Mixed Copying in Blogs: Evidence from Estonian-Russian Language Contacts

Anna Verschik
Tallinn University
annave@tlu.ee

Abstract

The paper applies the Code Copying Framework (CCF) to Estonian-Russian language contacts in Live Journal blogs. The nature of blogs (an asynchronous, more written-like genre) allows us to look into individual multilingual practices and to discover aspects of contact-induced change that are absent in oral communication (choice of script, rendition of other-language items etc). CCF distinguishes between global copying (akin to code-switching/borrowing in other frameworks), selective copying (phenomena in morphosyntax, semantics etc) and mixed copying. The latter means that one component of a complex item is a global copy and the other a selective one and occurs in multi-word items (compounds, constructions, analytic forms, idioms). Six types of mixed copies are analysed. It is argued that this type of copying requires closer attention because 1) it demonstrates what is perceived as a collocation or multiword unit by a multilingual user; 2) it contributes to the understanding of meaning (semantically specific components are likely to be copied globally; 3) it is in accordance with notions in cognitive linguistics (compositionality, blending).

Keywords

code-copying – blogs – language contacts – Estonian – Russian

1 Introduction

Personal blogs are a kind of private virtual space. Bloggers have the freedom to openly declare their language policy or, in any case, to choose the extent to which they adhere to current norms, language ideologies and general opinion.
on what the ‘right’ language use should be like. A ‘personal touch’ can be added also by combining two or more languages. It is not surprising, then, that the study of multilingualism on the internet has been dealing with pragmatic, sociolinguistic, attitudinal and cultural aspects. At the same time, the multilingual internet provides numerous examples of phenomena that are well known in contact linguistics (code-switching, convergence), and a contact linguistic approach would be desirable. For a scholar in contact linguistics, computer mediated communication (CMC) in general and blogs in particular can provide ample data on contact-induced innovation and change. Blogs allow collecting data in a relatively simple way; moreover, written texts are in their nature more deliberate than oral texts and the appearance of changes there hints at a greater degree of conventionalization (that is, systematic use of other language content words and patterns is more visible and cannot be dismissed just as a slip of the tongue or anecdotal occurrence; see also the discussion in Dorleijn (2016).

It has been observed in the literature that code-switching demonstrates how lexicon is linked to morphosyntax (Backus 2003, 2005; the notion of a composite Matrix Language by Myers-Scotton 2002: 22; different types of code-switching in the Triangular Model by Muysken, 1995, 2000). No doubt, blogs can provide valuable data for further research and, because electronic communication is less ephemeral than oral, highlight certain phenomena that have received little attention.

The following Example (1) comes from a Russian-language blog, written by an ethnic Russian from Estonia with Russian as L1. From his blog we can infer that he is probably in his thirties, has a higher education and is proficient in Estonian.

(1) У меня была карта персонального пангалдуса.

U menja byla karta personal'nogo pangandus-a

‘I had a card for personal banking’ (a type of service in an Estonian bank)

The item in bold is modelled on Estonian *personaalne pangadus* ‘personal banking’. The two languages have a common internationalism, the adjective *personal'nyj/personaalne* ‘personal’. One may suggest that the last item in (1) is a one word switch to Estonian and the modifying adjective is not relevant at all. However, this appears counterintuitive because the adjective is similar in the two languages. The next step would be to claim that the common internationalism facilitates switching, ‘triggering’ in the terms of Clyne (2003: 163–192). While this seems likely, this is not the whole story because the adjective and the noun constitute a term with a very specific composite meaning. There
is no corresponding term in monolingual Russian; one can coin something like персональное банковское обслуживание (personal'noje bankovskoje obsluživanije) ‘personal banking service’ but this is not a conventionalized equivalent. Estonian пандандус ‘banking’ may have several Russian equivalents, depending on a particular context. In other words, these items ‘belong together’ for speakers of Estonian and probably for Russians in Estonia as well. Although one of the components appears in its Russian shape and the other is Estonian, it would be reasonable to consider it as a whole.

Instances such as (1) have received relatively little attention in the literature as compared to code-switching and morphosyntactic convergence. The Code Copying Framework (hereafter CCF) by Lars Johanson (1992) provides a term for this, namely, mixed copying. The current article discusses CCF (Johanson, 1992, 1993, 1999) as applied to Estonian-Russian language contacts in blogs, with a special focus on mixed copies. I chose CCF for the following reasons: (1) it appears to be descriptively accurate and flexible (Verschik, 2008: Chapter 3) and (2) it views contact-induced processes and results at all levels of language within the same terminological framework and does not draw sharp divisions between contact-induced influence in lexicon (code-switching, lexical borrowing) and morphosyntax (see Backus 2005 on connections between code-switching and convergence; Backus and Verschik, 2012 on the advantages of a holistic approach to lexicon and morphosyntax).

In addition to the above mentioned reasons, I find the notion of mixed copying (Johanson, 1993: 215) helpful both descriptively and theoretically because it brings to attention examples like (1) that have more often than not escaped scholarly attention (with the exception of Haugen 1972: 85 who mentioned loan-blends but did not dedicate much space to them) and, second, helps understanding what elements ‘belong together’ in a speaker’s cognition.

My approach is qualitative, i.e., the aim is to find out what kind of language contact phenomena is present (for various kinds of Estonian impact on Russian in oral data see Verschik, 2008; Zabrodskaja, 2009). First of all, I will describe the relevance of CMC for contact linguistics. Then I will briefly discuss the situation of Russian-speakers in Estonia, showing that this is not a typical immigrant/minority community and the sociolinguistic background of the current Russian-Estonian bilinguals differs from that of well-researched communities in Western Europe (Turks in Germany, Moroccans in Holland etc). Then, after a short characterization of the data and blog environment Live Journal (lj) the importance of an individual language user for linguistic research will be elaborated upon with reference to blogs as a key to individual linguistic space. Afterwards, a short outline of CCF will be presented with the focus on mixed copies and a tentative typology thereof.
CMC and Contact Linguistics

During the last two decades, computer-mediated communication (hereafter CMC) has significantly influenced the way of life, the nature of communication as such via introducing new genres (blogs, forums, instant messaging etc), and brought about new aspects of linguistic phenomena that initially were observed and studied by sociolinguists in oral speech (first and foremost, code-switching). The body of research on CMC can be subdivided into monolingual and multilingual research. For instance, the comprehensive overview by Crystal (2001) dedicates very little space to multilingual communication. On the other hand, several scholars already focusing on multilingual communication (Androutsopoulos, 2006; Dorleijn and Nortier, 2009; Hinrichs, 2006 and many others) became interested in multilingual practices in various types of CMC. The collection of articles titled The Multilingual Internet (Danet and Hering, 2007) linked multilingualism in CMC to other areas of inquiry, such as language choice and gender on the net, identity and globalization, and writing systems.

In the above mentioned collection as well as in other scholarly articles on multilingual CMC, the focus is on pragmatics and functions of code-switching (Androutsopoulos, 2006; Hinrichs 2006), expression of identity, youth language and on-line (immigrant) communities (Androutsopoulos, 2006; Siebenhaar, 2006), ideology of different spelling systems (Koutsogiannis and Mitsikopoulou, 2003; Androutsopoulos, 2009).

However, surprisingly, very little has been written about structural properties of multilingual CMC (Dorleijn and Nortier, 2009). Being an important source of data for pragmatic, sociopsychological and attitudinal sides of multilingualism on the net, CMC also provides excellent data for contact linguistics, including not only code-switching but non-monolingual structural phenomena as well. For instance, it would be instructive to consider influential models, such as MLF (Myers-Scotton, 1993), the 4 M Model (Myers-Scotton, 2002), the Triangular Model (Muysken, 1995, 2000) and others. Estonian-Russian data on the net have been investigated to some extent (Oja, 2008 on code-switching in a popular Estonian teens’ portal, Verschik, 2010a on preliminary analysis of Estonian-Russian code-copying in Live Journal blog environment).

Emerging Russian-Estonian Bilingualism and the Data

Sociolinguistic Situation
Currently, ethnic Estonians constitute 68.7 % and Russians constitute 24.8 % of Estonia’s inhabitants, or 889,770 and 321,198 respectively out of total number of
1,294,455 (Population and Housing Census 2011). Importantly, most of them are descendants of the Soviet era migrants. After Estonia’s occupation and annexation by USSR (1940–1941, 1944–1991), the central authorities wished to tie the new union republic (together with the Baltic states of Latvia and Lithuania) with the Soviet empire and covertly encouraged migration of Russian-speakers from the rest of the USSR by means of creating new jobs and providing better housing facilities. From 1957 on, Russian was made an obligatory subject in USSR in all schools with a non-Russian language of instruction, while learning local languages was not compulsory in Russian-medium schools (see Rannut, 2007 for a detailed analysis of the Soviet language policy; see Pavlenko, 2008 and Verschik, 2010b for an overview of the post-Soviet sociolinguistic situation).

Unlike migrants in Europe who realized that they were settling in a different country and would have to acquire at least some command in the local language, the settlers in the Baltic republics did not conceptualize their moving to Estonia as emigration but rather as internal migration within the same country and did not bother to learn local languages. Needless to say, Estonians were obliged to learn Russian but informal interactions, friendships and marriages across the ethnolinguistic borders were extremely rare; the two communities became polarized, Estonians being bilingual to a certain degree and strongly identifying with the former independent state and Russians being monolingual. As Skutnabb-Kangas (1992) observed, Russians in the Baltic region were a “majoritized minority”, while the indigenous populations were “minoritized majorities”.

After the collapse of the Soviet Union and the regaining of the independence in 1991, the sociolinguistic situation in Estonia radically changed. Estonian became the sole official language (though this did not mean that public usage of other languages was prohibited or restricted) and instruction in Estonian became obligatory in Russian-medium schools. Technically speaking, the first generation of Russian-Estonian bilinguals was in fact the third generation of migrants who had Russian as their L1.

Observations during the last two decades indicate that Estonian has made its way in Russian-to-Russian communication, and a range of non-monolingual varieties has emerged (Verschik, 2008; Zabrodskaja, 2009). Sometimes younger Russians stress that they are not simply Russians but Estonia’s Russians; some go even further and state they do not care how “it should sound in proper Russian” because they live in a different reality. Such cases of a very clearly expressed position towards language and identity appear in blog entries as well.

3.2 The Data
The data comprises LJ blog entries during the period 2008-January 2012. All bloggers are ethnic Russian and have Russian as L1. The corpus consists of
Mixed copying in blogs

63,470 words in total. The entries have been created by three bloggers but it should be mentioned that contact linguistic phenomena are also present in numerous comments. It would be more appropriate to speak about a kind of social network rather than three isolated bloggers (see more on LJ blogs in the next section). The users are speakers of Russian as L1, some of them have a higher education or are students; all of them have at least some proficiency in Estonian, some of them are actually highly proficient (this topic comes up frequently in blog entries). All bloggers seem not to be restrained by monolingual norms and seem to enjoy their multilingualism, engaging in language play or discussing metalinguistic issues, such as the use of Estonian object cases etc. These users represent what has been conceptualized by social scientists in the Estonian sociocultural context as young urban dwellers (Vihalemm, 2002): young people with a sufficient proficiency in Estonian, residing in bigger urban centres such as Tallinn and Tartu, professionally successful and identifying themselves with the Estonian state. In other words, the users exemplify bilingual Russians in contrast to monolingual Russians in Estonia (Vihalemm and Masso, 2002).

Completely monolingual entries were not included into the corpus. At this stage, I also did not consider code alternations (Muysken, 2005), that is, longer monolingual stretches of Estonian within Russian posts for the reason that nothing “happens” there from the point of view of language contacts.

Blogs as a Window into Individual Use. Live Journal Blog Environment

As Puschmann (2013: 80) observes, blogs are highly variable forms of self-expression. The content and the format can be more or less personalized; an entry can be more or less standard-like; the addressee can be just anybody or friends only, and so on. The concept of audience design (developed by Bell) can be applied to blogs, too: the author conceptualizes the audience s/he is writing for, while realizing that others can read the entries, too (see the discussion in Puschmann, 2013: 85–86), unless the author chooses to limit the access.

The reason why the language use of an individual (whether monolingual or multilingual) should be studied, is the understanding that innovation starts precisely on the level of the individual user. Whether such an innovation becomes part of a person’s repertoire (or: entrenched) or becomes part of the community norms is a different question altogether. Several scholars (Matras, 2009: 310 ff; Blommaert and Backus, 2011) have recently called for a focus on the individual in multilingualism research rather than on the community. Indeed,
it is extremely difficult to provide a satisfying definition of speech community; communities may have more rigid or more fluid borders (focused vs. diffused in the terms of Le Page and Tabouret-Keller, 1985), and a person can view himself/herself as a member of several speech communities. Cook (2009) provides several definitions of language, such as: a system of mental representations, a group symbol, something in possession of a community; in the same vein, I propose yet another way of conceptualizing language: it is something in possession of an individual, because one is a native speaker of his/her idiolect only, and at the same time the only native speaker of one's idiolect (it is true that idiolects do overlap, which creates an image of a community of native speakers of language X or Y). Blommaert and Backus (2011) argue that the actual use of linguistic resources is highly personalized and utilize the concept of linguistic repertoire that refers to linguistic resources arranged autobiographically.

For the current discussion, it is important to note that blogs are not only a window into personal language use and ‘private language policy’, so to say, but also into the emerging multilingual practices. Matras (2009: 68) suggests that multilingual individuals are in a sense agents of influence: they produce contact-induced innovations that potentially may affect monolingual use as well (the same idea is supported by Thomason, 2001). Diary-like blogs (see Puschmann, 2013 for classification) are probably a better source than forums for looking into individual’s language use (because such blogs are individual spaces or platforms by definition and forums have many participants and are polylogic in nature) and, if language policy is openly declared or particular choices are explained by a given blogger, into metalinguistic awareness as well.

For some time, the Live Journal blog environment has been popular among Russian-speakers all over the world. Interestingly, Estonian L1-speakers prefer the so-called 'stand alone blogs' (Blogspot, Wordpress), while Live Journal combines the possibilities of keeping a classical diary-like blog with networking. Every user can make friends among other users and belong to communities. A user’s profile displays his or her friends and mutual friends, interests, place of residence etc. It is common to extensively comment and, if one likes an entry, to repost it in his/her blog. For a researcher, the information in the profile, especially about mutual friends, is extremely helpful because networks are explicit. Thus, while having general features of blogs as a type of CMC (asynchronous, no limit of space, mostly more written-like, monologue), at the same time Live Journal blogs have features characteristic of forums and social networks (communities, dialogue in comments).
5 The Code-Copying Framework

5.1 Copying and Types of Copies

CCF was at first formulated by Lars Johanson (1992) based on contacts of Turkic languages. In later versions, he turned to more general aspects of CCF (Johanson, 1999, 2002). Quite soon, CCF became known outside the Turkology scholarly communities and has since then been applied to a variety of language contact situations, mostly to the cases of L2 impact on L1: Verschik, 2004, 2008 on copying from Estonian into Russian in oral communication; Verschik, 2010a on possibilities of CCF application in CMC (blogs); Praakli (2009) to Finnish-Estonian bilingualism among recent Estonian immigrants; Vaba (2010) to English impact on Estonian in Skype communication; Paljasmaa (2012) on copying from French in blogs of Estonians living in France. In the recent contribution by Backus and Verschik (2012), CCF was a central component in the discussion on borrowability of bound morphemes.

“Code” stands for any variety. According to Johanson (2008: 62), “borrowing”, “mixing”, “switching”, “interference” are wrong metaphors because copies are not identical to the models, even if a copy preserves all properties of the original. Copies are not transplanted or transferred but they may have their own development trajectory in the new code, or are subject to code-internal development. There is always a model code and a basic code. Both L1 and L2 can play both roles. If a code user produces a monolingual stretch in another code, this is called code alternation (Muysken, 1995 uses the term in the same meaning). Other cases of code-interaction are considered as insertion: copies from the model code are inserted onto the basic code. Copying from L1 is called imposition and from L2 adoption.

Contact-induced influence on all levels of language is considered as copying. Thus, code A copies an item, pattern, rule, meaning etc from code B. The framework is not constraint-based: any kind of copying can happen. Note that copying of items and copying of meaning, pattern etc is still copying. Thus, morphosyntax, meaning and lexicon are considered within the same terminological framework. This allows us to observe connections between all levels of language. Although it is usually assumed that there is no strict division between, say, grammar and lexicon, in practice this is often forgotten (for instance, borrowability hierarchies deal mostly with various types of lexical items, see details in Backus and Verschik, 2012).

Every item has four types of properties: material (the phonetic shape of an item), semantic, combinational and frequential. One can copy all properties, which corresponds to “borrowing” or “code-switching” in other models.
Alternatively, it is also possible to copy only certain properties. The former case is called global copying and the latter selective copying.

Example (2) demonstrates a global copy from Estonian. The Estonian item *pass* ‘passport’ is transliterated into Russian (*пассь*) but note that transliteration is not an indicator of conventionalization in local Russian (see discussion in Verschik, 2010a).

(2) я свой пассь не стал менять на биометрический
Ja svoj pass’ ne stal menjat’ na biometritšeskij
‘so far I have not exchanged my passport for the biometrical one’
(Russian паспорт (passport) ‘passport’)

Selective copies may incorporate only one property (for instance, meaning) or several (meaning and combinatory rules, such as choice of case, preposition etc). Selective copying may occur in multiword items (fixed expressions, analytic forms). In example (3a) the blogger uses the Russian item прилавок (*prilavok*) ‘counter’ in the accusative and together with the preposition в (v) ‘into’ that implies internal local meaning. It is a selective copy of the properties of the Estonian lett ‘counter’ that is normally used in the illative, a directional internal local case (implying movement into something) after verbs with the meaning ‘to go’ (see 3b):

(3a) blog
Ид-ёшь в прилавок, а там уже очередь
Id-još’ v prilavok, a tam uže otšered’
go-2SG into counter:acc but there already queue
‘you go to the counter but there is already a queue there’

Estonian
Lähe-d letti, aga seal on juba saba
go-2SG counter:ill but there be:3 already queue
‘you go to the counter but there is already a queue there’

Compare with monolingual Russian in (3c):

(3c) Ид-ёшь к прилавк-у
Id-još’ k prilavk-u
go-2SG to counter-dat
‘you go to the counter’
In (3a), Estonian combinatory rules are copied (internal directional meaning expressed by the illative); such instances are also referred to as selective combinational copies. Cases of selective copying also include copying of meaning, word order, collocations, and so on.

Matras and Sakel (2007) provided a similar descriptive model where contact-induced change on all levels is viewed as the process of replication. Replication can be of two types: MAT (matter replication) and PAT (pattern replication), corresponding to Johanson’s global copying and selective copying respectively. However, in one respect these models differ. While Matras and Sakel’s models have MAT and PAT, covered by global and selective copying, CCF has mixed copying. Mixed copying occurs in a complex item (compounds, analytic verbs, fixed expressions, idioms, constructions, where components ‘belong together’ and have a composite meaning) when one component is copied globally and the other (or others) selectively.

To illustrate mixed copying, Johanson (1992: 183) mentions Uzbek yarimavtomat ‘half-automate’ modelled on Russian полуавтомат where yarim ‘half’ is Uzbek, avtomat < Russian автомат (avtomat) ‘automate’ is Russian and the combination is Russian. Also phraseological verbs are often copied from Persian into varieties of Turkic in the sense that the nominal component is Persian and the auxiliary Turkic: Tebriz Azerbaijani telefon vur- ‘to call’ (literally, ‘telephone strike’) from Persian telefon zadan ‘to call’ (Johanson, 1998: 334).

5.2 Mixed Copies in Live Journal Blogs

We now turn to a tentative classification of mixed copies. I will show that CMC provides additional details for analysis of mixed copies because CMC uses script and in certain fuzzy cases spelling helps to draw firm borders. It appears that not all mixed copies are alike.

Type 1. Let us start from more prototypical examples. In (4) Estonian säästu-kaart ‘saving card’ (saving gen + card) yields сээсту карточка:

(4) Сээсту карточка действует
Säästu kartotška dejstvujet
‘the saving card is valid’

Monolingual Russian has different ways of rendering the meaning, for instance, карт(очк)а клиента, literally, ‘client card’ (card + client gen). The common internationalism has different shapes in the two languages: Estonian kaart ‘card’ and Russian карта (karta) or, in this context another version of this
internationalism with a slightly different meaning карточка (kartočka). The combination of the items and their order (the modifier in genitive preceding the head in nominative) is Estonian. The fact that it is rendered as two separate orthographic words and that both are written in Cyrillic should not prevent from analysing this as a mixed copy because the item in question has a highly specific meaning (not just any kind of card but a particular saving card) and because of the complex meaning (säästukaart as a whole in the model code). In this type, one component is a Russian word or a common internationalism, provided that the latter differs sufficiently in two languages. Thus, it is possible to easily distinguish between the mixed copy and a conventionalized global copy säästukaart/сяэстука(a)rt (sjaéstuka(a)rt) that is also used both in oral speech and in СМС.

Type 2. Mixed copy involves a common internationalism with a fairly similar shape in both languages, which make the border between global and selective copying fuzzy.

In example (5a), a blogger renders Estonian isikukood ‘personal code’ (a rough equivalent of ‘social security number’ in USA) as исикукод (isiku- kood). The entry and the comments discuss what the last digit of the code means:

(5a) Для каждого исикукода она своя, в зависимости от цифр
Dlja každogo isikukoda ona svoja, v zavisimosti ot cifer
‘it is different for every personal code, depending on (other) digits’

Note that the transliteration preserves the double vowel. No doubt we deal with a global copy here. However, in (5b), the same lexical item is transliterated with only one o, which turns the second component at least orthographically into Russian код (kod) ‘code’.

(5b) Нашли моё имя и исикукод, но в графе адрес было пусто.
Našli moje imja i isikukod, no v grafe adres bylo pusto.
‘they found my name and personal code but the address box was not filled’

Russian lacks phonological distinction between long and short vowels and consonants, whereas duration is most distinctive in Estonian. Without going into details of the distinction between long and super-long vowels here, it is important to note that double vowel in spelling shows that the vowel is not short.
I cannot say whether some users systematically prefer one version for the other; one should consider the possibility of typos, too. But a realization like in (5b) means that it is a mixed copy where the second component is Russian.

Such borderline examples tend to appear where Estonian internationalism differs from its Russian equivalent only minimally (double vowel/consonant vs. single vowel/consonant and so on). The same happens in политсейакадемия (politsejakademija) ‘police academy’ < politseiaakadeemia, cf. Russian академия (akademija) and Estonian akadeemia ‘academy’; лишклас хулиган (lükluxxuligan), also rendered as лишклас хулиган (lüklux xuligan) ‘reckless driver/pedestrian’ (‘traffic’ + ‘hooligan’) < lüklushuligaan, cf. Russian хулиган (xuligan) and Estonian huligaan. I will return to the importance of spelling in the final section.

Type 3. The globally copied part is an abbreviation. An abbreviation can serve as a modifier component in Estonian compounds. This model of compounding has become productive in Russian of Russia as well under the influence of English (Kitajgorodskaja, 1996). Compounds like мобильная парковка < m-parkimine ‘parking with payment by mobile telephone’ where m- is an Estonian abbreviation from mobiil- ‘mobile, of mobile’, occur in oral speech (Verschik, 2008). In the following example (6), the common Estonian abbreviation е- ‘electronic’ (< elektrooniline, elektroon- ‘electronic’), as in e-kiri ‘e-mail message’, e-valitsus ‘e-government’, e-pilet ‘e-ticket’ etc appears as Russian э-. Monolingual Russian of Russia tends to use the adjective электронный ‘electronic’ in official registers but in casual speech and writing электронный э- or электронный эмейл (émejl) ‘e-mail’ (with some spelling variations) would be used, imitating more or less the English pronunciation.

(6) из всевозможных э-открыток, эта самая-самая
iz vsevozmožnyx é-otkrytok, éta samaja-samaja
‘of all kinds of e-cards, this is the best one’

Cf. Estonian e-kaart ‘e-card’, kaart ‘card, postcard’ and Russian открытка (otkrytka) ‘postcard’

Differently from -e, certain abbreviations are exclusively Estonian and do not have any international parallels. For instance, in (7) JOKK comes from the expression juriidiliselt on kõik korrektne/korras ‘legally everything is correct (but in reality this looks nasty)’ that has gained currency during the recent years at first as a colloquialism and then became conventionalized. As compounding in Estonian is extremely productive and compounds can be formed spontaneously (for an overview see Verschik, 2008: 119–120), it is possible to say
JOKK-värk (the latter component meaning ‘stuff, things’), JOKK-skeem ‘scheme, plot’ etc. The blogger writes:

(7) Хотели провернуть JOKK-сему
Xoteli provernut’ JOKK-sxem-u
‘(they/one) wanted to get through with a JOKK-scheme’

Note the rendition of the mixed copy. I believe that the Estonian spelling of the Estonian JOKK helps to avoid ambiguity: it looks emblematic and immediately recognizable to anyone familiar with the Estonian language and realities, whereas a hypothetic transliteration into Russian would look clumsy or funny. The blogger uses a hyphen that indicates the composite character of the expression.

Type 4. Border-line cases where copies appear in collocations that are similar to compounds but where the components are linked more loosely: proper name/toponym/brand + noun. It may be argued that proper names are highly specific and unique and thus cannot be analysed as copies. While I acknowledge that their status is unique and that they have a different kind of meaning compared to common nouns, they are nevertheless relevant for the current discussion at least for two reasons.

First, some proper nouns are derived from common nouns and are therefore unavoidable for bilinguals as well. Consider (8):

(8) Про Сыпрузе мост вообще молчу
Pro Sypruze most voobšše moltšu
‘I am not even mentioning the Sõpruse bridge’

Cf. Estonian Sõpruse sild ‘Sõpruse bridge’. The blogger is talking about traffic in Tartu in the vicinity of the mentioned bridge across the river Emajõgi. The proper name Sõpruse is the genitive form from Estonian sõprus ‘friendship’. Thus, anybody minimally familiar with Estonian would identify the noun and would recognize the genitive case. The collocation Sõpruse sild is structurally not different from Estonian compound nouns N gen + N nom like käekell ‘watch’ (hand gen + clock/watch nom), raamatukapp ‘bookcase’ (book gen + case nom) and so on.

Second, whether completely unique or derivable from a common noun, the Estonian component in the genitive precedes the selectively copied component in the nominative and, thus, replicates the Estonian word order. In
Russian, the modifier in an NP follows the head. While in Russian in general this word order has been gaining currency under the influence of English (also in collocations with proper names or with certain conventionalized copies from English like бизнес-встреча (biznes-vstreča) ‘business meeting’, see Kitajgorodskaja, 1996), one cannot exclude Estonian impact in examples like (8) where the context and realities are Estonian.

Type 5. Mixed copies in constructions (analytical verb constructions) and fixed expressions. The copying of Estonian particle verbs into Russian has been described earlier and the observation was made that if particle verbs are subject to mixed copying, then the particle would be copied globally (Verschik 2008: 146–147). Backus and Verschik (2012) took this further and assumed that this happens because it is the particle that specifies the meaning: учить (učit’) ‘to learn’ is more general and less specific than учить пähe (učit’ pähe) ‘to learn by heart’ (literally, ‘to learn into the head’), copied from Estonian пähe õppima ‘id.’ I have not found any mixed copies of particle verbs in the blogs. There are mixed copies of idiomatic expressions but, for the sake of brevity, I will not discuss them because they do not differ from those described in Verschik (2008: 146–147). Still, there was an occurrence of mixed copy in a different type of verb construction that deserves attention.

(9a) A моральный ущерб ... тупо оставил läbivaatamata
A moralnyj ušherb... tupo ostavil läbivaatamata
and moral damage stupidly left un-viewed
‘and the (application for) moral damage was stupidly dismissed [by the county court]’

Cf. Estonian jättis läbivaatamata ‘dismissed’ (literally, ‘left un-viewed’). The construction under consideration encompasses the so-called Abessive form of the supine (ma-infinitive) in Estonian that can be schematically presented as follows: jätma X-mata, literally, ‘to leave un-X-ed’. The meaning of the construction can be rendered in English as ‘to forget/neglect/ignore doing X’ or ‘to end up without having done X’. In Estonian, the construction can be derived from any verb, transitive or intransitive (jättis tulematu ‘did not show up’ from tulema ‘to come’).

The rendition in (9a) is a mixed copy, where Estonian jättis läbivaatamata ‘dismissed’ (‘left un-viewed’) yields оставил (ostavil) ‘left’ + läbivaatamata ‘unviewed’ (the Abessive form), the first part being a selective copy and the second a global copy.

In Russian, it is possible to use the verb оставить (ostavít’) ‘to leave’ in combination with the preposition без (bez) ‘without’ + N gen: оставить без
присмотра (оставить без присмотра) ‘to leave unattended’, оставить без удовлетворения (заявление) (оставить без удовлетворения) ‘to dismiss (an application)’, literally, ‘to leave without satisfaction’. However, unlike in Estonian, it is not possible to form a noun from every verb. Apparently, it is a matter of convention why оставить без рассмотрения (оставить без рассмотрения) (a full equivalent of jätta läbivaatamata) does not ‘sound right’ in monolingual Russian. Should such a rendition appear, it would be a selective copy. I assume that a mixed copy where the verb would be copied globally and the Abessive form selectively is unlikely because the meaning of the verb is rather general and its lexical meaning in this particular construction has faded, that is, it has grammaticalised, and because this is more schematic/less unique and a fixed, unchanging part of the construction, while the Abessive form can be derived from any verb and expresses the lexical (that is, more specific) meaning:

(9b) jättis  без  рассмотрения (?)
jättis  bez  rassmotrenija
left  without  viewing
‘dismissed’

To my mind, (9b) is rather unlikely to occur, although I would be cautious to formulate strict constraints. I will return to the discussion of meaning in the final section.

Type 6. The modifier takes on a Russian adjective derivational suffix in combination with Estonian stems/utilises existing Russian relative adjectives derived from common internationalisms; head either in Estonian or Russian, whereas the combination of items and meaning come from Estonian.

To the best of my knowledge, such constructions have not been dealt with in the literature, let alone considered as mixed copies.

The copies under consideration are based on Estonian compound nouns N GEN + N NOM or sometimes on adjective phrases, like (1). They are often built on common internationalisms, only that Estonian uses the genitive form and Russian a relative adjective. Consider (10a) and (10b):

(10a) mixed copy
май-ск-ий  йоокс
maj-sk-ij  jooks
‘May marathon’
Russian has so-called relative adjectives that are usually derived from nouns, like дерево: деревянный (derevo: derevjannyj) ‘tree: wooden’ etc. Quite frequently, monolingual translation equivalents of Estonian compounds into Russian yield an adjective phrase with relative adjectives (corresponding to Estonian modifier). In bilingual speech, unconventional relative adjectives may appear (see examples in Verschik, 2008: 130–131). In Example (10a) the existing relative adjective is utilised: май: майский (maj: majskij) ‘May: of May’. As I noted in the explanation of (1), it would be counterintuitive to analyse the Estonian component as a global copy and to disregard the collocation and its Estonian model where the components obviously ‘stick together’ and the same reasoning applies here as well.

This type also comprises instances where the relative adjective is formed either from an Estonian toponym, as in (11a) or from a common noun (12a).

(11a) mixed copy

| iru-sk-ii | hooldekodu |
| Iru- ADJ-NOM MASC ADJ | retirement home |

(11b) Estonian

| Iru | hoolde-kodu |
| Iru:GEN | care:GEN-HOME |

The toponym Iru has become a part of the name for this particular retirement home (Iru hooldekodu). Note that hooldekodu < hooldekodu ‘retirement home’ part is a compound in itself (hoolde ‘care:GEN’ + kudo ‘home’) but it is irrelevant for the current example because, from the copier’s point of view, so to say, it is unanalysed. The toponym is formed with the derivational suffix -ск (-sk-), which is a regular suffix, used to form adjective from both Russian and foreign toponyms: Псков: псковский (Pskov: pskovskij) ‘Pskov: of Pskov’ (Russian), Лондон: лондонский (London: londonskij) ‘London: of London’ (foreign). The adjective iru-sk-ii (iru-sk-ij) ‘of Iru’ follows the derivational model. Thus, it is something less usual than майский (maj-skij) ‘of May’ because the “Estonianness” of Iru is obvious to the Russian-speakers living in Estonia. The
globally copied Estonian component and the specificity of the given collocation (a name of an institution) suggest that the items are perceived as a semantic whole: the overall Estonian shape is recognizable (the Estonian toponym + Estonian noun) and it is the Russian suffix that turns it into a mixed copy.

Example (12a) is notable because the relative adjective is formed from the Estonian stem buss ‘bus’. Although one may argue that Russian has автобус (avtobus) whose second part is essentially the same stem, the derivational suffix -овск- (-ovsk-) used in (12a) is different from the suffix -н- (-n-), conventionally added to the stem автобус: автобусный (avtobus: avtobusnyj) (see Russian in 12b):

(12a) mixed copy

| бусс-овск-ие | полос-ы |
| buss-ovsk-ije | polos-y |
| bus-ADJ-nom PL | lane-NOM PL |

‘buss lanes’

(12b) Russian

| автобус-н-ые | полос-ы |
| avtobus-n-yje | polos-y |
| buss-ADJ-NOM PL | lane-NOM PL |

‘buss lanes’

Compare to Estonian in (12c):

(12c) buss-i raja-d

| buss-gen | lane-NOM PL |
| бусс-gen | lane-NOM PL |

‘bus lanes’

Judging from the context, I cannot altogether exclude that it was a deliberate creation on the part of the blogger and that (12a) has somewhat ironical undertones. It was a discussion on the changes in Tallinn traffic during the summer 2012 when more space was given to bus lanes at the expense of other vehicles and the current mayor of Tallinn was subject to criticism and jokes. The lexical item bussirada ‘bus lane’ or plural bussirajad ‘bus lanes’ was very much in everybody’s mind and speech.

To sum up, the following points concerning mixed copies in blogs should be developed further: composite meaning, frequency and spelling (choice of orthography). This will be done in the next section. The suggested typology of mixed copies is summarized in Table 1.
Table 1  Types of mixed copies

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>Mostly compound nouns; the most prototypical case, often involves common internationalism that, nevertheless, have a sufficiently different shape in two languages.</td>
<td>сфёсту-карточка (sjaéstu-kartotška) ‘saving card’ &lt; säästukaart</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cf. Rus kartotška and Est kaart</td>
</tr>
<tr>
<td>Type 2</td>
<td>Compounds with a common internationalism that has a very similar shape in two languages; a border-line case between global and mixed copying.</td>
<td>иускукод (isikukod)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘personal code’ &lt; isikukood</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cf. Rus kod and Est kood</td>
</tr>
<tr>
<td>Type 3</td>
<td>Globally copied component is an abbreviation</td>
<td>JOKK-схема (JOKK-sxema)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘legal but ethically questionable scheme’ &lt; JOKK-skeem</td>
</tr>
<tr>
<td>Type 4</td>
<td>Globally copied component is a proper name or place name; it “brings in” the word order of the model code</td>
<td>Сыпрузе мост (Sypruże most) ‘Sõpruse bridge’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt; Sõpruse sild</td>
</tr>
<tr>
<td>Type 5</td>
<td>Construction where semantically specific component is globally copied</td>
<td>оставили лёбваатамата (ostavili lëbvaatamata) ‘(they, one) disregarded, ignored’ &lt; jäteti lëbvaatama-ta (verbal construction jätma X-mata)</td>
</tr>
<tr>
<td>Type 6</td>
<td>Fixed expression, collocation etc where stems are global copies and derivational morphology comes from the basic code</td>
<td>майский йоокс (maj-sk-ij jooks) &lt; majjoks’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>May marathon’</td>
</tr>
</tbody>
</table>

6 Discussion and Conclusions

Backus (2012) calls for a more precise look into what he calls ‘complex insertions in code-switching data’. Such insertions include collocations such as adjective + noun, verb + object, idiomatic or semi-idiomatic expressions etc. Items described in the current article fit under this heading.

In Backus and Verschik (2012) it is argued that meaning is all-important in the discussion on borrowability (or, I would say, copiability): highly specific meaning (in content words) is likely to yield global copies, whereas general
meaning (functions of cases, for instance) is likely to be copied selectively. This also correlates with frequency: the more specific items are less frequent and items with a more general meaning (case markers, prepositions) are more frequent. In this sense, mixed copies may be helpful in the research on the correlation between meaning and degrees of copying: in the examples analysed above the component that makes the meaning more precise and specific (modifier in compound nouns/adjective in adjective + noun collocations) is usually copied globally. Thus, säästukaart ‘saving card’ is more precise than just kaart ‘card’; hence, the probability is high that the Estonian item will end up either as global copy or as a mixed copy where säästu- ‘saving’ (genitive) is copied globally.

Indeed, some of examples with common internationalisms are borderline cases where it is difficult to distinguish between global copying and selective copying (код ~ код ‘code’, cf. Estonian kood ‘code’) but examples with abbreviations as modifiers and verb construction in example (9a) are rather clear. Type 6 where stems may be both Russian and Estonian always has Russian derivational suffixes, whose meaning by definition is more general than that of content words. Thus, the evidence is in accordance with the hypothesis on the correlation between type of meaning and degree of copying.

Another question is that of variation in degree of copying. Although content words are likely to be globally copied, their selective copying is not impossible (consider, for instance, copying of meaning: Estonian panema ‘to put, to place (both vertically and horizontally)’ > Estonian Russian поставить (postavit’) with the same (broadened) meaning, while in monolingual Russian it means only ‘to place vertically’). I have not considered copying of idiomatic expressions here because these are not very numerous in my data but I have to mention that variation is possible there as well: global copying in some instances and selective copying (loan translation) in genres that imply monolingual use (that is, no overt use of obviously foreign lexical items) but also mixed copying (see Verschik, 2008: 102). Interestingly, Johanson (1998: 334) observes variation in the copying of the same expression: it exists in Turkic as a global copy from Persian rūz be rūz ‘day to day’ and as a mixed copy gün be gün, where gün is ‘day’ in Turkic and be ‘to’ is Persian. In that case the question is what kind of variation is possible: what parts can or cannot be copied globally and to what extent this depends on the kind of a particular complex item (for instance, how loosely are the components bound, whether one of them can be easily replaced etc). As I showed, a hypothetical Example (9b) jättis без рассмотрения (bez rass-motrenija) ‘dismissed’ where the verb with a more general meaning ‘to leave’ is copied globally and the Abessive form selectively seems counter-intuitive.

Another issue to be discussed is spelling and orthography of mixed copies. Written texts provide additional material for a kind of analysis that cannot be
done with oral multilingual speech. In the in-between examples with common internationalisms, it is often impossible to draw a precise line between global copies and mixed copies because a Russian-speaker can have a Russian accent in Estonian that will, so to say, neutralise the internationalism in question (i.e., these will sound more Russian-like not because of a conscious adaptation to monolingual norms but because the speaker is unable to produce them in any other way). In writing, however, the use of an Estonian version (albeit transliterated into the Cyrillic script) and the use of a Russian version can be distinguished.

One may say then that this is simply a kind of graphemic adaptation to Russian but it would be too general a claim that internationalisms in mixed copies always undergo a graphemic adaptation to Russian (recall examples (5a) and (5b)). While it is beyond doubt that commonalities, including common internationalisms, facilitate copying, we deal with multi-word items here and the point is that language users perceive them as a whole rather than as isolated components (whether more or less adapted).

I am rather cautious to claim that rendition in Cyrillic gives evidence of an item’s conventionalization. The reason for using Cyrillic throughout may be that a user does not bother to switch to the Roman keyboard and back. However, when a user does bother to switch, like in (7) and in (10a), this signals his/her awareness that an item ‘belongs to another language’, so to say.

All in all, I suggest that when mixed copies appear in written-like CMC genres, this may be indicative of entrenchment (becoming a routine practice) on the level of individual users (disregard of chosen alphabet). The reason for this suggestion is that writing, even in highly personalized environments such as blogs where adherence to current (monolingual) norms is not necessary, is more premeditated than speech. While mixed copies in my data appear more seldom than global copies, they nevertheless are present and call for a closer look.

Compared to my earlier research (Verschik, 2008), I have found the same patterns of mixed copying (except for particle verbs that I have not discovered in blogs, another question to ponder): (1) compound nouns, (2) constructions, idioms, fixed expressions, (3) semantically specific nouns + word order in NP.

Within copied compounds, there is a type where the modifier takes Russian adjective derivational suffixes, while the head is Estonian. I have not encountered such copies from Estonian in oral speech but this, of course, does not mean they do not exist.

On a more general level, complex items have become a prominent topic in cognitive linguistics (and, of course, construction grammar). For instance Barlow (2000) speaks of blending as a fundamental cognitive process that
affects language (the rise of collocations, constructions, fixed expressions etc). In the same spirit, Langacker (2000) discusses constructions and compositional meaning. Data from multilingualism research and cases of mixed copies in particular are in accordance with these ideas: a multilingual language user probably does not perceive isolated components, but the composite nature of a multi-word item/blend/lexical chunk in the model code, and reproduces it in the basic code. What kind of variation in the degree of copying is possible and what components are more likely to be globally/selectively copied is a matter for further investigation on a broader empirical scale. The presence of mixed copies in CMC gives evidence of internalization thereof at least at the individual level.

References


