) constructed example

*Pričem po exitus-planeerimis-seadus-u* (building-planning-act-DAT)
Example (4a) demonstrates that the lack of Russian inflectional morphology does not depend on the choice of the script: in both Examples (2) and (4a) Estonian nouns are rendered in Cyrillic.

One may assume that a reason for non-integration may be the fact that ehitusplaneerimisseadus is a compound noun comprising three stems: ehitus ‘building’ + planerimis- ‘planning’ (contracted genitive form) + seadus ‘law’, while in Russian compounding is not as productive as in Estonian and thus long compounds may be perceived somehow differently (i.e., the Estonian derivation is transparent to the speakers/bloggers who are proficient in Estonian). The assumption requires further investigation concerning the status of compounds but note that also linnavalitsus in (2) is a compound (linna city:GEN + valitsus government) that takes on Russian genitive singular ending.

Larisa Leisiö (2001a) who has investigated gender assignment to Finnish nouns in the Russian matrix notes that proficient bilinguals do not necessarily need the same integration strategies as monolinguals. Her investigation provides a relevant parallel to that of ours because Finnish is closely related to Estonian and is structurally similar. Although we are rather dealing with integration in general rather than with gender assignment in particular, this observation appears relevant here, too. Indeed, a bilingual will not face difficulties understanding Example (2a) and the like.

Type 3: Instances of CS where there is no morphological integration, but where the syntactic position indeed does not require the addition of case markers.

This applies to the Nominative singular, and, in certain noun classes, to the accusative where that form is identical to that of the nominative.

(5) Oral
kakoj lehekülg v raamat-?
what page in book
‘what page in the book?’
Cf. Estonian lehekülg ‘page’

The agreeing pronoun kakoj ‘which’ shows that the noun is assigned masculine gender but no case ending is required because it is nominative singular. The noun raamat ‘book’ is a different instance (absence of the locative marker), analogical to the one in Example (3).

Since nominative and accusative, that is, subject and direct object cases, appear frequently, there are plenty of instances like in Example (5). In Example (6), Estonian varjupaik ‘shelter’ can be analysed as a 2nd declension noun and the
preposition (pro) ‘about’ calls for the accusative, which, in this declension class, has zero ending:

(6) Blog
ja kak raz nedavno slušala peredaču pro varjupaik-ø (about shelter-ACC)
‘just recently I heard a broadcast about a shelter [for victims of domestic violence]’
(comment to Blogger 2)

Compared to oral speech, blog entries provide more possibilities for the syntactic position of nominative, namely, in titles. For instance, Blogger 1 and Blogger 2 oftentimes have Estonian-language titles for otherwise predominantly monolingual Russian blog entries. That is, an isolated Estonian noun or NP is used in a specific function, such as a reference to generally known facts, realities, cultural contexts, etc., or for emphasis. Clearly, isolated Estonian nouns in the role of titles are specific to written communication (both conventional writing and blogging), while such a use is impossible in oral communication. Example (7) consists of an isolated noun and Example (8) of an isolated NP:

(7) Kultuurierinevused
‘cultural differences’

(8) Krooni ajal
‘in the era of Estonian kroons’
(refers to the time before Estonia’s transition to Euro)

**Type 4:** Morphological integration of the Estonian noun only for gender, indicated by agreeing adjectives, demonstrative pronouns, verb in singular past tense, etc., but not for case.

Consider Example (9), where we have feminine gender indicated with singular past tense, adjective and possessive pronoun:

(9) Oral
Im by-l-a nužn-a moj-a allkiri-ø
them:DAT was-PAST-FEM needed-FEM my-FEM NOM signature-ø
‘they needed my signature’

The gender is assigned semantically, based on the Russian equivalent podpis ‘signature’, which is feminine and belongs to the 3rd declension. The Estonian
noun allkiri ‘signature’ does not fit into Russian declension classes but the position requires nominative, so no case marking is needed. We will return to semantic and morphologic gender assignment below.

The same type is present in blogs (Example 10):

(10) Blog

trebovat’ uplaty dolga nov-omu vylusal’daja-ø (new-DAT creditor-ø)
‘to demand the payment to the new creditor’
(Blogger 2)

Theoretically, the noun võlausaldaja ‘creditor’ in (19) can be reinterpreted as belonging to the 1st declension (feminine and masculine nouns ending with -a) but the expected dative ending -e is lacking. Theoretically, something like vylausaldaj-e (creditor-DAT) can be expected. The agreeing adjective novo-mu ‘to the new’ (new-DAT) indicates the masculine gender.

According to Leisiö (2001a), there are two strategies of gender assignment: morphologic (based on formal criteria, i.e., into what declension class a given noun fits) and semantic (based on Russian translation equivalent or hyperonym). In her research, proficient bilinguals tended to assign gender rather morphologically than semantically. See Example (11):

(11) Blog

poves’ svoj ametijuhend
‘put on the display your job description’
(comments to Blogger 2)

The pronoun svoj (svoj) ‘own’ indicates masculine. Indeed, the noun amet-i-juhend (profession-GEN + instruction) ends with a consonant and would thus be classifiable as a Russian 2nd declension masculine noun. Based on the translational equivalent, the gender assignment would be different, as Russian dolžnostnaja instrukcjija, literally, ‘professional instruction’, is feminine (1st declension). The syntactic position calls for the accusative, whose marker is zero in the 2nd declension, hence no case marking added on.

As the current research is not quantitative, we cannot say what gender assignment strategy prevails, but our educated guess is that both are present and gender assignment is not predictable in a straightforward way.

It should be noted, however, that in certain instances it is impossible to determine the gender assignment strategy. It happens if both semantic and morphological criteria yield the same result. In other words, an Estonian noun is analyzed as belonging to the same declension as the Russian equivalent/
hyperonym. In Example (11) there is a discrepancy between the two (morphologically masculine but semantically feminine). Instances of coincidence of both criteria would be considered under ambivalent cases.

**Type 5:** In some situations the assignment of gender (and sometimes the addition of a case marker too) is impossible if the Estonian noun is in the plural.

In the case of Example (12) it is impossible to determine what gender piletipunktid could be, because it is used in the nominative plural (as the ending -d clearly indicates). Combined with its grammatical role of the genitive case, this presupposes that it could be potentially used as a masculine noun piletipunkt-a, literally, ‘at a ticket sales point’, but the form is used in plural and, thus, no information on gender assignment is available.

(12) Oral

\[
\textit{my vstretilis’ s nimi u piletipunktid} \quad \text{‘we met with them near ticket point-PL’}
\]

In Example (13), the Estonian noun tantsuõhtud ‘dancing parties’ (literally, ‘dancing evenings’) is in the nominative plural, while syntactically the accusative is expectable. On the other hand, plural accusative forms are identical to those of the nominative in inanimate nouns: the Russian translation equivalent večer tancev, literally, ‘evening of dances’, yields nominative and accusative plural večer-a tancev. Therefore, one can say that the Estonian nominative plural form is used as if it were a Russian nominative/accusative form.

(13) Blog

\[
\textit{tam ješčje i tantsuyxtu-d provodjatsja} \quad \text{‘also dancing parties are being held there’ (Blogger 2)}
\]

The Estonian nominative plural appears also in syntactic positions where, according to both Russian and Estonian monolingual grammars, neither nominative nor accusative plural is called for. Consider Example (14) where Estonian noun menetluskulud ‘procedure expenses’ is in the nominative plural, lacking the expected genitive plural marker -ov, as in the hypothetical menetluskulud-ov, cf. Russian sudebn-yx rasxod-ov (procedural-GEN PL expense-GEN PL).
Interestingly, the forms of Estonian nominative plural and Estonian stem with the Russian nominative plural ending can occur in the very same utterance, as in (15):

(15) Blog

\[\text{snačala oplatu ix menetluskulud} \]
\(\text{'in the beginning the compensation of their procedure expenses'}\)

Type 6: Two ambivalent cases:

6.1 The gender cannot be determined on the basis of the ending of an unintegrated Estonian noun (for example because agreeing adjectives or pronouns are absent).

6.2 The final vowel of an Estonian stem can theoretically be interpreted as a case marker. However, indicators of its role, such as agreeing adjectives or past tense singular verbs, are absent. Such constructions occur with the locative case, which is -е for the 1st and 2nd declension classes.

These types were neither present in the blog data nor in the oral data; however, one should not make strong claims that these types are absent or not possible in principle. Zabrodskaja (2009:370–372) has considered such examples in a different corpus. As contact-linguistic investigation of blogs is just starting, it may very well be that more data are needed.

In addition to these types, there are several types not considered in Zabrodskaja (2009) because, technically speaking, there is no morphosyntactic integration into the Russian matrix to speak of. Nevertheless, such instances should be looked into because they illustrate what happens to code-switched Estonian nouns in general.

Type 7: An inserted Estonian noun has Estonian case marking, although the base language is Russian. These, in Myers-Scotton’s (1993, 1997) terminology, are called EL islands.

In Example (16) two Estonian nouns are inserted into the Russian matrix in accordance with Estonian grammar. The first one, tööd, is in the partitive, from töö 'work', which is normal in a negative sentence, so in monolingual Estonian it would be se-da töö-d, ‘this-PART work-PART’. Russian does not have the
partitive; instead, genitive would be required in this position. Thus, we see an EL-

island whose form is chosen according to Estonian grammar rules, although the

ML is Russian. The second noun _esmaspäev-a_ ‘Monday-GEN’ or ‘Monday-PART’

is an ambiguous example and might be interpreted as being an inserted Estonian

EL island in the genitive or partitive because, in the Estonian case system, these

two cases have the same ending -a or an Estonian noun _esmaspäev_ might be

interpreted in terms of Russian morphology as belonging to the 2\textsuperscript{nd} declension

class (masculine noun with zero-ending), similar to, for example, _veter_ ‘wind’.

One more possible explanation of the example could be that the ending -a comes

from the correspondent Russian form _ponedel’nik-a_ ‘Monday-GEN’. It might be

used not only on a morphological basis (like in the case of _veter_ ‘wind’), but on

analogy with the Russian equivalent.

(16) Oral

Étu töö-d mne ne sdelat’ do _esmaspäev-a_

This-ACC work-PART I:DAT not to do till Monday-GEN/PART

‘I cannot do this work till Monday’

Such instances can occur in blogs as well. Consider (17a). In Estonian, all

numerals except üks ‘one’ govern nouns in the partitive singular. In Russian, there

is no such case and, depending on a particular numeral, case and number vary.

The numeral три (tri) ‘three’ calls for genitive singular: три книги (tri knigi)

‘three books’, cf. nominative singular книга (kniga) ‘book’. In Estonian, ‘three

books’ would be kolm raamatut, cf. nominative singular raamat ‘book’. What we

see in (17a) is the Russian numeral три (tri) ‘three’ followed by the Estonian

partitive singular form majaveini, nominative singular majavein ‘house wine’:

(17a) Blog

три majaveini

calyx tri maja-vein-i (house-wine-PART)

‘as much as three (kinds of) house wine’

In principle, the Estonian noun can be interpreted as belonging to the 2\textsuperscript{nd}
declension and, hypothetically, can receive genitive marker -a, as in (17b). What is

more, Russian has a number of conventionalized lexical borrowing with the

component -vejn (-vejn) such as in портвеjn (portvejn) ‘port’, глинтвеjn (glinntejn)

‘mulled wine’. However, this is not the case in (17b).

(17b) constructed example

три majavein-a

calyx tri maja-vein-a (house-wine-GEN)

‘as much as three (kinds of) house wine’
Examples (16) and (17a) contradict the prediction of MLF that only one language, the ML, supplies the morphosyntactic structure (Myers-Scotton, 2005). An ample overview and discussion based on Estonian-Russian CS data are provided by Zabrodskaja (2009:48-49). The mentioned example cannot be analyzed in the terms of monolingual varieties. It demonstrates again that structural compatibility with Russian (or, generally speaking, ML) grammatical classes cannot serve as a predictor of addition of inflectional morphology.

Type 8: Morphological and semantic gender assignment cannot be distinguished because the two coincide: both the gender based on the shape of the Estonian noun and the gender of the Russian equivalent would be the same.

Consider Example (18), where the Estonian noun _virsik_ ‘peach’ fits into the 2nd declension class (masculine nouns with zero ending in the nominative singular) and its Russian equivalent _persik_ belongs to the same class. In inanimate nouns of this class, the nominative is identical to the accusative (zero ending).

(18) Oral
_Tebje dat’ tak-oj malen’k-ij viirsi-k-o?_ 
To you give such-MASC ACC small- MASC ACC peach-ACC
‘would you like me to give you a small peach?’

Cf. Estonian _virsik_ ‘peach’

As can be seen, Estonian _virsik_ ‘peach’ and Russian _persik_ are quite similar materially (in fact, both are conventionalized borrowings of German _Pfirsich_ ‘peach’). Even if a speaker is not aware of the etymology, the similarity is quite obvious and facilitates the switch (see Clyne, 2003:162–175 on facilitation). However, Type 8 is not limited to the instances of materially similar nouns in both languages (i.e., borrowings from the same variety and common internationalisms). The next Example (19) illustrates the point.

In Example (19), the Estonian noun _võlaõigusseadus_ (obligation:GEN-right-law) ‘law of obligation acts’ ends with a consonant and is interpreted as a Russian masculine 2nd declension noun. Its Russian equivalent is rather lengthy: _zakon o dolgovykh objazatel’stvax_, but the head of this NP is the noun _zakon_ (zakon) ‘law, act’, which also belongs to that declension class, and the locative case ending _-e_ is added accordingly.

(19) Blog
_ž vyl-ajgus-seadus-e_ in obligation-right-law-LOC
‘in the law of obligation acts’
(Blogger 3)

Thus, although a noun can take on a case marker, it cannot be determined whether the gender is assigned semantically or morphologically.

6 Script choice and morphosyntactic integration

Looking at the data from blogs, it would be unreasonable to claim that the rendition of Estonian-language items in the Russian script is synonymous to their conventionalization. Choosing the Russian script may occur for reasons unrelated to the status of the Estonian-language items, for instance, when a blogger does not bother to switch keyboards. What is more, at times longer monolingual stretches in Estonian (alternational CS) may appear in Cyrillic. It is highly doubtful that alternational switches such as longer utterances and even whole paragraphs can be conventionalized. Therefore, script choice is not evidence of assignment to a particular monolingual variety.

Although it is too early to make generalizations concerning orthography and its link to morphosyntactic shape of a code-switch item, some preliminary considerations can be formulated. Language use is indeed much less restricted in blogs than, say, in public signage (although this may vary depending on the ‘seriousness’ of a particular sign). It can be assumed that choice of script would be a predictor of morphosyntactic integration: Russian transliteration enables addition of case marker and the retention of the Estonian original spelling prevents it. In his study on English nouns in Russian signs in the USA, Angermeyer (2005) shows that when English-language items were retained in the Latin script, no Russian inflectional morphology was added on.

However, the state of things as described by Angermeyer is not universal for all contact situations where languages in question employ different scripts. While it indeed sounds reasonable that transliteration may facilitate morphosyntactic integration, this is not necessarily hindered by the retention of the original script. Consider Example (20):

(20)

\[
\begin{array}{c}
\text{Politseiamet} \quad XXX \\
\text{Politseiamet’a} \quad adres XXX
\end{array}
\]

‘the Police Department address is XXX’
(Blogger 2)

The syntactic position calls for the genitive. The noun Politseiamet ‘Police Department’ is treated as a 2nd declension noun and the genitive marker -a is
added. Notably, the Estonian stem is separated with an apostrophe from the case marker. Thus, the blogger demonstrates his awareness of the grammars of both languages.

Example (21a) shows that transliteration does not imply the addition of case markers (the very same can be concluded based on Example [4a]). The preposition po (po) ‘according’ calls for the dative case, and kiirmenetlus ‘expedited processing’ fits into the 2nd declension. Despite the transliteration and structural compatibility, the expected dative case marker -u is not added:

(21a)

zaplatit’ počti 4 štuki po maksekjasu kiirmenetlus
‘to pay almost 4 grand according to expedited processing of payment order’

Compare this to the theoretically possible (21b):

(21b) constructed example

po maksekjasu kiirmenetlus-u
‘according to expedited processing of payment order’

It remains to be seen whether the fact that there is a genitive NP (with the modifier maksekäsu ‘payment order’ [GEN]) plays a role and that multiword items behave differently from single nouns as far as integration is concerned.

7 Discussion and conclusions

Our data demonstrate that code-switched constituents may, but do not have to be, ‘well-formed’ grammatically. As Bullock and Toribio (2009:190) state, ‘the component languages of a bilingual engaged in CS should not be expected to be invariable, monolingual-like systems but, instead, these languages – used independently or in conjunction, as in CS – can provide us with evidence of a rich repertoire of bilingual forms’. Neither in oral speech nor in blogs are Estonian nouns always fully integrated into the Russian matrix, and rather are often at odds with MLF model principles and show a considerable degree of compromise, incompatible with two monolingual grammars. Russian-Estonian bilingual data show that CS triggers convergence, which, in its turn, can trigger a switch (see also Backus, 2005).

Both oral and blog CS data exhibit more or less similar patterns of (non-)integration. Further research will show whether there are significant
quantitative difference. The presence of different types of integration in both sets of data is exemplified in Table 5.

Table 5. Types of integration in oral and blog data.

<table>
<thead>
<tr>
<th>Types (1–6 according to Zabrodskaja 2009)</th>
<th>Gender and case assignment</th>
<th>Oral data</th>
<th>Blogs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full morphological integration</td>
<td>+ case, + gender</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>No integration in the cases where integration is syntactically needed (Estonian Nom Sg in whatever syntactic position)</td>
<td>– case, – gender, no agreeing adjectives, etc.</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Integration is syntactically not necessary</td>
<td>Nom Sg, Nom Sg = Acc Sg in 2nd declension</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Partial integration (no case marker, gender assignment either semantic or morphological)</td>
<td>– case, +gender agreeing adjectives, etc.)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Estonian Nom Pl in all positions</td>
<td>– case, – gender</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gender assignment impossible (no agreeing adjectives), case ending is lacking</td>
<td>– case, – gender</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Estonian final vowel can be interpreted as a case marker (-e can be reinterpreted as Russian Loc Sg in some classes)</td>
<td>? case ? gender</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Estonian stem gets Estonian case marker; if there are no agreeing adjectives, etc., gender assignment impossible</td>
<td>? gender; Est case</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Morphological and semantic gender coincide, hence no gender assignment; optional case marking</td>
<td>? gender, (?) case</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

According to Johanson (2002:13), ‘[a]s a rule…the more morphological classes and inflectional patterns language A contains, the more extensive are its modification strategies’. Russian-Estonian intra-sentential and constituent-internal or word-internal CS (or insertion) prevail in the corpora, especially in the case of semantically specific items. In our analysis, attention is paid to intra-sentential switches because this is where Russian and Estonian grammars come into contact. CS in blogs exhibits the same patterns of integration. Contrary to what one would expect in blog data, two different alphabets do not prevent integration.
Morphological integration (gender and/or case assignment, or case only in plural contexts) of Estonian nouns into the Russian matrix depends to an extent on structural factors (i.e., whether the noun in question fits into any of the Russian noun declension classes) but mostly on individual preferences, types of discourse, oral or written modality, etc. On the whole, Estonian-Russian CS in blogs confirms observations made by several scholars. Leisiö’s (2001a) argument that bilinguals do not require the same mechanisms of integration as monolinguals has been mentioned. Matras (2009:113–114) gives some attention to CS vs. borrowing and morphosyntactic integration, and concludes that there is a certain fuzzy area for bilinguals: an item can be used frequently but preserve its original shape, whereas an item that is used only occasionally can be fully integrated. Gardner-Chloros (2009:12) believes that integration cannot be predicted in strict terms based on formal criteria.

Acknowledgments

The research leading to these results was supported by Project MJD96 ‘Transfer of morphosyntactic patterns in the Estonian-Russian contact setting’ (2010–2013) and Project EKKM09-85 under grant agreement ‘Vene-eesti ja inglise-eesti koodivahetuse ja koodikopeerimise korpusse koostamine ja haldamine’ [Russian-Estonian and English-Estonian code-switching and code-copying corpora creation and management].

About the authors

Anastassia Zabrodskaja is Professor of Estonian as a Second Language at the Institute of Estonian Language and Culture at Tallinn University, Estonia. She also works as a Senior Research Fellow in Sociolinguistics at the Institute of Estonian and General Linguistics at the University of Tartu, Estonia. Her scholarly interests include language contact, dynamics and change in Estonia and ethnolinguistic vitality.

Anna Verschik is Professor of General Linguistics at the Institute of Estonian Language and Culture at Tallinn University, Estonia. Her main research interests are language contacts, contact-induced language change and the sociolinguistics of the Baltic region.
References


