

I (2016)

# Philologia Estonica

T A L L I N N E N S I S

**KEELEKONTAKTIDE JA  
MITMEKEELSUSE KEELELISED,  
SOTSIAALSED JA KOGNITIIVSED  
ASPEKTID**

*Linguistic, Social and Cognitive  
Aspects of Language Contacts and  
Multilingualism*

Tallinna Ülikooli Kirjastus  
Tallinn 2016

Philologia Estonica Tallinnensis 1 (2016)

Keelekontaktide ja mitmekeelsuse keelelised, sotsiaalsed ja kognitiivsed aspektid  
Linguistic, social and cognitive aspects of language contacts and multilingualism

Philologia Estonica Tallinnensis eelkäija on Tallinna Ülikooli Eesti Keele  
ja Kultuuri Instituudi toimetised (ilmus 2004–2015, ISSN 1736-8804)

Toimetuskolleegium / Advisory Board Lars Gunnar Larsson (Uppsala),  
Maisa Martin (Jyväskylä), Kaili Müürisepp (Tartu), Helle Metslang (Tartu),  
Meelis Mihkla (Tallinn), Renate Pajusalu (Tartu), Helena Sulkala (Oulu),  
Urmas Sutrop (Tartu), Maria Voeikova (Viin/Sankt-Peterburg), Cornelius  
Hasselblatt (Groningen), Epp Annus (Eesti Kirjandusmuuseum / Ohio  
State University)

Peatoimetaja: Reili Argus

Toimetaja: Anna Verschik

Ingliseelsete artiklite keeleteoimetaja: Colm James Doyle

Eestikeelsete artiklite keeleteoimetaja: Victoria Parmas

Autoriõigus: Tallinna Ülikooli Kirjastus ja autorid, 2016

ISSN 2504-6616 (trükis)

ISSN 2504-6624 (võrguväljaanne)

Tallinna Ülikooli Kirjastus

Narva mnt 25

10120 Tallinn

[www.tlupress.com](http://www.tlupress.com)

Trükk: Pakett

# SISUKORD

## **Anna Verschik**

Eessõna . . . . .	5
Introduction . . . . .	11

## **Daria Bahtina-Jantsikene, Ad Backus**

Limited common ground, unlimited communicative success: an experimental study into Lingua Receptiva using Estonian and Russian . . . . .	17
<i>Piiratud ühisteadmised, piiramatud suhtlusvõimalused: eesti-vene retseptiivse kakskeelsuse eksperimentaaluuring . . . . .</i>	<i>37</i>

## **Jim Hlavac**

Code-switching, lexico-grammatical features and loan translation: data from a large Macedonian-English corpus. . . . .	38
<i>Koodivahetus, leksikaalgrammatilised omadused ja tõlkelaenud: andmed suurest makedoonia-inglise korpusest . . . . .</i>	<i>61</i>

## **Elīna Joenurma**

Eesti-läti koodikopeerimine: adaptatsioon ja impostsioon . . . . .	62
<i>Estonian-Latvian code-copying: adoption and imposition . . . . .</i>	<i>78</i>

## **Helin Kask**

English-Estonian code-copying in Estonian blogs . . . . .	80
<i>Inglise-eesti koodikopeerimine blogides . . . . .</i>	<i>102</i>

**Lea Meriläinen, Helka Riionheimo,****Päivi Kuusi and Hanna Lantto**

Loan translations as a language contact phenomenon:

Crossing the boundaries between contact linguistics, second language acquisition research and translation studies . . . . . 104

*Tõlkelaenud kontaktlingvistilise nähtusena:**ületades kontaktlingvistika, teise keele omandamise**ja tõlketeaduse piire . . . . . 125***Kristiina Praakli**

Eesti-soome koodivahetuse mitu nägu Facebooki

vestluste näitel. . . . . 126

*The multiple faces of Estonian-Finnish code-switching**seen from Facebook conversations . . . . . 152***Anette Ross**

Estonian Lotfitka Romani and its contact languages . . . . . 154

*Eestis kõneldava roma keele Lotfitka murre ja selle**kontaktkeeled. . . . . 173***Virve-Anneli Vihman**

Code-switching in emergent grammars: Verb marking

in bilingual children's speech . . . . . 175

*Koodivahetus arenevas grammatikas: verbi markeerimine**kakskeelsete laste kõnes . . . . . 199*

# EESSÕNA

Anna Verschik

Tallinna ülikool

Selle aasta kogumiku teema on „Keelekontaktide ja mitmekeelsuse keelelised, sotsiaalsed ja kognitiivsed aspektid“. Ükski keel ei eksisteeri teistest keeltest isoleeritult ja nii-öelda kontaktivabalt; iseasi on, kas see on uurija jaoks perifeerne asjaolu või vastupidi, uurimise põhiobjekt.

Kogumiku autorkond on rahvusvaheline, on nii tunnustatud kui algajaid teadlasi (magistrante, doktorante). Peaaegu kõik artiklid käsitlevad kontaktsituatsioone, kus üks osaline on eesti keel.

Teemade ring on päris lai. Lea Meriläinen, Helka Riionheimo, Päivi Kuusi ja Hanna Lantto artikkel on ülevaade teooriatest, mis ühel või teisel määral käsitlevad tõlkelaene. Tõlkelaenudele makeedoonia-inglise kakskeelses kõnes läheneb kontaktlingvistika seisukohalt ka Jim Hlavac. Virve Vihmani ja Jim Hlavaci artiklid testivad kontaktlingvistikas tuntud teoreetilisi mudeleid ja tõdevad järjekordselt, et universaalseid piiranguid pole olemas ning et keelekontaktide protsessis sünnivad uuendused, mis ei järgi kummagi keele ükskeelset grammatikat. Anette Ross käsitleb Eestis elavate romade murret (Lotfitka) teiste roma keelekujude kontekstis ja üldisel keelekontaktide intensiivsuse skaalal (Thomason and Kaufman 1988).

Kaks artiklit keskenduvad mitmekeelsele virtuaalsele suhtlusele. Kristiina Praakli uurib Soomes elavate eestlaste suhtlusrühma mitmekeelse pragmaatika seisukohalt ning Helin Kask arutleb inglise keele mõju üle eesti moeblogides.

Koodikopeerimismudelit (Johanson 1993) kasutab Helin Kask just mainitud artiklis, aga ka Elina Joenurma eesti-läti kakskeelse kõne uurimiseks. Viimases pööratakse tähelepanu mõlemasuunalisele mõjule.

Daria Bahtina-Jantsikene ja Ad Backus käsitlevad eesti-vene retseptiivset kakskeelsust. Erinevalt teistest autoritest on nende artiklis tegu eksperimentaalse metodoloogiaga. Võib öelda, et retseptiivse kakskeelsuse teema kerkib üles ka Kristiina Praakli artiklis, kuna ta uurib kahe lähedase sugulaskeelega kontakte.

### Lähemalt artiklitest

Ad Backus ja Daria Bahtina-Jantsikene käsitlevad retseptiivset mitmekeelsust (*lingua receptiva* ehk LaRa) eesti-vene suhtluses. Retseptiivne mitmekeelsus tähendab seda, et kumbki osapool kõneleb oma keeles või vajaduse korral mugandab seda ning et kumbki osapool suudab enam-vähem mõista, mida teine omas keeles räägib (lähemalt vt Rehbein, ten Thije and Verschik 2012). Selline suhtlemine ei piirdu ainult lähedaste sugulaskeelte paariga, vaid on võimalik ka mittedugulaskeelte puhul, kui kummalgi kaasvestlejal on teise keele passiivne oskus. Eksperimendi andmete põhjal selgus, et teise keele oskus ei ole alati ainuotsustav ning et edukas suhtlus on võimalik ka piiratud keeleoskuse puhul. See võib tähendada seda, et keelejuhid on hinnanud oma keeleoskust liiga tagasihoidlikult, aga ka seda, et kommunikatsioon ei sõltu ainuüksi keeleoskusest, vaid sõltub ka nn meta-kommunikatiivsetest strateegiatest.

Jim Hlavac räägib makedoonia-inglise koodivahetusest Austraalias, pöörates tähelepanu eelkõige kergverbile. Kergverb kui analüütiliste verbikonstruktsioonidega osa (inglise keeles *light verb* ehk *dummy verb*) eksisteerib makedoonia keeles marginaalselt (keelekorraldajad on pidanud vajalikuks see analüütiline konstruktsioon kirjakeelest välja rookida), kuid olemasolev võimalus „tuleb appi“, kui on vaja integreerida inglise verbe. Tavaliselt on tuumaks mõni üldise tähendusega verb, nagu tegema, saada vms. Nõnda tekib konstruktsioon, nagu *walking praješ* ‘jalutad’ (tegijanimi verbist *walk* ‘kõndima’ + makedoonia verb tähendusega ‘tegema’ vastavas isikus, arvus ja ajas). Pole selge, kas mudeliks on marginaalselt eksisteeriv makedoonia konstruktsioon või mõni inglise püsiväljend,

nagu *to do shopping* 'sisseoste tegema', *to have lunch* 'lantši sööma' > makedoonia *ima lunch* (ima 'omama').

Elina Joenurma keskendub eesti-läti kakskeelse keelejuhi kõnele. Tasakaalus kakskeelsuse tingimustes ei saa öelda, kumb keel on nii-öelda tugevam, kuigi kronoloogiline pilt on selge (eesti keel on omandatud esimesena). Kuigi mõju kahesuunalisus ei ole kontaktlingvistikas uudne asi, on täiskasvanud kakskeelsete uurimustes tavaliselt rõhuasetus kas  $K1 > K2$  (teise keele omandamise uurimuses) või  $K2 > K1$  (enamik kontaktlingvistilist kirjandust). Siin aga pöörab autor tähelepanu mõlemale suunale, läti-eesti ja eesti-läti, kasutades koodikopeerimismudelit (Johanson 1993). Selgub, et see, mida ja kuidas kopeeritakse, on mõlema suuna puhul küllalt sarnane. Erinevusi võib leida diskursuspragmaatiliste sõnade kopeerimisel: pigem kopeeritakse eesti keelest läti keelde kui vastupidi. See võib tähendada muuhulgas seda, et keelejuhi jaoks on eesti keel pragmaatiliselt domineeriv (Matras 1998 terminoloogias).

Helin Kask uurib inglise-eesti koodikopeerimist eestlaste moeblogides. Inglise keele positsioonist ja mõjust Eestis on seni kirjutatud pigem makrotasandil, toetudes küsitlustele. Artikli keskmes on aga kontaktlingvistilised aspektid. Enamik koopiaid inglise keelest on täielikud koopiaid (sõnad, püsiühendid, väljendid). See on oodatav ja arusaadav, sest leksikaalne ja semantiline mõju avaldub keelekontaktide varajases staadiumis. Samas on palju ka tekstitüübit: mitmekeelsus on moeblogides pigem norm, blogi on monoloogiline formaat (erinevalt jututoast, foorumist vms) ja seetõttu ei ole isegi pikemad lõigud teises keeles võimatud. Inglise elemendid pole aga alati eesti grammatika kohaselt mugandatud. Lisaks semantiliselt spetsiifilisele (valdkonna omasele) sõnavara kopeerimisele esineb hulgaliselt inglise diskursuspragmaatilisi sõnu (*well, anyway* vms). Viimased esinevad ka noorema põlvkonna suulises kõnes ning ei ole valdkonna ega tekstitüübi spetsiifilised. Valikulisi koopiaid (mallide ja konstruktsioonide kopeerimist) on vähe (2% juhtumitest), kusjuures segakoopiaid (*epic-kiire* taolisi näiteid) on rohkem (7%).

Lea Meriläinen, Helka Riionheimo, Päivi Kuusi ja Hanna Lantto, vaatlevad, kuidas eri distsipliinid (teise keele omandamise uurin-gud, kontaktlingvistika ja tõlketeadus) käsitlevad tõlkelaene. Kõne-alused distsipliinid kasutavad päris erinevat metakeelt ja suhtumine tõlkelaenudesse erineb, kuigi tõlkelaenamise kognitiivseid aluseid nähakse samamoodi. Teise keele omandamise seisukohalt nähakse tõlkelaene eelkõige esimese keele mõjuna. Selle tagajärjel produtsee-ritakse vorme, mis sihtkeeles puuduvad (nn *non-target forms*). Kon-taktlingvistika käsitleb tõlkelaene neutraalselt kui emma-kumma keele mõju all tekkinud uuendust. Tõlketeaduses on teema üllatavalt marginaalne. Tõlkelaene käsitletakse neologismidena. Kõigi kolme puhul on ühine, et tõlkelaen ei ole pelgalt leksikaalne nähtus, vaid võib sisse tuua ka morfosüntaktilisi ja semantilisi uuendusi.

Kristiina Praakli analüüsib materjali, mida ta on kogunud Soomes elavate eestlaste Facebooki rühma leheküljelt. Võrreldes teiste artiklitega, on see kõige vähem keelesüsteemikeskne. Põhirõhk on kakskeelse suhtluse pragmaatikal. Teemaalgataja keelevelik on üht-lasi vihje kommenteerijatele. Antud virtuaalkeskonnas, nagu suu-lises kõneski, on refereerimine (tsiteerimine, teiste isikute sõnavõt-tudele viitamine/ümberjutustus) koodivahetuse põhifunktsiooniks. Koodivahetus on rühmas pigem kirjutamata norm kui erand, siiski väga pikki soomekeelseid löike ei taheta aktsepteerida. Kui eestla-sed omavahel kirjutavad pikalt soome keeles, tekitab see paljudes imestust ja kutsub esile metalingvistilisi kommentaare keeleveliku kohta. Koodivahetus on sujuv ja peaaegu ei mõjuta morfosüntaksi (vrd näiteks Virve Vihmani ja Helin Kase artiklitega).

Anette Ross uurib Eestis elavate romade keele suhteid teiste roma keelevariantide ja muude mõjukeeltega (vene, läti, eesti). Artikkel paigutab Eestis elavate romade Lotfitka keelevariandi Tho-masoni ja Kaufmani (1988) laenatavuse skaalasse. Keele säilitamise puhul algavad muutused sõnavarast, järgmises staadiumis on või-malik tähenduste ja mõningate struktuuride laenamine (sõnajärg, rektsioon, intonatsioon vms) jne. Eestis elavad romad on enamasti pärit Lätist ning kõnelevad nn Lotfitka (Läti) murret, mis kuulub



kirderoma murderühma (koos Venemaa romade murde Xaladytka, Poolas kõneldava Polska romaga jne). Lotfitkale on ajalooliselt mõju avaldanud läti ja vene keel. Samuti elab Eestis teatud hulk Xaladytka variandi kõnelejaid. Praeguseks on läti keele mõju nõrgenenud ja eestikeelses keskkonnas on uue mõjukeelena üles kerkinud eesti keel. On märgata eesti keele leksikaalset mõju, aga ka tähenduste laenamist. Mitmed leksikaalsed laenud säilitavad eesti foneeme. Kontaktid eesti keelega paiknevad Thomasoni ja Kaufmani skaalal esimese (juhuslik kontakt) ja teise (intensiivsem kontakt) staadiumi vahel.

Virve-Anneli Vihman mõtestab lahti kontaktlingvistikas laialt tuntud mudelid (maatrikskeeleraamistiku mudel ehk MKR ja selle täiendatud versioon, nn 4M mudel) ja püstitab küsimuse, kas keelekontaktid toimuvad ühe osalause sees või ka mujal. Ta järeldab, et mudelid töötavad siis, kui maatrikskeel on selgelt määratav, aga alati see nõnda pole. Üks sisestatud verb võib muuta kogu lause grammatikat, nt ütluses *doesn't täida soovid* 'ei täida soove' peaks MKR järgi eestikeelne osa käituma eesti grammatikale vastavalt (oodatav oleks mitmuse osastav), aga nii ei juhtu. Seega ei saa alati öelda, kumb keel annab ette grammatilise raamistiku. Seega mõlemad keeled võivad dikteerida kakskeelse kõne grammatikat. Sarnaseid tähelepanekuid on ka teiste keelepaaride kohta (Auer and Muhamedova 2005).

Toimetaja loodab, et selle kogumikuga õnnestub kontaktlingvistika probleemistik lugejale lähemale tuua ja näidata, et kõik autorid osalevad ühises rahvusvahelises diskussioonis ja kasutavad sama metakeelt, olenemata sellest, mis keelte materjaliga nad töötavad.

## KIRJANDUS

- Auer, Peter; Muhamedova, Raihan 2005. 'Embedded language' and 'matrix language' insertional language mixing: Some problematic cases. *Italian Journal of Linguistics / Rivista di linguistica* 17 (1), 35–54 (special issue, ed. G. Berruto).
- Johanson, Lars 1993. Code-copying in immigrant Turkish. – Immigrant languages in Europe. Eds. G. Extra, L. Verhoeven. Clevedon, Philadelphia, Adelaide, 197–221.
- Matras, Yaron 1998. Utterance modifiers and universals of grammatical borrowing – *Linguistics* 36 (2), 281–331.
- Rehbein, Jochen; ten Thije, Jan D.; Verschik, Anna 2012. Remarks on the quintessence of receptive multilingualism. Eds. Jan D. ten Thije, Jochen Rehbein, Anna Verschik. *Lingua Receptiva. International Journal for Bilingualism* 16 (3), 248–264.
- Thomason, Sarah Grey; Kaufman, Terence 1998. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.

# INTRODUCTION

**Anna Verschik**

Tallinn University

The topic of the current yearbook is „Linguistic, social and cognitive aspects of language contacts and multilingualism“. No language exists in isolation nor in a contact-free environment; another matter is whether this knowledge is peripheral for a researcher or, on the contrary, is the focus of his/her research.

The authors of the current article collection come from different countries, and include both acknowledged scholars in the field of contact linguistics as well as beginners (MA and PhD students). Almost all the papers deal with linguistic situations where Estonian is involved.

The range of topics is rather broad. For instance, Lea Meriläinen, Helka Riionheimo, Päivi Kuusi and Hanna Lantto provide an overview of theories that explore loan translations to a greater or lesser extent. The topic of loan translations is considered from a contact linguistic perspective by Jim Hlavac. The papers by Virve-Anneli Vihman and Jim Hlavac test some well-known contact linguistic theoretical models and arrive at the conclusion that universal constraints on language contacts do not exist and that in the process innovations emerge that are not „well-formed“ from the point of view of two separate monolingual grammars. Anette Ross describes the Lotfitka variety spoken by the majority of Roma in Estonia against the background of other Romani varieties and considering the scale of borrowability proposed by Thomason and Kaufman (1988).

Two articles explore multilingual communication on the internet: Kristiina Praakli describes practices of communication in a Facebook community of Estonians living in Finland from a pragmatic perspective, while Helin Kask considers English impact in Estonian fashion blogs.

Two articles employ a code-copying framework (Johanson 1993): the above mentioned article by Kask and a study on Estonian-Latvian bilingual speech by Elina Joenurma. The latter focuses on bidirectionality of impact.

Daria Bahtina-Jantsikene and Ad Backus investigate Estonian-Russian receptive bilingualism. Differently from the rest of the authors, their research is based on experimental methodology. One may say that the topic of receptive bilingualism arises also in Praakli's article because she focuses on the contact of two closely related languages, Estonian and Finnish.

### Brief summary of the papers

Ad Backus and Daria Bahtina-Jantsikene describe receptive multilingualism (*lingua receptiva* or LaRa) in Estonian-Russian communication. Receptive multilingualism is a mode of communication in which each participant uses his/her language and adjusts language use if needed. Thus, all participants are able to understand to an extent what is being said (for more details see Rehbein, ten Thije and Verschik 2012). This mode of communication is not limited to closely related varieties but is also possible if both participants have at least a passive command of each other's variety. As communication in this experimental setting demonstrated, proficiency in the co-participant's language is not the only decisive factor that contributes to successful communication, and that achievement of communicative goals is possible also if proficiency is limited. This may mean that the informants are too critical in their assessment of their proficiency and/or that success depends on so-called metacommunicative strategies as well.

Jim Hlavac describes Macedonian-English code-switching in Australia with a special focus on light verbs. Light verbs exist in Macedonian, albeit marginally (language planners considered it their business to purge them from the standard language). Yet, the tendency comes in handy when there is a need to integrate an English

verb into a Macedonian grammatical matrix. Thus, constructions emerge such as *walking praješ* 'you walk' (the gerund form of the verb *walk* + Macedonian verb 'to do' in the appropriate person, number and tense). It is not entirely clear whether the marginally existing possibility in Macedonian has served as a model or whether English fixed expressions such as *to do shopping*, *to have lunch* and the like have resulted in Australian Macedonian *ima lunch*.

Elina Joenurma focuses on the speech of an Estonian-Latvian bilingual informant. In the situation of balanced bilingualism it is not possible to determine which of the two languages is dominant, although the chronology of acquisition is clear enough (Estonian is L1). Although bidirectionality of impact is not unknown in contact linguistics, in research on adult bilingualism the stress is either on L1 > L2 impact (SLA research) or L2 > L1 (most contact linguistic research). The author considers impact in both directions and employs a code-copying framework (Johanson 1993). What gets copied and to what degree is not very different in Estonian to Latvian and in Latvian to Estonian copying. The difference is in copying of discourse pragmatic particles, where the direction of copying is from Estonian to Latvian. This may mean that, in the terms of Matras (1998), Estonian is pragmatically dominant language for the informant.

The contribution by Helin Kask deals with English-Estonian code-copying in online fashion blogs. So far the position of English in Estonia has been mostly described and analysed from a macrosociolinguistic perspective (based on surveys, etc.). The article concentrates on contact linguistic aspects of the English influence. Most of the copies are global copies (one or multi-word lexical items, idioms). This is rather expected and understandable because lexical and semantic impact appears in early stages of language contact. Yet much depends on the genre of text in question: multilingualism is a norm in Estonian fashion blogs. A blog is a monological format (different from chats or forums) and, therefore, even longer stretches in another language are possible. English lexical items are not always

entirely integrated into Estonian grammar. In addition to semantic-specific lexical items (fashion terms), English lexical impact is visible in discourse pragmatic particles (e.g. *well*, *anyway*). The latter are present in oral speech of young Estonians as well and are not specific to any topic or text type. In the data, selective copies are rather rare (2 % of all copies) and mixed copies (for instance, *epic-kiire* ‘epic fast’) are slightly more frequent (7 % of occurrences).

Lea Meriläinen, Helka Riionheimo, Päivi Kuusi and Hanna Lantto provide a picture on how different linguistic disciplines view loan translations. The disciplines in question (SLA, contact linguistics, translation studies) use rather different metalanguage, although all three agree on the cognitive basis of loan translations. SLA primarily sees loan translations as L1 impact. This impact is responsible for production of non-target forms. Contact linguistics sees loan translations in a neutral light as innovations appearing as a result of impact from either language. Surprisingly, the topic is somewhat marginal in translation studies, where loan translations are viewed as neologisms. All three disciplines agree that loan translations are not a merely lexical phenomenon but that they can also introduce morphosyntactic and semantic innovations.

Kristiina Praakli analyses data from a Facebook group of Estonians residing in Finland. Compared to the other articles, this one is less concentrated on linguistic structures. The author instead chooses a pragmatic approach to bilingual communication. Language choice by topic starter is also a hint to commentators. Like in oral communication, one of the main functions of code-switching in this virtual environment is the reference function (quotation, rendition of other people’s speech, digest). Code-switching is an unwritten norm in the community, yet very long stretches in Finnish do not seem to be acceptable. The fact that some Estonians choose to communicate to each other in Finnish is surprising to others and leads to meta-linguistic comments on language choice. Code-switching is smooth and does not affect morphosyntax (but see Kask, Vihman in this volume).

Anette Ross describes Estonian Lotfitka, a variety used by Roma people in Estonia with reference to the other Romani and non-Romani varieties (such as Russian, Latvian, Estonian) that have impacted it. The language is placed into the context of the borrowability scale proposed by Thomason and Kaufman (1988). In the situation of language maintenance, changes start from the lexicon and later borrowing of meaning and non-core structural elements (word order, intonation, argument structure, etc.) becomes possible. The Roma in Estonia have mostly arrived from Latvia and their variety belongs to the North-Eastern family of Romani varieties (together with Russian Romani (Xaladytka), the Polish variety of Polska Roma, etc.). There exist some Xaladytka speakers in Estonia as well. Historically, Lotfitka has been impacted by Russian and Latvian. As of today, Latvian impact on Estonian Lotfitka has weakened and Estonian has a growing influence. There is evidence of Estonian lexical and semantic impact. New lexical items preserve Estonian phonemes. In the framework of Thomason and Kaufman, contacts between Estonian and Lotfitka are between stage one (casual contact) and stage two (more intensive contact).

Virve-Anneli Vihman tests highly influential models such as MLF and the M4 model and questions whether language contacts occur within a clause or also outside it. She concludes that the models are accurate in situations where the matrix language is clear, but this is not always the case. For instance, an inserted English verb may affect the grammar of the entire clause: in *doesn't täida soovid* 'does not fulfil wishes' the Estonian verb phrase is not well-formed from the point of view of Estonian monolingual grammar and instead the partitive plural (*soove* 'wishes') should appear; yet, the form *soovid* is in nominative plural. Thus, both languages can contribute to the grammar of bilingual speech. Similar observations have been made about other language pairs as well (Auer and Muhamedova 2005).

It is my hope that the collection will familiarise readers with contact linguistic research and will successfully demonstrate that

all authors participate in an international discussion and share theoretical metalanguage, no matter what language pair(s) they investigate.

#### REFERENCES

- Auer, Peter and Muhamedova, Raihan 2005. 'Embedded language' and 'matrix language' insertional language mixing: Some problematic cases. *Italian Journal of Linguistics / Rivista di linguistica* 17 (1), 35–54 (special issue, ed. G. Berruto).
- Johanson, Lars 1993. Code-copying in immigrant Turkish. – *Immigrant languages in Europe*. (Eds.) G. Extra, L. Verhoeven. Clevedon, Philadelphia, Adelaide, 197–221.
- Matras, Yaron 1998. Utterance modifiers and universals of grammatical borrowing – *Linguistics* 36 (2), 281–331.
- Rehbein, Jochen, ten Thije, Jan D. and Verschik, Anna 2012. Remarks on the quintessence of receptive multilingualism. In ten Thije, J. D.; Rehbein, J. and Verschik, A. (Eds.) *Lingua Receptiva. International Journal for Bilingualism* 16 (3), 248–264
- Thomason, Sarah Grey and Kaufman, Terence 1998. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.



# LIMITED COMMON GROUND, UNLIMITED COMMUNICATIVE SUCCESS: AN EXPERIMENTAL STUDY INTO LINGUA RECEPTIVA USING ESTONIAN AND RUSSIAN

Daria Bahtina-Jantsikene

University of Helsinki

Ad Backus

Tilburg University

**Abstract.** Previous research on Estonian-Russian interaction suggests that *lingua receptiva* (LaRa), or Receptive Multilingualism, has the potential to create, boost or restore common ground, or mutual understanding, in situations where common ground is jeopardized. This mode is characterized by the simultaneous use of multiple languages as interlocutors each speak their mother tongue and count on the receptive skills of the other. Alignment in the LaRa mode was tested in a series of experiments. The use of various meta-communicative strategies provided insight into the mechanisms behind LaRa. Their distribution was influenced somewhat by L2 proficiency and exposure to multilingual communicative situations. Interestingly, higher L2 knowledge was not a prerequisite for success. Moreover, it was the composition of the dyad rather than the characteristics of the individuals that had predictive power regarding communicative success.

**Keywords:** lingua receptiva, meta-communicative devices, L2 proficiency, multilingualism, common ground, Estonian, Russian

## Introduction

Communicative success can be conceptualized as the degree to which interlocutors manage to reach common ground (Clark 1996). At the most atomic level of a communicative exchange, two alternating turns between two people engaged in a dialogue, a speaker aims

to get a message across and the hearer is supposed to understand that message. If the latter indeed 'gets it', the two partners may be said to have reached maximal common ground, as they both now know the same thing. In designing a message, a speaker takes into account an educated guess about the state of knowledge the hearer possesses before the exchange. This is easier if there is already a lot of common ground to begin with. In such cases, the hearer will understand most things without problem, including not just the words and grammatical patterns that the speaker uses, but also what's behind all the stylistic choices, the nods and head turns, the intonation patterns, etc. This is why communication with friends or family members about familiar topics in familiar settings tends to be so fast and effortless. At the other extreme are unfamiliar communicative settings. Lack of familiarity can result from at least two sources. The interlocutors might not know each other well, or the communicative task may be unfamiliar. Lack of familiarity produces stressful situations characterized by low degrees of common ground. Communicative success needs effort.

This article reports on an experimental study in which pre-existing common ground was kept small, to see how people would handle this situation and what strategies they would use to nevertheless ensure successful communication. Specifically, language choice was manipulated. Bilingual speakers of Estonian and Russian were instructed to only use their mother tongue in communication with mother tongue speakers of the other language, a pattern known as *Lingua Receptiva* (LaRa) or *Receptive Multilingualism*. This way of communicating, while conventional in a number of multilingual settings around the world, was virtually unknown to the participants. Some of the results were somewhat surprising, as it turned out that limitations in pre-existing common ground sometimes had a beneficial effect for the conversational task, essentially because it induced people to help each other more.

The first section will discuss the role of common ground in communication in general and the way in which it was kept small in the

present study. The subsequent sections present the methodology and the results, focusing on the abovementioned surprises. The final section explores the implications of these findings for future studies.

## Common ground

The importance of ‘common ground’ for communication was emphasized by Clark (1996), as the central concept in his theory of communication (and also taken up by Pickering, Garrod 2004, Tomasello 2008, Croft 2009, and others). In communication, people strive towards maximizing the common ground between them, and the inherent puzzle of communication is how this is achieved given that two interlocutors can never know for sure how much common ground they have between them.

Common ground is greater the more familiar the communicative setting is, and familiarity is high when interlocutors know each other well, and the setting is like many other settings the interlocutors have experienced in the past. Often, these two sources of familiarity go together, for example when a group of close friends engage in small talk on their weekly night out. However, close friends too may sometimes engage in a conversation type they don’t experience together often, for example when one of them suddenly has to explain a dramatic development at his job (e.g. lay-offs) despite the organizational features of his workplace having never before been a topic of conversation in the group. Other examples would include an academic professor telling his/her non-academic cousin about his research when he/she has never done such a thing before, or colleagues running into each other in town, and realizing they have so far only talked about task-related issues at work and never exchanged small talk.

In psycholinguistics, common ground is often conceptualized as alignment between speakers. Research tends to focus on ‘deep’ cognitive processes that are beyond the conscious attention of language users (Pickering, Garrod 2004). Primary evidence for unconscious or

automatic processing is the existence of priming effects in dialogue. If Speaker A uses a particular word or construction, the chances that Interlocutor B will use it too increase, without any evidence that this was because of a conscious act of imitation or because Speaker A somehow urged him to. Alignment can exist at various levels, and Pickering, Garrod (2004) argue for a system in which alignment on one level induces alignment at 'higher' levels, so that priming at lexical and syntactic levels contributes to speaker and hearer mentally representing the same proposition (which they call alignment of 'situation models'). At the same time, however, alignment can also be achieved through more conscious efforts, for example by actually checking whether the hearer has understood what was said. Also, if communication partners are interested in harmony, understanding and conviviality, they might imitate and accommodate to each other for social reasons. A desire for cementing the bond between speaker and hearer increases the chance of mutual understanding because it increases the wish to align. The present study focuses on such conscious ways of facilitating alignment. The participants in our study had to overtly attempt to bring about alignment at the propositional level, and one of our prime objectives was to investigate the means by which they did that.

We particularly focus on how interlocutors overcame the limitation of not being allowed to use the other's first language, and how they used 'meta-communicative devices' (henceforth: 'MCDs') to do this. MCDs and similar concepts have been discussed in various schools in pragmatics, interactional sociolinguistics and conversational analysis, and have consequently been referred to with various terms (e.g. 'hearer-oriented strategies' in Functional Pragmatics, cf. Rehbein et al. 2012). Non-automatic alignment is achieved on the basis of conscious moves that speakers and hearers make in conversation. As these serve to regulate the communication itself, rather than convey referential content, they are 'meta'-communicative. Examples include explicit negotiation about the communicative task at hand, or explicit attention to how a particular word should be

understood in the current context. Further details will be given in the 'Method' section.

### The current study

Common ground is greatest when people who know each other well engage in conversational settings that are routine. In order to study how people manage to build common ground in unfamiliar conversational settings, an experiment was designed in which participants engaged in a communicative task, a so-called 'maze task'. The task was characterized by four features designed to lower the level of familiarity.

First, the interlocutors did not know each other. As a result, they could not rely on the historically built communicative common ground that close friends share. Not having extensive communicative history makes it harder to understand each other's intentions. Second, all participants had a different native language from that of their assigned interlocutor (though they knew the language of the other as an L2, to varying degrees). If your interlocutor speaks the same language as you, you can rely on extensive linguistic common ground, especially if you also share similar social and geographical backgrounds. If you are paired with an L2 speaker, on the other hand, you cannot know as well whether the words and grammatical constructions you employ are known by him or her. Third, as mentioned, they each had to use their own native language during the communicative task they carried out. As participants will not be familiar with this way of communicating, they could not rely extensively on communicative routines that were familiar to them. LaRa is generally assumed to be possible as long as the languages are mutually intelligible, as in much of Scandinavia. However, there is a logical second possibility: speakers can also understand each other if they have *learned* each other's languages. In that case, the native languages of the two interlocutors do not have to be mutually intelligible. This is what applies to our case, as our participants

are bilingual in Russian and Estonian (see the section ‘Methodology’). One important consequence when LaRa is dependent on mastery of a second language (‘Acquired LaRa’) is that you can never be sure *how much* the partner understands. Few people learn a second language so perfectly that they can pass for native speakers, so one always has to work with less than perfect understanding. Finally, the dialogue participants had to interact through Skype, without video, rather than face-to-face. This meant there was relatively poor physical common ground. Nonverbal features such as gestures and body language, which people normally use in meaningful ways in face-to-face communication, could not be used. We were interested in seeing how pairs of participants would try to reach common ground in this less than ideal situation.

The study employed a maze task (cf. Pickering, Garrod 2004). As experimental designs go, maze tasks lead to reasonably natural conversation while allowing for considerable experimental control over the language use that participants will produce. We were particularly interested in whether participants would compensate for the lack of common ground by an increased use of MCDs.

## Methodology

The study we report on had participants carry out a maze game in pairs, through the medium of Skype.

## Participants

As it was important to keep pre-existing common ground relatively small, we paired interlocutors who did not know each other, except for the pilot run, in which participants were recruited from the same company. The members of each pair were native speakers of different languages, one Estonian and one Russian, except for ten monolingual ‘control’ pairs. As all participants were residents of Estonia, they were bilingual to some extent, but varied in the degree to which

they knew the other participant's L1<sup>1</sup>. The participants were pre-selected so that there were comparable numbers of dyads with specific L2 proficiency combinations (both advanced L2-speakers, both limited, mixed).

The 96 participants were grouped into 38 pairs that communicated in the LaRa mode (the 'experimental group'), five pairs that communicated in Russian only and five that communicated in Estonian only (the 'control groups'). Even though the latter two groups communicated in their mother tongue, they were part of the same bilingual Russian-Estonian speech community as the participants in the experimental condition. All participants were told that their interlocutor had at least some command of their L1.

All participants were adults (age range 22–56), two thirds were female, and all had an educational background that included higher education (some were undergraduates and the rest had completed their studies). They were paid for their participation and were informed that the study looked into multilingual practices, but no further details about its purpose were provided.

## Materials and Procedure

As mentioned, the study was designed to limit pre-existing common ground between participants. That was one reason why they were required to communicate with someone who was speaking a language that was an L2 for them. LaRa has been studied either from a historical perspective (e.g., Braunmüller 2007) or in naturalistic settings (e.g. Beerkens 2010). This study was the first that explored

---

<sup>1</sup> Due to historical developments in Estonia over the past 50 years, the respective L2 proficiency is highly asymmetrical across the two speech communities as well as across generations. According to the Estonian Integration Monitoring report from 2015, younger generations of Russian-speakers are increasingly more proficient in Estonian whereas it is mostly older generations of native Estonian speakers who have an active command of Russian as L2 (yet, the majority claims at least passive knowledge of Russian regardless of the age group).

this communication mode in an experimentally controlled environment (see Blees et al 2014 for a similar study). In addition, the study was set up in such a way that people were expected to put effort into maximizing the resources they have in their repertoire. The maze game is not easy: participants need to coordinate their efforts, and the task is virtually impossible to complete without the use of MCDs, needed as they are for checking whether the participants are both on the same wavelength. In order to gain enough information on participants' linguistic and sociolinguistic abilities, they were given a sociolinguistic questionnaire and a second language proficiency test to complete.

The sociolinguistic questionnaire was presented in the mother tongue of the participant and covered self-reported language proficiencies, the extent to which the participant had experience with multilingual situations, and the participant's attitudes towards Estonian and Russian speakers. Self-assessed L2 proficiency was used to pair participants into dyads with specific combinations.

L2 proficiency was further tested through a cloze-test using a set of L2 texts. We opted for a flexible grading system which allowed the inclusion of the partial knowledge that is crucial in LaRa dialogues: responses that included orthographic mistakes or stylistically poor choices were not discarded, but lowered the scores.

The pairs had to jointly conduct the maze task through telephone-like conversations on Skype. In such tasks, participants have to find their way around on a fictional street plan; in our case they had to find each other's location on a very abstract map designed specially to reduce the level of familiarity. It excluded landmarks, place names or any other references that would make the task simple. One extra measure that was taken to increase the difficulty of the task was that there were small differences between the fictional maps of each speaker in a dyad. It was assumed that resolving this task would require active discussion and efforts to reach mutual understanding. The time allocated for completing the task was limited to ten minutes and success was defined as complete, partial or



zero based on whether they managed to locate both, one or no position on the map.

All conversations were transcribed with EXMARaLDA free software tools (Schmidt, Wörner 2009) and then annotated using a unified coding system. The parameters marked during annotation, as far as relevant to the present article, included time to completion, level of success, number of utterances per dyad and per interlocutor, the number of the meta-communicative devices used and their distribution, and the total percentage of interaction that was in the LaRa mode. The results were subjected to statistical analysis, and are reported in full in Bahtina-Jantsikene (2013).

In this paper, we first provide some of these results and then focus on the use of MCDs. While similar categories and terms exist in related fields, we based our classification on the traditions in functional pragmatics and psycholinguistic alignment. We defined four hierarchically ordered meta-communicative devices: MCD1 is any form of explicit negotiation about the communicative task at hand, aiming for achieving common ground about the goal of communication, e.g. ‘what are we doing here?’ Assuming this higher-order alignment has been established, MCD2 is used to make sure the specifics of the current communicative task are shared, e.g. ‘do you see X on your map?’ in the case of our maze task. Even when overall and specific goals are known by both interlocutors, they may not possess sufficient knowledge of all linguistic forms that are used (including codeswitching). MCD3 helps create this knowledge by focusing the other’s attention on specific linguistic forms (e.g. ‘what does that word mean?’). Finally, MCD4 can feature any of the above but has the overall goal of explicitly checking for shared understanding (‘confirmation checks’, e.g. ‘do you follow?’).

The hypothesis was that there would be a trade-off between degree of pre-existing common ground and the use of MCDs: the less one can expect the hearer to understand what one is saying (i.e. the lower the hearer’s proficiency in your native language), the more extensive the use of MCDs would be. One might also expect

particular types of MCDs to be especially relevant in particular constellations of linguistic proficiency of the dyad, and perhaps also be sensitive to other features, such as the dyad's experience with multilingual communication. However, we also remained open to the idea that MCDs are used simply for phatic reasons, since it is also conducive to communicative success if the communication partners establish positive rapport.

## Results

The general results were published in Bahtina-Jantsikene (2013) and will only briefly be summarized here, so see also Bahtina et al. (2013). The analysis here will focus mostly on the use of MCDs, in line with our research question about the degree to which participants in a communicative exchange help each other reach understanding when common ground can be presumed to be small. The first sub-section summarizes the main findings of the whole study; the second section deals with MCDs and their distribution.

### Main findings

We ran a series of tests aimed at understanding the relation between L2 composition of the dyad and their success rate, also taking into account the degree to which participants said they had experience with multilingual communicative contexts. Table 1 contains a selection of results: we report here only significant differences. For that reason, not all dyad types are represented in this Table or in Table 2 in the next subsection. If for a particular dyad type (e.g. HL; i.e. one member with High and one with Low L2 proficiency) there was no significant result for a particular outcome (e.g. reaching zero success), it is not included. The results show a surprisingly low direct impact of high L2 proficiency; one would instead expect the higher degree of common ground that this entails to have a beneficial effect on communicative efficiency. Pairs in which both partners had high

L2 proficiency were not faster than pairs in which L2 proficiency was distributed more unevenly. Perhaps even more surprising, the monolingual pairs (the ‘control groups’) were not statistically faster at completing the maze game than the bilingual pairs. These results suggest that the degree of pre-existing common ground is not necessarily predictive of communicative success, at least not when measured in terms of shared linguistic means. Advanced command of a second language did not guarantee mutual understanding. In fact, dyads with mixed L2 proficiencies had higher success rates on average.

**Table 1:** Statistically significant effects of linguistic proficiency on speed and success rates (HH = dyad with high L2 proficiency levels; HL = dyad with mixed L2 proficiency levels)

Measure of success	Differences that reach significance
Speed of finding points on the map (in seconds)	<p>HH dyads find 1st point on the map slower than Monolingual pairs, <math>t(24)=2.65</math>, <math>p=0.01</math></p> <p>HL dyads find one point fast, <math>F(1,36)=4.28</math>, <math>p=0.05</math></p> <p>HL dyads find both points fast, <math>F(1,31)=5.11</math>, <math>p=0.03</math></p>
Level of success (zero – partial – full)	<p>HL dyads reach partial rather than full success, <math>F(1, 36)=4.05</math>, <math>p=0.05</math></p> <p>HL dyads with higher Russian L2 reach partial rather than full success, <math>F(1, 36)=5.50</math>, <math>p=0.02</math></p> <p>HL and LL dyads reach higher level of success, Fisher’s <math>2*2</math>, <math>p=0.01</math></p>

LaRa dyads in which both interlocutors had higher L2 scores were significantly slower than their monolingual counterparts. This suggests that in this multilingual mode better command of the other’s language can even be an obstacle. Completely against expectation, pairs in which one of the participants reported *lower* L2 proficiency corresponded with *higher* success rates and *faster* completion of the

task. One explanation would be that the partner of a participant with low L2 skills may have been extra careful to make him- or herself understood. The latter point would be in accordance with the general expectation that whenever trouble is to be expected in communication people will try harder to establish common ground, for example by using more MCDs.

### MCD use

Next, we look at the use of meta-communicative devices, specifically focusing on whether their use correlates with L2 proficiency and exposure to multilingual situations. One might expect that experienced multilingual communicators make judicious use of MCDs to enhance understanding. We also expected these MCDs to particularly be of the type where it is checked whether the interlocutor has understood ('confirmation checks'). The results show that this is indeed the case, but with a twist. Dyads in which both interlocutors had high L2 proficiency ('HH') and lots of experience with bilingual communication, showed significantly higher MCD use. Yet, in pairs with extensive experience with bilingual communication confirmation checks (MCD 4) were used much less. Further examination of the data revealed pronounced differences in the *distribution* of MCDs: L2 proficiency and degree of experience with multilingual situations affected the choice of MCD type. Table 2 illustrates the relevant results.

MCD use was highest for speakers who reported high exposure to the L2 and who were in pairs in which both speakers reported high L2 proficiency. This suggests that MCD use comes naturally to speakers who often communicate in bilingual settings. A likely explanation is that such speakers have learned through experience that intercultural communication is more successful if one uses MCDs. Interestingly, among the dyads that reached full success L1 Russian speakers with high exposure to L2 Estonian used MCDs significantly *less* often. These findings suggest that the impact of

**Table 2:** Statistically significant effects of MCD use on task success and of background factors (L2 proficiency combination, experience with multilingual situations, and attitudes towards respective L2 and its speakers) on MCD use

MCD type	Distribution / Success level
Overall use of MCDs	<p>HH dyads use more MCDs than HL, <math>r(16)=0.51</math>, <math>p=0.04</math></p> <p>Subjects with L1 Russian and high exposure to Estonian use fewer MCDs and reach higher success, <math>r(16)=-0.54</math>, <math>p=0.03</math></p> <p>Subjects with L1 Estonian and positive attitude to Russian use fewer MCDs and reach higher success, <math>r(16)=-0.60</math>, <math>p=0.01</math></p>
MCD1	<p>Faster at finding the 1st point on the map, <math>r(26)=4.1</math>, <math>p&lt;0.05</math></p> <p>Faster at finding the 2nd point on the map, <math>r(23)=6.4</math>, <math>p&lt;0.01</math></p> <p>Faster at finding both points on the map, <math>r(33)=3.5</math>, <math>p&lt;0.05</math></p> <p>HL dyads with higher L2 exposure use more MCD1, <math>r(22)=0.49</math>, <math>p=0.02</math></p> <p>HL dyads with subjects with L1 Estonian and higher exposure to Russian use more MCD1, <math>r(22)=0.64</math>, <math>p&lt;0.01</math></p>
MCD2	<p>HH dyads use more MCD2 than HL, <math>F(1,36)=5.83</math>, <math>p=0.02</math></p> <p>Subjects with L1 Estonian and high exposure to Russian use more MCD2 and reach higher success, <math>r(16)=.49</math>, <math>p=0.05</math></p>
MCD3	<p>Subjects with lower L2 tend to use more MCD3, <math>r(38)=-0.22</math>, <math>p=0.06</math></p> <p>Subjects with L1 Russian and lower L2 tend to use more MCD3, <math>r(38)=-0.31</math>, <math>p=0.06</math></p> <p>HH dyads with positive attitudes to respective L2 use more MCD3, <math>r(16)=0.53</math>, <math>p=0.04</math></p> <p>HH dyads with subjects with L1 Russian and positive attitudes to L2 use more MCD3, <math>r(16)=0.56</math>, <math>p=0.02</math></p>
MCD4	<p>HL dyads with subjects with L1 Estonian and higher exposure to L2 use fewer MCD4, <math>r(22)=-0.50</math>, <math>p=0.02</math></p>
MCDs in phases: Before finding first point vs after	<p>HH dyads use more MCDs in Phase 1 than HL dyads, <math>F(1,24)=7.25</math>, <math>p=0.01</math></p> <p>HH dyads use more MCDs in Phase 1 than in Phase 2, <math>t(9)=2.91</math>, <math>p=0.02</math></p> <p>More MCDs in Phase 2 in full success dyads, <math>F(1, 24)=4.60</math>, <math>p=0.4</math></p> <p>More MCDs in Phase 1 in slow dyads, <math>r(26)=-0.82</math>, <math>p&lt;0.01</math></p> <p>More MCDs in Phase 2 in fast dyads, <math>r(26)=0.73</math>, <math>p&lt;0.01</math></p>

exposure and L2 proficiency on task success and on MCD use is not straightforward (but see Discussion).

Using MCD1, i.e. conducting explicit negotiation about the goals of the communicative event, correlated positively with the speed of task completion. Extensive previous experience with bilingual communication appears to stimulate the use of this type of discourse technique. There was no correlation with L2 proficiency, however, only with amount of experience, and this holds especially for people whose L1 is Estonian.

MCD2 is a more difficult device to use, as it requires sophisticated vocabulary to check with the interlocutor whether the layout and orientation of the maps speaker and hearer have in front of them match. It was indeed used more by pairs that should have the least trouble understanding one another: pairs with higher L2 proficiencies ('HH'). More bilingual experience also corresponded with an increased use of this MCD type as did higher chance of success in the experiment, at least for Estonian L1 speakers.

While MCDs 1 and 2 are tightly related to the task at hand, and could be said to be triggered by its unfamiliar nature, the other two MCDs involve strategies people employ every day in conversations with friends, colleagues, officials and strangers. MCD3 was characterized as 'creating linguistic common ground', and includes all meta-talk as well as code-switching. Though it did not correlate with higher success rates in the task or with going faster through it, this device was used both for functional reasons (by less proficient L2 speakers) and presumably phatic ones (by participants with high L2 and self-reported positive attitude to the L2). What this suggests is that when speakers feel understanding is compromised by a perceived lack of sufficient L2 skills on the part of the interlocutor, MCD3 serves as a means to compensate. An example comes from a dyad with an Estonian speaker who was not so fluent in Russian. She informs her interlocutor about a break in communication (Est: 'ee... ma ei tea seda sõna vist', Eng: 'ehm... guess I do not know this word') and the Russian-speaking participant provides a translation

of the key word (Est: 'ülesse', Eng: 'up'). Once the issue is resolved, they continue in their respective L1. Phatic use was in evidence when sufficient L2 knowledge made it safe to talk normally: one HH dyad upon reaching the point on the map that was manipulated to look different for the two interlocutors realized that they had misaligned and the Russian-speaking participant commented in Estonian 'juba mingi jama!' ('already some mess!'), before continuing in Russian, both of them laughing.

MCD4, characterized as a confirmation check, can be viewed as the simplest mechanism to restore understanding in challenging, unfamiliar situations. Little can be said about its use in the data, at least as far as correlations with background characteristics, speed and success rates are concerned. The only significant result suggests that this strategy is used less when pairs have higher experience with bilingual communication, especially in dyads with mixed L2 proficiency. This 'parroting' device could be assumed to compensate for a lack of (perceived) initial common ground, but it becomes less relevant when extensive multilingual experience provides more confidence in one's intercultural abilities.

As for development during the execution of the task, HL pairs, i.e. pairs in which one speaker had low L2 proficiency, tended to use more MCDs, especially MCD3, as the task progressed. In contrast, HH pairs tapered off their use of MCDs. Concerning the effect of MCD use on creating mutual understanding, dyads that continued using MCDs in the second phase, were remarkably better both in terms of task completion and speed in getting there.

## Discussion

Perhaps the first result to highlight is that the LaRa mode of communication, despite participants' lack of experience with it, did not keep people from communicating meaningfully and successfully. It has its limitations, though, as can be seen by the trouble it caused pairs in which both members spoke the other language well. This is

probably because in such situations few people will naturally choose to practice LaRa, unless they already have a communicative history with their interlocutor in which this mode has been established.

It is of course puzzling that pairs which reported low or even zero L2 proficiency still managed to communicate successfully in the LaRa mode. There are various possible explanations. Presumably, some participants underestimated their L2 proficiency. Second, it might be the case that the lexical and grammatical skills tested in standard proficiency tests, such as the cloze test we used, do not suffice. Communicative competence is the ability to interact, rather than just talk, but such skills were not tested. Finally, our experimental design may have induced some particular effects not readily found in daily life. First, the difficulty of the communicative task encouraged people with limited grammatical and lexical competence in L2 to make full use of other aspects of their communicative competence, including the judicious use of MCDs as speakers and elicitation of MCDs as hearers. There is some evidence in our data that pairs that could expect communicative trouble worked harder at avoiding it, while pairs who assumed they could rely on the high L2 proficiency of each partner, appeared to underestimate the need to attend to each other's degree of understanding. This interactive nature of competence also explains why the individual characteristics of the subjects – those of a speaker rather than of a dyad – had very little predictive power when we looked at overall success rates or at MCD use. Second, as one reviewer pointed out to us, the maze task constrains the sorts of meaning any utterance may be conveying. The circumscribed nature of the task in effect increases the common ground to higher levels than would be the case in ordinary communicative settings.

MCDs catch the moment where interlocutors interrupt or slow down the conversation to avoid possible misunderstanding. The question is what prompts speakers to make more or less extensive use of MCDs. The decision to use them needs to be interpreted as part of a larger complex of behavioral characteristics, including



lessons from past experience, attitudes to interlocutor and language, the nature of the communicative task, amongst probably many other aspects. Past experience determines which discourse patterns are entrenched in a speaker's mind. Someone who has often been in intercultural communicative settings will have seen how problems with understanding emerge, have practiced ways of avoiding misunderstanding, and have developed ideas about what works and what does not. Experienced intercultural communicators are likely to have developed sophisticated insights into the common sense notion that you need to help interlocutors who don't speak your language well.

In multilingual settings, codeswitching often functions as a contextualization cue, assisting both speakers and hearers in elucidating pragmatic meaning. Often, it has the general phatic effect of cementing the bond of community among bilinguals. In the experimental setting, codeswitching was effectively banned, as speakers were supposed to speak their L1 throughout. Nevertheless, various instances were attested, and they often appeared to serve phatic functions. This use of codeswitching could be interpreted as a general MCD: it makes the conversation more pleasant because it allows speakers to meet the other halfway. At the same time, specific cases of codeswitching may be used simply to clear up a very local problem of understanding, for example if a particular L1 word is not understood by the partner and the speaker therefore inserts the L2 equivalent. On an even more general level, codeswitching is just a special case of 'languaging': using whatever resources that seem to get the communicative job done.

When it comes to other factors of influence, extensive experience with bilingual situations and having a positive attitude towards the L2 and its speakers repeatedly showed their impact. These features seemed to help speakers overcome the disadvantages of low L2 proficiency and in some cases helped such pairs in achieving higher levels of success in the maze task. In this respect, communicative competence, in addition to having its source in linguistic knowledge

and interactional experience, also includes sociolinguistic components and is affected by interpersonal and intercultural attitudes.

The degree to which MCD use evolved during the Skype conversations suggests a close connection between MCD selection and proficiency. Recall that pairs in which one speaker had low L2 proficiency tended to use more MCD3 as the task progressed, while HH pairs reduced their use of MCDs, and the one they used most was MCD2, which indeed requires higher linguistic skill. This suggests that the pairs who could expect communication breakdowns tried harder to actively construct common ground. Obviously, if your proficiency is higher, you are also better able to use MCDs, but the data also show that they are very useful tools exactly when proficiency is low. Interestingly, such extensive use of MCDs also seems to lead to improved results for task completion, perhaps as a by-product of the conscious need to increase common ground.

### Future research

The study answered some questions but elicited new ones. A number of factors were found to impact communication in unexpected ways. However, further exploration of these requires more than a post hoc analysis. MCD use was found to differ depending on the linguistic composition of the pairs and participants' experience with bilingual situations in general. In currently on-going follow-up research, we systematically vary several types of familiarity: L2 proficiency, degree of multilingual experience, and experience with particular communicative settings. We also vary the communicative tasks, since some of the results we obtained may have been caused by particular features of the maze task.

In the new study interlocutors familiar with each other engage with each other in situations they know well since they have experienced them many times before, but also in contexts that do not resemble their typical communicative constellations. Other factors that are manipulated include the extent of experience that

communicators have with the skills needed to conduct the conversation: especially whether they are native speakers, how good their L2 skills are, and to what extent they have experience with multilingual settings. The data reported on in this paper suggest that when familiarity is low, judicious use of meta-communicative devices becomes more important. This may also hold for related strategies not investigated yet, such as the use of communicative routines, phonological reduction, and joint construction of discourse. It will also be investigated what the effect is of the presence of a third participant. An uninformed third interlocutor instructed to jump in whenever there is a sign of communication breakdown is used to analyze how (potential) misunderstanding is detected, prevented or fixed in a multi-party conversation. The results of this kind of research may be relevant in the development of training modules for intercultural communication and other linguistic and interactional constellations that are prone to miscommunication due to lack of familiarity.

#### REFERENCES

- Bahtina-Jantsikene, Daria 2013. *Mind Your Languages: Lingua Receptiva in Estonian-Russian Communication*. LOT dissertation series 338. Utrecht University.
- Bahtina, Daria; ten Thije Jan D.; and Wijnen, Frank 2013. Combining cognitive and interactive approaches to Lingua Receptiva. *International Journal of Multilingualism* 10 (2), 159–180.
- Beerkens, Roos 2010. *Receptive Multilingualism as a Language Mode in the Dutch-German Border Area*. Münster: Waxmann Verlag.
- Blees, Gerda; Mak, Willem; and ten Thije, Jan (2014). English as a lingua franca versus lingua receptiva in problem-solving conversations between Dutch and German students. *Applied Linguistics Review*, 5 (1), 173–193.
- Braunmüller, Kurt 2007. Receptive Multilingualism in Northern Europe in the Middle Ages: A description of a scenario. In J. ten Thije and L. Zeevaert, eds. *Receptive Multilingualism: Linguistic Analyses, Language Policies and Didactic Concepts*, 25–49. Amsterdam: John Benjamins.
- Clark, Herbert H. 1996. *Using Language*. Cambridge University Press.

- Croft, William 2009. *Toward a Social Cognitive Linguistics*. In V. Evans and S. Pourcel, eds. *New Directions in Cognitive Linguistics*, 395–420. Amsterdam: John Benjamins.
- Estonian Integration Monitoring 2015. Carried out by the Institute of Baltic Studies, Tallinn University and Praxis Center for Policy Studies. Available at <http://www.kul.ee/sites/kulminn/files/6peatykk.pdf>
- Pickering, Martin J.; and Simon Garrod 2004. *Toward a mechanistic psychology of dialogue*. *Behavioral and Brain Sciences* 27, 169–225.
- Rehbein, J., ten Thije, J.D. and Verschik, A. (2012). *Remarks on the Quintessence of Receptive Multilingualism*. In ten Thije, J. D.; Rehbein, J. and Verschik, A. (Eds.) *Lingua Receptiva*. *International Journal for Bilingualism* 16 (3), 248–264.
- Schmidt, Thomas; Wörner, Kai (2009). *EXMARaLDA – Creating, sharing and analyzing spoken language corpora for pragmatic research*. *Pragmatics* 19 (4), 565–582.
- Tomasello, Michael 2008. *Origins of Human Communication*. MIT Press.

RESÜMEE

## PIIRATUD ÜHISTEADMISED, PIIRAMATUD SUHTLUSVÕIMALUSED: EESTI-VENE RETSEPTIIVSE KAKSKEELSUSE EKSPERIMENTAALUURING

Eesti-vene suhtluse uuringu põhjal võib järeldada, et retseptiivne mitmekeelsus (*lingua receptiva* ehk LaRa) on üks mitmekeelse kommunikatsiooni viise, millel on potentsiaali nn ühise aluse (ingl. *common ground*) ehk kooskõlastatud vestluse loomiseks, võimendamiseks ning taastamiseks olukordades, kus ühiselt jagatud teadmised ei garanteeri sujuvat suhtlemist. *Lingua receptiva* on mitmekeelne suhtlemisviis, kus kaasvestlejad räägivad oma emakeelt ning püüavad üksteise mõistmiseni jõuda, tuginedes oma võõr- või teise keele oskustele.

Psühholingvistilist joondumist (ingl. *psycholinguistic alignment*) kui protsessi on vaadeldud eksperimentaalsetes olukordades. Erinevate meta-kommunikatiivsete strateegiate kasutamine annab ettekujutuse protsessidest, mis on *lingua receptiva* taga. Tulemuste põhjal on selgitatud, et nende strateegiate valik oli osaliselt tingitud võõrkeele valdamistasemest ja varasematest mitmekeelsuse kogemustest, kuid eksperimendi tulemus ei sõltunud otseselt keeleoskustasemest. Enamgi veel, dialoogi tulemuse prognoosimiseks on ainuüksi vestluspartnerite individuaalsete omaduste teadmisest vähe, olulisem on teada dialoogi partnerite ühiseid omadusi.

# CODE-SWITCHING, LEXICO-GRAMMATICAL FEATURES AND LOAN TRANSLATION: DATA FROM A LARGE MACEDONIAN-ENGLISH CORPUS

Jim Hlavac

Monash University

**Abstract.** This paper looks at the speech of bilinguals, in this case a group of 103 Macedonian-Australians and focuses on code-switching in a lexical sense, e.g. embedding, insertion, alternation, referring to lexical material that is transferred across languages, regardless of length. Two- or multi-item constructions as a code-switching related phenomenon are investigated, especially the role of light-verb constructions. Lexico-grammatical and lexico-semantic influence (via loan translation) is examined to see whether imported lexical items co-occur with and appear to precipitate structural innovations.

**Keywords:** code-switching, grammaticalisation, loan translation, light verbs, Macedonian

## 1. Contact linguistics and code-switching

This paper examines excerpts of speech from 103 bilinguals and focuses firstly on instances of code-switching, here understood as the transfer of lexemes across languages. The statistical incidence and frequency of code-switching is presented, along with examples to illustrate the type of English-origin content-referential items that are found in some people's Australian English-Macedonian speech. Code-switching can be looked at as a phenomenon that refers primarily to the importation of lexical items, and (a high number of) code-switched lexemes alone have historically often been the main

focus of descriptions of language contact situations. Code-switching need not co-occur with or appear to be related to other features of bilingual speech, such as syntactic, phonological or semantic features that may differ from those found in the speech of monolinguals. There may, however, be connections in a statistical and a proximal sense between the presence of other-language lexemes and grammatical structures that bear input from two languages. The phenomenon of inserting other-language items into one's speech, regardless of how habitualised these may be, is unlikely to be restricted to single words only. When conversing with others who share ability in their languages, bilinguals may transfer single words or groups of words, or transfer phrases or collocations as calques (cf. Grosjean's (2001) notion of modes and sociolinguistic features that condition a speaker's level of activation of one or both languages). The latter are also known as 'loan translations' i.e. constructions that bear „any use of morphemes in Language A that is the result of the literal translation of one or more elements in a semantically equivalent expression in Language B“ (Backus, Dorleijn 2009: 77). This phenomenon is the focus of the second part of this paper.

The boundary between code-switching of lexemes and loan translation is, however, fluid: constructions that are transferred from another language may be 'completely' rendered via recipient language morphemes, but they also may be only partially rendered, wherein some elements remain 'untranslated'. Such constructions could be labelled 'partial loan translations' or 'lexico-semantic constructions with code-switching'. In regard to the latter, the term 'code-switching' (whether hyphenated or not) has, for many, become a hypernym that encompasses *all* types of cross-linguistic transfer – lexical, semantic, syntactic, phonological, and pragmatic – e.g. „Codeswitching (CS) refers to language use that consists of material from two or more language varieties at any level from the discourse to the clause“ (Jake, Myers-Scotton 2009: 207). Other researchers distinguish (lexical) code-switching from the latter, which they label

‘semantic transference’ (cf. Clyne 2003) or ‘selective code copying’ of semantic features (cf. Johanson 1998).

As stated, this paper looks firstly at code-switching as a lexical phenomenon and secondly at the incidence of lexico-grammatical and lexico-semantic constructions that may be patterned on models transferred from the other language. These constructions may appear as a sub-set of lexical code-switching where constructions are made up of items from both languages, i.e. a verb phrase in which one or more items may be code-switched across languages, while other elements in the same verb phrase are rendered via recipient language morphemes, though the pattern is source-language influenced. The type of constructions that are most studied are those consisting of a recipient language ‘light’ verb (hereafter: ‘LV’) and a source language object. The choice of the LV that is ‘pressed into service’ is of interest, inasmuch as an existing LV is the one chosen (or another one) and if the LV in such instances appears to be used in a way congruent to its function hitherto. In some cases it appears that a source language collocation is the basis for the form of the partial loan translations. In other cases, a particular type of ‘do+OBJ’ may assume such regularity that this pattern spreads further, leading to other verbs being used in this way.

My examination of code-switching *and* loan translation (including possible structural innovation) in this same paper does not suggest that I believe that a causative link exists between the two. In contact linguistics, there is the well-known example of speakers of Urdu, Marathi and Kannaḍa in Kupwar, amongst whom code-switching is almost unknown as it is a socially condemned practice, but amongst whom structural change and convergence between the languages of three groups has occurred (Gumperz, Wilson 1971). On the other hand, Friedman’s (1994, 1995, 2009) work on Balkan languages shows that along with structural convergence between Macedonian, Albanian, Bulgarian, Greek, Turkish and Romany in particular, there is much evidence that code-switching was also a widespread phenomenon amongst speakers of these languages,



even if a causal link between convergence and code-switching is not overtly made. Myers-Scotton (2002) and Backus (2005) provide more detailed accounts of the relationship that can exist between code-switching and structural change.

Returning to loan translation, sometimes the semantic features transferred from a source-language model result in little more than 'a new expression' that does not violate syntax, as Otheguy (1993: 23) maintains that the phrase [VERB + *para atrás*] in American Spanish varieties still „exploits inherent Spanish possibilities“ even though it is probably modelled on the English [VERB + *back*]. Most loan translations may be fixed combinations of two words or morphemes inserted into the recipient language with no structural consequences beyond the distribution of the words or morphemes involved. What can also happen is that the loan translation can take the form of an unconventional combination of morphemes, and with it, a structural innovation occurs. Such an innovation is suggestive of structural change, that is, an on-going process active amongst more than a mere handful of speakers that could, at least theoretically, be ascertained through longitudinal observation. Another thing is, however, important here: a structural innovation such as that mentioned above came about as a largely accidental by-product of loan translation, and the boundary between loan translation and structural innovation may be overlapping.

This paper is descriptive and presents the frequency and incidence of code-switching data. The size of the corpus allows some discussion on propagation, i.e. cross-speaker habitualisation of form. This, in turn, allows speculation on whether what appear as individual occurrences can be conceived of as 'systematically present' features, i.e. to be innovations that are well on the road to becoming examples of *structural change* as a diachronic feature. Lastly, loan translation is looked at, together with the incidence of LV constructions, as a phenomenon associated with structural change. This paper is informed by Myers-Scotton's (2002) approach to language contact phenomena, and one of her recent works on verb

constructions (Myers-Scotton, Jake 2014). Further, this paper draws on Backus and Dorleijn's (2009) and Backus's (2009) work on loan translations and LVs respectively. The legacy of Friedman's work on language change in Macedonian, from both a synchronic and diachronic perspective, is also discernible in this paper.

## 2. Data sample and profile of informants

This paper is based on the speech of bilinguals whose heritage language, Macedonian, is an immigrant, transposed one in predominantly Anglophone Australia. The 103 bilinguals are mainly first-generation (73), while just over a quarter (27) are Australian-born second-generation speakers. Amongst the first-generation speakers the majority is made up of those born in Aegean Macedonia (northern Greece), who speak a non-standard variety of Macedonian as their L1, who had little or no formal schooling in this language (or in Greek, which only a small number speak as an L2), and who acquired English after emigration to Australia as young adults in the 1950s and 1960s. The other group of first-generation Macedonian-speakers is from today's Republic of Macedonia, and these emigrants received formal instruction in their first language, Macedonian, and sometimes also in English. They emigrated to Australia from the mid-1960s to the mid-1980s.

The second-generation speakers are English-dominant and the variety of Macedonian that they speak is strongly reflective of the vernacular of their parents; only seven received formal instruction in Macedonian in Australia. All informants are residents of Melbourne, a city of 4.2 million people. According to the latest census figures from 2011, approx. 30,000 Melburnians reported that their 'language spoken at home' was Macedonian. A further 30,000 people in Melbourne, mostly second- or third-generation speakers, and those who reported another language as their home language, have proficiency in Macedonian. There is a concentration of Macedonian-speakers across some parts of Melbourne, with two areas having a

concentration level of those born in the Republic of Macedonia of over 10% (ABS 2012).

The data on which this paper is based come from video- or audio-recorded interviews between the informants, either individually or in small groups, conducted not by the present author, but with a fieldworker and data collector, who is himself a second-generation Australian-Macedonian. Informants were usually interviewed in their own homes and the corpus of available data consists of an approx. 10-minute period of each of the interviews that was transcribed by the field-worker<sup>2</sup>.

The data sample consists of 95,028 words (or tokens that are lexical items, excluding non-lexicalised forms such as ‘uh-huh’) and 3244 turns from the 103 informants. The ‘unmarked’ or predominant language of the recorded interviews was Macedonian. However, as the interviewer himself was an Australian-born, Macedonian-English bilingual known to most informants, the informants were free to code-switch between English and Macedonian. Table 1 below sets out features of the sample relevant to an analysis of code-switching.

There are some differences in the incidence of use of monolingual or bilingual speech between the two generations, and the number of turns that the informants produced that were either monolingual Macedonian, or consisting of both languages (through code-switching), or even monolingual English. On average, 21% of turns contain code-switches. Examples of code-switches are provided with literal glosses and information on grammatical categories. A free translation, together with the following information on the informant are also provided: informant number (e.g. Inf. 25) and generational membership (e.g. Gen.1).

---

<sup>2</sup> All data were gained through a large-scale project on the language of Macedonian-speakers in Melbourne which included video and audio-taped recordings. Permission to conduct this research on human informants was granted by Monash University Human Research Ethics Committee, Project No. CF10/2366 – 2010001346 on 11 October 2010. A link to a selection of the videos, some of which are sub-titled, can be found at AMHRC (2014).

**Table I:** Data on the 76 first-generation and 27 second-generation informants and the number of words, turns and code-switches across the sample.

	First Generation	Second Generation	Total
No. of informants	76	27	103
Ave words per person	951	841	923
Total words	72,297	22,731	95,028
Ave. no. of turns per informant	29	38	31
Total turns	2,215	1,029	3,244
Total Mac. monolingual turns	1,838	667	2,505
Percentage of turns that are Mac. monolingual	83 %	65 %	77 %
Total turns containing code-switches	373	306	679
Percentage of turns containing code-switches	17 %	30 %	21%
Total Eng. monolingual turns	4	56	60
No. of Eng. code-switches / code-switched items	758	708	1,466
Ave. no. of code-switches per informant	11	26	14

The informants' speech is represented here according to its form: the vast majority of the informants are speakers of south-western dialects of Macedonian originating from the Lerin (Gk. Flórina), Kostur (Gk. Kastoriá) and Drama (Gk. Dráma) regions of the provinces of western and central Macedonia in northern Greece (Aegean Macedonia) and from the Bitola region of the Republic of Macedonia. Readers familiar with Macedonian will recognise forms that vary from standard Macedonian<sup>3</sup>. Macedonian is written in the Cyrillic alphabet, but the examples here are given in Roman-script

<sup>3</sup> South-west Macedonian dialects spoken in Aegean Macedonia in the Lerin and Kostur regions have a number of features that distinguish them from Standard Macedonian. Amongst these are: devoicing of some consonants in medial and not only final position - *vapsuvafme* vs. Standard *vapsuvavme* 'dye 1.PL.IMPERF'; loss of consonants in intervocalic position - *ojme* vs. Standard *odime* 'go 1.PL.PRES.'; loss of consonants in final position - *ka:* vs. Standard *kak* 'how'; changes in consonant clusters - *šo* vs. Standard *što* 'what', etc., cf. Friedman (1993) and Vidoeski (2005).

transliterations. See Appendix One for a guide to the transliteration convention employed. The data sample is part of a project on the speech of Australian-Macedonians and aspects of their language use. A sociolinguistic description of this sample of informants is provided in Hlavac (2016).

### 3. Code-switching

Code-switching data here are presented in general terms to provide a brief insight into the frequency and type of code-switches that are found in the sample. First, a break-up of instances of code-switching is made according to the position of the code-switch within a clause, or at a clause boundary. Examples of all three different types are given below. Table 2 below sets out in statistical terms the frequency of different types of code-switching according to position vis-à-vis clause boundary.

**Table 2:** Categories and numbers of code-switches (single-item and multiple item)

	Single items	Multiple items	Total
Extra-clausal code-switching	884	79	963
Intra-clausal code-switching	307	95	412
Inter-clausal code-switching	44	47	91
Total	1235	221	1466

As Table 2 shows, most code-switches (66%) are extra-clausal code-switches, i.e. code-switches that occur at a clause boundary, and which function usually as discourse-specific devices such as *you know* rather than as content-referential items. It is this quality that distinguishes them from inter-clausal switches. In the first place, English-origin items in the Macedonian speech of the informants usually occur in positions where insertion or embedding is morpho-syntactically less constrained, i.e. at clause boundaries. There may also be features about the English items that are inserted

into Macedonian speech to indicate that they may commonly be discourse-pragmatic items that otherwise occur at clause boundaries in English and that their use in Macedonian also coincides with their placement at Macedonian clause boundaries. These items are usually single-item or compound-item code-switches such as *well*, *yeah*, *alright*. Some of these are contained in (1) below:

- (1) **Yeah**, vo, vo Srbija, ne to'ku vo Makedonija. Po.. pojketo  
**Yeah**, in, in Serbia, NEG so much in Macedonia. Mo.. more  
 godini vo Srbija gi pominafme, **yeah**.  
*years in Serbia themACC.PL spend1.PL.IMPERF, yeah.*
- I tamu jajca gi vapsuvafme, takvi raboti.  
*And there eggs themACC.PL dye1.PL.IMPERF, such things.*
- Orajt..** i so čupinata, **really...** pojké kaj  
**Alright..** and with girls+DEF.ART **really...** more at  
 familijata. **Yeah**, so familijata si s..  
*family+DEF.ART. Yeah, with family+DEF.ART REFLEX.PRON*
- yeah..** Veligden go slavevme sekoja  
**yeah..** Easter itACC.SG celebrate1.PL.IMPERF each  
 godina... **yeah, yeah**.  
*year... yeah, yeah.*
- Yeah**, in, in Serbia, not so much in Macedonia. We spent mo.. more  
 years in Serbia, **yeah**. And we used to dye eggs there, things like that.  
**Alright..** and with the girls, **really...** more with the family. **Yeah**,  
 with the family we ourselves, our.. **yeah**. We celebrated Easter every  
 year... **yeah, yeah**. (Inf. 89, Gen.1)

In example (1) above, English-origin *yeah* appears six times. Its dictionary-entry function is that of an affirmative which is the function it fulfils at the start of the turn. *Yeah* is, however, poly-functional, here acting as: a pause-filler (second instance), both affirmative and pause-filler (third instance), utterance terminator (fourth instance)

and then turn terminator (fifth and sixth instances). The point is that the function of *yeah* is independent of the morpho-syntactic grid and lexical-referential content of the clauses that otherwise make up example (1) and it is used as a discourse marker with various functions (cf. Hlavac 2006). The two other extra-clausal switches in example (1) above are *orajt* ('alright') as an evaluative or summative marker of the activities that the informant engaged in over Easter, and later, *really* is employed as an amplifier, and possibly also as a pause-filler. Although *really* appears as an adverb that occurs internally in an otherwise Macedonian-language clause, its discourse function is largely separate to the lexical-referential content of the clause, and can therefore be considered to be 'outside' it.

The second-most frequent type of code-switching is intra-clausal code-switching. Example (2) contains instances of this:

- (2) ... vo **office** rabotam, um, i gledam, um, za pulam za  
 ... in **office** work1.SG, um, and look1.SG, um, for look1.SG for  
 fabrikata što praje. Ah gi gledam za  
 factory+DET REL.PRON do3.SG. Ah themACC.PL. look1.SG for  
**sales** što praat i treba kako, nekako **balance** da  
**sales** REL.PRON do3.PL and need3.SG how, somehow **balance** COMP  
 prajime od **sales** i um fabrikata, ama i, i dosadno  
 do1.PL from **sales** and um factory+DET but and, and boring  
 mi e.  
 me.OBJ(IMPERS.SUBJECT] be3.SG  
 Ah, um, I work in, in [an]**office**, um, and I see, um, I look at how the  
 factory is doing/going. Ah, I look at the **sales** they are making and  
 I have to like, somehow we have to do a **balance** on the basis of the  
**sales** and, um, the factory, but it's also boring for me. (Inf. 83, Gen.2)

In example (2) above, English-origin items that are not phonologically integrated occur in this second-generation speaker's speech, when recounting aspects of his workplace. As stated, most intra-clausal code-switches are single words, and the prominence of nouns,

with a thematic link to the informant's largely English-speaking workplace, is clear. What is also apparent in example (2) is that the English insertions have a lexical-referential function only. The code-switches to English do not have any function at a discourse level as they do not 'do' anything other than refer to English concepts.

The least frequent type of code-switching is inter-clausal. Example (3) contains instances of this:

- (3) *sega* *rabotam* ... *oh, I work for Aussie Post, Australia Post.*  
*now work* 1.SG.... *oh, I work for Aussie Post, Australia Post.*
- Ah... *jas sum transport and I'm a van driver* što  
*Ah... I be1.SG transport and I'm a van driver* REL.PRON
- drive a van and I do,* *ka se veli?..*  
*drive a van and I do,* *how REFLEX.PRON say3.SG*
- company mail pick-up, so, so* *ako, ako, ako imaš*  
*company mail pick-up, so, so if, if, if have2.SG*
- fabrika i imaš pisma, ti imaš contract with*  
*factory and have2.SG letters, you have2.SG contract with*
- Aussie Post i nie ojme da zemame pismata and we*  
*Aussie Post and we come1.PL COMP take1.PL letters+DET and we*
- take 'em off ya every day i setne we go back to the depot i*  
*take 'em off ya every day and then we go back to the depot and*
- setne toa odi out.*  
*then it go3.SG. out.*
- Now I work ... oh, I work for Aussie Post, Australia Post. Ah... I am*  
*transport and I'm a van driver that drive a van and I do, how do*  
*you say it?.. company mail pick-up, so, so if, if, if you have a factory*  
*[business] and you have letters, you have [a] contract with Aussie*  
*Post and we come and pick up the letters and we take 'em off ya*  
*every day and then we go back to the depot and then that goes out.*  
 (Inf. 63, Gen.2)



There are some intra-clausal switches as well in example (3), but the most prominent feature of this informant's turn are the clause-length descriptions of his duties at work. While English words and morphemes outnumber Macedonian ones in this turn, Macedonian discourse markers frame most of the turn, and conjunctions such as *ako* ('if'), *i* ('and') and the sequential marker *setne* ('then') occur as more macro-level discourse 'signposts' and are supplied mostly from Macedonian. Further, the English verb *drive* remains in its un conjugated 'bare form', although the controlling subject is 3.SG. and the tense PRES.SIMPLE., i.e. one would expect 3.SG. *drives*.

To summarise the main features of the examples above, Example (1) is an instance of code-switching of English forms whose incidence can be accounted for by the pragmatic function that they perform. In fact, it may be possible to speak of something further happening here: the incidence of *orajt* ('alright') and *really* above is characteristic of Australian-English more so than Macedonian pragmatics. This suggests that what may be happening here is not only a replacement of Macedonian discourse markers with English ones, but the adoption of Australian-English pragmatic norms in function *and in form*. I make this observation about pragmatic features here, but do not further look at them as a feature of particular interest in this paper.

In example (2) above, two English insertions, *sales* and *balance*, co-occur with the Macedonian verb *praje*<sup>4</sup> ('to do'). The verb *praje* controls these English items as objects, in the sense of 'make sales' and 'do a balance'. It is of little surprise that items relating to this speaker's work duties are supplied from English as these forms, typically content-referential ones such as nouns, are well documented in other studies on bilingual speech (Clyne 2003, Verschik 2008). In relation to a large Croatian-English bilingual corpus, Hlavac

---

<sup>4</sup> Macedonian does not have an infinitive, and the 'base' form of a verb (including the one used in dictionary entries) is the 3.SG.PRES. form. The form *praje* here is a non-standard form for standard Macedonian *pravi* 'to do'.

(2003: 93) reported that 686 (71%) of the 962 English single-item code-switches were nouns. Myers-Scotton and Jake (2014: 511) provided a conceptual account of the differential role that nouns play, compared to verbs, such that the former „do not project information about syntactic and argument structure that is included in the abstract structure of finite verbs“. As verbs play a central role in determining relations within the clause through valency and case assignment this role can be an obstacle in facilitating their cross-language transfer in a way that does not apply to other grammatical categories. The choice of verb that co-occurs with an imported English item will be returned to below.

In example (3), the last two words in that example are a Macedonian verb *odi* ('goes') and the English adverb *out*. The meaning of the last two words is 'go out' in the sense of 'leave with no specific destination', as letters and parcels are sent out from a postal despatch centre to various destinations. It appears that this construction is based on the English phrasal verb *go out*. There are equivalent Macedonian constructions available:

(3a) ... i     setne   toa             odi     nadvor  
       ... *and then it*NEUT.SG *go*3.SG *outside*  
       '... and then it goes outside'

(3b) ... i     setne   toa             izleguva  
       ... *and then it*NEUT.SG *leave/exit*3.SG  
       '... and then it leaves/exits'

It seems, however, at least in the mental lexicon of this speaker, that none of the possibilities offered by (3a) or (3b) are able to capture his sense of 'goes out'. (It can be presumed these equivalents could have been produced by this speaker as they are sentences commonly heard from second-generation speakers.) In short, the English phrasal verb expression is availed of and appears here with one part of the loan translated, while the other remains untranslated.

#### 4. Loan translation phrases and light verbs

As discussed, nouns, alongside discourse markers, are the most likely candidates for insertional code-switching. In this corpus, over 80% of the 307 single-item intra-clausal are nouns. Amongst those that I allocate to this group are English *-ing* forms, i.e. verb-origin substantivisations. Example (4) below contains an *-ing* form twice, *cycling*, initially in an inter-clausal code-switch, then in an intra-clausal one:

- (4) *šetam*                    **with** *familijata*    *i ah*        **with friends and also cycling.**  
*walk*<sub>1.SG.PRES</sub> **with** *family*+DET *and ah* **with friends and also cycling.**
- sakam*                    *nogu*        **cycling** *so*        **bajsiko.**  
*like*<sub>1.SG.PRES</sub>        *much*        **cycling** *with*        **bicycle**+DET
- ‘I walk **with** the family and, ah, **with friends and also cycling.** I like **cycling** a lot by **bicycle.**’ (Inf. 85, Gen.2)

Here, *cycling* is referred to as an activity, i.e. as a gerund or verb form that functions as a noun. The second instance of *cycling* follows the Macedonian verb, *sakam* (‘I like’) and this structure is a transparent *to like* + OBJ one, in which *to like* is a full verb and *cycling* is its object. There are no examples of *cycle* as a code-switched verb attracting Macedonian verbal morphology (e.g. *cyclam* [seiklam] 1.SG.PRES. ‘I cycle’), nor are there congruent analytic constructions with this verb such as *I do cycling*. There are, however, 39 instances of imported *-ing* forms from English, many of which co-occur with the Macedonian verb *praje*, which is employed in such constructions as an LV.

A short digression on *praje* and its use as an LV is warranted here. Firstly, *praje* (and its standard form *pravi*) equates to ‘to make’ in its primary meaning, and ‘to do’ in its secondary meaning, e.g. *Taa praje nered* ‘She make<sub>3.SG</sub> mess’ = ‘She is making a mess’ and *Šo praješ?* ‘What do<sub>2.SG</sub>’ = ‘What are you doing?’. Historically, it was this verb (along with a synonym *čini* ‘to do’) that performed the function of an LV, often in combination with Turkisms that entered

Macedonian during the Ottoman occupation of the Balkans, e.g. *praje aber* [do3.SG + Tk. *haber* ‘news’] meaning ‘to inform’ (cf. Jašar-Nasteva 1962/1963). Codification of Macedonian in the twentieth century downgraded the LV construction to non-standard or low-register speech. However, the construction remained a part of most speakers’ vernaculars, and particularly amongst those in northern Greece, whose Macedonian-speakers remained dislocated from the development of a literary standard<sup>5</sup>. Thus, incidence of *praje* as a grammaticalised DO-verb in this immigrant language contact setting cannot be conclusively attributed to the contact situation (and the influence of the English LV *do*) or to speakers’ employment of the most frequent Macedonian verb as an ‘attractive’ candidate for LV constructions, as there are diachronic data to show it existed previously in homeland varieties of Macedonian. What is of interest for contact linguists is to observe how dispersed the use of *praje* with English *-ing* forms is, and to see if this innovation appears to be widely propagated enough to be considered structural change. Here, *praje* as a grammaticalised LV performs a function in Macedonian similar to that performed by *yapmak* (or *etmek*) in immigrant Turkish (cf. Backus 2009: 307-339) or *tegeme* or *saama* as Estonian grammaticalised verbs in Russian-Estonian speech (Verschik 2008: 137, 149).

Across this data sample, there are 515 instances of *praje* (including all its conjugational and tense forms), which makes this by far the most frequent verb in the sample. Further, *praje* co-occurs with 14 of the 39 English *-ing* forms. Below are three examples:

---

<sup>5</sup> Pontic Greek, spoken by Pontian refugees settled in Aegean Macedonia, also features a DO-verb construction *chtízo* (‘I build’), probably modelled on Turkish *yapmak* (Van Hasselt 1972). Those speakers from northern Greece may have had contact with speakers of this variety of Greek, as well as with speakers of Modern Greek, that features compounds with an LV, *kano* (‘I do’), such as *kano psonia* (‘I do shopping’), alongside *psonizo* (‘I shop’). In an émigré setting, *kano*, as an LV with code-switched objects from English, occurs in the speech of Greek-English bilinguals, e.g. *kano jogging* (‘I do jogging’) and *kano zapping* (‘I zap’), (Edwards, Gardner-Chloros 2007: 77).

- (5) si imaja i torbite... luĝeto  
 REFLEX.PRON. *have*<sub>3.PL.PRES</sub> *and bags* + DET... *people*+DET  
**shoppingto**<sup>6</sup> si praat.  
**shopping**+DET REFLEX.PRON *do*<sub>3.PL.PRES</sub>  
 ‘They also had the bags as well... the people are doing the(ir) shopping.’ (Inf. 89, Gen.1)
- (6) ... odam na **gym... boxing** praam, ah i isto taka  
 ... *go*<sub>1.SG.PRES</sub> to **gym... boxing** *do*<sub>1.SG.PRES</sub>. *ah and also so*  
 plivam.  
*swim*<sub>1.SG.PRES</sub>  
 ‘I go to the gym... I do boxing, ah and I swim as well.’ (Inf. 45, Gen.2)
- (7) **walking** praješe?  
**walking** *do*<sub>2.SG.IMPERF</sub>  
 ‘Did you do (any) walking?’ (Inf. 97, Gen.2)

Examples (5) to (7) above show the co-occurrence of *praje* and an English *-ing* form, in both cases as an object of the LV. The question of why this construction occurs needs to be asked. Does it appear that a speaker wishes to employ an English-origin verb form, but does not do so, and instead uses a Macedonian LV, with a substantivised form of the English verb in order to conform to Macedonian syntax? Or is the speaker employing the English *-ing* form as the form that s/he would use anyway in an equivalent English construction – *The people are **doing** their shopping; I **do** boxing; Did you **do** (any) walking?* – and these are loan translations that feature *praje* as a translation of the English LV *do*, which just happens to be the semantic equivalent of the Macedonian LV, *praje*? This question is

<sup>6</sup> The form *shoppingto* ‘shopping+DET’ (= ‘the shopping’) is of interest. The phonotactic form and consonantal ending of *shopping* would usually attract the SG.MASC determiner ‘-ot’. Instead, the speaker employed the SG.NEUT determiner ‘-to’, perhaps influenced by the form of the Macedonian gerund equivalent, *kupuvanje* (‘shopping’), which is neuter, yielding *kupuvanjeteto* (‘kupuvanje+DET’).

hard to answer. In response to the first question, in the data there are examples of some English-origin verbs that are morphologically integrated into Macedonian, attracting verb conjugation markers:

- (8) celi tua se, izbegaa, ama pa se  
*all*MASC.PL *here* *be*<sub>3</sub>.PL.PRES, *leave*<sub>3</sub>.PL.AOR *but* *then* REFLEX.PRON
- razberame so edni, ringuvame  
*understand*<sub>1</sub>.PL.PRES with *some*<sub>3</sub>.MASC.PL *ring*<sub>1</sub>.PL.ITERATIVE.PRES
- pojke so kanadejci, tamu se moite.  
*more* with *Canadians*, *there* *be*<sub>3</sub>.PL *my*MASC.PL.+DET
- ‘They’re all here. They left, but we contact some of them, we ring the Canadians mostly, that’s where mine are.’ (Inf. 57, Gen.1)

The number of English-origin verbs that occur in the corpus, most of them morphologically integrated into Macedonian, is 11. This is fewer than the 39 *-ing* forms. In general, a non-finite construction to convey English-origin verbs in Macedonian is a preferred strategy (cf. Myers-Scotton, Jake 2014). Those English-origin verbs with Macedonian morphological markers that occur in the sample tend to appear in the speech of older migrants from northern Greece and that of their children, and appear to be habitualised borrowings. This view is based on the author’s long-term contact with Macedonian-speakers aside from the forms found in this linguistic sample, e.g. *juzam sé* ‘*use*<sub>1</sub>.SG *everything*’ = ‘I use everything’; *ne se filvam dobro* ‘NEG. REFLEX.PRON. *feel*<sub>1</sub>.SG *well*’ = ‘I don’t feel well’.

In relation to the second question posed above, it is hard to discount the influence of English *do* as a model, as *do* can, alongside its role as a LV with OBJ. *-ing*, now be combined with other common nouns, e.g. ‘Let’s *do* lunch’; ‘I’m *doing* summer’. The distinction between a code-switched English *-ing* form with LV *praje* and what looks like a loan translation is blurred in instances when the English equivalent contains desemanticised *do*. At least one example of *praje* + *-ing* that is clearly modelled on an English construction is the following:

- (9) *toj praje cleaning*  
*he do3.SG.PRES cleaning*  
 ‘He does cleaning’ or ‘He is a cleaner’. (Inf. 97, Gen.2)

This utterance was made by an informant about another person and the job he performs. A tendency to avoid (perhaps less well known) nouns relating to occupational groups may account for this, such that constructions containing the noun forms such as *toj e čistač* (‘he is [a] cleaner’) are less common. A similar avoidance amongst other Macedonian-speakers in Australia for nominal terms to refer to occupations was recorded by Stewart (1995: 22), ‘popravam karo’i’ *fix1.SG cars* = ‘I am a mechanic’; ‘rabotam vo banka’ *work1.SG in bank* = ‘I am a bank teller’. Loan translation as the causation factor in the following construction is clear where *ima* ‘to have’ is being employed in a way resembling its use in equivalent English constructions.

Use of *have* in English as a light or heavily desemanticised verb, e.g. ‘to have a rest’, ‘to have dinner’, appears to be responsible for the transfer of this role to the Macedonian *ima* in the following example:

- (10) ... *imame lunch*  
 ... *have1.PL.PRES lunch*  
 ‘... we’re having lunch.’ (Inf. 97, Gen.2)

This use of *ima* as an LV, analogous to the use of English *have*, is unknown in homeland Macedonian. (In homeland Macedonian *jajeme* ‘eat1.PL.PRES.’ or *ručame* ‘eat lunch1.PL.PRES.’ would be used.) This appears as a nascent grammaticalisation, at least in the idiolect of this speaker, of *ima* as an LV similar to that which is known for *praje*. Stewart (1995: 22) in her corpus from over 20 years ago, recorded a similar example: ‘*Ķe imam šaoa*’ ‘FUT.MARKER *have1.SG.PRES shower*’ = ‘I will have a shower’.

## 5. Conclusion

In this paper I have focussed on examples of code-switching in a large bilingual corpus, and presented, in terms of frequency and position vis-à-vis clause boundary, the incidence of different types of code-switching. Extra-clausal code-switches, i.e. affirmatives and high-frequency discourse markers from English make up by far the largest group. This suggests, at least in statistical terms, that pragmatic particles are highly transferable, not only due to their ambient – and therefore perhaps more penetrative – use, but also because they are usually syntactically independent of the grammatical and argument structure projected by other constituents. Intra-clausal code-switching is the second most common form of code-switching found, and within this type of code-switching, nouns are the most prominent group. The content-referential information of an English-origin noun may take the form of a single-item code-switch. The forms that bear other-language content-referential information may be employed as discrete items that have clear-cut boundaries and which have no further effect on the grammatical and semantic structure of utterances other than being lexical insertions. While there may be fewer barriers to the transfer of single forms that are discrete items – as the frequency of discourse markers shows us – it is perhaps counter-intuitive to postulate that the possibility or amenability of transferring content-referential information would be restricted to single lexemes only. Single lexemes may themselves be part of larger constructions.

There are features specific to the language supplying code-switched nouns, here the sub-set of English substantivisations that are *-ing* forms. The development of *-ing* forms is a result of the *-ing* suffix being productive in terms of grammaticalisation in a manner that goes beyond ‘continuousness’ for verb forms, extending to ‘noun-formation’, i.e. gerunds, a form with which some verbs *must* collocate (e.g. *to mind + -ing*) and a component in other grammaticalisations, such as *to be going to* as FUTURE MARKER [+ INTENTION].



When the English-origin *-ing* (noun) form is inserted into Macedonian speech the lexico-grammatical features of the *-ing* can be 'carried with it', such that this conditions the verb form supplied from the recipient language. In some of the examples provided this frequently coincides with what is also the choice of forms in the equivalent English construction, namely *do + -ing*. It is hard to attribute the incidence of *praje + -ing* in a conclusive way to either the lexico-grammatical features of *-ing* as an OBJ. in LV constructions or to the lexico-semantic features of the *do + -ing*, which is a loan translation from English. Both influences appear to coalesce. Given the frequency of *-ing* forms in English and the prevalence of English as a contributing language in contact situations, it is instructive to look at its incidence in bilingual corpora. Of particular interest are the choice of recipient language forms that surround *-ing* and the possibility that English-origin grammatical and semantic information may co-determine these chosen forms.

#### Appendix One. Transliteration of Macedonian into Roman-script graphemes.

Macedonian speech is transliterated with the following Roman-script letters, including letters with diacritic marks.

Cyrillic	а	б	в	г	д	ѓ	е	ж	з	с	и	ј	к	л	љ	м
Roman	a	b	v	g	d	ǰ	e	ž	z	dz	i	j	k	l	lj	m

  

Cyrillic	н	њ	о	п	р	с	т	ќ	у	ф	х	ц	ч	џ	ш
Roman	n	ñ	o	p	r	s	t	ќ	u	f	h	c	č	dž	š

## REFERENCES

- ABS [Australian Bureau of Statistics] 2012. Data analysis of 2011 Census. <http://www.abs.gov.au/websitedbs/censushome.nsf/home/Data> (12.04.2016)
- AMHRC [Australian Macedonian Human Rights Committee] 2014. The Macedonian Speech of First and Second Generation Australian Macedonians: A Joint Project between the AMHRC and Dr Jim Hlavac of Monash University <http://www.macedonianhr.org.au/contents/163> (12.04.2016)
- Backus, Ad 2005. Codeswitching and language change: One thing leads to another – *International Journal of Bilingualism* 9(3–4), 307–340.
- Backus, Ad 2009. Codeswitching as one piece of the puzzle of language change: The case of Turkish *yapmak*. In: Isurin, Ludmila; Winford, Donald; de Bot, Kees. (Eds.) *Multidisciplinary Approaches to Code Switching*. Amsterdam/Philadelphia: John Benjamins. 307–336.
- Backus, Ad; Dorleijn, Margreet 2009. Loan translations versus code-switching. In: Bullock, Barbara; Toribio, Almeida Jacqueline. (Eds.) *The Cambridge Handbook of Linguistics Code-switching*. Cambridge, UK: Cambridge University Press. 75–94.
- Clyne, Michael 2003. *Dynamics of Language Contact*. Cambridge UK: Cambridge University Press.
- Edwards, Malcolm; Gardner-Chloros, Penelope 2007. Compound verbs in codeswitching: Bilinguals making do? – *International Journal of Bilingualism* 11(1), 73–91.
- Friedman, Victor 1993. Dialect Variation and Questions of Standardization in Macedonia: Macedonian, Albanian and Romani – *Zbornik za filologiju i lingvistiku*, 36(2), 7–35.
- Friedman, Victor 1994. Turkisms in a comparative Balkan context. In: *Septième congrès international d'études du sud-est européen*. (Thessalonique, 29 août – 4 septembre 1994) Athens. 521–543.
- Friedman, Victor 1995. Persistence and change in Ottoman patterns of code-switching in the Republic of Macedonia: nostalgia, duress and language shift in Southeastern Europe. In: *Summer School Code-Switching and Language Contact*. Network on Code-Switching and Language Contact. Ljouwert/Leeuwarden, 14–17 September 1994. *Fryske Akademy*. 58–67.

- Friedman, Victor 2009. Turkish infinitives in Balkan Romani: from codeswitching to paradigm shift – Балканско езикознание / Linguistique balkanique XLVIII (1-2), 27–32.
- Grosjean, François 2001. The bilingual's language modes. In: Nicol, Janet. One mind, two languages: Bilingual language processing. Oxford and Malden, MA: Blackwell. 1–22.
- Gumperz, John; Wilson John 1971. Convergence and creolization: a case from the Indo-Aryan/Dravidian border in India. In: Hymes, Dell (Ed.) Pidginization and creolization of languages. Cambridge, UK: Cambridge University Press, 151–167.
- Hlavac, Jim 2003 Second-generation speech. Lexicon, code-switching and morpho-syntax of Croatian-English bilinguals. Bern: Peter Lang.
- Hlavac, Jim 2006 Bilingual discourse markers: Evidence from Croatian-English code-switching – Journal of Pragmatics 38, 1870–1900.
- Hlavac, Jim 2016. Three generations, two countries of origin, one speech community: Australian-Macedonians and their language(s). Munich: Verlag Otto Sagner.
- Jake, Janice; Myers-Scotton, Carol 2009. Which language? Participation potentials across lexical categories in codeswitching. In: Isurin, Ludmila; Winford, Donald; de Bot, Kees. (Eds.) Multidisciplinary Approaches to Code Switching. Amsterdam/Philadelphia: John Benjamins. 307–336.
- Jašar-Nasteva, Olivera 1962/1963 Makedonski kalki od turskiot jazik – Makedonski Jazik 13-14, 109–172.
- Johanson, Lars 1998. Frame-changing code-copying in immigrant varieties. In: Extra, Guus & Verhoeven, Ludo (Eds.) Bilingualism and migration. Berlin: Mouton de Gruyter. 247–260.
- Myers-Scotton, Carol 2002 Contact Linguistics. Bilingual Encounters and Grammatical Outcomes. Oxford: Oxford University Press.
- Myers-Scotton, Carol; Jake, Janice 2014. Nonfinite verbs and negotiating bilingualism in codeswitching: Implications for a language production model – Bilingualism: Language and Cognition 17(3), 511–525.
- Otheguy Ricardo 1993. A reconsideration of the notion of loan translation in the analysis of U.S. Spanish In: Roca, Ana; Lipski, John (Eds.) Spanish in the United States: Linguistic contact and diversity. Berlin/New York: Mouton de Gruyter, 21–45.

- Stewart, Betty 1995. Some observations on the Speech of Bilingual Macedonians. In: Bivell, Victor (Ed.) *Macedonian Agenda*. Sydney: Pollitecon. 15–26.
- Van Hasselt, F. G. 1972. Idiomatic parallels in Turkish and Modern Greek. In: Bakker, Willem; van Gemert Arnold; Aerts, Willem (Eds.) *Studia Byzantina et Neohellenica Neerlandica* Leiden: Brill. 270–280.
- Verschik, Anna 2008. *Emerging Bilingual Speech. From Monolingualism to Code-Copying*. London/New York: Continuum.
- Vidoeski, Božidar 2005. *Dialects of Macedonian* [Translated by Paul Foster]. Bloomington, IND: Slavica.

RESÜMEE

## KOODIVAHETUS, LEKSIKAALGRAMMATILISED OMADUSED JA TÕLKELAENUD: ANDMED SUUREST MAKEDOONIA-INGLISE KORPUSEST

Artikkel käsitleb 103 makedoonia päritolu austraallase kakskeelset kõnet ja keskendub koodivahetusele leksikaalses mõttes (sisestus, vaheldus, viitamine teistest keeltest ülekantud leksikaalsele materjalile olenemata teise keele löikude pikkusest). Kahe- või mitmelemendilisi konstruktsioone vaadeldakse kui koodivahetusega seotud nähtust, erilise tähelepanu all on nn kergverbi konstruktsioonid. Uuritakse leksikaalgrammatilist ja -semantilist mõju, mis tekib tõlkelaenude kaudu, selleks et teada saada, kas imporditud leksikaalsed elemendid esinevad samal ajal struktuuriliste uuendustega ja soodustavad nende tekkimist.

# EESTI-LÄTI KOODIKOPEERIMINE: ADAPTSIOON JA IMPOSITSIOON

Elina Joenurma

Tallinna ülikool

**Ülevaade.** Artiklis käsitlen kontaktist johtuvaid keelemuutusi eesti-läti kakskeelse isiku keelekasutuses sotsiolingvistikulisest lähenemisnurgast. Artikkel tugineb minu Tallinna Ülikoolis 2016. aastal kaitstud magistritööle „Eesti-läti kakskeelse isiku keelekasutus“. Keeleainestiku kogusin välitöö käigus ning kogutud materjali analüüsimiseks kasutan Lars Johanson'i koodikopeerimise mudelit, kuna see on paindlik ning käsitleb sõnavaras ja grammatikas toimuvat ühtses terminoloogias. Samuti saab seda mudelit rakendada kopeerimise uurimiseks mõlemas suunas, nii adaptatsioon  $K2 > K1$  kui ka impositsioon  $K1 > K2$ . Kuna sellist eesti-läti kakskeelsust ei ole varem uuritud, ei ole võimalik sõnastada ka väga konkreetseid hüpoteese. Küll aga võib eeldada varasemate kontaktlingvistika tööde põhjal, et ülekaalus on täielikud koopiad ning valikulisi ja segakoopiaid on täielike koopiatega võrreldes vähe.

**Märksõnad:** koodikopeerimine, adaptatsioon, impositsioon, eesti, läti

## Sissejuhatus

Eestlasi ja lätlasi kui naabreid seob pikk, aastasadu kestnud ühine ajalugu, esiajaloolised kokkupuuted ulatuvad aga mitme aastatuhande taha (Vaba 2010a: 361). Ka läti-eesti keelekontaktidel on pikk ajalugu ning mõlemas keeles on selle jälgi. Seni on rohkem uuritud eesti-läti ja läti-eesti vastastikust sõnavara laenamist (Vaba 1997, 2010a, 2010b). Näiteks toob Vaba (2010b: 210) välja, et eesti lätlaenulises sõnavaras domineerib koduse eluga seotud sõnavara (näiteks joogid-söögid, rõivad). Samuti on rohkem tähelepanu pöördunud eestlaste ajaloole Lätis ja lätlaste ajaloole Eestis ning eestlaste ja lätlaste elulugudele (Mela 2007, 2013, Utno 2007, 2008).

Keelekontaktid ei piirdu ainult sõnavara laenamisega ning sünkrooniliselt ei ole eesti-läti kakskeelsust seni uuritud. Eesti-läti kakskeelsusest ei saa rääkida kui makrotasandil levinud nähtusest, küll aga esineb niisugust kakskeelsust individuaalselt ja segaperedes.

Minu eesmärk oli uurida kontaktist johtuvaid keelemuutusi eesti-läti kakskeelse isiku keelekasutuses sotsiolingvistilisest lähenemisnurgast. Eesti-läti kakskeelse kõne uurimiseks püstitasin kolm uurimisküsimust.

1. Kas koodikopeerimine toimub mõlema keele suunas?
2. Milliseid koopiaaliike keelejuhi kõnes esineb?
3. Kui koodikopeerimine on mõlemapoolne, kas koopiatega liigid erinevad olenevalt kopeerimissuunast?

Püstitatud uurimisküsimustele vastuste leidmiseks kasutan empiirilise keelematerjali analüüsimiseks Lars Johansonini koodikopeerimise mudelit, kuna see on paindlik ning käsitleb sõnavaras ja grammatikas toimuvat ühtses terminoloogias (Joenuurma 2016: 4–6).

Kuna sellist materjali ei ole varem uuritud, ei ole võimalik sõnastada väga konkreetseid hüpoteese. Küll aga eeldan varasemate kontaktlingvistiliste tööde põhjal, (näiteks Roosileht 2013, Igav 2013, Paljasmaa 2012, Vaba 2010), et ülekaalus on täielikud koopiad ning valikulisi ja segakoopiad on täielike koopiatega võrreldes vähe.

Artikkel koosneb kuuest osast. Esimeses osas annan ülevaate andmete kogumisest ning keelejuhi taustast. Teises osas tutvustan Lars Johansonini koodikopeerimise mudelit ja selle mudeli koopiaaliike. Kolmandas ja neljandas osas käsitlen adaptatsiooni ja impositiooni. Uuritud materjal esines ka koodivahetust, millest annan lühikese ülevaate viiendas osas. Artikli kuuendas osas esitan ülevaate uurimuse tulemustest.

## 1. Andmed ja keelejuht

Minu eesmärk oli uurida kas Eestis elava lätlase või Lätis elava eestlase kakskeelset keelekasutust. Keelejuhi otsimisel oli oluline kriteerium, et ta oleks elanud ühes või teises riigis vähemalt nii kaua, et

räägiks soravalt nii läti kui ka eesti keelt. Keelejuht ei pidanud tingimata olema eesti-läti kakskeelne keeleteadaja lapsepõlvest saati (Joenurma 2016: 8).

Kuna keelejuhi leidmine ei olnud väga kerge, siis otsustasin töötada keelejuhiga, kelle kohta oli teada, et ta on kakskeelne, ja kes oli intervjuuerimisega nõus. Keelematerjali kogusin välitöö käigus, intervjuuerides 2013. aasta veebruaris ning 2015. aastal 14. juunist kuni 20. septembrini Lätis Ainaži linnas elavat 1935. aastal sündinud prouat. Keelejuhi esimene abikaasa oli eestlane ja nad elasid Eestis. Tema mõlema lapse emakeel on eesti keel, kuna nad on Eestis elanud ja koolis käinud. Teine abikaasa oli tal Läti eestlane ning keelejuht kolis 40-aastasena Lätti. Nende kodune keel oli eesti keel, aga nad oskasid ka läti keelt. Oluline on see, et keelejuht on omandanud keelekuju, mida kasutatakse tema vahetus ümbruses. Keelejuhti ei ole kunagi õpetatud õpiku järgi, vaid ta on läti keelt õppinud töö ning kohalike elanikega suheldes. Seega läti keele variant, mida keelejuht valdab, on kohalik murre, mis kuulub liivipäraste murrete hulka (*Tāmnieku dialekts*). Ei saa küll väita, et Ainažis räägitakse ainult murdes, kuid see murre ei ole kadunud. Keelejuht oskab kindlasti passiivselt ka läti kirjakeelt, kuna ta loeb Läti ajalehti ja ajakirju (Joenurma 2016: 10–12).

Intervjuusid viisin läbi nii eesti kui ka läti keeles. Intervjuusid lindistasin keelejuhi kodus ning ühe intervjuu pikkus oli 50–60 minutit. Kokku viisin läbi kaheksa intervjuud. Vestlust alustasin poolstruktureeritud intervjuuna, küsisin eluloolisi andmeid ja keeleteadustavade kohta. Sellele järgnes vabas vormis vestlus, mis oligi minu eesmärk, et koguda keeleainestikku, mis oleks lähedane argivestlusele ning võimaldaks koguda vajalikus matus keeleainestikku. Lisaks kõne salvestamisele tegin ka märkmeid iga intervjuu kohta, näiteks millest me rääkisime, kuidas vestlus kujunes.

Vestlustes käsitlesime erinevaid teemasid. Rääkisime keelejuhi elust, tema Lätti elama asumisest, perekonnast, keeleteadustavast perekonnas, tööst, reisimisest, tervisest, naabritest, argipäeva toimetamisest, poliitikast, kohalikest elanikest, sõbrannadest,



sugulastest ja kassist, keda keelejuht peab perekonnaliikmeks ning väga oluliseks (Joenurma 2016: 12).

Välitööl peab alati arvesse võtma uuritava ja uurija omavahelisi suhteid. Näiteks mina ja keelejuht elame samas linnas ja tunneme kohalikku kultuuri ning inimesi, samuti oleme mõlemad eesti-läti kakskeelsed keelekasutajad. Minu arvates aitas see kujundada intervjuud argivestluseks, kuna keelejuhil ja minul oli kerge vestlust suunata ja rääkida teemadest, millega oleme mõlemad tuttavad. Seega võib väita, et oli kergem saavutada püstitatud eesmärk ehk salvestada tavalist argivestlust. Kuigi alati peab arvesse võtma, et lindistamisel võib keelejuht oma keelekasutust muuta, tundub mulle siiski, et minu keelejuht ei pööranud lindistamisele väga suurt tähelepanu, kuid ma ei saa väita, et lindistamine ja minu juuresolek keelejuhi keelekasutust üldse ei mõjutanud (Joenurma 2016: 9–12).

## 2. Koodikopeerimise mudel

Koodikopeerimise mudelit on Eestis edukalt kasutatud ka varem. Lars Johanson'i koodikopeerimise mudeli abil on eesti-vene keelekontakte uurinud Anna Verschik (2006, 2007, 2008, 2010, 2014), samuti on mudelit rakendanud Kristiina Praakli (2009), Reet Igav (2013), Veronika Paljasma (2012), Helin Roosileht (2013), Marja Vaba (2010).

Koodikopeerimise mudeli põhimõte on keele elementide, näiteks leksikaalsete üksuste ja grammatiliste mallide, kopeerimine ühest keelest teise. Keelt, mille elemente sisestatakse, nimetatakse mudelkoodiks ning keelt, millesse elemente sisestatakse, nimetatakse põhikoodiks. Koodikopeerimise mudelis eristab Johanson täielikke koopiaid (*global copy*), valikulisi koopiaid (*selective copy*) ning segakoopiaid (*mixed copy*). Koopiatel võib olla nelja tüüpi omadusi – materiaalsed (hääldus, aktsent, rõhk ning intonatsioon), kombinatoorsed (sõnajärg, rektsioon, fraasi struktuur), semantilised (sõna või väljendi tähendus) ning sageduslikud (keeleelemendi esinemissagedus). Koopialiigid on omavahel seotud ning erinevad

vaid kopeerimise mahu poolest. Kopeerida võib kas kõiki omadusi täies mahus või ainult osa omadusi (Johanson 2002: 288–292).

Täielik koopia koodikopeerimise mudelis tähendab, et mudelkoodi üksus kopeeritakse tervikuna, kaasa arvatud selle kõik materiaalsed, semantilised, kombinatoorsed ning sageduslikud omadused (Johanson 2002: 291–292). Valikuline koopia tähendab, et kopeeritakse üht või mitut, kuid mitte kõiki omadusi. Mudelkoodist kopeeritakse põhikoodi valikuliselt üksikuid materiaalseid, kombinatoorseid, semantilisi ja sageduslikke omadusi. Segakoopiad kombineerivad mõlemaid elemente nii mudelkoodist kui ka põhikoodist. Vähemalt üks element peab olema täielik koopia ja teine element on valikuline (Johanson 1999: 45).

Lars Johanson koodikopeerimise mudeli valisin ka seetõttu, et seda on võimalik rakendada mõlemas kopeerimissuunas. Seega nimetab Johanson adaptatsiooniks seda, kui kopeeritakse sotsiolingvistiliselt domineerivast keelest B domineeritavasse keelde A, kus mudelkood toob muutusi põhikoodi. Impositsiooniks nimetab Johanson seda, kui kopeeritakse domineeritavast keelest A domineerivasse keelde B, kus põhikood toob muutusi mudelkoodi (Johanson 2002: 291).

Käesolevas artiklis on tegemist nii adaptatsiooniga ( $K_2 > K_1$  ehk läti keelest eesti keelde) kui ka impositsiooniga ( $K_1 > K_2$  ehk eesti keelest läti keelde). Kuna aga keelejuht räägib soravalt mõlemat keelt ja kohati on koodikopeerimine üsna tihe, siis on keelte omandamise järjekord nummerdatud kronoloogiliselt (eesti keel on esimesena ( $K_1$ ) ja läti keel on teisena ( $K_2$ ) omandatud), aga ei kirjelda keeleoskustaset (Joenurma 2016: 18).

### 3. Adaptatsioon ehk kopeerimine läti keelest eesti keelde

Täielikku kopeerimist soodustab sageli elemendi atraktiivsus. Mõned elemendid on atraktiivsemad kui teised, aga atraktiivsust ei saa vaadata kui üldist terminit, kuna see sõltub konkreetsest kontaktsituatsioonist (Backus, Verschik 2012: 134). Atraktiivsuse

määravad keelelemendi omadused, nagu spetsiifiline tähendus ehk semantiline spetsiifilisus. Näiteks läti keelest eesti keelde kopeerib keelejuht perekonnaga seotud sõnu, nagu läti *māsica*, liivipärane murre *māsica*, eesti 'täditütär/onutütär', liivipärane murre *mās*, läti *māsa*, eesti 'õde', läti *mazdēls*, eesti 'pojapoeg', liivipärane murre *onkul*, läti *onkulis*, eesti 'onu'. Perekonnaga seotud sõnade kopeerimist võib soodustada see, et keelejuht on palju rääkinud oma perekonnast läti keeles ning need on juba muutunud harjumuspärasteks sõnadeks nagu näidetes 1 ja 2 (Joenurma 2016: 28).

- (1) Tegi **māsica** nime peale maja  
(murdes *māsica*, läti *māsica*, eesti 'täditütär/onutütär')
- (2) **Mās** helistab!  
(murdes *mās*, läti *māsa*, eesti 'õde')

Kopeerimist võivad soodustada ka sarnasused, näiteks sama sõnajärg, ühisest allikast pärit leksikaalsed elemendid, näiteks ühised võõrsõnad (Verschik 2007: 367, 2008: 90). Eesti ja läti keeles on palju laene ühisest allikast, nimelt saksa keele variantidest, kõige rohkem alamsaksa keelest (Joenurma 2016: 29).

Näites 3 esineb sõna *korsten*, mis on nii läti kui ka eesti keelde laenatud alamsaksa sõnast *schorstēn*. Läti kirjakeeles on sõna *skurstenis*, kuid keelejuhi kõnes on kadunud sõna lõpp *-is*.

- (3) **Skursten** on ära.  
(murdes *skursten*, läti *skurstenis*, eesti 'korsten')

Näites 4 esineb sõna *knapi* 'napilt'. Sõna napp on laenatud alamsaksa sõnast *knap(pe)*. Mõlema sõna sarnasus võis kopeerimist soosida.

- (4) Aga ta oli nii tulnud **knapi** (.) et oleks aeda ära noh (.) riivinud.  
(läti *knapi*, eesti 'napilt')

Keelejuhi kõnes esineb nimisõnade morfosüntaktiline integreerumine, mis ei toimu küll alati. Integreerumine tähendab mudelkoodi elementide kohandamist K1 fonoloogilise, morfoloogilise ning süntaktilise struktuuriga (Praagli 2009: 116). Näites 5 esineb sõna

*jauneid* ‘uusi’. Keelejuht räägib sugulasest, kes on talle riideid toonud. Ta nimetab toodud riideid kaltsudeks, kuna need on talle liiga suured. Läti nimisõna *jaun-* tüvele on keelejuht lisanud *-e-*, et saaks moodustada eesti keele mitmuse osastavat *jaun-e-id*. Läti kirjakeeles *jaunas* ‘uued’ on mitmuse nimetavas käändes. Diskursusmarkerist *tur* ‘seal’ kirjutan näites 8.

- (5) Sellel on ka tead **tur** (.) siia tõi mulle **jauneid** kaltsusid tead.  
(läti *jaunas*, eesti ‘uued’)

Kopeerimist ei ole alati võimalik selgitada. Näiteks tavaliselt öeldakse, et nn kinnised klassid, kuhu kuuluvad grammatilised sõnad, kaasa arvatud asesõna, ei ole hõlpsasti kopeeritavad, aga samas ei ole see ka võimatu (Backus, Verschik 2012). Näites 6 esineb määratlev asesõna *abi divi* ‘mõlemad kaks’. Selle asesõna kopeerimist on raske seletada.

- (6) Ta saatis selle (...) lapsed ära (..) ja siis me **abi divi** jõime.  
(läti *abi divi*, eesti ‘mõlemad kaks’)

Samuti võib kopeerimise põhjus olla pragmaatiline olulisus. Kui mõni element on pragmaatiliselt oluline, näiteks väljendab kõneleja suhtumist või ütluste omavahelist suhet, siis on selle kopeerimine tõenäoline (Backus, Verschik 2012: 142). Sellised elemendid on diskursusmarkerid, mida kopeeritakse tavaliselt täielikult (Joenurma 2016: 22).

Yaron Matrasi (2009: 193) järgi asuvad just diskursus-pragmaatilised sõnad laenatavuse hierarhia tipus. Küll aga kopeeritakse diskursusmarkereid vähe läti keelest eesti keelde. Diskursusmarkerite liigitamisel kasutan Suzanne Wertheimi (2003) klassifikatsiooni, mille järgi esineb keelejuhi kõnes rinnastav konjunktsioon (*un* ‘ja’) näites 7, metakommentaar (adverb *tur* ‘seal’, asesõna *kas* ‘mis’) näites 8 ja hinnangusõna (adverb *nē* ‘ei’). Ei esinenud aga ühtegi suhtlusperformatiivi. See võib olla sellepärast, et kuna enamasti rääkis keelejuht ehk esines monoloog, siis suhtlussituatsioon ei eeldanudki suhtlusperformatiivide kasutamist (näiteks *tere, head aega, tänan*

jne). Seega on suhtlusperformatiivide esinemine tingitud suhtlus-situatsioonist (Joenurma 2016: 33). Näites 8 olevast sõnast *jauneid* kirjutasin näites 5.

- (7) Aga siis ta tuli ütlema et tüdrukud lapsed on haiged ja see jääb ära see jääb kas siis on teisipäev või kolmapäev **un un** panin uue selle tead.

(läti *un*, eesti 'ja')

- (8) Sellel on ka tead **tur** (.) siia tõi mulle **jauneid** kaltsusid tead.

(läti *tur*, eesti 'seal')

Valikulist kopeerimist esineb vähem kui täielikku kopeerimist ning kopeeritakse materiaalseid ja semantilisi omadusi. Näites 9 esineb sõna *analīsid* 'analüüsid' Läti keeles on sõna ainsuse nimetavas *analīze* ning mitmuse nimetavas *analīzes* sama tähendusega nagu eesti keeles 'analüüs', mis tähendab 'aine koostise osalist või täielikku kindlakstegemist'. Tüvi on tajutav *analīūs-* või *analīz-*. Eesti keeles oleks 'tehakse analüüse' mitmuse osastavas käändes, läti keeles *taisa analīzes* 'tehakse analüüsid' mitmuse nimetavas. Sõna on ühine internatsionalism, mis on mõlemas keeles sarnane, aga mitte identne. Eesti keeles kasutatakse küll mitmuse osastavat, aga läti keeles mitmuse nimetavat. Keelejuht kasutab tüve lätipärasel kujul, lisab eesti mitmuse osastava tunnuse *-sid*, aga mall on jälle läti oma (nimetav kääne). Seega esineb läti tüvi (ehk käändelõppu pole) ning eesti mitmuse nimetav: *analīs-id*. Siin on juhtum, kus sõnavara ja morfosüntaks ei ole selgelt eristatavad.

- (9) Seal tehakse **analīsīd**.

(läti *analīzes*, eesti 'analüüsid')

Huvitaval kombel esinevad ka vahepealsed juhtumid ning on raske määrata, mis koopiaaliiki nad kuuluvad. Selliseid juhtumeid on ainult üksikuid. Näites 10 esineb sõna *pruutgan* tähendusega 'peigmees'. Eesti keelde on sõna *pruut*, läti *brūte* laenatud alamsaksa sõnast *brūt* ning läti keele *brūtgāns*, eesti 'peigmees' on laenatud alamsaksa sõnast *brūdegam* (Karulis 1992: 150). Eesti keeles tähendab *pruut*

naist, aga läti keeles on *brūtgāns* mees. Sõnast *brūdegam* on see läti keeles mugandus. Materiaalne sarnasus soodustab kopeerimist, kuigi tähendus kummaski keeles erineb. Seega keelejuht kopeerib läti sõnakuju, kuna see on sarnane, aga tähendus on eesti oma.

- (10) Ja tuleb jälle tagasi (...) tal on ka noh **pruutgan** tead (.)  
(läti *brūtgāns*, eesti 'peigmees')

#### 4. Impositsioon ehk kopeerimine eesti keelest läti keelde

Nagu juba eespool kirjutasin, ei integreerita täielikult kopeeritud nimisõnu alati morfosüntaktiliselt. Näiteks näites 11 esineb sõna *nutitelefon*. Läti kõnekeeles kasutatakse ka sõna *smārtfons*, mis on pärit inglise sõnast *smartphone*, aga läti kirjakeeles on *viedtālrūnis*. Selles kõnesituatsioonis võib keelejuhi jaoks olla sõna *nutitelefon* atraktiivsem sõna, kuna tema perekonna suhtluskeel on eesti keel ja ta teab, millisest telefonist räägitakse. Võimalik on ka olukord, et keelejuht ei ole teadlik, kuidas sõna *nutitelefon* läti keeles on, kuna tal ei ole olnud lätikeelse sõnaga kokkupuudet. Sõna *smārtfons* on kasutusel olnud suhteliselt lühikest aega ja on võimalik, et sõna ei olegi keelejuhi sõnavaras. Läti keeles on rektsioon *vajag* + akusatiiv *vajag nutitelefon-u*. Ka *to* on akusatiiv. Siit on näha, et puudub ühildumine demonstratiivpronoomeni ja nimisõna vahel. Nutitelefon on jäetud nimetavasse käändesse ning morfosüntaktiline integratsioon puudub.

- (11) Vajag **to nutitelefon**.  
'Vaja seda nutitelefoni'

Näidetes 12 ja 13 esineb sõna *kāula*. Sõna *kāula* tähendab kõnekeeles 'väiksem, viletsam laev, alus, paat või auto; logu'. Näites 12 räägib keelejuht, millistes toidupoodides ta käib. Selles näites kasutab ta sõna *kāula* tähenduses 'ratastel kott'. Kuna tal on raske liikuda ja ta ei saa raskeid asju tõsta, siis on talle ostetud ratastel kott, millega poes käia.

(12) Tā, ka man tas **kāula** būs atpakaļ, tad es var site iet.

‘Siis, kui mul see kāula on tagasi toodud, siis ma saan sinna minna’

Näites 13 kasutab keelejuht sõna *kāula* ‘auto’ tähenduses. Keelejuht räägib, et poeg oli ostnud uue auto, aga et pojapoeg on nii pikk ja ta ei mahu väiksesse autosse sisse.

(13) Man brāls bij līdzi un sak: tāda maza **kāula** viņi ju netiks ieksa nemaz tur.

‘Mul oli vend kaasas ja ta ütleb: nii väiksesse kāulasse nad ei saa ju sisse üldse’

Nagu adaptatsioonis kopeeritakse ka impostsioonis sarnaseid leksi-kaalseid elemente. Näites 14 esineb sõna *poliitik*. Läti keeles on vaste *politiķis*. Küsin keelejuhi käest, mida ta arvab poliitikast. Keelejuht vastab naljatades, et ta on ise parem poliitik. Kuna sõnad on eesti ja läti keeles sarnased, tunnetab keelejuht, et läti keeles on natuke teistsugune sõna, ja lisab korrates sõnale *poliitik* läti meessoost tunnuse *-s*. Läti keeles on *politiķis* (meessoost) / *politiķe* (naissoost).

(14) EJ: Kā ar politiku?

‘Kuidas on poliitikaga?’

KJ: Es pats ir labaks **poliitik**

‘mina olen ise parem poliitik’

EJ: [Naerab]

KJ: **Poliitiks** vien plāpašan ir **noh**.

‘Poliitik üks lobisemine on noh’

Võrreldes adaptatsiooniga, esineb impostsioonis rohkesti diskursusmarkereid. Diskursusmarkeritena esinevad alistav (*et*), rinnastav (*ja*) ning vastandav konjunktsioon (*aga, vōi*) näidetes 15 ja 16, metakommentaar ja deiksis (*ju, oota, oot-oot, vaata, vat, tead, noh*), hinnangusõnad (*ai jumal, ei, vāāks, nurr, auh, uha-uha*). Suhtlusperformatiive aga ei esine, kuna ilmselt vestlussituatsioon neid ei soodustanud.

(15) Ja būt zinajusi **aga** tu nezin.

‘Kui oleks teadnud aga sa ei tea

(16) Tas bij ari tas vien (.) dakter **või** kas tas tur bij un jā (.) jā un tā (.)  
atsakas un es parakst un viss.

‘See oli ka üks arst või kes ta seal oli ja jaa ja nii keeldun ja ma all-  
kirjastan ja kõik’

Näites 17 esineb partikkel *oot-oot*. *Oota* on justkui stoppmärk, millega võib sisse juhatada pausi, seletusküsimuse, kõrvaljärjendi või ka teemamuutuse (Keevallik 2001: 143). Keelejuht räägib enda maja omanikust ja mina küsin, kus maja omanik nüüd elab. Keelejuht võtab aega partikliga *oot-oot* ja kohe kordab sama läti keeles – *pagaid* ‘oota’. Selline kordamine annab talle rohkem mõtlemisaega.

(17) **Oot-oot** pagaid (.) uz Igaun robeža uz Igaunī puse Latvija.

‘Oot-oot oota Eesti piiril Eesti piiril Läti poolel’

Näites 18 esineb onomatopoeetiline sõna *nurra*. Keelejuhile on kass väga oluline ja kassiga räägib ta ainult eesti keeles. Näites räägib keelejuht, et ta oli haiglas ja kui ta koju tagasi tuli, siis oli kass väga rõõmus olnud ja kogu aeg keelejuhile sülle ja voodisse magama läinud. Läti keeles jälgendatakse kassi ketravat hääliksust *murr-murr*. Läti keeles on tegusõna *murrāt* ‘nurruma’ ning *kaķis murrā* ‘kass nurub’. Võimalik, et keelejuht lisas sõnale *nurr* lõpu *-a* sõnast *murrā*.

(18) Sākuma, kad es nāk vairak nekas nebij, ka tikai klēpi guļ man blakus un tikai nāk site un **nurra, nurra**.

‘Alguses, kui ma tulin siis rohkem midagi muud ei olnud, et ainult magab süles mu kõrval ja ainult tuleb siia ja nurr-nurr’

Valikulisi koopiaid on kordades vähem kui täielikke koopiaid. Näites 19 esineb läti keelses vestluses konjunktsioon *või*. Keelejuht kasutab *või* sellises funktsioonis, nagu seda eesti keeles ei kasutata. Läti keeles saab konjunktsiooniga *vai* moodustada küsilause, aga eesti keeles *või* niisuguses funktsioonis ei kasutata, vaid küsilausest alustab sõna *kas*. Keelejuht on võtnud eesti *või* materiaalsed omadused, aga kasutab vastava läti keele konjunktsiooni funktsiooni.



(19) Un mäs tülit (.) mäs ari bij klāt tur un un un ta pras **või** (.) **või** vini var maksat tas?

‘Ja õde kohe tuli ja küsib kas ta saab ära maksta?’

Nii nagu valikulisi koopiaid on ka segakoopiaid vähem kui täielikke koopiaid. Näites 20 esineb segakopia *neerus vēz*. Keelejuht räägib oma vennast, kellel oli neeruvähk. Läti keeles *neerud* on *nieres* ja *vāhk* on *vēzis* ning läti keeles öeldakse haiguse kohta *nieru vēzis* nagu ka eesti keeles ‘neeruvähk’. Keelejuht tunnetab, et läti keeles on eesti sõna *neerud* sarnane, ning annab pausiga endale mõtlemisaega ja proovib veel kahel korral leida õiget läti sõna. Keelejuht moodustab segakopia, ühildades eesti *neerus* ja läti *vēz*.

(20) Un viņi nomir vinim bij **neer** (...) **neer vēz neerus vēz**.

‘Ja ta suri ja tal oli neeruvähk’

Ka impositsioonis esineb ebaselge juhtum, mille puhul on raske määrata, mis koopialiiki see kuulub. Näites 21 esineb sõna *narret* ‘nerrot’. Eesti ja läti keelde on see laenatud kas alamsaksa keelest *narre* või saksa keelest *narren*. Eesti sõnast *narrima* ja läti sõnast *nerrot* on moodustatud sõna *narret*. Eesti tüvele *narr-* lisatakse läti kirjakeeles infinitiivi lõpp *-ot*, keelejuht on selle asemel aga kasutanud lõppu *-et*. Sõnaga *durns* mõtleb keelejuht ‘loll’. *Durn-am* esineb daativis, mis on üks laialdasi liivipärase murde tunnuseid. Kirjakeeles on *durnu* akusatiivis.

(21) Durnam var **narret**.

‘Lolli võib narrida’

## 5. Koodivaheldus

Johanson eristab oma mudelis koodikopeerimist ja koodivaheldust (*code-alternation*). Koodivaheldus tähendab selles mudelis terveid lõike teises koodis (näiteks osalauseid ja lauseid) (Johanson 1999: 39, 2002: 287).

Ka kogutud keeleainestikus esineb mõni koodivahelduse näide. Koodivaheldus esineb mõlema keele suunas nii lausesiseselt kui ka

lauseväliselt. Ühel korral toimub koodivaheldus kolme keele vahel: läti, vene ja eesti. Keelejuhi kõnes esineb koodivaheldus sageli just kellegi refereerimisel.

Näites 22 on lätikeelne osa paksus kirjas. Selles näites esineb koodivaheldus lausesiseselt ning toimub kahe keele vaheldus: eesti-läti-eesti. Selles näites räägib keelejuht eesti keeles oma kassist, kellega ta räägib ainult eesti keeles. Siin on hästi tihe kopeerimine. Keelejuht hakkab rääkima kassist ning läheb üle läti keelele, aga siis peatub ning tajub, et on läti keelele üle läinud, ning parandab ennast eesti keeles ja läheb taas üle läti keelele sõnaga *ir* 'on' ning alles siis jätkab eesti keeles.

(22) Aga meil siin **tas kaķ iet un kaķ** [rõhutab] (.) kass **ir** aga tema vist ei saa kätte.

'Aga meil siin see kass läheb ja kass kass on aga tema vist ei saa kätte'

## 6. Tulemused

Keelematerjali analüüsidest selgus, et kopeerimine toimub mõlemas suunas ehk läti keelest eesti keelde ja eesti keelest läti keelde. Nagu oli oodata, on mõlema kopeerimissuuna puhul ülekaalus täielikud koopiad. Samuti on mõlemas suunas kopeeritud täielikult enamasti nimisõnu; verbe ja teisi sõnaliike on vähe. Kuigi on arvatud, et asesõnad ei ole hõlpsasti kopeeritavad, esineb siiski ühel korral ka asesõna kopeerimist. Ka kopeerimise põhjused mõlemas suunas on peaaegu samasugused. Kopeeritakse semantiliselt spetsiifilisi elemente ning materiaalselt sarnaseid elemente (ühised laenud alamsaksa keelest). Adaptiooni puhul kopeeritakse ka sõnu, mis on seotud keelejuhi perekonnaga ja mida on harjumus kopeerida läti keelest eesti keelde. Täielikult kopeeritud nimisõnu ei integreerita alati morfosüntaktiliselt (s.o tüvele ei lisatud käändelõppe).

Diskursusmarkereid kopeeritakse mõlemas suunas. Adaptiooni puhul (läti keelest eesti keelde) on siiski vähem pragmaatilisi sõnu. Impositsiooni puhul kopeeritakse rohkesti konjunktsioone,

hinnangusõnu ja metakommentaare, interjektsioone. Suhtlusperformatiive ei esine kummaski kopeerimissuunas, kuna ilmselt suhtlussituatsioon ei soosinud seda. Vastavalt Matrasi (2009: 193) kontseptsioonile, mille järgi asuvad diskursus-pragmatilised sõnad laenatavuse hierarhia tipus ning pragmaatilisi sõnu laenatakse pragmaatiliselt domineerivast keelest, mis võib olla ka sotsiolingvistiliselt domineeriv keel, võib ettevaatlikult oletada, et keelejuhi pragmaatiliselt domineeriv keel on eesti keel, kuna rohkesti kopeeritakse diskursusmarkereid just eesti keelest läti keelde.

Valikulisi koopiaid ja segakoopiaid on kordades vähem kui täielikke koopiaid. Mõlemas suunas esineb üksikuid näiteid materiaalsete ja semantiliste omaduste kopeerimise kohta. Segakoopiaid esineb vähe ning vaid impositsiooni puhul, kuid adaptatsioonis ei esine ühtegi.

Mõlemas kopeerimissuunas esineb ka paar ebaselget juhtumit, mille puhul on raske määrata, mis koopialiiki need kuuluvad.

Keelejuhi kõnes esineb koodivaheldust mõlema keele suunas nii lausesiseselt kui ka lauseväliselt sageli just kellegi refereerimisel.

Nagu juba eespool kirjutasin, ei ole eesti-läti kakskeelsust siiani sünkroniliselt uuritud, kuid seda oleks vaja rohkem uurida. Selleks oleks vaja koguda rohkem keeleainestikku ning keskenduda nii eesti-läti ja läti-eesti üksikisiku keelekasutusele kui ka segaperekondade keelekasutusele. Minu arvates võimaldaks selline kakskeelsuse uurimine teha põhjalikumaid järeldusi, milline on kakskeelse isiku keelekasutus osates kahte naaberriigi keelt.

## KIRJANDUS

- Backus, Ad; Verschik, Anna 2012. Copiability of (bound) morphology. – Copies versus Cognates in Bound Morphology. Eds. Lars Johanson, Martine Robbeets. Boston, Leiden: BRILL Academic Publishers, 123–149.
- Igav, Reet 2013. Inglise-eesti koodikopeerimine Facebooki vestlustes. Magistritöö. Tallinna Ülikool.
- Joenurma, Elina 2016. Eesti-läti kakskeelse isiku keelekasutus. Magistritöö. Tallinna Ülikool.

- Johanson, Lars 1999. The Dynamics of code-copying in language encounters. – Language encounters across time and space. Eds. Bernt Brendemoen, Elizabeth Lanza and Else Ryen. Oslo: Novus Press, 39–62.
- Johanson, Lars 2002. Contact-induced change in a code-copying framework. – Language change: The interplay of internal, external and extralinguistic factors. (Contribution to the Sociology of Language, 86.) Eds. Mari C. Jones, Edith Esch. Berlin, New York: Mouton de Gruyter, 285–313.
- Karulis, Konstantīns 1992. Latviešu etimoloģijas vārdnīca I. Zinātniskais red. N. Andrejevs. Rīga: Avots.
- Keevallik, Leelo 2001. Tracing grammaticalization of oota ‘wait’ in Estonian conversation. – Papers in Estonian Cognitive Linguistics. Eds. Ilona Tragel. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 2. Tartu, 119–144.
- Matras, Yaron 2009. Language contact. – Cambridge Textbooks in Linguistics. New York: Cambridge University Press.
- Mela, Marjo 2007. Lāti eestlased: ajalugu, keel ja kultuur. Tallinn: Eesti Keele Sihtasutus.
- Mela, Marjo 2013. Eesti lätlased. Tallinn: Nornberg & Co.
- Paljasma, Veronika 2012. Prantsuse-eesti koodikopeerimine blogides. Magistritöö. Tallinna Ülikool.
- Praakli, Kristiina 2009. Esimese põlvkonna Soome eestlaste kakskeelne keelekasutus ja koodikopeerimine. Doktoriväitekiri. Tartu: Tartu Ülikooli kirjastus.
- Roosileht, Helin 2013. Inglise-eesti koodikopeerimine blogides. Magistritöö. Tallinna Ülikool.
- Utno, Leili 2007. Eestlased Lätis I. Toim. Karl Kruus. Tallinn: Välis-Eesti.
- Utno, Leili 2008. Eestlased Lätis II. Toim. Karl Kruus. Tallinn: Välis-Eesti.
- Vaba, Lembit 1997. Uurimusi Läti-Eesti keelesuhetest. Tallinn-Helsinki: Eesti keele instituut/ Tampereen Yliopiston suomen kielen ja yleisen kielitieteen laitos.
- Vaba, Lembit 2010a. Eestlased ja eesti keel Lätis: Liepna ja Balvi Eesti siirdlaskonna tekke-, arengu- ja hääbumislugu ühe suguvõsa perepärilmuste põhjal. – Eestlased ja eesti keel välismaal. Koost ja toim Kristiina Praakli, Jüri Viikberg. Tallinn: Eesti Keele Sihtasutus, 361–381.

- Vaba, Lembit 2010b. Kuidas läti-eesti keelekontakt on mõjutanud eesti murdekeele grammatikat ja sõnamoodustust. – Emakeele Seltsi aastaraamat 56. Tallinn: Teaduste Akadeemia Kirjastus, 204–246.
- Vaba, Marja 2010. Inglise-eesti koodikopeerimisest Skype'i Tallinna kontori kahe vestlusgrupi näitel. Magistritöö. Tallinna Ülikool.
- Verschik, Anna 2006. Koodikopeerimise mudel: tõlgendades keelekontakte mikrotasandil. – Teoreetiline keeleteadus Eestis II. Toim Ilona Tragel, Haldur Õim. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised. Tartu: Tartu Ülikooli Kirjastus, 97–111.
- Verschik, Anna 2007. Keelekontaktid, laenatavus ja verbi kopeerimine eestivene keelevariandis. – Keel ja Kirjandus 5, 357–377.
- Verschik, Anna 2008. Emerging Bilingual Speech: from monolingualism to code-copying. London: Continuum Publishers.
- Verschik, Anna 2010. Russian-Estonian code-copying in Live Journal blogs: a preliminary overview. – Instrumentarium of Linguistics: Sociolinguistic Approaches to Non-Standard Russian. Slavica Helsingiensia 40. Eds. Arto Mustajoki, Ekaterina Protassova, Nikolai Vakhtin. Helsinki, 355–365.
- Verschik, Anna 2014. Conjunctions in early Yiddish-Lithuanian Bilingualism: heritage language and contact linguistic perspectives. – Language contacts at the crossroads of disciplines. Eds. Heli Paulasto, Lea Meriläinen, Helka Riionheimo, Maria Kok. Cambridge Scholars Publishing, 33–58.
- Wertheim, Suzanne 2003. Linguistic purism, language shift, and contact-induced change in Tatar. University of California Berkeley dissertation.

## SUMMARY

ESTONIAN-LATVIAN CODE-COPYING:  
ADOPTION AND IMPOSITION

The aim of the current article is to give an overview of Estonian-Latvian code-copying based on naturalistic language data. The article is based on my MA thesis, 'Eesti-läti kakskeelse isiku keelekasutus' (*Estonian-Latvian Bilingual Speech*). I have used Lars Johanson's code-copying framework to analyze the data. In the code-copying framework code-copying means that items and patterns from the model code are copied into the basic code. My aim was to observe copying in both directions and this framework is applicable in both copying directions. The copying process L1 > L2 is called 'imposition' (in this case Estonian > Latvian) and L2 > L1 copying is referred to as 'adoption' (in this case Latvian > Estonian). Code-copying can be classified as global, selective or mixed copying, according to the degree of copying involved.

The data were collected in February 2013 and from 14 June to 20 September 2015 by recording the natural conversation of an ethnic Estonian informant (b. 1935) who has lived in Ainaži, Latvia for almost 40 years. She is fluent in both Estonian and Latvian. The length of a conversation session is 50–60 min.

The results show that code-copying occurs in both directions and global copies were used the most of all the types of copying. The most copied type of content words were nouns, but there were also some adverb and verb copies. The reasons for copying globally in both directions were semantic specificity, pragmatic prominence and the similar origin of some words both in Estonian and Latvian.

Very few discourse markers were copied in adoption, with more occurring in imposition. Discourse markers were used to direct, stress and express emotions.

There were very few selective and mixed copies, regardless of direction. No mixed copies occurred in adoption. Code alternation occurred in both directions.

Until now Estonian-Latvian bilingualism has not been studied by modern contact-linguistic research and using naturalistic language data. Usage of this method would provide a basis for further research of Estonian-Latvian bilingual speech.

# ENGLISH-ESTONIAN CODE-COPYING IN ESTONIAN BLOGS

Helin Kask

Tallinn University

**Abstract.** The article discusses English-language impact on Estonian in online blogs. The data comprise blog entries from 15 Estonian fashion, beauty and lifestyle blogs from 2012 to February 2016. The corpus consists of 283 post entries (141,480 words (tokens)). The research showed that there are far fewer selective copies than global copies, probably because selective copies appear at more advanced stages of acquisition. Global copies are usually copied due to semantic specificity, for example fashion lexis (e.g. *look*, etc.). They are also copied due to their importance at the discourse level, for example *god*, *oh well*, etc.

Often the English *something is done by someone* and *have*-constructions are copied, as these are already habitualized in Estonian. There is also evidence of semantic copying, which causes changes in content or usage of unusual expressions under English influence.

There are also mixed copies, where global copies are used as lexical cores in copied combinational patterns (e.g. Eng. 'epic' used to form an adjective). A mixed copy often represents a transitory stage between a global copy and a selective copy; this might be the case in this research, as there were more mixed copies than selective ones. This might be because selective copies appear at more advanced stages of acquisition.

**Keywords:** code-copying, blogs, language contacts, English, Estonian

## 1. Introduction

After Estonia regained its independence in 1991 English-Estonian language contact intensified and several English loans have been established in Estonian (Leemets 2002: 41). English-language impact on Estonian and English-Estonian language contact have not been



investigated, apart from studies on conventionalized lexical borrowings in Standard Estonian (Leemets 2003) and a handful of MA theses written within one or another contact linguistics theoretical framework and exploring the impact of English in internet communication (Igav 2013, Roosileht 2013, Vaba 2010).

Multilingual communication on the internet is of interest to a wide range of scholars, for instance, to researchers on language in virtual communities, innovative language use and language contacts (Androutsopoulos 2012, 2013b, 2015, Danet and Herring 2007, Herring 2012, Hinrichs 2006, Jaworska 2014, Dorleijn and Nortier 2009 and many others). The current article focuses on English impact on Estonian in beauty, fashion and lifestyle online blogs, where English has become a *lingua franca* and consequently there is greater exposure to English. The objective of the paper is to discuss why English is used and what types of copies are used the most. From the existing literature it is apparent that the main reasons for code-copying are semantic specificity and importance on a pragmatic level. I am going to show that the reason for copying depends on the type of copy in question.

Due to the rise of what is called 'superdiversity' (Vertovec 2007), there is a need to describe many new concepts; usually they are first described in English. Therefore, many English words and expressions are used in Estonian, some of which are copied, while others are adapted. The use of English is very common in the fields that develop and change rapidly, e.g., information technology. Contact with English is, therefore, both extensive (i.e., English appears in different types of text and oral communication, in which many are proficient) and intensive.

Over the years the importance of English has increased; it is considered as one of the principal *lingua francas*. Thirty-eight percent (38%) of Estonians claim ability to speak English; the figure being higher among the younger generation (Kruusvall 2015: 77). According to Eurobarometer, 97% of Estonian students had learned English as part of their general studies, and in 2014 64% of Estonians

of secondary school age claimed their level in English was fluent or excellent (Ehala 2015: 195, based on MeeMa 2014). It is not, therefore, surprising that Estonian students consider English as the most important subject at school (Tammemägi, Ehala 2012: 249). It is also very common (for younger people) to be exposed to English-language entertainment and popular culture and use English in social networking online on a daily basis. The aim of this article is to describe how English affects the language use of ‘bloggers’ (blog authors) who live in Estonia but use English to read the English language blogs of other bloggers to keep abreast of developments and trends and how this affects their language use on their own Estonian-language (see also Soler-Carbonell 2014; Siiner and Vihalemm 2011).

The article is organized as follows. First, blogging and language use in blogs are discussed. After that data and informants are described. This is followed by an overview of the code-copying model that is used to analyse the data, which is analysed both quantitatively and qualitatively; distinctive or unique examples of all types of copies are presented and analysed. Lastly, a discussion is had and conclusions are presented.

## 2. Blogging and language use in blogs

A blog is a web journal that consists of posts in chronological order. A blog post can vary in length from that of a short note to an essay. Posts are usually written according to a schedule, for example every day or once a week. Blog posts are asynchronous monologs – the posts can be read days or even years after they are written. Usually, a blog is run by one person alone and its appearance reflects the personality of the blogger (Crystal 2007: 15, 240).

A blog can be about sports, a hobby, handicraft, a political view, etc., i.e. there are different reasons for writing a blog. For example, some blogs reflect radical views and it might be difficult to release articles about that topic in the mainstream media, so a blog is used

to impart those ideas. Some blogs are run by a political party, a department of a university, a music store to inform their readers about the latest news, etc. However, mostly blogs are written by one person to express their ideas, everyday life and interests. According to Thurlow (et al 2004: 133), blogging might also have a communicative function – nowadays when people travel and move abroad blogging is a way to keep in touch with family or to communicate and socialise with fellow expats.

When analysing the language used in blogs it must be taken into account that bloggers might not always know who read their blogs. Due to this it differs from a normal conversation, where interlocutors have prior knowledge about each other and where it is possible to consider the command of language of the interlocutors or to agree on using a certain language.

According to Crystal (2007: 15–16), the language of blogs displays the process of writing in its naked, unedited form. The Internet is fluid, people are more creative, exploring different expressions and using fresh combinations of elements and a blogger can manifest her/his own language policy. Because of this it is possible to explore multilingual language use and contact phenomena that do not appear in oral use, for example orthography and the use of new constituent order patterns or phrase structures. Blogs are therefore a good resource for the study of language use as they contain the material in naked form (see also Igav 2013, Roosileht 2013, Vaba 2010).

### 3. Data and informants

In this article 15 fashion, beauty and lifestyle blogs were analysed. Only public blogs were analysed. According to a recent study carried out by IPREX (see Kulper 2015) fashion is one of the main topics in Estonian blogs. The majority of Estonian fashion bloggers write in Estonian (albeit with copies from English), although some blogs have many or all entries in two languages (parallel texts that present more or less the same information). However, the text that appears

second is not a precise translation of the first. All bloggers in the present study are female, likely due to the topic of the blogs. They are 19–30 years old and all reside in Estonia, mostly in Tallinn. They all have revealed at least their first name.

The blog posts analysed were written in 2012–2016. Every blog entry was saved as a separate file in the Notepad++ program, and language contact phenomena were annotated. These data form a corpus that consists of 283 files and 141,480 words (tokens) in total. There are 1,977 instances of code-copying (CC) – on the average there are 7 instances of CC in every blog post. The frequency of posting varies: some bloggers write every day, some once or twice a week. On average, there are 500 words per blog post, with a range of from just few sentences up to 1,000 words. In Table 1 below the number of blog posts each blogger wrote and the number of tokens and CC instances that occurred across all the posts are presented.

**Table 1:** Number of CC instances

Bloggers	Entries	Tokens	CC instances
Blogger 1	32	12,835	138
Blogger 2	4	1,756	31
Blogger 3	20	12,116	178
Blogger 4	55	26,217	302
Blogger 5	33	24,268	229
Blogger 6	11	3,034	40
Blogger 7	1	486	2
Blogger 8	21	17,146	302
Blogger 9	4	2,218	26
Blogger 10	16	6,826	88
Blogger 11	49	13,744	274
Blogger 12	12	1,701	64
Blogger 13	4	601	14
Blogger 14	15	5,936	112
Blogger 15	6	12,596	177
TOTAL	283	141,480	1,977

Most likely bloggers use English to read the English language blogs of other bloggers to keep abreast of fashion and beauty, but the use of English in blogs varies. Although 70–80% of Estonians in the age of 20–29 claim they speak English fluently or excellently (Ehala 2015: 195), it might be assumed that bloggers (subconsciously) limit the use of English because their aim is to write a monolingual text so the readers would understand it. However, considering the specific topic of the blogs it is likely that blog readers are also interested in fashion and beauty and are familiar with this kind of vocabulary.

#### 4. Code-copying in blogs

The data were analysed qualitatively to find out the reasons for presence of English-Estonian CC. For this Johanson's code-copying model was used. In Estonia this model has been previously used by Verschik for Russian-Estonian CC (2014, 2011), Praakli for Finnish-Estonian CC (2009), Vaba for English-Estonian CC (2010), Paljasma for French-Estonian CC (2012), Igav for English-Estonian CC (2013), Pere for French-Estonian CC (2014), Joenurma for Latvian-Estonian CC (2015) and by the present author for English-Estonian CC (2013).

The notion of code-copying is that linguistic elements (units and patterns) of language or code are copied to another. Copies of elements from a foreign model code are inserted into a basic code; usually the basic code is the speaker's first language (L1) and the model code is the second language (L2) (Johanson 2002: 289).

This model focuses on intraclausal CC. The clausal level is chosen, since it is easier to decide whether the basic code is A or B. In some cases it is difficult to define the basic code, since many linguistic elements of the model code are used – a clause may still be A-coded in spite of heavy amount of B-code elements. As a result, purely quantitative criteria are not sufficiently decisive when determining the basic code (Johanson 1993: 199). In this paper the basic code is Estonian and the model code is English.

In this model the asymmetrical dominance relations of the codes are taken into account. Usually elements of a sociolinguistically dominating code B are copied into the dominated code A. For example the dominating code may be associated with power or status among the speakers of the dominated code. However, the relation between dominating and dominated codes may vary considerably according to the contact situation. In this framework it is not relevant whether the codes are genetically related, the codes may be languages, geographical dialects, sociolects, etc. (Johanson 2002: 289–290).

In the context of the blogs analysed in this paper, the bloggers copy elements from the sociolinguistically dominating or ‘strong’ code English to the dominated or ‘weak’ code Estonian. This is the case of adoption; traditionally this is referred to as ‘borrowing’, etc. Vice versa, where speakers use their own variety in the dominating code B, this is called ‘imposition’. Although bloggers may find that English and Estonian are equally dominant for them, English may be sociolinguistically dominant due to the fact that they receive information about the topic they blog about in English.

According to Johanson (2002: 292), all linguistic elements have four types of properties – material, semantical, combinational and frequential. All types of copies are closely related, differing only in terms of the scope of copying.

Material properties are sound features, accent patterns, phonotactic patterns, including pronunciation, accent, stress, intonation. Verschik (2008: 171) adds graphic properties, meaning that in written data orthography should also be taken into account. For example, sometimes the orthography of the basic code is retained, consider *sniikpiik* ‘sneak peek’ and *khuul* ‘cool’, where English words are written as if they were Estonian. In these cases only material properties are copied, but not graphic properties. The choice of orthography might be a compromise or indicative of creativity (Androutsopoulos 2009, Koutsogiannis and Mitsikopoulou 2003) or it could be an example of language play or a joke (Vaba 2010: 64).

Semantic copying means that denotative or connotative content elements of the model code units serve as models and are copied onto units of the basic code. For example, with *see on kõndimiskaugusel* ('it is within walking distance') only the meaning and collocation of the English expression are copied. This is a word-for-word rendition (referred to as 'loan translation' by some researchers); the idiomatic Estonian equivalent is *see on kiviviske kaugusel* 'it's in a stone's throw'.

Combinational properties are word order, phrase structure, government, etc. Due to English influence a progressive construction (*on tegemas* 'is doing') is used in Estonian (Metslang 2006: 719). Often English word order 'something is done by someone' is used, for example *olen toredate inimeste poolt ümbritsetud* 'I'm surrounded by nice people'; idiomatic Estonian requires a different word order (*mind ümbritsevad toredad inimesed* 'nice people surround me'). The use of such constructions is not limited to blogs alone, but is to be found universally in Estonian.

Frequential properties means that the occurrence of a linguistic element might increase or decrease due to the influence of the model code. For example, everyday language use shows that the use of word *omama* 'have' has increased in Estonian and is in daily use.

According to the CC framework, all types of copies are closely related and differ only in terms of their scope. Copying of all properties would result in a global copy; if only certain properties are copied, the result would be a selective copy. In addition, there are mixed copies. A mixed copy is a multiple word item which is copied in a way such that some of its elements are global copies and some selective copies (Johanson 2002: 291–292).

#### 4.1. GLOBAL COPIES

A global copy means that all the properties of the model code unit are copied, in other words, the unit of the model code is copied as a whole, including its form and functions. It can comprise one or

more words, be morphemically simple or complex, and belong to open or closed word class.

Typical global copies tend to have a relatively specific meaning and due to their specific meaning these model code elements are attractive for copying. Model code elements are semantically specific because there is unlikely to exist a native element that matches its semantics completely or because the element marks something unique or is important to the informant; it can also mark contact with new culture or field (Backus, Verschik 2012: 19; Praakli 2009: 94; Backus 2001: 128). In this article it was found that the main reason for copying is filling in of lexical gaps. Basic vocabulary was not copied and remained in Estonian, e.g. *king* 'shoe', *kleit* 'dress', etc.

One reason for copying is importance on the pragmatic level. According to Matras (1998: 281), bilinguals prefer the discourse markers of the pragmatically dominant language. Matras suggests that their prominence at the discourse level is responsible for their copiability. Backus and Verschik (2012) also note that if an element is perceptually salient, for example because it occupies a special focus position in the clause, it becomes extra noticeable and this increases the element's copiability.

Due to semantic specificity mostly nouns and noun phrases were globally copied, but also some adjectives and verbs, and mostly English orthography was used. A number of bloggers used italics or quotation marks to mark the English words, thus demonstrating linguistic awareness.

- (1) ta on kõige õigem inimene kollektsiooni 70ndate *boho chic vibe*'i edastama  
'She is the right person to communicate the 70's *boho chic vibe* of the collection'
- (2) theBalm Mary-Lou Manizer *highlighter* on üks parimaid turul  
'theBalm Lou Manizer *highlighter* is one of the best on the market'



As mentioned above, in some cases the global copy was written in Estonian orthography.

- (3) Ja siin on ka **sniikpiik** ühest kleidist.  
‘And here is a **sneak peek** of one dress.’

There are mostly global copies that mark fashion lexis. Nowadays English has become the *lingua franca* in the fashion world; the descriptions of styles, materials, silhouettes, etc., for instance, are in English, even when the native tongue of the designers is not English.

- (4) Sügise üks märksõnu on **layering**.  
‘One of the keywords for fall is **layering**.’
- (5) soe kuid siiski sügisene **look**  
‘Warm, yet autumnal **look**’
- (6) Kas sul on mõni eeskuju/ **style icon**, kelle stiili imetled?  
‘Do you have a role model / **style icon**, whose style you admire?’

Copies from English are also used a lot when writing about beauty products. Bloggers often follow YouTube beauty bloggers and order cosmetics and products from online shops, so they are used to naming these products in English.

- (7) enne kui paned peale mis tahes meigi, tasub kasutada ka **primerit**  
‘Before you use any kind of makeup it is good to use a **primer**’
- (8) Läike kohta pigment väga head, kuid ei ole päris **liquid lipstick**.  
‘The pigment is quite good for a lip gloss, but they’re not **liquid lipsticks**.’
- (9) esmalt ostsin **stippling brush**’i, siis põsepunapintsli  
‘First I bought a **stippling brush**, then a blush brush.’

Adjectives are globally copied to express informants emotions, feelings, attitudes. In the below examples the adjectives copied have a really strong expressive connotation. In example (12) the adjective might be attractive for copying because the song the blogger is referring to is in English.

- (10) Ma ei venita enam pikemalt, aga siiski enne kui lugema hakkate siis jagan teiega ka oma viimase ajal kõige lemmikumat laulu, mis on lihtsalt **mesmerizing**.

'I'm not going to stall any longer, but before you start reading I'm going to share my favourite song with you, it's just **mesmerizing**.'

In example (11) the informant used *obsessed*, which marks powerful emotion. One of the reasons for copying might also be that there isn't an exact equivalent or construction in Estonian. To express the same idea, a different construction would be used (*see on mu kinnisidee* 'this is my obsession').

- (11) jah ma olen jätkuvalt **obsessed** mustvalgetest fotodest :D  
'Yes, I'm still **obsessed** with black and white photos :D'

The names of different shades of colours are also copied. In Estonian there are few prefixes to mark different shades of colours: mostly *hele* 'light', *tume* 'dark', *erk* 'bright' are used. In English the colour palette is much wider.

- (12) Sametist saapad (**burgungy** värvi)  
'Velvet boots (in **burgundy**)'

- (13) enne nendesse süvenemist näitan ära selle talve põhivärvid:  
**oxblood red** ja **military green**  
'Before moving on I introduce the main colours of this winter:  
**oxblood red** and **military green**.'

Verbs that are associated with certain websites are copied often. These are usually verbs that are used in blogging environments and on Facebook.

In example (14) the English verb *share* is copied. Lots of web sites offer the possibility of sharing the content of that web site with your friends through social networks, for example Facebook. So even though there is a short and semantical equivalent in Estonian informants prefer the English word. The reason for this might be that the word *share* is associated with a certain subculture and informants

are already familiar with this word; it is also a part of the semantic frame of online culture.

- (14) ja muidugi sundis see video otsima mind ülesse meie täispika video..ja **sheerin** jälle:)  
 ‘And of course this video made me look up our full-length video .. and I **share** it again :)’

The same is with the verb *follow*.

- (15) Võite mind julgelt seal [Instagramis] **follow**’ida.  
 ‘You can easily **follow** me there [in Instagram].’

In example (16) the word *like* is copied. This verb has strong associations with Facebook, so again it refers to a specific subculture. In Estonian the word-for-word rendition (*sulle peab meeldima* ‘you have to like’) is not used, so to express the same idea, a different expression and construction would be used (*„vajuta nuppu meeldib“* ‘press the like-button’). So here English *like* is used due to the lack of an Estonian equivalent construction and to achieve linguistic economy (Praakli 2009).

- (16) Pead **like**’ima LET HER SPEAKi facebooki lehte.  
 ‘You have to **like** the LET HER SPEAK Facebook page.’

Several phrasal verbs were globally copied. These are semantically specific because there isn’t an equivalent in Estonian that matches its semantics completely. Also the reason for using these phrasal verbs might be to achieve some kind of linguistic economy, because the English verb is short and fitting. This has also been noticed in Estonian-Russian (Verschik 2008: 135; 2007) and Estonian-Finnish contact situations (Praakli 2009: 149–156).

- (17) Nii, et *cozy night* saunaga **coming up** :)  
 ‘So cozy night with sauna **coming up** :)’
- (18) Kui kellelgi on mingeid soove, millist postitust ta tahab näha siis **speak up** ja teen ära selle:)  
 ‘If anybody has a wish, what kind of posts s/he wants to see, then **speak up** and I’ll do it :)’

- (19) Igastahes *help me out* sel teemal.  
 ‘Anyway, **help me out** in this.’

As stated above, discourse markers from the pragmatically dominant language are preferred. Global copies included particles, interjections, adverbs and interrogative words, but no conjunctions were copied. Igav (2013: 38–54) also reached the same results.

In examples (20) and (21) the English adverb is globally copied; the reason might be that the informant thinks that the English word emphasizes the meaning more than the Estonian word.

- (20) Mitte küll täpselt samad, **obviously**, leidsin ma need prillid ebay’st!  
 ‘Not exactly the same, **obviously**, I found these glasses on ebay!’

- (21) saan kohtuda reaalselt kõige-kõige huvitavamate persoonidega keda ma üldse suudan ette kujutada ja **ofc** külastada Eo „inkubatooreid“  
 ‘I get to meet people who are the most interesting I can imagine and **ofc** [of course] I can visit Eo incubators.’

The English interjection *god* was globally copied several times.

- (22) **God**, ma mäletan neid näägutamisi „pane kurk kinni, müts pähe, sukapüksid jalga“ haha.  
 ‘**God**, I remember nagging „wear a scarf, wear a hat, put on your tights“ haha.’
- (23) **God**, ma elan kaugel.  
 ‘**God**, I live far.’

In examples (24) and (25) *oh well* is used. The meaning of the word *well* depends on the context. According to Aijmer and Simon-Vandenberg (2006: 1123) *well* can be used to express very different emotions, such as surprise or reluctance, also politeness and hesitation. The reason for copying might also be that there isn’t an exact equivalent in Estonian and also for the pragmatic impact.

(24) **Oh well**, ütlen otse, väga halb on.

‘**Oh well**, I’ll say it straight out, it’s very bad.’

(25) unustasin isegi teile häid jõule soovida. **Oh well**...

‘I even forgot to wish you merry Christmas. **Oh well**...’

Although according to Matras (1998: 293, 311), conjunctions are also usually copied, in my data there were no such examples; Igav (2013) similarly found no examples in her data. Paljasma (2012: 49) suggests that this might depend on the intensity of language contacts. However, the use of foreign conjunctions in Estonian has been noted: in Spanish-Estonian language contact in Argentina (Jürgenson 2010: 254) and in Portuguese-Estonian language contact in Brazil (Jürisson 2010: 298).

#### 4.2. SELECTIVE COPIES

Selective copying means that only one or some, but not all properties are copied. The similarity of basic and model code elements can favour copying. According to Johanson (1993: 211) selective copying is used more when speakers are more proficient in the model code. The differences between basic and model code elements are not that salient; for example, if there are no English words present, it may seem at first glance that the construction is (wholly) from the basic code, so it might not be easy to decide that the element is from the model code.

As bloggers use English to a great extent on an everyday basis, it can be assumed that the main reason for selective copying is frequency. Bilinguals who use model code often have a good command of that language’s grammatical patterns, especially the ones that are needed frequently and so they become productive. Frequency makes foreign patterns, word order, etc. attractive for copying.

I mentioned above that the *have*-construction and the word order *something is done by someone* are used in Estonian often due to English influences. These constructions are already conventionalized in Estonian.

(26) **oman** üht 200 eurost acne salli

‘**I have** an acne scarf that cost 200 euros.’

(27) ma olen täitsa olemas, elus, Tallinnas, hea tervise korras nüüd ja superõnnelik+ **toredate inimeste poolt ümbritsetud**

‘I’m here, I’m alive, in Tallinn, in good health now and super happy + **I’m surrounded by nice people.**’

There are few examples of copying semantic and combinational properties.

In English the word *walk* has several meanings; Estonian has different words to mark types of walks. In example (28) the informant translates the meaning word-for-word whereas in Estonian a different phrasing would be used – *see on kiviviske kaugusel* ‘it’s in a stone’s throw’.

(28) veel palju parem on kohale minna, asub PTI Group firma kontorist **kõndimiskaugusel**.

‘It’s much better to go there personally, it is **within walking distance** from PTI Group’s office.’

In example (29) the English phrase structure is used. In Estonian this construction should be in the genitive: denim-Sg-GEN + white lace-Sg-GEN or simple white shirt-Sg-GEN combination-Sg-NOM, but in English these attributes are in the nominative.

(29) Ärme alustagi **teksta + valge pits või lihtne valge särk kombinatsioonist**, see töötab mu meelest alati ja imeliselt

‘Don’t even get me started with **the combination of denim + white lace or simple white shirt**, I think it always works amazingly.’

#### 4.3. MIXED COPIES

Mixed copies are copies where one part is copied globally due to semantic specificity and other parts are copied selectively, mainly combinational. Mostly these are compound nouns and fixed word

combinations. Johanson (1993: 215) suggests that mixed copies might be a transitory stage between global and selective copies.

In my data mixed copies were mostly compound nouns, although there were some adjectives, too.

(30) neid näete juba üsna pea **outfiti postitustes**

‘You will see them soon in an **outfit posts**.’

(31) ja need on mul neljandad (viierendad?) **clip-in salgud**

‘And these are my fourth (fifth?) **clip-in hair extensions**.’

(32) Õrnas **nude-roosas** toonis mahukas käekott unikaalse disaini ja ilusate detailidega.

‘A capacious handbag in **nude-pink** with unique design and beautiful details’

(33) Esiteks, mis oleks selline **pool casual**, aga võib kontsa ka alla lükata, kui äri vaja teha.

‘First, this would be **half-casual**, but for business meeting you can wear heels.’

There is an interesting mixed copy in the next example. A traditional meaning of the word *epic* is *heroic*. Nowadays it can also be used, especially in the language of young people, as a prefix, meaning ‘huge and powerful’. In this example it is also used to emphasise the meaning.

(34) See nädal oli **epic-kiire** ja seda päris mitmel põhjusel.

‘This week was **epic-busy** and that was for several reasons.’

To find out what types of copies prevail in my data, the corpus was analysed quantitatively with the program Corpus Stat. The analyse showed that global copies prevail – there were 1,772 global copies (90.5%), 148 mixed copies (7.5%) and only 39 selective copies (2%). On the average there were 500 words and 7 instances of CC in a blog post.

## 5. Discussion and conclusions

In this article English-Estonian code-copying in Estonians blogs was analysed. Blog posts were used because the language of blogs displays the process of writing in its naked, unedited form. Fifteen (15) fashion, beauty and lifestyle blogs run by Estonians were studied. The blog posts were written in 2012–2016.

For qualitative analysis of the data Johanson's code-copying model was used. This model helps us to flexibly describe and analyse context-dependant language change. This model subdivides copies into global, selective and mixed copying, depending on whether all or only certain properties of an element are copied.

This paper focused on how many and what type of copies were employed and what the reasons for code-copying (potentially) were. A quantitative analysis showed that global copies prevailed – 90.5% of the copies are copied globally, 7.5% were mixed copies and 2% were selective copies. The reason why there were so few selective copies might be that selective copies are a transitory stage between global and mixed copies. When informants are more proficient in the model code, they utilise more selective copies.

The main reason for global copying is semantic specificity or pragmatic meaning. Mostly nouns and noun phrases were copied globally, the reasons being a lack of an equivalent in Estonian, the model code element marking something unique or important to the informant or contact with new culture or field. As English is the *lingua franca* in fashion, the topic might also favour global copies.

Adjectives were copied to express strong emotions and feelings; a separate category was copying colour terminology. Verbs that were copied are mostly associated with certain websites, for example Facebook, and blogging sites.

Due to the importance at the pragmatic level discourse particles were copied; this includes particles, adverbs, interjections and interrogatives. Discourse particles were used to intensify the meaning, to



express surprise and emotions. There were no examples of copying of conjunctions (Igav 2013 had similar findings).

The main reason for selective copying is frequency – when informants use the model language often they consequently develop a good command of that language's grammatical patterns and so they start using those patterns in the basic code; selective copies appear, therefore, when the informant has used the model code for some time. In my data, there were few selective copies. Mostly *have*-construction and *something is done by someone* were copied; however, these constructions are conventionalized in Estonian and are often also used on a daily basis in other domains.

In my data there were more mixed copies than selective copies. As the mixed copies were compound nouns, where one part was copied globally, it suggests that mixed copies are a transitory stage. In mixed copies the global copying appeared due to semantic specificity and selectively combinational properties were copied.

For further conclusions, oral speech should be investigated. Based on everyday experience, English-Estonian bilingual speech is rather common; however, no research on English impact in oral speech has yet been undertaken. Although blogs are highly personalized, oral speech is more immediate and spontaneous and might also reveal interesting linguistic behaviour. Furthermore, data about the informants' proficiency in English and interviews with the informants might provide some thoughtful insight on the reasons for copying.

## REFERENCES

- Aijmer, Karin; Anne-Marie Simon-Vandenberg 2006. The discourse particle *well* and its equivalents in Swedish and Dutch. – *Linguistics* 41-6, pp. 1123–1161.
- Androutsopoulos, Jannis. 2006. Mehrsprachigkeit im deutschen Internet: Sprachwahl und Sprachwechsel. – *Ethno-Portalen*. P. Schlobinski, (Hg.), Von \*hdl\* bis \*cul8r\*. Sprache und Kommunikation in den Neuen Medien, Mannheim: Dudenverlag (= Thema Deutsch, Band 7), pp. 172–196.

- Backus, Ad 2001. The role of semantic specificity in insertional codeswitching: evidence from Dutch-Turkish. – Codeswitching Worldwide II. Ed. R. Jacobson. Berlin, New York: Mouton de Gruyter, pp. 125–154.
- Backus, Ad; Anna Verschik 2012. Copiability of (bound) morphology. – Copies versus Cognates in Bound Morphology. Volume 2. Eds. L. Johanson, M. Robbeets. Boston, Leiden: BRILL, pp. 1–32.
- Blommaert, Jan 2010. The Sociolinguistics of Globalization. Cambridge: Cambridge University Press.
- Crystal, David 2007. Language and the Internet. Cambridge, New York: Cambridge University Press.
- Ehala, Martin 2015. Sustainability of Estonian language. – Estonian Human Development Report 2014/2015. (Ed.) R. Vetik. Tallinn: SA Eesti Koostöökoda, pp. 191–198.
- Eurostat 2016. Foreign language learning statistics. [http://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign\\_language\\_learning\\_statistics](http://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign_language_learning_statistics). (11.04.2016)
- Fanshawe, Simon; Dhananjayan Sriskandarajah 2010. You can't put me in a box. – Super-diversity and the End of Identity Politics in Britain. Institute for Public Policy Research.
- Fraser, Bruce 1996. Pragmatic markers. – Pragmatics, nr 6(2), pp. 167–190.
- Igav, Reet 2013. Inglise-eesti koodikopeerimine Facebooki vestlustes [English-Estonian Code-Copying in Facebook Conversations]. Master's Thesis. Tallinn University.
- Joenuur, Elina 2016. Eesti-Läti kakskeelse isiku keelekasutus [Estonian-Latvian Bilingual Speech]. Master's Thesis. Tallinn University.
- Johanson, Lars 1993. Code-copying in immigrant Turkish. – Immigrant Languages in Europe. (Eds.) G. Extra, L. Verhoeven. Clevedon, Philadelphia, Adelaide, pp. 197–221.
- Johanson, Lars 2002. Contact-induced linguistic change in a code-copying framework. – Language Change: The Interplay of Internal, External and Extra-linguistic Factors. Eds. M. C. Jones, E. Esch. (Contributions to the Sociology of Language, 86.) Berlin: Mouton de Gruyter, pp. 285–313.
- Jürgenson, Aivar 2010. Eestlased ja eesti keel Argentiinas. – Eestlased ja eesti keel välismaal. [Estonians and the Estonian language in Argentina. – Estonians and the Estonian Language Abroad.] Eds. Kristiina Praakli ja Jüri Viikberg. Tallinn: Eesti Keele Sihtasutus, pp. 239–264.

- Jürgenson, Aivar; Jürisson, Sander 2010. Eestlased ja eesti keel Brasiilias. – Eestlased ja eesti keel välismaal. [Estonians and the Estonian language in Brazil. – Estonians and the Estonian Language Abroad.] Eds. Kristiina Praakli ja Jüri Viikberg. Tallinn: Eesti Keele Sihtasutus, pp. 285–304.
- Koutsogiannis, Dimitris; Mitsikopoulou, Bessie 2003. Greeklish and Greekness: Trends and discourses of „Glocalness“. – Computer Mediated Communication 9 (1), [http://jcmc.indiana.edu/vol9/issue1/kouts\\_mits.html](http://jcmc.indiana.edu/vol9/issue1/kouts_mits.html).
- Kruusvall, Jüri 2015. Keelteoskus ja keelte praktiline kasutamine. – Eesti ühiskonna lõimumismonitooring 2015. [Language skills and the practical use of languages. – Monitoring of Integration in Estonian Society in 2015.] In: Kristina Kallas, Raivo Vetik, Jüri Kruusvall, Ellu Saar, Jelena Helemäe, Laura Kirss, Cenely Leppik, Külliki Seppel, Kats Kivistik, Pille Ubakivi-Hadachi. Tartu: Haridus- ja Teadusministeerium.
- Kulper, Kersti 2015. Eesti keskmine blogija: 29-aastane naine, kes kirjutab toidust, moest ja ilust. [The average blogger in Estonia: a 29 year-old-woman who writes about food, fashion and beauty.] – Postimees, <http://naine24.postimees.ee/3152991/eesti-keskmine-blogija-29-aastane-naine-kes-kirjutab-toidust-moest-ja-ilust>.
- Leemets, Tiina 2002. Mullivann, peenema nimega jaccusi. [How to name a jacuzzi in Estonian.] – Oma Keel, 1, pp. 41–46.
- Leemets, Tiina 2003. Inglise laenud sajandivahetuse eesti keeles. [English borrowings in Estonian at the turn of the century.] – Keel ja Kirjandus, 8, pp. 571–584.
- Matras, Yaron 1998. Utterance modifiers and universals of grammatical borrowing. – Linguistics 36-2. Berlin: Mouton de Gruyter, pp. 281–331.
- Matras, Yaron 2009. Language Contact. Cambridge, New York: Cambridge University Press.
- MeeMa 2014. Keeleoskus ja võõrkeelte kasutamine Eestis eestlaste ja mitte-eestlaste seas. – Mina. Maailm. Meedia. TÜ Ühiskonnateaduste Instituudi sotsioloogiline uuring. [Language skills and the use of foreign languages in Estonia among Estonians and non-Estonians. – Me. The World. The Media. A Sociological Study by the Institute of Social Studies of the University of Tartu.] Tartu: Saar Poll.
- Mestlang, Helle 2006. Predikaat ajastut kogemas. [Predicate – the experienter of the era.] – Keel ja Kirjandus, 9, pp. 714–727.

- Paljasma, Veronika 2012. Prantsuse-eesti koodikopeerimine blogides [French-Estonian Code-Copying in Blogs]. Master's Thesis. Tallinn University.
- Pere, Kärt-Katrin 2014. Koodikopeerimine Pariisi eestlaste suulistes vestlustes [Code-Copying in Oral Conversations of Parisian Estonians]. Master's Thesis. Tallinn University.
- Praakli, Kristiina 2009. Esimese põlvkonna Soome eestlaste kakskeelne keelekasutus ja koodikopeerimine. [Bilingual Language Use by First Generation Estonians in Finland and Code-Copying.] Doctoral Thesis. Tartu: Tartu Ülikooli Kirjastus.
- Roosileht, Helin 2013. Inglise-eesti koodikopeerimine blogides [English-Estonian Code-Copying in Blogs]. Master's Thesis. Tallinn University.
- Soler-Carbonell, Josep 2014. University language policy and language choice among PhD graduates in Estonia: The (unbalanced) interplay between English and Estonian. – *Multilingua*, 33(3–4), pp. 413–436.
- Tammemägi, Anni; Ehala, Martin 2012. Eesti õpilaste keelehoiakud 2011. aastal. [Language attitudes of Estonian School students in 2011] – *Keel ja Kirjandus*, 4, pp. 241–260.
- Thurlow, Crispin; Lengel, Laura B.; Tomic, Alice 2004. *Computer Mediated Communication: Social Interaction and the Internet*. London: SAGE.
- Vaba, Marja 2010. Inglise-eesti koodikopeerimisest Tallinna Skype'i kontori kahe vestlusgrupi näitel [English-Estonian Code-Copying on the Example of Two Chat Groups at Skype Tallinn Office]. Master's Thesis. Tallinn University.
- Vihalemm, Triin; Siiner, Maarja 2011. Balti riikide lõimumis- ja keelepoliitika Euroopa Liidu kontekstis. – Eesti inimarengu aruanne 2010/2011. [Language and integration policies of the Baltic states in the EU. – Estonian Human Development Report 2010/2011.] Eesti Koostöö Kogu, pp. 118–121.
- Verschik, Anna 2007. Keelekontaktid, laenatavus ja verbi kopeerimine eestivene keelevariandis. [Copying of verbs in Estonia's Russian.] – *Keel ja Kirjandus*, 5, pp. 357–377.
- Verschik, Anna 2008. *Emerging Bilingual Speech: from Monolingualism to Code-copying*. London, New York: Continuum.
- Verschik, Anna 2011. Russian-Estonian code-copying in Live Journal blogs: A preliminary overview. – *Slavica Helsingiensia*, pp. 355–365.

- Verschik, Anna 2014. Estonian-Russian code-copying in Russian-language blogs: Language change and a new kind of linguistic awareness. – Negotiating Linguistic Identity. (Ed.) Vihman, Virve Anneli & Praakli, Kristiina. Oxford, Bern, Berlin, Brussels, Frankfurt am Main, New York: Peter Lang Europäischer Verlag der Wissenschaften. (Nationalism Across the Globe; 14), pp. 59–87.
- Vertovec, Steven 2007. Super-diversity and its implications. – Ethnic and Racial Studies 29(6), pp. 1024–1054.
- Wertheim, Suzanne 2003. Discourse pragmatics as a means of contact-induced change. – Selected Proceedings of the CLIC-LISO Ninth Annual Conference on Language, Interaction and Culture, pp. 1–18. <http://www.suzannewertheim.com/publications/CLIC-LISO.pdf> (06.02.2013)

## INGLISE-EESTI KOODIKOPEERIMINE BLOGIDES

Artiklis kirjeldatakse inglise-eesti koodikopeerimist eestlaste blogides. Uurimismaterjaliks on blogisissekanded, sest selle teksti on keelejuht ise toimetamata kujul kirja pannud, kajastades enda jaoks olulisi sündmusi. Kokku uuriti 15 Eestis elava eestlase blogi. Uuritud blogisissekanded on kirjutatud aastatel 2012–2016.

Koodikopeerimise uurimiseks kasutati Johansoni koodikopeerimise mudelit, selles mudelis on kõik koopialiigid omavahel tiheidalt seotud ning kopeeritud omaduste poolest eristatakse täielikke, valikulisi ning segakoopiaid. Enim esines täielikke koopiaid (90,5%), vähem segakoopiaid (7,5%) ja valikulisi koopiaid (2%).

Täieliku kopeerimise põhjus on enamasti mudelkoodi elemendi semantiline spetsiifilisus või pragmaatiline olulisus, samuti keeleühiku kultuuriline spetsiifilisus ning keelelise lihtsuse ja ökonoomia saavutamine. Täielikult kopeeriti enamasti nimisõnu ja nimisõna fraase, vähem ka omadus- ja tegusõnu. Nende sõnade semantiline spetsiifilisus seisneb näiteks selles, et põhikoodis ei leidu vastet või ei ole vaste piisavalt täpne. Näiteks kopeerivad blogijad täielikult värvuste ja riietusesemetega seotud ning verbe, mis üldjuhul seonduvad mingite suhtlusvõrgustike või blogikeskkonnaga.

Pragmaatilise olulisuse tõttu kopeeriti diskursuspartikleid, mille roll on enamasti kommunikatsiooni suunamine. Käesoleva töö jaoks koostatud korpuses esines nii partikleid, adverbe, interjektsioone kui ka küsisõnu. Korpuses ei esinenud näiteid konjunktsioonide kopeerimise kohta, kuid ei ole võimalik kindalt öelda, miks neid ei kopeeritud. Samade tulemusteni jõudis oma uurimuses ka Reet Igav (2013: 37). Diskursuspartiklite kopeerimise eesmärk oli suhtlemist tõhusamaks muuta.

Valikulise kopeerimise peamine põhjus on keeleühiku kasutusagedus mudelkoodis. Uurimuses esines valikulisi koopiaid vähe.

Valikuliste koopiate seas esines palju kombinatoorsete ja sageduslike omaduste kopeerimist, enim *poolt*-konstruktsiooni ja sõna „omama“ kopeerimist, mis on aga eesti ühiskeeles juba konventionaliseerunud. Valikuliste koopiate vähesest esinemissagedusest võib järeldada, et keelejuhid ei kasuta mudelkoodi sellisel tasemel, et kasutaksid selle konstruktsioone ja malli produktiivselt ka põhikoodis.

Korpuses esines ka mitu segakoopiat, mida on peetud täielikelt koopiatelt valikulistele üleminekuperioodiks. Seda kinnitas ka koopialiikide esinemisstatistika – segakoopiaid esines veidi rohkem kui valikulisi koopiaid. Korpuses esinenud segakoopiaid on liitnimisõnad, mille üks osa on kopeeritud täielikult ning ülejäänud valikuliselt. Segakoopiate esinemise põhjuseid on raske kindlaks määrata, üks põhjus võib olla täielikult kopeeritud keeleühiku semantiline spetsiifilisus.

# LOAN TRANSLATIONS AS A LANGUAGE CONTACT PHENOMENON: CROSSING THE BOUNDARIES BETWEEN CONTACT LINGUISTICS, SECOND LANGUAGE ACQUISITION RESEARCH AND TRANSLATION STUDIES

Lea Meriläinen, Helka Riionheimo,  
Päivi Kuusi and Hanna Lantto  
University of Eastern Finland

**Abstract.** This article provides a review of loan translations as a language contact phenomenon from the perspectives of contact linguistics, second language acquisition (SLA) research and translation studies (TS). We discuss both similarities and differences in the ways in which loan translations are conceptualized across these three disciplines. The discussion highlights a common cognitive basis underlying bilingual language use, SLA and translation, while at the same time the prevailing attitudes to loan translations in these disciplines reveal differing underlying ideologies. This study is a contribution towards broadening the scope of language contact studies to cover related disciplines that examine similar phenomena.<sup>1</sup>

**Keywords:** loan translation, calque, cross-linguistic influence, language contact, translation, second language learning, bilingual processing

---

<sup>1</sup> This research was supported by the research projects *Kontu* (financed by the Emil Aaltonen Foundation), *Kiännä* (financed by the Kone Foundation) and the University of Eastern Finland. We would like to thank two anonymous reviewers for their valuable comments.



## I. Introduction

Loan translations (or calques) are a well-recognized phenomenon in contact linguistic literature. They are generally defined as „words or phrases that are reproduced as literal translations from one language into another“ (Backus, Dorleijn 2009: 75). Examples of loan translations abound in earlier research on language contact effects in bilingual communities, but the process of loan translation may also occur in any context where an individual needs to learn or use another language, such as in classroom-based acquisition of a foreign language and translation. The starting point for the present article is the authors' observation that while loan translations have been studied within the fields of contact linguistics, second language acquisition (SLA) research and translation studies (TS), their findings and theoretical approaches have not yet been systematically brought together. Earlier contact linguistic studies have primarily treated contact-induced features, including loan translations, as a community-level phenomenon, although they essentially originate from the cognitive processes of a bilingual individual (see Weinreich 1974 [1953]: 1, Matras 2009: 3, 5). It is therefore well justified to expand the scope of language contact studies to cover related disciplines that examine language contact effects at the level of an individual, in this case, SLA and TS.

This article provides a review of loan translations as occurring in bilingual language use, SLA and translation. We discuss both similarities and differences emerging from earlier literature on loan translations in these three contexts of language contact. The aim of this article is to enhance the emerging line of dialogue between contact linguistics, SLA and TS, which have long remained separate paradigms (in line with Paulasto *et al.* eds. 2014 and Riionheimo *et al.* eds. 2014). Due to space limitations, this article does not attempt to provide a conclusive summary of the phenomenon; rather it focuses on those aspects of loan translations that connect these three disciplines.

The outline of the article is as follows: Sections 2–4 discuss the notion of loan translations in the fields of contact linguistics, SLA and TS<sup>2</sup>, covering historical perspectives and developments, cognitive dimensions and the emerging ideological differences. Each of these sections provides a somewhat different viewpoint of the phenomenon, which reflects differing terminological conventions and underlying ideologies across these three disciplines. Section 5 draws these perspectives together and discusses the benefits of bringing together different disciplines in the study of language contact effects.

## 2. Loan translations in contact linguistics

Loan translation is an old term originating from the historical linguistics of the early 1900s (see e.g. Weinreich 1974 [1953]: 48–49 and the references therein). The phenomenon is defined as „ad hoc word-for-word or morpheme-per-morpheme translations from one language into another“ (Aikhenvald 2006: 24) or „a type of interference in which word or sentence structure is transferred without actual morphemes“ (Thomason 2001: 260). Two aspects are central here: 1) what is transferred from the model language to the recipient language is the semantic content, not the actual phonological form and 2) the process involves translating, i.e., replacing the words or morphemes of the model language with their equivalents in the recipient language. Loan translation is thus viewed as a covert form of cross-linguistic influence in contrast to overt borrowing of phonological substance. Traditionally, loan translations were described

---

<sup>2</sup> A note should be made on the differing use of terminology across these disciplines to refer to the languages in contact. Contact linguistics employs a variety of terms, such as *donor* or *model language* (the language giving the influence) and *borrowing* or *recipient* or *replica language* (the language receiving the influence); here we adopt the terms *model* and *recipient language*. SLA generally uses the terms *first language (L1)* to refer to the learner’s mother tongue and *second/foreign language (L2)* (also *target language*) for the language the learner is attempting to learn. In TS, the terms *source language* and *target language* are generally used to refer to the language that is being translated from and the language being translated to respectively.

as a type of lexical borrowing (e.g., in the seminal works of Haugen 1972 [1950] and Weinreich 1974 [1953]). The most prototypical examples of loan translations found in earlier literature are compound nouns, but it has been recognized that loan translations may also involve idioms, other phrasal expressions or longer fixed units such as proverbs (e.g. Weinreich 1974 [1953]: 50).

Towards the end of the 20<sup>th</sup> century, the focus of contact linguistics shifted from narrowly defined (primarily lexical) phenomena onto code-switching and other broader manifestations of cross-linguistic influence. As a consequence, loan translations became largely neglected in contact linguistic discussion (for an overview, see Backus, Dorleijn 2009). Along with the increased interest of language contact research in the structural or grammatical outcomes of contact, however, loan translations have re-emerged under a new disguise, and the notion has been expanded to cover instances of structural influence. The phenomenon of loan translation is embedded in some recent, comprehensive models capturing language contact effects. For example, in the code-copying framework (Johanson 1998, 2002, Verschik 2008), loan translations are placed among code-switching and morphosyntactic influence. The model distinguishes between two basic types of cross-linguistic influence: *global copying*, where material (loanwords and insertional code-switches) is transferred from the model language to the recipient language, and *selective copying*, where the model language expression is copied only partially. There are two subtypes of selective copying that involve phenomena traditionally labelled as loan translations: *semantic copying* (i.e., copying the semantic features of a model code; Verschik 2008: 63) and *combinational copying* (i.e., replication of a word combination, such as a phrasal structure, word-internal morphemic pattern or a morpho-syntactic combination; Johanson 2002: 15). Of special interest to this article is the connection between the code-copying framework and translating; Verschik (2008: 113–114) describes translation as „selective copying par excellence“, which

may function as a pathway for semantic copying between the contacting languages.

In another model, Matras and Sakel (2007; see also Matras 2009: 240–243) capture a language processing mechanism which they term *pattern replication*, i.e., a form of cross-linguistic influence in which elements of the recipient language are rearranged in line with the semantic and grammatical patterns of the model language. Pattern replication covers all forms of semantic and structural influence, including loan translation. Matras and Sakel propose the mechanism of *pivot-matching*:

We suggest that it involves identifying a structure that plays a pivotal role in the model construction, and matching it with a structure in the replica language, to which a similar, pivotal role is assigned in a new, replica construction. [...] The replica construction evolves around the new pivot in a way that generally respects various constraints of the replica language. (Matras, Sakel 2007: 830.)

The pivot-matching model allows for subjective and creative decisions by bilingual speakers in the process of selecting and combining the elements of the replica structure, which accounts for the fact that there is not always a neat one-to-one-correspondence between the model construction and the replica construction. This model thus stretches the notion of loan translation to cover instances which are not literal translations of the model structure.

The most detailed approach yet taken to loan translations is provided by Backus and Dorleijn (2009, see also Backus 2010), who discuss the difficulty of defining loan translations and teasing them apart from other contact-induced phenomena such as code-switching, lexical borrowing, semantic extensions and structural or grammatical borrowing. They view loan translation as a synchronic process closely related to both structural interference and (insertional) code-switching. They argue that the process underlying loan translation and structural or grammatical interference is the same

although the items involved may be different; in the case of content words, it is common to use the term *loan translation* whereas non-content words are typically classified under *interference* (Backus 2010: 235–236). Backus and Dorleijn (2009) propose a typology of loan translations based on the specificity of the meaning of the linguistic elements involved: 1) content morphemes (one-word semantic extensions, prototypical two-word expressions and multi-word translations such as phrases), 2) function morphemes, 3) grammatical morphemes and 4) discourse patterns. Thus, Backus and Dorleijn (2009) quite radically expand the boundaries of loan translation as a contact mechanism. Loan translation and code-switching are viewed as related phenomena because they are both motivated by the wish of a bilingual speaker „to say something in a base language in the way it is said in the other language“ (Backus, Dorleijn 2009: 90), the difference being that in code-switching the speaker uses overt ingredients from the other language and in loan translation the expression is translated.

The above discussed studies address the process by which loan translations emerge in bilingual speech, and highlight multilingual speakers' creative and innovative potential. The role of the individual is significant in two respects. First, many researchers emphasise that the trigger for loan translation is „bilingual speakers' need to express the same thoughts in two languages“ (Sasse 1990: 32, according to Ross 2007: 132; see also Backus, Dorleijn 2009: 90; Backus 2010: 239; Matras, Sakel 2007: 832; Matras 2009: 240–241). Multilingual speakers are seen as creative actors who use their full linguistic repertoire, and loan translations are one manifestation of their innovative agency. Secondly, the process of loan translation is in practice possible only if there are semantically corresponding words or forms in the model language and the replica language (see e.g., Matras, Sakel 2007: 234; Backus, Dorleijn 2009: 90). This correspondence is referred to as an *equivalence*, *translation equivalence* or *interlingual identification* (the latter term originating from Weinreich 1974 [1953]), and the mental or cognitive connections of

bilingual speakers are viewed as subjective and based on perceived equivalents (Johanson 2002: 57, Backus & Dorleijn 2009: 90). This interest in the mental connections and subjective perceptions of the bilingual individual is also shared by SLA research.

### 3. Loan translations in second language acquisition research

In the field of SLA research, loan translations are treated as a form of lexicosemantic transfer<sup>3</sup> from the learners' L1 into the L2 (e.g., Ringbom 1987, 2007; James 1998; Jarvis 2009). The notion of loan translation has been taken over from contact linguistic literature (e.g., Weinreich 1974 [1953]), and it refers to instances where „semantic properties of one item [are] transferred in a combination of lexical items“ (Ringbom 1987: 117). Unlike in language contact literature, where such forms are considered lexical innovations or loan words, in the context of (often classroom based) foreign language acquisition such forms are viewed as errors that break the conventions of the target language. Several works have indeed examined loan translations as a category of lexical errors, along with, e.g., semantic extensions (i.e., extension of meaning based on the semantic range of an L1 word), substitutions (or borrowings; i.e., an L1 word is used in L2 in an unmodified form) and relexifications (or coinages; i.e., L1 word is tailored to the structure of the L2) (e.g., Ringbom 1987, 2007, James 1998, Meriläinen 2010). The following discussion focuses on two recent works, Ringbom (2007) and Jarvis (2009), which are relevant to this article in the sense that they address the nature of transfer underlying loan translations. As SLA research views transfer as an individual-level cognitive phenomenon, these works shed light on the cognitive processes and psycholinguistic variables that explain loan translations as well as other types of lexical non-target forms.

Ringbom (2007) examines the effect of cross-linguistic similarity in foreign language learning through evidence from

---

<sup>3</sup> The term *transfer* is used interchangeably with *cross-linguistic influence*.

Finnish-speaking and Swedish-speaking learners of English in Finland. His work incorporates loan translations as one category of lexical errors. Ringbom (2007) brings up two important points about loan translations: 1) the type of transfer that underlies them and 2) how they are affected by cross-linguistic similarity. According to Ringbom (2007: 54–58), loan translations are a manifestation of *system transfer*, i.e., the transfer of abstract principles of organizing information, as opposed to *item transfer*, in which learners establish simplified one-to-one relationships between L1 and L2 items (e.g., words, sounds, letters, morphemes). While item transfer requires some degree of formal similarity between the languages, system transfer does not; learners tend to assume that abstract systems (e.g., semantics, pragmatics) are similar in different languages (*perceived vs. assumed similarity*, Ringbom 2007: 24–26). What is transferred in the case of loan translations are abstract lexical procedures. Based on extensive evidence from L1 Finnish and L1 Swedish learner populations, Ringbom (1987, 2007) has shown that lexical transfer generally manifests itself in the form of loan translations and semantic extensions when learners' L1 and L2 are typologically distant (Finnish and English), while in the case of related languages (Swedish and English) learners' errors most often involve words that are formally similar but have different meanings or functions in L1 and L2 (system transfer vs. item transfer; for examples, the reader is referred to Ringbom 2007). The source for loan translations is usually the learners' L1 because system transfer requires native-like or very advanced proficiency in the source language (Ringbom 2007: 86–87). Item transfer, on the other hand, may take place from any language that the learner has perceived to be (formally) similar to the target language (e.g., from L2 to L3).

Jarvis (2009) provides a comprehensive account of different types of lexical transfer in the light of current thinking concerning the structure of the bilingual mental lexicon. He differentiates between *lexemic* and *lemmatic* transfer; the former refers to phonological and graphemic L1 influence while the latter encompasses

semantic and syntactic properties of words. This division is based on two distinct levels of lexical entry in the mental lexicon; lexemes (i.e., form-related properties) and lemmas (i.e., semantic and syntactic information) (see Jarvis 2009: 100–102). Lexemic transfer may manifest itself in the form of borrowings, coinages and deceptive cognates, while lemmatic transfer results in semantic extensions, loan translations, collocational transfer and subcategorization transfer (i.e., the choice of a complement to accompany a particular headword). Lemmatic transfer thus covers semantic, collocational, morphological and syntactic constraints on words, which cannot always be strictly separated from one another. As discussed in Jarvis (2009: 114–115), loan translations may involve simple compound words as well as more complex constructions and collocational constraints. This is evident in the phrase *spend cat's days* (lead an easy life), where a Finnish learner has transferred a Finnish idiom (*cat's days*) as well as a collocating word (*spend*) into English (Meriläinen 2010: 125). Loan translations are in essence similar to transferred idioms and other types of transferred fixed expressions in that „what is transferred is a blueprint for organizing multiple forms (words and morphemes) together in specific orders and within specific syntactic constructions in order to allow them to convey a specific intended message“ (Jarvis 2009: 115). This resembles the central idea presented in Backus and Dorleijn (2009) and Backus (2010) according to which loan translations involve not just separate lexical items, but any units that convey a particular meaning. These units may cover lexical items as well as grammatical elements, which are not strictly separated in the L2 speaker's mind (cf. Jarvis 2009).

From the review of earlier literature, it becomes evident that loan translations do not receive as much attention amongst the SLA community as they used to, largely because they tend to be equated with learner errors. Due to a shift of emphasis from error analysis to a more holistic analysis of learner language (see, e.g., Ellis, Barkhuizen 2005), and from grammar-translation oriented teaching methods to communicativeness (see, e.g., Richards, Rogers 2001), errors do



not occupy a very central position in current thinking on language learning and teaching. However, as a natural product of bilingual language use, loan translations deserve more attention in the SLA context. The recent cognitive orientation of SLA research has a great deal to contribute to the study of loan translations in different types of language contact settings. Furthermore, contact linguistic works emphasising the creative mixing of bilingual resources, including loan translating (cf. Matras, Sakel 2007, Matras 2009) resonate with the plurilingual ideology that has recently emerged in the discussion on language education policies and the goals of language teaching. This is evident, for example, in the language education policy guidelines by the Council of Europe (2007), which encourage the acquisition of different languages/varieties to differing degrees, and the flexible use of different communicative resources in different situations, including simultaneous use of different languages/varieties (i.e., code-switching). Instead of errors, loan translations could be approached as an effective foreign language production strategy, which can be employed in certain communicative contexts along with code-switching. The practical application and wider acceptance of these ideas at the classroom level is yet to be seen, but it appears evident that we are moving away from a monolingual norm and from viewing native speaker competence as the implicit target of foreign language learning. This might, sooner or later, change the way loan translations are perceived in the context of SLA.

#### 4. Loan translation in Translation Studies

Perhaps surprisingly, loan translation as an object of research occupies a marginal position in Translation Studies<sup>4</sup>. In most studies on the subject, loan translation is not approached on its own but as one of the local (small-scale) translation strategies, typically used by

---

<sup>4</sup> In the present article, the discipline of Translation Studies is taken to include Interpreting Studies as well.

professional translators for solving a particular type of translation problem such as realia (culture-bound words), allusions, or terminology (see, for example, Vlahov and Florin 2009 [1980], Leppihalme 1997, Šarčević 1985). These phenomena pose problems for translators because they often lack natural equivalents – equivalents that exist in the target language prior to translating (on natural equivalence, see Pym 2010: 12ff.). Consequently, loan translation is perceived as a lexical phenomenon that produces neologisms in the target language such as *kick sled* in English for the Finnish realia *potkukelkka* (Leppihalme 2011: 129). The need for a neologism serves as a justification for resorting to loan translation: it is considered an accepted translation strategy when there is a lexical gap in the target language (Pym 2010: 14). This means, basically, that what is expected from a professional translator is information retrieval – finding a pre-existing equivalent term – rather than lexical innovation in the form of loan translation. Cabré (2010: 364), for example, suggests limiting the use of neologisms to cases where „all the possibilities of finding a *real* term have been exhausted“ (emphasis original).

In a classic treatment by Vinay and Darbelnet (1995 [1958]), translation strategies (or procedures, as Vinay and Darbelnet call them) form a continuum from the most literal to the most free or creative, and loan translation is located at the literal end of this continuum, together with loan words and literal translation (Vinay, Darbelnet 1995 [1958]: 85–86; Pym 2010: 13). The difference between calques and literal translation is not explicitly discussed, but it seems to be one of degree and not of kind: calques are translations of separate elements, whether lexical or structural, whereas literal translation concerns whole sentences (Vinay, Darbelnet 1995 [1958]: 85–88). A similar distinction is displayed in discussions on loan translation and interference. Loan translation is defined, for example, by Leppihalme (2011: 129) as „a word-for-word translation resulting in a target-language neologism“, while interference is described in Lamberger-Felber and Schneider (2008: 217) as „a projection of characteristics of the source text into the target text“. The reference to

neologisms suggests a lexical approach, while „characteristics of the source text“ point to a somewhat broader phenomenon. The distinction, however, is not clear-cut. In studies on interference, the subcategory of lexical interference is often discussed (see, for example, Toury 1985, Franco Aixelá 2009), but mostly with no reference to its relationship with loan translation.

The overlap between the notions of loan translation and interference probably results from the fact that in actual translations, lexical and syntactic source-language influence frequently co-occur. For example, Musacchio (2005) reports on economic articles translated from English into Italian, where the influence of English is present not only in lexical borrowings, but involves loan translation, compound term formation, phraseology, syntactic constructs and the use of cohesive links. Similarly, in Interpreting Studies, Lamberger-Felber and Schneider (2008) find evidence of both lexical and morphosyntactic interference in simultaneous interpreting. The terminological overlap is enhanced by the fact that the terms *calquing* and *interference* have both been used to designate similar phenomena on the level of discourse. Toury (1985: 8) discusses the possibility of interference on the discourse level, pointing out that a translation with no interference on lexical and syntactic levels can still mirror the textual model of the source text. For instance, a translation of a cooking recipe may, on the whole, mirror the source-language rather than the target-language organisation for cooking recipes. Bennett (2011) refers to a similar phenomenon as *calquing* on the level of discourse, claiming that in many non-English-speaking countries scientific discourse does not result from evolution of the domestic discourse system, but has been calqued from the English model (Bennet 2011: 190).

Another distinction between loan translation and interference concerns their intentionality. Loan translations are generally considered deliberate choices on the part of the translator, whereas interference is often perceived as an unintended source-language influence on translation (Chesterman 1997: 94). Typically, this

influence is deemed undesirable, because it is likely to make the translation sound unnatural and compromise fluency (for a short overview of this negative view, see, for example, Lamberger-Felber, Schneider 2008: 217). However, interference can also result from a translator's conscious choice to opt for a literal translation strategy<sup>5</sup>. Here, the notion of strategy refers to the translator's overall (macro-level) approach to translation, which may be either form-based (literal translation) or meaning-based (free translation)<sup>6</sup>. Form-based translation involves keeping close to both the lexical and the syntactic features of the source text; Maier (2011: 76) describes it as a tendency to „follow the words and the surface syntax of the source text more closely than when working under a meaning-based approach“.

Form-based approach to translation is often thought typical of untrained translators or interpreters ('regular' multilinguals) rather than professionals (see Maier 2011: 76–78). However, there is evidence that both novice and expert translators actually resort to form-based, word-for-word translation (Tirkkonen-Condit 2005; Englund-Dimitrova 2005). The difference lies in the fact that professional translators monitor their own performance in order to detect renderings that are not acceptable or appropriate and find other solutions (Tirkkonen-Condit 2005: 407–408), and this kind of monitoring and control is considered an essential part of their competence (PACTE 2003: 58). For a professional translator, literal translation is an intermediary step that helps to distribute cognitive effort during the translation process (Englund-Dimitrova 2005: 234) and a carefully considered strategic choice when employed in the final output

---

<sup>5</sup> As a manifestation of a deliberate strategy of literal, foreignising translation, interference has its ardent advocates as well, the most prominent in recent years being Lawrence Venuti (1998).

<sup>6</sup> The dichotomy has been alternately referred to as domesticating vs. foreignising translation, source-language orientated vs. target-language orientated translation, overt vs. covert translation or formal vs. dynamic equivalence, to name just a few. For a more thorough discussion, see Chesterman (1997: 9ff.).

(Maier 2011: 78). For non-professional translators, it seems to be the default tendency (*ibid.*). In simultaneous interpreting, interference is a frequent phenomenon even in the output of trained professionals (Lamberger-Felber, Schneider 2008: 232), but this only corroborates the role of monitoring and control in avoiding interference: due to extreme time pressure, there are less possibilities for corrections in interpreting, which makes it especially liable to interference (Lauterbach, Pöchhacker 2015: 194).

This control and monitoring of one's own work is in line with Toury's (2012 [1995]: 313) laws of translational behaviour which suggest that accomplished translators are less prone to interference. However, apart from these cognitive factors, Toury's laws also take into account the whole socio-cultural situation surrounding them. Specifically, interference is predicted to be more common (and better tolerated) in translations made from a dominant, prestigious language into a language that is, in some sense, weaker than the source language (Toury 2012 [1995]: 314). This helps to explain why professional translators are not free from a tendency to calque, despite having been trained to use loan translations sparingly. The socio-cultural tolerance of visible influence from a prestigious language makes loan translation a more accepted strategy in both professional and non-professional translation.

All in all, the attitude towards both loan translation and interference in TS has been largely negative. However, while interference is mostly seen as the translator's failure to keep the source-language and the target-language codes apart, loan translation is generally accepted as a strategic choice in the case of a lexical gap in the target language. For example, in legal translation, translators are often required to maintain formal equivalence, which makes loan translation an appealing option (Harvey 2002: 180). However, even in legal translation, the attitudes have been changing in recent years, with fidelity being re-defined as „achieving an equivalent impact on the target reader“ (*ibid.*). It is precisely this need to get the message

through to ordinary (lay) readers in legal translation that renders loan translation problematic (Šarčević 1985: 130).

In sum, the arguments presented against loan translation in TS are plenty. In both general and special field translation, the risk of producing *faux amis* (false cognates) is often mentioned in relation to loan translation (Vinay, Darbelnet 1995 [1958]: 85, Šarčević 1985: 129). In translating allusions in literary texts, Leppihalme (1997) found loan translation (referred to as *minimum change*) to be a common translation strategy, even though in the case of allusions a word-for-word translation can hardly be expected to get the full meaning across. Loan translation results in „a literal translation, without regard to connotative or contextual meaning“ (Leppihalme 1997: 84), thus leaving something essential untranslated. Perhaps the most striking example of the negative attitude towards loan translation is found in Bennett (2011: 195), who refers to the calquing of English-language model of scientific discourse as „cultural colonization“ and explains it with reference to unequal power relations between cultures (ibid. 199). Bennett’s argument brings us back to Toury’s law of interference. A form-based approach to translation, including loan translation, is not only a matter of cognitive act: in translations, tolerance of interference is also socio-culturally conditioned (Toury 2012 [1995]: 311).

## 5. Discussion

This article has attempted to demonstrate that loan translations provide a fertile meeting ground for cross-disciplinary dialogue between contact linguistics, SLA research and TS, and that they are worthy of further empirical research. Based on the preceding review of the literature, it appears evident that loan translations, whether produced by bilingual/multilingual speakers, foreign language learners or translators, are in essence a similar phenomenon with a similar cognitive basis, notwithstanding the considerable dissimilarities between these different groups of language users. Regardless

of the context in which they occur, loan translations are a manifestation of individual-level cross-linguistic influence and have a similar underlying motivation; as suggested by Backus and Dorleijn (2009: 90), „the basis of any loan translation is an urge that a bilingual feels, consciously or not, to say something in a base language in a way that it is said in the other language“. The idea that loan translations extend beyond lexical units into morphological, syntactic or discourse domains receives support across all these three disciplines, which further highlights their common origin in the bilingual cognitive network, where different levels of language are intertwined (cf. the notion of lemmatic transfer by Jarvis 2009).

A major difference between these disciplines lies, however, in the ways in which the innovations resulting from the process of loan translation are perceived amongst scholars. Contact linguistic literature treats loan translations in a relatively neutral manner as one type of linguistic innovation in bilingual communities. In the context of second language learning and teaching, where native-like foreign language competence has long been viewed as an implicit goal, loan translations are considered errors that break target language conventions. As the goals set for professional translator training are even higher, the ability to suppress and control for effects of cross-linguistic influence, including loan translation, is considered an integral part of a professional translator's competence. Formal education, metalinguistic awareness and differing underlying ideologies regarding adequate linguistic competence thus emerge as differentiating factors between these fields.

Another perspective that we find relevant in the study of loan translations is that of TS. The term loan translation in itself implies that their use involves translation from one language into another. This is also highlighted in Backus (2010: 239), who states that „the mechanism suggested to underlie all contact-induced change in which the source of the change is cross-linguistic influence [...] is translation“. However, it remains unclear what exactly 'translation' means in this context. Translation as a mechanism for language

contact has not been given much prominence to date in contact linguistics literature (see, e.g., Kranich *et al.* 2011, Kolehmainen 2013), although it may be a more pervasive phenomenon than previously assumed (Kolehmainen *et al.* 2015). This is where TS has a contribution to make, especially research into non-professional translation (for an overview, see Antonini 2011); translational actions performed by untrained bilingual individuals may be influential in that they provide a channel through which innovations spread from one language or language variety into another. Loan translations are an example of this; although their exact source is often difficult to identify, it is likely that they are often introduced into the language not only by professional translators (whether as strategic choices or as occasional failures to keep the codes apart), but also by ordinary language users who are familiar with another language.

All in all, through this discussion we hope to have demonstrated that loan translations are anything but a marginal phenomenon. We therefore agree with Backus and Dorleijn (2009: 76) who state that the real frequency of loan translations depends on how you define them. The combined perspectives from contact linguistics, SLA and TS support the idea that loan translations form an interlocking system with other language contact effects, which deserve to be examined as a whole and with evidence from neighbouring disciplines.

#### REFERENCES

- Aikhenvald, Alexandra Y. 2006. Grammars in contact: A cross-linguistic perspective. – Grammars in contact: A cross-linguistic typology. Eds Alexandra Y. Aikhenvald, R. M. W. Dixon. Oxford: Oxford University Press, 1–66.
- Antonini, Rachele 2011. Natural translator and interpreter. – Handbook of Translation Studies, Volume 2. Eds Yves Gambier, Luc van Doorslaer. Philadelphia, PA, USA: John Benjamins, 102–104.
- Backus, Ad; Margreet Dorleijn 2009. Loan translations versus code-switching. – The Cambridge Handbook of Linguistic Code-switching. Eds



- Barbara E. Bullock, Almeida Jacqueline Toribio. Cambridge: Cambridge University Press, 75–93.
- Backus, Ad 2010. The role of codeswitching, loan translation and interference in the emergence of an immigrant variety of Turkish. – *Working Papers in Corpus-based Linguistics and Language Education* 5, 225–241.
- Bennett, Karen 2011. The scientific revolution and its repercussions on the translation of technical discourse. – *The Translator* 17:2, 189–210.
- Cabré, Maria Teresa 2010. Terminology and translation. – *Handbook of Translation Studies*, Vol. 1. Eds Yves Gambier, Luc van Doorslaer. Amsterdam: John Benjamins, 356–365.
- Chesterman, Andrew 1997. *Memes of Translation. The Spread of Ideas in Translation Theory*. Amsterdam: John Benjamins.
- Council of Europe 2007. *Guide for the Development of Language Education Policies in Europe: From Linguistic Diversity to Plurilingual Education*. Available at: [www.coe.int/lang](http://www.coe.int/lang) (accessed May 2016)
- Ellis, Rod; Gary Barkhuizen 2005. *Analysing Learner Language*. Oxford: Oxford University Press.
- Englund-Dimitrova, Birgitta 2005. *Expertise and Explication in the Translation Process*. Amsterdam: John Benjamins.
- Franco Aixelá, Javier 2009. An overview of interference in scientific and technical translation. – *The Journal of Specialised Translation* 11, 75–88.
- Harvey, Malcolm 2002. What's so special about legal translation? – *Meta: journal des traducteurs / Meta: Translators' Journal* 47:2, 177–185.
- Haugen, Einar 1972 [1950]. The analysis of linguistic borrowing. – *Studies by Einar Haugen. Presented on the Occasion of his 65th Birthday, April 19, 1971*. Ed. Evelyn S. Firchow. Tübingen: Walter de Gruyter, 161–185.
- James, Carl 1998. *Errors in Language Learning and Use*. New York: Longman.
- Jarvis, Scott 2009. Lexical transfer. – *The Bilingual Mental Lexicon. Interdisciplinary Approaches*. Ed. by Anita Pavlenko. Bristol: Multilingual Matters.
- Johanson, Lars 1998. Frame-changing code-copying in immigrant varieties. – *Bilingualism and Migration*. Eds Guus Extra, Ludo Verhoeven. London and New York, Routledge, 247–260.
- Johanson, Lars 2002. *Structural Factors in Turkic Language Contacts*. Richmond: Curzon.

- Kolehmainen, Leena 2013. Käännöstiede ja kontaktilingvistiikka kohtaavat [Translation studies and contact linguistics meet]. – Kielten vertailun metodiikka [The methodology of language comparison]. Eds Leena Kolehmainen, Matti Miestamo, Taru Nordlund. Helsinki: Finnish Literature Society, 420–460.
- Kolehmainen, Leena; Kaisa Koskinen; Helka Riionheimo 2015. Arjen näkymätön kääntäminen. Translatorisen toiminnan jatkumot. [Invisible everyday translation: the continuums of translatorial action]. – *Virittäjä* 3/2015, 372–400.
- Kranich, Svenja; Viktor Becher; Steffen Höder 2011. A tentative typology of translation-induced language change. – *Multilingual Discourse Production. Diachronic and Synchronic Perspectives*. Eds Svenja Kranich, Viktor Becher, Steffen Höder, Juliane House. Amsterdam: John Benjamins, 11–43.
- Lamberger-Felber, Heike; Julia Schneider 2008. Linguistic interference in simultaneous interpreting with text: A case study. – *Efforts and Models in Interpreting and Translation Research. A Tribute to Daniel Gile*. Eds Gyde Hansen, Andrew Chesterman, Heidrun Gerzymisch-Arbogast. Amsterdam: John Benjamins, 215–236.
- Lauterbach, Eike; Franz Pöchhacker 2015. Interference. – *Routledge Encyclopaedia of Interpreting Studies*. Ed. Franz Pöchhacker. New York: Routledge, 194–195.
- Leppihalme, Ritva 1997. Culture Bumps. An Empirical Approach to the Translation of Allusions. *Topics in Translation* 10. Clevedon: Multilingual Matters.
- Leppihalme, Ritva 2011. Realia. – *Handbook of Translation Studies*. Eds Luc van Doorslaer, Yves Gambier. Amsterdam: John Benjamins, 126–130.
- Maier, Robert M. 2011. Towards a psycholinguistic model of translation processes: Directionality in natural translation. – *Interpreting Naturally: A Tribute to Brian Harris*. Eds María Jesús Blasco Mayor, María Amparo Jimenez Ivars. Bern: Peter Lang, 67–102.
- Matras, Yaron 2009. *Language Contact*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Matras, Yaron; Jeanette Sakel 2007. Investigating the mechanisms of pattern replication in language convergence. – *Studies in Language* 31:4, 829–865.

- Meriläinen, Lea 2010. Language transfer in the written English of Finnish students. Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology, no 9. Joensuu: University of Eastern Finland. Available at: [http://epublications.uef.fi/pub/urn\\_isbn\\_978-952-61-0230-6/](http://epublications.uef.fi/pub/urn_isbn_978-952-61-0230-6/)
- Musacchio, Maria Teresa 2005. The influence of English on Italian: The case of translations of economic articles. – In and Out of English: For Better, for Worse. Eds Gunilla M. Anderman, Margaret Rogers. Clevedon: Multilingual Matters, 71–96.
- PACTE 2003. Building a translation competence model. – Triangulating Translation. Perspectives in Process-Oriented Research. Ed. Fabio Alves. Amsterdam: John Benjamins, 43–66.
- Paulasto, Heli; Lea Meriläinen; Helka Riionheimo; Maria Kok (eds) 2014. Language Contacts at the Crossroads of Disciplines. Newcastle-upon-Tyne: Cambridge Scholars Publishing.
- Pym, Anthony 2010. Exploring Translation Theories. London: Routledge.
- Richards, Jack C.; Theodore S. Rodgers 2001. Approaches and Methods in Language Teaching: A Description and Analysis. 2<sup>nd</sup> Edition. Cambridge: Cambridge University Press.
- Riionheimo, Helka; Minna Kumpulainen; Lea Meriläinen; Pirkko Muikku-Werner (eds) 2014. Virittäjä 3/2014. Special issue: Kielikontaktit tutkimussuuntausten risteyksessä [Language contacts at the crossroads of disciplines].
- Ringbom, Håkan 1987. The Role of the First Language in Foreign Language Learning. Clevedon: Multilingual Matters.
- Ringbom, Håkan 2007. Cross-linguistic Similarity in Foreign Language Learning. Clevedon: Multilingual Matters.
- Ross, Malcolm 2007. Calquing and metatypy. – Journal of language contact – THELMA 1, 116–143.
- Šarčević, Susan 1985. Translation of culture-bound terms in laws. – *Multilingua* 4–3, 127–133.
- Sasse, Hans-Jurgen 1990. Language decay and contact-induced change: similarities and differences. – *Arbeitspapier (Institut für Sprachwissenschaft, Universität zu Köln)* 12 (Neue Folge), 30–56.
- Thomason, Sarah G. 2001. Language Contact: An Introduction. Washington: Georgetown University Press.

- Tirkkonen-Condit, Sonja 2005. The monitor model revisited: Evidence from process research. – *Meta: journal des traducteurs / Meta: Translators' Journal* 50:2, 405–414.
- Toury, Gideon 1985. Aspects of translating into minority languages from the point of view of translation studies. – *Multilingua* 4, 3–10.
- Toury, Gideon 2012 [1995]. *Descriptive Translation Studies – and Beyond*. Revised edition. Amsterdam: John Benjamins.
- Venuti, Lawrence 1998. *The Scandals of Translation. Towards an Ethics of Difference*. London and New York: Routledge.
- Verschik, Anna 2008. *Emerging Bilingual Speech – from Monolingualism to Code-copying*. London: Continuum.
- Vinay, Jean-Paul; Darbelnet, Jean 1995 [1958]. *Comparative Stylistics of French and English: A Methodology for Translation*. Translated into English by Juan C. Sager & M. J. Hamel. Amsterdam: John Benjamins.
- Vlahov, Sergeï; Florin, Sider 2006 [1969]. *Neperevodimoe v perevode [Untranslatable in Translation]*. Moscow: Valent.
- Weinreich, Uriel 1974 [1953]. *Languages in contact: Findings and problems*. The Hague: Mouton.

## RESÜMEE

### TÕLKELAENUD KONTAKTLINGVISTILISE NÄHTUSENA: ÜLETADES KONTAKTLINGVISTIKA, TEISE KEELE OMANDAMISE JA TÕLKETEADUSE PIIRE

Artiklis käsitletakse tõlkelaene keelekontaktide, teise keele omandamise uurimise ja tõlketeaduse seisukohalt. Arutatakse kolme distsipliini lähenemiste sarnasusi ja erinevusi tõlkelaenude käsitlemisel. Arutelu rõhutab, et kõik kolm näevad samamoodi tõlkelaenude kognitiivset alust, samas valdav suhtumine tõlkelaenudesse näitab eri distsipliinide erinevaid ideoloogiaid. Uurimuse eesmärk on laiendada kontaktlingvistilise käsitluse perspektiive võrdluse kaudu teiste aladega, mis käsitlevad samu nähtusi.

# EESTI-SOOME KOODIVAHETUSE MITU NÄGU FACEBOOKI VESTLUSTE NÄITEL

Kristiina Praakli

Tartu ülikool

**Ülevaade.** Artikli keskmes on mitmehäälne diskursus. Põhitähelepanu lasub koodivahetuse funktsioonidel ning tähenduse loomisel Soomega seotud eesti keele kõnelejate Facebooki suhtluses. Lähtun koodivahetuse defineerimisel selle suhteliselt laiaast määratlusest, mõistes koodivahetuse all eri keelte elementide vahelduvat kasutust suhtluses. Samas on koodivahetus heteroglossiline vahend, mille all mõistetakse eri häälte ning diskursuste paljusust. Selles ühinevad suhtluses osalejate hääled, sellega taasluuakse, edastatakse või esitletakse tähendust, mis on omakorda seotud varasemate diskursustega. Seega ei tähenda koodivahetus ainult öeldu edastamist, vaid selle abil toimub ka tähenduse loomine. Ainetiku diskursiivne analüüs näitab, et koodivahetus on nii funktsioon, vestluse kontekstile viitaja kui ka rühmadevaheliste piiride tõmbaja.

**Märksõnad:** koodivahetus, mitmehäälsus, tähenduse loomine, eesti keel, soome keel

## I. Taust ja teoreetiline raamistik

Soome keel mängib kümnete tuhandete eesti keele kõnelejate elus üha suuremat rolli. Eesti keel on järjest enam kuuldav ja nähtav Soome linnamaastikel nii tava-suhtluses kui ka pildilis-sõnalises keeles (toidukauplused, välireklaamid) või institutsionaalses suhtluses (riigiasutuste eestikeelsed veebilehed). Eesti keele ja selle kõnelejate jõulist sisenemist Soome ja soome keeleruumi tähistavad kõige ilmekamalt Eesti sotsiaaldemokraatliku erakonna valimisreklaamid 2015. a kevadel Helsingi linnapildis sõnumiga „Ootame sind tagasi“. Soome statistikaameti andmeil elab Soomes 48 087 inimest, kelle emakeel on eesti keel (Tilastokeskus 2016). Olgu võrdlusena öeldud,

et Eesti taasiseseisvumiseelselt oli eestikeelseid elanikke Soomes 1394, sajandivahetusel 10 176, eelmise kümnendi lõpus 28 493 (Tilastokeskus 2016). Eestikeelse elanikkonna kujunemist on mõjutanud Eesti ühinemine Euroopa Liiduga (2004) ning olukord Eesti tööturul. Kogukonna kujunemisel artiklis ma pikemalt ei peatu, neid teemasid on süvitsi käsitlenud mitmed uurijad (nt Jakobson jt 2012, Praakli 2009, 2010, Koreinik, Praakli ilmmas).

Praegune riigipiirideülene maailm on oluliselt muutnud arusaama suhtlusruumist. Keelekeskkonnad on laienenud virtuaalsfääridesse (Ehala jt 2014) ning üha suurem osa suhtlusest toimub veebiavarustes. Rände ja eeskätt hargmaisuse kontekstis tähendab see traditsiooniliste suhtlusstruktuuride (nt seltsiõhtud, omakeelne kultuuriline-klubiline tegevus jne) asendumist või paralleelselt nendega osalemist virtuaalsetes suhtlusruumides. Nii on ka Soomega seotud eesti keele kõnelejate põhilised suhtluskanalid arvukad veebivõrgustikud, mida ainuüksi Facebookis on paarkümmend: väiksemates paarsada, suurimas umbes kolmkümmend tuhat liiget.

Üleilmastumine, võrkeelte parem kättesaadavus ja keeleoskuse übermõtestamine on avardanud mitmekeelsuse tähendust ja mõistmist. See nähtub selgesti ka suhtlusuuringute rõhuasetuste muutumises ning uute vaatenurkade esiletõus, kus üksikkõneleja keeleliste vahendite kogu nähakse senisest laiemalt, keskendudes küsimusele, milleks kõneleja kõiki tema keelises repertuaaris olevaid vahendeid kasutab. Kontseptsioonid, nagu *polylingualism*, *polylinguaging* (Jørgensen 2008, Jørgensen jt 2011), *metrolinguaging* (Otsuji, Pennycook 2010) või *translanguaging* (Garcia, Wei 2014), signaliseerivad fookuse liikumist keelesüsteemide struktuuriliselt ühinemiselt mitmekeelsele kõnelejale ja tema tegevustele suhtluses. Neid ühendab ka kriitiline vaade keelele kui ideoloogilisele konstruktsioonile (detailne ülevaade Androutsopoulos 2013: 1–3, Garcia, Wei 2014, Jørgensen 2008, Jørgensen et al 2011, Lehtonen 2015, Otsuji, Pennycook 2010).

Siinse artikli keskmes on küsimus koodivahetuse funktsioonidest. Koodivahetuse definitsioone, teoreetilisi mudeleid ning üldse

arusaamu mitmekeelsest suhtlusest on rohkesti (detailne ülevaade vt Verschik 2008: 2–24). Selles artiklis ma koodivahetuse mõistestikul, mudelitel, võrdlustel ning ühisosade otsinguil ei peatu, vaid keskendun kirjaliku ainekliku esitlusele ning sellele, milliseid lisavõimalusi mitme keele rakendamine suhtluses osalejatele pakub. Ka mõistet koodivahetus kasutan, viitamaks mis tahes üleminekutele, kus eesti keel soome keelega vahetub.

Üks esimesi koodivahetust vestluspragmaatilisest vaatenurgast avavaid teoreetilisi käsitlusi pärineb John Gumperzilt (1982). Gumperzi tööd on omakorda edasi arendanud ning nendega dialoogi astunud mitmed teised uurijad, eeskätt Peter Auer, kelle mitmekeelse suhtluse mudelid ning funktsioonide käsitluse teoreetiline raamistik on omakorda olnud vundamendiks teistele (vt detailne ülevaade Auer 1998: 1–28). Võõras pole teema ka eesti-soome koodivahetuse uurijatele (vt nt Frick 2010, Praakli 2009). Senised uurimused on avardanud eesti-soome mitmekeelsuse mõistmist nii struktuurilistest aspektidest – eeskätt Helka Riionheimo (2007, 2009, 2011, 2013a) ja Maria Fricki uurimused (2009, 2013, ka Riionheimo, Frick 2014) – kui ka koodivahetuse pragmaatikast (vt Frick 2009, Praakli 2009, Härmävaara, Frick ilmumas). Eesti-soome koodivahetust on valdavalt käsitletud suuliste aineklike, vähem kirjalike näitel (Frick 2010). Soomega seotud eesti keele kõnelejate elektroonilises suhtluses avalduvat koodivahetust pole autorile teadaolevalt varem käsitletud.

## 2. Keeleainestik, keelejuhid ja analüüsimeetod

Artikli keeleainestik pärineb Facebookist. Facebook on 2004. aastal Ameerika Ühendriikides loodud veebipõhine suhtlusvõrgustik, millel on umbes 1,3 miljardit kasutajat. Eestis on Facebooki kasutajaid 590 000 (Hansalu 2015), Soomes 2,4 miljonit (Kärkkäinen 2015). Facebook pakub suhtluseks erinevaid platvorme: kasutaja isiklik profiil, grupid (*groups*) ja lehed (*pages*). Gruppide suhtlust iseloomustab üks-mitmele suhtlus. Vestluse algatamiseks postitab grupi liige oma sõnumi rühma avalikule ajajoonel (n-ö seinale), kus



see on kohe nähtav ning avatud edasiseks vestluseks kõikidele grupi liikmetele. Seega on ka kirjaliku suhtluse puhul tegu grupi liikmete ühiselt produtseeritud vestlustega.

Artikli ainestik hõlmab grupi liikmete postitusi 2015. ja 2016. aastal. Tekstikogu koosneb 421 tekstist. Artiklis võtan vaatluse alla 16 näidet. Kõik keelenäited on muudetud anonüümseks ning iga-sugune isikustamist võimaldav teave on keelenäidetest eemaldatud. Artikli autor on üks grupi (passiivne) liige. Vestlustemaatiliselt liigituvad postitused valdavalt olmeteemade hulka, hõlmates kõikvõimalikke teemasid, mis elukohariiki vahetanul ette võivad tulla (elukoha üürimine, tööpakkumised, töötasud, sotsiaalteenused, juriidiline nõustamine, ost-müük-vahetus jne). Vähem on kultuuri-spetsiifilisi ja päevapoliitilisi arutelusid.

Vaatluse all oleva suhtlusgrupi liikmed on üldjuhul noored täiskasvanud, kes kõnelevad emakeelena enamasti eesti keelt, mitmed ka vene või soome keelt. Rühma liikmete taust on heterogeenne, kuid neid ühendavad sarnased kogemused lähiajaloost ning ühine keele- ja kultuuritaust. Enamik neist on esimest korda soome keelega kokku puutunud täiskasvanuna, sestap on ka nende soome keele oskuse tase ebaühtlane; paljudel grupi liikmetel ei pruugi soome keelega veel ka pikemat suhtluskogemust olla. Samas on soome keel rühma liikmete üks ühiseid koode ning koodivahetuse rakendamine ei eeldagi teise kontaktkeele valdamist. Kõneleja on suuteline saavutama vestluseesmärgid ka siis, kui tal on teises keeles minimaalsed oskused (Auer 2013: 460). Viimast kinnitab ka selle artikli keeleainestik, millest nähtub, et koodivahetus on võimalik ka piiratud või sootuks olematu soome keele oskuse korral (vt näited 7 ja 8).

Kuigi varases kontaktsituatsioonis või vähese mitmekeelsuskogemusega kõnelejate puhul on koodivahetus pigem perifeerne, toob ainestik nähtavale individuaalsete koodivahetusmuustrite paljususe. Kirjalik suhtlusruum peegeldab ka eesti netisuhtlusele oma-seid jooni, nagu eriortograafia (nt grafeemide õ ja ü asendamine õ või y-ga), suulise kõne hääldust imiteeriv kirjapilt, suulise kõne foneetilised, leksikaalsed ja morfoloogilised jooned, laen- ja slängisõnad,

emotikonid jne. Kirjaliku veebisuhtluse kohatine anonüümsus (alati pole teada, kes mingi kasutajanime taga peitub) võimaldab keelekasutajal keelt loovamalt kasutada, mis ei pruugi suulises kõnes tingimata aktsepteeritud olla.

Veebiainestiku analüüsimisel rakendan mikrosotsiolingvistika kvalitatiivset analüüsimeetodit, võttes appi ka vestlusanalüüsi vahendid. Keskendun suhtluses osalejate keelelistele valikutele ja funktsioonidele vestlustes ning tähendusloomele. Põhiküsimus lasub sellel, kuidas mitmekeelne diskursus luuakse ning mida soome keelega kirjalikus suhtlusruumis (Facebook) tehakse. Artikkel on jätk autori varasemale uurimistöele (Praakli 2009, 2010, 2014).

### 3. Keelevalikud ja koodivahetuse mallid kirjalikus suhtlusruumis

Avamaks koodivahetuse olemust, võtan esmalt vaatluse alla grupi liikmete keelelised valikud ning koodivahetuse põhilised mallid kirjalikus suhtlusruumis. Grupi postitused ning ühtlasi vestluste algatamised – laiemas mõttes vestluskeele valikud – toimuvad peaaegu eranditult eesti keeles, harva soome, eesti-soome või vene keeles. Tekstiline info edastatakse sageli fotode või linkide vormis, edasist vestlust algatamata. Soome keel postituste tervikliku keelevalikuna on aktuaalne soomekeelsete uudiste, kultuuriteadete, reklaamide jne edastamise puhul, sest need eeldavad sõnumi täpsust ning autentsuse säilitamist. Samas toimivad need kontekstivihjena, markeerides vestluse toimumise aega, ruumi ning vestluses osalejate seotust Soomega. Keelevalik annab vestlusesse astujatele ja seal osalejatele vihje, et soome keel on rühma liikmete üks ühiseid koode, mille kasutus on üldjuhul (aga mitte alati) aktsepteeritud. Kakskeelsed postitused viitavad ka võimalikele adressaatidele ning asjaolule, et postituse vastuvõtjateks võivad olla nii eesti kui soome keele kõnelejad. Seega näidatakse keele valikuga, et sõnum on suunatud mõlemale rühmale. Siinkohal tulebki appi koodivahetus: kõneleja näitab keele valikuga, kellele sõnum on suunatud.

Näites 1 edastabki teemaalgataja võrgustiku liikmetele Eesti Vabariigi presidendi kakskeelse tervituse Soome iseseisvuspäeva puhul. Näites 2 algatab teine teemaalgataja soomekeelse uudislingi postitusega eesti päritolu sisserännanute (Soldiers of Odini liikmed) patrullimisest Soome tänavatel.

- (1) A: *Oikein hyvää itsenäisyyspäivää rakkaalle naapurillemme Suomelle ja kaikille suomalaisille! Soovin Soome Vabariigile ja kõikidele soomlastele õnne tänase iseseisvuspäeva puhul!*

‘A: Väga head iseseisvuspäeva meie armsale naabritele Soomele ja kõikidele soomlastele.’

- (2) A: *Virolaiset näyttävät mallia ja vartioivat Helsingin kaduilla.*

‘Eestlased näitavad eeskju ja patrullivad Helsingi tänavatel.’

Postitusele järgneb link soomekeelse uudisega *Virolaiset vartioivat Helsingin kaduilla*. ‘Eestlased patrullivad Helsingi tänavatel’.

Kuigi on selge, et koodivahetus on võrgustiku liikmete seas pigem tavapärane keekekasutuse viis, siis alati pikemaid üleminekuid soome keelele siiski ei aktsepteerita (vt näide 7).

Koodivahetuste struktuurist lähtuvalt on eesti-soome kirjallikus suhtluses avalduvad üleminekud valdavalt ühesõnalised (substantiivid, partiklid, verbid) ja esinevad vestluse põhikeeles (eesti keeles) kas fonoloogiliselt ja/või morfosüntaktiliselt integreeritult (näide 3) või integreerumata kujul, st säilitades terviklikult soome keele struktuuri (näide 4). Koodivahetus toimub peaaegu eranditult leksikaalsel tasandil, hõlmates valdavalt institutsionaalset sõnavara (*maistraatti, poliisi*), tänamis-, palumis-, tervitus- ja hüvastijätuvormeleid, interjektsioone (*heippa, kiitos, moi* jne) ning vandesõnu (voi *perkele*; mitä helvettiä jne). Ümberlülitumised on pigem perifeersed ja ulatuvad leksikaalselt tasandilt harva sügavamale morfosüntaktilisele tasandile.

Soome struktuuri terviklikku säilimist koodivahetuse protsessis olen valinud illustreerima näite 3, mis kirjeldab tüüpilist ühesubstantiivilist ümberlülitumist. Kõneleja esitab grupi liikmetele küsimuse, kas erakorralisse vastuvõttu minnes tuleb tasuda

visiiditasu. Küsimuse olulisimat infot kannab substantiiv *päivystys* (erakorraline vastuvõtt), mille küsimuse esitaja esitab kui selles vestluskontekstis relevantseima soome keeles. Kuigi kõneleja keelelise repertuaari kuuluvad sel teemal ilmselt ka eestikeelsed vahendid (erakorraline meditsiin, EMO jne), võimaldab grupi liikmetega jagatav ühine kood ning tegutsemine ühises soomekeelses suhtlusruumis tarvitada soome keele pakutavaid vahendeid. Koodivahetuse morfoloogilise struktuuri mõttes järgib küsimuse esitaja terviklikult soome illatiivi struktuuri (*päivystykseen*).

- (3) *Hei! Rumal küsimus vb. Kui lähen **päivystykseen** kas maksan ka selle nn polikliinikumaksu?*  
soome *päivystys* 'erakorraline vastuvõtt'

Koodivahetuse teadlik rakendamine on ka kõneleja üks võimalusi edastada kultuurispetsiifilist teavet, mille asendamine omakeelse vastega pole ekvivalendi puudumise tõttu alati otstarbekas ega pruugi mõistesisu ikkagi avada. Näites 4 esitab teemaalgataja grupi liikmetele küsimuse ning saab sellele kaks asjakohast vastust. Neist esimene, C, annab teemaalgatajale pikema eestikeelse vastuse Soome koolilõputraditsioonide kohta, kasutamata selleks kordagi soome keelt. Vestlusesse sekkuv D vastab aga ühesõnalise üleminekuga soome keelele – *penkkarit!* –, tuues vastusega sisse vestluse toimumise kultuuriruumi.

- (4) A: *Miks on Mechelinkatu ja ka mujal täis maha pillatud komme nii et kõik urvitavad korjama?*  
B: *Mingi kommionu on liikvel :)*  
C: Abiturentidel hakkab eksamihooaeg ja selline on Soomlaste komme. Enne eksamihooaega tehakse üks suur sõit autodega ja visatakse maiust.  
**D: Penkkarit!**  
soome *penkkarit*, *penkinpainajaiset* '12. klasside õpilastel õppetundide lõppemist ning eksamiperioodi algust tähistav pidu'

Soome keele häälikulise ja grammatilise struktuuri terviklik ülekandmine vestluse põhikeelde on koodivahetuse protsessis eesti keele kõnelejate kirjalikes vestlustes pigem marginaalne, markeerides niisugustel juhtudel teadlikku keelelist käitumist diskursusstrateegilistel eesmärkidel. Kahe morfoloogiarikka lähisugulaskeele vahetus kontaktsituatsioonis on koodivahetuse üheks iseloomulikuks mustriks selles osalevate ühikute (enamasti substantiivid ja verbid) häälikuline kohandamine kõneleja emakeele päraseks ning elementide morfoloogilis-süntaktilises käitumises lähtumine vestluse põhikeele (eesti keel) struktuurist. Soome elementide integreerimine põhikeelde võib variatsioonide- ja kõnelejakesksete variantide rohke olla ka ühe ja sama elemendi lõikes, nagu näeme siinses artiklis substantiivi *henkilötunnus* kasutuse puhul (vt näited 9 ja 10).

Näited 4 ja 5 illustreerivadki soome ühikute sihtkeelepäraseks kohandamise fonoloogilisi ja morfoloogilisi protsesse. Vestlustemaatilisel on mõlema postituse sisu ja eesmärk rühma liikmetelt teabe küsimine või nende teavitamine formaalsetest protsessidest Soome kolimisel. Leksikaalsetest elementidest sisaldavad mõlemad postitused institutsionaalset sõnavara markeerivaid vahendeid, mis on ühtlasi ainsad koodivahetuses osalevad elemendid. Esimesena vaatluse alla tulevas näites näeme soome ühikute esinemist nominatiivi vormides. Substantiividest kolm – *poliisi* (politsei, politseijaoskond), *maistraatti* (elanikeregister) ja *vertoimisto* (maksuamet) – on häälikulises mõttes läbinud mitmed eesti keele päraseks kohandamise protsessid, nagu lõpukadu (vokaallõpuliste substantiivide lühenemine, vrd *poliisi* vs *poliis*, *maistraatti* vs *maistraat*) ja kaksikkonsonantide lühenemine (*maistraatti* vs *maistraat*). Eesti keele kõnelejate suulises kõnes maksuameti tähenduses esinev *vero* markeerib liitsubstantiivi lühenemist ning põhitähenduse ülekannet üksiksubstantiivile *vero*. Siinkohal nähtub selge erinevus soome ühiskeelega, kus keelend *vero* tähendab maksu või lõivu, kuid mitte maksuametit. Asutuse nime tähistavad soome keeles *vertoimisto* või *verovirasto*. Soome rahvapensioniameti (*Kansaneläkelaitos*)

lühendatud nimekuju *KELA* on samas vormis ja tähenduses kasutusel ka Soomega seotud eestlaste kirjalikus ja suulises suhtluses.

- (4) Nimelt kadunud kõik paberid seoses soome kolimisega - **poliis, maistraat, vero, kela**.  
soome *poliisi* 'politsei, politseinik, politseijaoskond', *maistraatti* 'elanike register', *vero* 'maks, lõiv', *KELA* 'rahvapensioniamet'

Eelmises näites illustreeritud elementide morfosüntaktilist integreerumist kirjeldab näide 5. Koodivahetuses osalevate keelendite morfosüntaktiline käitumine pakub mitmeid tõlgendusvõimalusi. Leksikaalsest aspektist vaadatuna on maksuameti tähenduses esinev *vero* moodustatud liitsubstantiivi lühendamise teel, kus asutuse põhitegevusele ning institutsioonile viitab liitsõna esikomponent *vero* (maks). Kõneleja järgib eesti sõnajärge ja inessiivi kasutust soome tüvele eesti inessiivi tunnuse *-s* aglutinatiivse liitmise teel. Analoogset morfoloogilist integreerumist näeme ka politseijaoskonda tähistava substantiivi *poliisi* puhul. Keelendi esinemisvorm pakub kaks tõlgendusvõimalust: kõneleja järgib soome kõnekeelele omast lühenenud tunnusega inessiivi kasutust (*-s* vs *-ssa/-ssä*), mis on lähedane inessiivi kasutusele eesti keeles, või tuleb kõneleja keeliste valikute puhul tõenäolisemaks pidada analoogiapõhist lähtumist eesti inessiivi kasutusest (politseis), kuna soome keel eeldab siinkohal väliskohakäänete kasutust kas vormide *poliisilaitoksella* või *poliisilla* kujul. Eesti keele morfoloogilisest struktuurist lähtub ka *kela* kasutus, mis esineb eesti lõputa genitiivis (vrd sm *kelan*), eesti keele pärane on ka lause sõnajärg. Liitsubstantiivi *henkilötunnus* tarvitust võib vaadata nii leksikaalsest kui morfoloogilisest aspektist. Esimesel juhul ei saa välistada kummagi kontaktkeele leksikaalsete vahendite põhjal kakskeelse kompromissvormi moodustumist, morfoloogilisest aspektist vaadatuna võib substantiivi kasutust tõlgendada analoogiapõhise ülekandena eesti malli alusel (vrd sm *tunnus : tunnuksen : tunnusta : tunnuksset*; ek *tunnus : tunnuse : tunnust : tunnused*).

- (5) A: /.../ *Tulime soome ilma igasuguste paberiajamisteta, kohe koos mööbli ja asjadega. Saime korteriomanikuga kokku, kirjutasime lepingu õues, laadisime asjad tuppa ja läksime sama õhtu tööle. Esimesel vabal päeval käisime **veros**. Saime ajutised **henkilötunnused** ja siis töotasime-elasime aasta soomes ilma **poliisis** käimata ja **kela** asju ajamata.*  
soome *poliisi* 'politsei, politseinik, politseijaoskond', *maistraatti* 'elanike register', *vero* 'maks, lõiv', KELA 'rahvapensioniamet'

Sama fraasi ulatuses võivad koodivahetuses osaleda nii põhikeele kui ka lähtekeele struktuuri järgivad üleminekud (näide 6). Koodivahetuse mitmekeelsust markeerivas ning selle referatiivset funktsiooni täitvas üleminekus näeme nii soome elemendi eesti keelde integreeritud kasutust (eesti lõputa genitiiv kirumisväljendi *saatana* puhul) kui ka soome keele tervikliku morfoloogilise struktuuri säilimist substantiivi *virolaiset* mitmuse nominatiivi kasutuses. Mitmehäälsuse aspektist vaatlen näidet artikli lõpuosas.

- (6) A: *Käisime mehega just poliisis end siia elama registreerimas. Esialgu vaatas ametnik kohe pilguga „saatana virolaiset“.* /.../  
Soome *saatanan virolaiset* 'pagana, kuradi eestlased'

Kuigi on selge, et koodivahetus on võrgustiku liikmete seas pigem tavapärane keelekasutuse viis, pole üleminekud soome keelele alati aktsepteeritud, nagu illustreerib näide 7.

- (7) A: *Head aega töölisklass, tere tulemast pagulased*  
B: *Suomessa pidetään somaleita luotettavimpinä kuin eestiläisiä.1990luvulla, kun alkoivat tulla Suomeen niin suomalaisille luvatiin että eivät vie suomalaisten työpaikkoja ja ne ovat sanansa pitäneet.*  
'Soomes peetakse somaale usaldusväärsemateks kui eestlasi. 1990ndatel, kui hakkasid Soome tulema, siis soomlastele lubati, et nad ei võta soomlaste töökohti ära ja nad on oma sõna pidanud.'  
C: *Suomalaiset on sen verran laiskoja, että sen takia virolaiset „vie“ työpaikkoja. Jos nei eivät olisi laiskoja ja työnlaatu olisi melkein sama ku virolaisten, sit ei olisi hätää.*

‘Soomlased on sel määral laisad, et selle pärast eestlased “võtavad“ töökohad ära. Kui nad ei oleks laisad ja nende töökvaliteet oleks peaaegu sama nagu eestlastel, siis ei oleks häda.’

**B: *B, superlaiskoja suomalainen ei viitsi tehdä työtä alle 25 euroa verokirjalla (tunti) minkä tyhmit tekee 5 euro mustana mutta koko päivän.***

‘B, superlaisk soomlane ei soovi teha tööd vähem kui 25 eurot koos maksudega (tund), mille lollid teevad ära viie euro eest mustalt, kuid kogu päeva.’

**D: *Miks te soome keeles kirjutate omavahel? On äge?***

**E: *Eputavad***

**F: *А нам соровно...***

‘A meil suva ...’

Vestluse algataja postitab sisserännet puudutava uudise ning avab vestluse sõnadega *head aega töölisklass, tere tulemast pagulased*. Avapostitusele järgneb kahe vestlusesse astuja vahel pikem soomekeelne arutelu eestlaste-soomlaste tööefektiivsuse teemadel. B ja C vahelisse soomekeelsesesse diskussiooni sekkub järgmises voorus D, kes teeb märkuse soome keele kasutamise ja selle võimalike põhjuste kohta: *Miks te soome keeles kirjutate omavahel? On äge?* D-le sekundeerib ühesõnalise vastusega E, kes pakub D esitatud küsimusele vastuseks, et *eputavad*. Arutelu lõpetab vestlusesse sekkuv F, kes tsiteerib lauset Vene filmimuusika klassikast *А нам все равно*.

Rühma liikmete heterogeenne taust nähtub ka nende soome keele oskusest, mis on ühtlasi mitme vestluse põhiteema. Grupisised vestlused soome keele oskuse teemadel viitavad selgelt asjaolule, et kuigi soome keelt võib kasutada ka ebasümmeetrilise oskuse puhul vägagi erinevalt, pole soome keele mis tahes kasutusviis aktsepteeritud: soome keele väärakasutamise üle heidetakse vähem või rohkem varjatud kujul nalja, tehakse keelelisi parandusi, vaieldakse õigekeelsusnormide üle.

Näites 8 algatab vestluse sel hetkel veel Eestis elanud eestlanna, kes esitab grupi liikmetele küsimuse perega Soome kolimise kohta.



Vestluses osaleb paarkümmend eesti keele kõnelejat, teemaalgataja küsimus kogub 139 vastust. Arutelu lõpuosas tõstatub põhiteema kõrvale küsimus soome keele oskusest. Rõhuasetuse muutumine johtub ühe vestluses osaleja viisist kasutada soome substantiivi *henkilötunnus* (isikukood) variandis *engelitunnus*, mis assotsieerub vestluses osalejatele sõnaga inglitanus (sm *enkeli* 'ingel'). Originaalpostituse täpne sõnastus pole enam tuvastatav, sest postitaja on selle eemaldanud.

Substantiivi *henkilötunnus* tavapäraest variatsioonidest (nt *hengelotunnus*, *hengilötunnus*, *henkar* jne) eesti keele kõnelejate kasutuses selgelt eristuv keelend pälvib kohe grupi liikmete tähelepanu. On ilmne, et substantiivi vestlusesse toonud grupi liikme soome keele oskus on veel vähene või on kõneleja omandanud substantiivi puhul väär vormi, kandes keelendile üle eesti suulisele kõnele tüüpilised häälikulised omadused – sõnaalguse h-konso-nandi mittehääldamine ja sõnasisene häälikute kadu – , mis omakorda kanduvad üle kirjalikku keelde. Resultaadiks on vorm, mis lahknub substantiivi normipärasest kasutusest soome keeles nii häälikulise kuju kui ka semantiliste omaduste poolest. Substantiivi hääldust imiteerida püüdev kõneleja annab sellele kirjalikus kasutuses tahtmatult uue, humoorika tähenduse, kõneldes isikukoodi asemel hoopiski inglitanusest. Kõneleja olematu või vähene soome keele oskus ei võimalda aru saada ega osaleda teiste vestlusesse sekkujate keelemängust ja naljatamisest, mille keelendi väärkasutus ja substantiivile tahtmatult antud lisatähendus kaasa toovad.

(9) B: *engelitunnus* :D, ikka *henkilötunnus*

D: sõna "*engelitunnus*" toob kananaha selga :D /.../

C: kust seda *engelitunnust* saab? :D surnuaiast?

D: kui tiibu seljas pole ☹ siis surnuaiast.

Soome *henkilötunnus* 'isikukood'

Kui esimese näite puhul saab soome keelest piiritõmbamise vahend keelt valdavate ja mitteoskava(te) grupi liikmete vahel, lubades soome keele oskajatel inglitanust edukalt rakendada ka keelemängus,

andmata keelendi esmakasutajale kordagi selget vihjet selle kohta, millise lisatähenduse eesti kõnekeelele tuginev hääldusmall sõnale annab, siis näite 9 puhul antakse postituse algatajale selged juhtnöörid soome keele õigeks kasutamiseks. Vestluse algatanud A esitab detailse ülevaate Soome elanikuks registreerimiseks vajaliku dokumentatsiooni kohta. Kuigi postitus sisaldab mitmeid ümberlülitumisi soome keelele (*käisime poliisis* ja *maistradis ära, kuu algusepoole poliisis käisime*) ning soome keele mittenormipärast kasutust, ei pälvi need võrgustiku liikmete tähelepanu. Tähelepanutõmbajaks saab taaskord (nagu ka näites 8) substantiivi *henkilötunnus* eripärane kasutus – näites 10 kujul *hengelotunnused* –, mis algatab vestluse soome keele oskuse ja kasutuse kohta.

(10) A: *Tere, käisime siis täna poliisis ja maistradis ära aga meie väga suureks üllatuseks selgub, et Eestis välja antud abielutunnistus ja lapse sünnitunnistus ei loe, need tuleb lasta kas inglise või soome keelde tõlkida ning notariaalselt kinnitada. Kui esimene korda kuu algusepoole poliisis käisime polnud sellest juttu. **Hengelotunnused** saame kätte alles siis kui tõlgitud ja notariaalselt kinnitatud tunnistused esitame.*

C: *Õige! Dokumentid peab olema ametliku tõlkija poolt tõlgitud ja notari kinnitatud. Ja mitte **hengelotunnus**, vaid **henkilötunnus***

A: *Mis puutub õigesse hääldusesse ja kirjaipilti siis olen soomes mõned nädalad ja puuudub soome keele oskus veel.*

D: *See **hengelotunnus** on 100x parem kui **inglitunnus** :D*

E: *Keegi ei nori keeleoskamatusse pärast. See on lihtsalt tüüpiline sõna, mida ikka väänatakse ja käänatakse omamoodi. Vigadest õpitakse ja vast nüüd jääb meelde.*

*sm **henkilötunnus** 'isikukood'*

Kuigi varases kontaktsituatsioonis on soome elementide kasutus pigem juhuslik ning ebareeglipärane, ei saa välistada, et mõned soome substantiivid on grupi liikmete kirjalikus kasutuses (ja ehk ka suulises kõnes) sel määral konventsionaliseerunud, et neid võidakse tajuda laensõnadena, mis asendavad eestikeelset

omasõna ning mille kasutamine on kogukonnas norm. Sotsiaalmeedia (Facebooki) suhtlusgruppide kirjaliku keelekasutuse näitel võib selliste keelendite hulka lugeda järgmised: vero (täheenduses 'maksuamet'), maistraat (täheenduses 'rahvastikuregister') ja poliis (täheenduses 'politsei') või ka siinses artiklis vaatluse alla tulev alko-laen (veebi vahendusel alkoholi „laenamine“ ehk illegaalne alkoholi müük). Detailsema ülevaate Soomes elavate eesti keele kõnelejade suulisest keelekasutusest leiab Kristiina Praakli väitekirjast (2009) ning artiklitest (2010, 2014).

#### 4. Refereerimine koodivahetuse funktsioonina

Järgnevalt vaatlen, kuidas kasutavad eesti keele kõnelejad koodivahetust eri funktsioonides. Koodivahetuse funktsioonide käsitluste tuntuimad teoreetikud on John J. Gumperz (1982) ning Peter Auer (1995, 1998). Eesti keeleteadlastest on koodivahetuse funktsioonidest detailse ülevaate andnud Anastassia Zabrodskaja (2006). Auer (1995) eristab koodivahetuse järgmised põhifunktsioonid, nagu refereerimine, vestluses osalejaskonna muutumine, vahelepõimed-kõrvalmärkused, öeldu kordamine, rolli- ja teemavahetus, naljad ja keelemängud ning topikaliseatsioon (lähemalt Praakli 2009: 102 ja sealsed viited). Mitmed loetletud funktsioonid, nagu näiteks refereeringud, naljad ja keelemängud, aga ka formaalsused (tervitused, pöördumised ja hüvastijätud) ning kultuurispetsiifiliste tekstide edastamine, nähtuvad ka selle artikli ainestikus, millest detailsemalt võtan vaatluse alla refereeringud.

Nii nagu ka suulise kõne puhul rakendatakse koodivahetust ennekõike refereerimise eesmärgil. Refereeritud kõne all mõistetakse ümberlülitumist teisele „häälele“, mis representeerib teist, kõneleja enda omast erinevat häält. Refereerimise abil tuuakse vestlusesse „kolmas“ hääle, millega edastatakse, esitletakse ning taastoodetakse selles situatsioonis varasemad diskursused (Praakli 2009: 103 ja sealsed viited). Refereeringute avaldumist kirjeldan näidetes 11 ja 12, pikemalt peatun teemal artikli viimases peatükis.

Koodivahetusena avalduvates refereeringutes on üks enim esinev ütlus *maassa maan tavalla* (käitu maal selle maa kombel), millega viidatakse soovitusel järgida Soomes olles sealseid kombeid, traditsioone, kirjutamata reegleid jne. Näites 11 tekitab grupi liikmete seas pikema arutelu kortermaja stendil olev kiri pidevalt remonti tegeva korterielaniku aadressil. Kuigi kirja aadressaati pole välja toodud, asuvad vestluses osalenud kohe seisukohale, et kriitika puudutab majas elavaid eestlasi.

(11) A: *Jumala aus tekst.... pange silt üles, korraldage talgud ja ongi kõik hästi....;*)

B: *Tyyppiline eestlane, kes ei arvesta naabritega, tehakse remonti ka peale 20.00, kui soomlane paneb lapsi magama /.../*

C: *Selline tekst ei ole mingi üllatus, meil ka läbi elatud. Tundub tõesti olevat suisa tavapärane, et otse ütlemist välditakse. See on ehk üks suurim komistuskivi siin maal. Tasub aegsasti arvesse võtta ja siiski „**maassa maan tavalla**“ toimida...;) tähendab, muudame ennast, mitte teisi...*

Soome *maassa maan tavalla* 'käitu maal selle maa kombel'

Analoogset üleminekut näeme ka näites 12, kus postituse autorile valmistab muret kultuurinormide mittetundmine (küsimus: *kas Soomes tuleb esimesse klassi minnes õpetajale lilli viia?*). Postituse autorile antud soovitus järgida riigis selle riigi kombeid ehk *maassa maan tavalla* edastatakse jällegi soome keeles. Ka siin tõstatub küsimus, miks teeb kõneleja seda soome keeles, kui tema keelelise repertuaari kuuluvad ka eesti keele vahendid. Vestluse lõpetab tõdemine (E), et Soomes viiakse lilli vaid sünnipäevaks ja matustele.

(12) A: *Öelge palun, kas Soomes on kombeks I klassi õpetajatele kooli alguses lilli viia? Eestis on, aga siinseid kombeid selle kohapealt ei tunne.*

B: *Ei tarvitse.☺*

C: *Poiss on su ju eestlane! Järgi meie tavasi ja kombeid ☺ Mis vahet sel on kas teised viivad või ei vii ☺*

D: *Minuarust lillede viimine on kyl kena komme. Aga eks need soomlased austavad ikka kui elatakse „maassa maan tavalla“ ja kui siin kombeks pole nii muidugi natuke naljakas.*

E: *Soomes viiakse lill vaid sünnipäevaks ja matustele, muid variante ei ole.*

Soome *maassa maan tavalla* 'käitu maal selle maa kombel'

Refereerimisel ning selle eri tähendustel, sh sotsiaalse tähenduse loomel, peatun detailsemalt artikli viimases peatükis.

Soome keelt võidakse vajadusel rakendada ka poeetilises ja/või humoorikas võtmes (näide 13). Üks grupi aktivistidest teeb võrgustiku liikmetele ettepaneku lasta fantaasial lennata ning kirjutada ühiselt luuletus. Järgmises näites näemegi, kuidas eesti keele kõnelejad rakendavad soome keele vahendeid poeetilise keelele omaseid vormivõtteid riimi loomisel. Luulevormis avalduv koodivahetus on mitmetahulisem kui muud artiklis vaatluse all olnud näited, kuna need annavad lugejale rohkesti kontekstivihjeid öeldu tõlgendamise kohta, tähistavad aega ja ruumi ning paotavad ukse hargmaisesse elukorraldusse, kirjeldades piirideülese töörande eri aspekte.

(13) A: *õhtust. igav on ja teeks õhtu huvitavaks. nimelt kuna siin on kõik nii andekad kommenteerivad siis kasutame seda ära ja laseme fantaasiat rakendada ka natuke asjalikumalt. vaatame mis kokku tuleb. teema siis fb ja seame riime ritta.. /.../ alustan siis ja iga järgmine kommenteerija lisab oma salmi.*

B: Tõusen yles silmis sära,

kohe lähen tööle ära.

Õhtul koju tagasi,

miskipärast kõik päevad sedasi.

Võtsin võlgu tulles Soome,

kogu palga sõpradega maha joome.

Segutegemisel aitab internet,

kiirteel peatab mind ment.

Sakko mulle kirjutab,

mina rõõmsalt meenutan.

murest aitab alko laina,  
 kohe nyd ma kaasua painaa  
 Soome sakko 'trahv'; alko laina 'alkoholi laen'; kaasua painaa  
 'annan gaasi'

Nagu näeme, rakendab postitaja soome keele vahendeid ka riimiloomes: murest aitab *alko laina*, kohe nyd ma *kaasua painaa* (murest aitab alkoholi laenamine, kohe nüüd ma gaasi annan). Koodivahetus põhineb soome keele foneetilisel materjalil (*laina-painaa*), kuigi teksti autor ei järgi koodivahetuses tingimata soome keele morfoloogilist struktuuri (vrd sm *kaasua painaa*). Koodivahetuses nähtub ka soome keele kasutamine rühmasisesena koodina, mis hoolimata soome keele vahendite kasutusest ei pruugi olla soome keelt emakeelena kõnelejatele konteksti tundmata mõistetav. Nimelt mõistetakse eestikeelse fraasi „alkolaen“ all Eestist toodud alkoholi ebaseaduslikku müüki, mis toimub alkoholi laenamise nime all. Nii seda kui eelpool nimetatud substantiive *poliisi*, *maistraati* ja *vero* võib pidada konventsionaliseerunud laensõnadeks Soomega seotud eestlaste keeles.

Nagu näidetest nähtub, toimub üleminek soome keelele paljudel põhjustel. Küllap võivad mõningad ümberlülitumised johtuda ka asjaolust, et kõneleja emakeeles puuduvad mingis teemavaldkonnas edukaks kommunikatsiooniks vajalikud keelelised vahendid, kuid see pole esmatähtis. Olulisem näib olevat tõsiasi, et soome keel on grupi liikmete ühine suhtluskood, mistõttu on soome keele kasutamine ootuspärane ning aktsepteeritud ja mis teeb võimalikuks info edastamise võrgustiku liikmetele nende ühises, üldjuhul kõigile mõistetavas koodis. Ühesõnalised substantiivid toimivad edukalt ka rühmade eristaja ja markeerijana, tõmmates piiri meie ja teiste vahele, on asukohariigi- ja kultuurispetsiifilised, viitavad vestluse toimumise asukohale ning kontekstile laiemas mõttes. Samas näivad mitmed keelendid olevat konventsionaliseerunud või kannavad võrgustiku liikmete vahelises suhtluses sotsiaalses mõttes olulist tähendust. Teemat käsitlen pikemalt artikli viimases peatükis.

## 5. Koodivahetuse mitmehäälsus

Koodivahetus on heteroglossiline vahend, mille all mõistetakse eri häälte ning diskursuste paljusust. Selles ühinevad vestluses osalejate hääled, sellega taasluuakse, edastatakse või esitletakse tähendust, mis on omakorda seotud varasemate diskursustega. Seega ei tähenda koodivahetus ainult öeldu edastamist, vaid selle abil toimub ka tähenduse loomine.

Vaatluse all olevast ainestikust nähtub, et kõnelejad edastavad keeleliste vahendite abil soomlaste võimalikku suhtumist eesti keele kõnelejate aadressil. Üks selline keelekasutuse sotsiaalselt tähendust kandev keelend on soome kõnekeeles levinud kirumisväljend *saa-tanan virolaiset* (kuradi eestlased, saatana eestlased). Keelend on eestlaste omavahelistes vestlustes aktiivselt kasutuses, seda rakendatakse eestlastele kui rühmale viitamise vahendina, millega edastatakse enamusrühma arvumusi või suhtumist vähemusrühma suhtes. Vale pole ilmselt ka väita, et keelend ja selle variatsioonid (*satana, saadana* jne) kannavad eesti keele kõnelejate jaoks spetsiifilist tähendust, mida kasutades viidatakse endale kui rühmale, kõneldes samal ajal soomlaste häälega. Keelendi kasutuses nähtub ka teatav ironia enamusrühma aadressil.

Koodivahetuses avalduvat mitmehäälsust olen valinud kirjeldama näited 14 ja 15. Näites 14 kirjeldab vestluse algatanud eestlanna vahejuhtumit ühistranspordis: alkoholihoobes soomlannale avaliku korra rikkumise kohta märkuse teinuna oli ta sunnitud kuulama soomlanna kommentaare eestlaste aadressil.

(14) A: *Hei, kaasmaalased!!! Reedene õhtu, sõidan koju peale tööd. Bussis kõrval istub purjus soome naine, kes hakkab ülbitsema. Mina loen, kuid ülbitsemine käib pinda ja palun oma rahu bussis. Mille peale hakkab karjuma, et olen **virolais vördjas** ja mingi tagasi kopikate eest töötama kodumaale. /.../ Kuidas Teie kaitsete ennast sellistel juhtudel?*

*/.../*

D: *ma kutsuks poliisi, nad ei tohi ka Soomes teisi tylitada. Ja ma ei hakkaks yldse seda loba kuulama, vaid teeks kohe kõne.*

B: *Parem olla Virolainene vördjas kuis Kodimainene – asotsiaal*

C: *Sama kutsuks politsei neid nad kardavad kui veel purjus*

D: *Õige! Maassa maan tavalla!!!*

E: *üttele et saatana kehitysvammainen mene sairaalaan ja hoitaa omia onkelmia sielä!*

F: *Dai boh kui yritab minna saatana virosse odavat õlut ostma /.../*

G: *Inimesed ma olen Teie üle uhke:nii huumorit, kaastannut ja tuge!!!-) Happy to be „virolais vördjas“*

*/.../*

E: *Olen kahjuks koos õega sellises olukorras olnud. Nimelt olime Koivukyla koera-aedikus, kus üks koertest oli agressiivne ja kippus mu koerale kallale. /.../ Palusime sel meesterahval oma koer ära kutsuda, mille peale ta aint irvitas rövedalt. Ja hakkas kena tekst pihta - see koer s88b su ära, saatana virolaine. /.../*

Soome *virolais*, *virolainen* vördjas 'eesti, eestlasest vördjas'; *kotimainen* 'kohalik'; *maassa maan tavalla* 'käitu maal selle maa kombel'; *saatanan Viro* 'kuradi, pagana Eesti'; *saatanan virolainen* 'kuradi, pagana eestlane'

Vene *dai boh* 'annaks jumal'; inglise *happy to be* 'õnnelik olla'

Postitus algab autori eestikeelse pöördumisega võrgustiku liikmete poole (*Hei, kaasmaalased!!!*) ning jätkub vahejuhtumi kirjeldusega eesti keeles (aeg, tegevused, osalised, olukorra kirjeldus jne). Postituse viimases lauses esitab ta grupi liikmetele küsimuse, *kuidas end sellistes olukordades kaitsta*. Nagu võime postitusest näha, esitab autor vahejuhtumi üldise kirjelduse eesti keeles, vahetades koodi tema jaoks juhtumi olulisima sõnumi edastamisel, viitega rühmale: mille peale hakkab karjuma, et *olen virolais vördjas*. Koodi vahetades esitab autor soomekeelsena tema jaoks sõnumi olulisima informatsiooni osa, markeerides koodivahetuse abil konflikti teise osapoole, enamusrühma esindaja arvamust Soomes elavate eestlaste suhtes. Soome keeles toimunud vahejuhtumit kirjeldades edastab postituse



autor soomekeelsena vaid rahvust puudutava osa (*virolais* 'eesti'), vahejuhtumi muid aspekte taasesitleb autor eesti keeles. Järelikult kannab soome keele kasutamine selles vestluskontekstis postituse autori jaoks vahejuhtumi olulisimat sõnumit, mille ta edastab refereeringu vormis soome keeles. Tõsi, refereering on vaid osaline, koodivahetus ei hõlma sõnumi vahest kõige olulisimat keelendit – millist soome keele vandesõna soomlanna eestlaste aadressil täpselt kasutas, edastades vestluse selle osa eesti keeles (värdjas). Vaadeldes koodivahetuse struktuuri lingvistilises mõttes, näeme, et tegu on fraasitasandil toimuva tsiteeringuga, kus soomekeelsena esitatakse rahvuse soomekeelne nimetus (*virolais*), pejoratiivse vandesõna esitab autor eestikeelsena. Nagu postitusest ilmekalt nähtub, pole seda tüüpi koodivahetuste puhul tegu ei vestlusstrateegia, leksikaalse lünga täitmise vms, vaid teadliku ümberlülitumisega soome keelele, eesmärgiga edastada soomekeelsena, autentsel terviklikul kujul kolmanda isiku öeldu.

Postituse autori koodivahetus ei jää selles paarikümne osalejaga vestluses ainsaks soome keele kasutuseks. Autorile sekundeerib kümmekeelne võrgustiku liiget, kelle puhul võib koodivahetuse rakendamises täheldada analoogset mustrit, kus etnilistele rühmadele viitamisel rakendatakse soome keele vahendeid. Mitmel juhul võib täheldada ka refereeringute kihistumist, kus kõneleja refereerib samaaegselt nii postituse autorit kui ka algsõnumit. Asjakohaseks näiteks on siin vestlusesse lülitunud grupi liikme B postitus, kelle koodivahetuses ühinevad nii postituse autori kui ka konflikti algatanud soomlanna "häälled": parem olla *virolainene* värdjas kui *kodimainene* – asotsiaal. Üleminekud toimivad selgelt rühma tähistajana, markeerides eri rühmi ja nende taustu soome keele vahenditega: *virolainene* (sm *virolainen*) versus *kodimainene* (sm *kotimainen* 'kodumaa, kodumaine') ehk eestlane kui sisserännanu ning soomlane kui põliselanik. Refereering on tähelepanuväärne ka keelelises mõttes. Üleminek soome keelele ei järgi kummagi keele kirjalikke norme, kahe keele vahendite baasil moodustub kolmas keelekuju. See, kas tegu on soome keele teadliku väärkasutusega või viitabki

kõneleja soome keele kujule, jääb paraku teadmata. Samas ei saa välistada ka seda, et kõneleja viitab soome adjektiivisufiksi *-nen* teadliku väärkasutusega vestlust soomepärasemaks muutes nii eri rühmadele kui ka nende soome keele oskuse tasemele (eestlased on sisserännanud ega oska isegi soome keelt).

Enamus- ja vähemusrühma vahele tõmmatakse soome keelt kasutades piirid veel mitmes postituses. Analoogset üleminekut näeme ka F-i postituses. Tema keelises esituses kandub vandesõnaga *saatana* edastatav pejoratiivne tähendus rahvuselt üle ka riigile (Dai boh kui yritab minna saatana virosse odavat õlut ostma). Vestlusesse sekkub postitaja G, kes avaldab enda postitusega poolehoidu eeskõnelejatele ning tähistab soome keelele ülemineku jutumärkidega: *Happy to be "virolais vördjas"*. Soome keele vahendeid kasutades viitab rahvusele veel ka postitaja E, kes kirjeldab temale osaks saanud vahejuhtumit soome kodanikuga, edastades soomekeelsena vestluse olulisima osa, viitega taaskord rahvusele: *Ja hakkas kena tekst pihhta - see koer s88b su ära, saatana virolaine*.

Analoogne soome keele kasutus nähtub ka kahes järgmises näitelauses, kus rahvusele viidates kasutatakse taaskord soome keele vahendeid. Kaks järgmist vestlust erinevad eespool kirjeldatust selle poolest, et siin pole tegu konkreetse kolmanda isiku kõne refereerimisega (esimese näite puhul taasesitas postituse autor konkreetset vahejuhtumit, olles üks juhtumi osapooltest), vaid koodivahetuse kaudu edastatakse enamusrühma oletatavat arvamust või hinnangut eesti vähemusrühma aadressil, mida tehakse taaskord keelendi *saatana* vahendusel. Järgmises näites kirjeldab postituse autor elukoha registreerimise protseduuri ametiasutuses, edastades kõneleja hinnangul talle osaks saanud negatiivse kogemuse ametnikust, kes olevat vaadanud teda kui eesti päritolu inimest halvustava pilguga „*saatana virolaiset*“.

- (15) A: *Käisime mehega just poliisis end siia elama registreerimas. Esi-  
algu vaatas ametnik kohe pilguga „saatana virolaiset“ . /.../  
soome saatanan virolaiset ‘kuradi, pagana eestlased’*

Ka selles vestluskatkes toimub koodivahetus teadlikult ning esineb markeeritud kujul, mida kirjalikus vestluses markeerivad eri „hääle“ eristajatena toimivad jutumärgid, mis toovad samaaegselt sisse nii oma kui ka võõra hääle, kuid mis lubavad kõnelejal ennast öeldust distantseerida (à la see on mitte minu, vaid soomlaste arvamus eestlaste kohta).

Viimane näide erineb eelmistest selle poolest, et postituses kirjeldatud tegevuses osalejateks (hernevargus) on eranditult eesti keele kõnelejad. Vestluse algataja postitab gruppi iroonilises võtmes kommentaari kaasmaalaste aadressil: Tervitan kõiki kaasmaalasi, kes praegu mul majalähedal herne raksus on. Postitust kommenteerivad rühma liikmed esitavad postituse autorile enda arvamusel hernevargil käinute rahvuse või muude etniliste tunnuste põhjal (originaalpostitus on pikem kui siinne vestluskatke). Nii arvab postitaja B, et hernepõllul käijateks olid ehk moslemid, mille kohta postituse autor vastab koodi vahetades, et ei olnud, käijad olid täitsa *satana virolaised*. Ka siin pole koodivahetus enam pelk kahe keele vaheldumine, vaid teadlik soome keele vahendite rakendamine, millega kõneleja viitab eestlastele samu soome keeles levinud pejoratiivseid vahendeid kasutades.

(16) A: *Tervitan kõiki kaasmaalasi, kes praegu mul majalähedal herne raksus on :)))*

B: *Need on moslemid*

A: *Ei ei ☺ Täitsa **satana virolaised** olid ... on... raske nüüd juba öelda ... võib olla juba järgmised.*

Soome *saatanan virolaiset* 'kuradi, pagana eestlased'

## Kokkuvõte

Artiklis oli vaatluse all Soomega eesti keele kõnelejate koodivahetus Facebooki-diskursuste näitel. Vaatluse all olnud keeleainestiku valguses võib koodivahetuse kohta esialgselt öelda järgmist. Soome keele rakendamine vestlustes on nii funktsioon, vestluse kontekstile viitaja, kõneleja identiteedi tähistaja, rühmade eristaja

kui ka rühmadevaheliste piiride tõmbaja. Mitmekeelsetes suhtlusolukordades saab koodivahetusest ka sotsiaalne akt, mis kannab, edastab, toodab ja taastoodab tähendusi. Koodivahetus on ühtlasi mitme häälega kõnelemine, milles põimuvad kõnelejate kultuurilis-keelelised kogemused. Võrgustiku liikmete puhul on selgelt näha, et koodivahetus on aktsepteeritud keelelise käitumise viis ning koodivahetuse kasutamine on norm. Seda ka olukorras, kus kõnelejal puudub igasugune soome keele oskus, mis pole aga takistuseks kommunikatsioonis edukalt toimimiseks. Samas on eesti keele kõnelejate kirjalikus suhtluses nähtav variatiivsus ning ebasümmeetria soome keele kasutuses (laiemas mõttes soome keele oskuses) üks grupi liikmete positsioneerimise ning staatuse näitamise vahendeid. Selles võib näha ka teatavat jõu ja võimu vahendit, eri rühmade (eestlased ühel- ja teiselpool Soome lahte) ja/või eluolukordade eristajat.

#### KIRJANDUS

- Androutsopoulos, Jannis 2013. Networked multilingualism: Some language practices on Facebook and their implications. – *International Journal of Bilingualism*, 1–21.
- Auer, Peter 1995. The pragmatics of code-switching: a sequential approach. – *One speaker, Two languages: Cross-Disciplinary Perspectives on Code-Switching*. Eds. Lesley Milroy, Pieter Muysken. Cambridge: Cambridge University Press, 115–135.
- Auer, Peter 1998 (ed.). *Code-Switching in Conversation: language, interaction and identity*. London: Routledge.
- Auer, Peter 2013. Code-switching/mixing. – *The SAGE Handbook of Sociolinguistics*. Edits. Ruth Wodak, Barbara Johnstone, Paul Kerswill. SAGE Publications Ltd, 460–478.
- Gumperz, John J. 1982. *Discourse Strategies*. Cambridge University Press.
- Ehala, Martin; Koreinik, Kadri; Praakli, Kristiina; Siiner, Maarja 2014. Kuidas uurida keele kestlikkust? – *Keel ja Kirjandus* 7, 489–508.
- Frick, Maria 2009. Morphological integration of Estonian lexical elements in a Finnish language base during the first years of immigration. – *Journal of Linguistic and Intercultural Education* 1, 81–95.

- Frick, Maria 2010. Suomi-viro koodinvaihto Facebookissa. – Lähivõrdlusi. Lähivertailuja 20. Toim. Eve Mikone, Kirsti Siitonen, Maria-Maren Sopper. Tallinn: Eesti keele sihtasutus, 49–67.
- Frick, Maria. 2013. Emergent bilingual constructions: Finnish-Estonian codeswitching in interaction. Doctoral dissertation. Helsinki: University of Helsinki.
- García, Ofelia; Wei, Li 2014. Translanguaging: Language, Bilingualism and Education. Basingstoke: Palgrave Macmillan.
- Hansalu, Kadri 2015. Eestis on 590 000 Facebooki kasutajat. <http://majandus24.postimees.ee/3415827/eestis-on-590-000-facebooki-kasutajat> (24.04.2016)
- Härmävaara, Hanna-Ilona; Frick, Maria. Handling linguistic asymmetries by bilingual punning in conversations among speakers of cognate languages. – Crossing Languages to Play with Words: Interdisciplinary Perspectives. Eds. Sebastian Knospe, Alexander Onysko, Maik Goth. Berlin: DeGruyter. [ilmumas].
- Jakobson, Mari-Liis; Järvinen-Alenius, Pauliina; Pitkänen, Pirkko; Ruutsoo, Rein; Keski-Hirvelä, Elisa; Kalev, Leif 2012. The Emergence of Estonian-Finnish Transnational Space. – Migration and Transformation: Multi-Level Analysis of Migrant Transnationalism. Eds. Pirkko Pitkänen, Ahmet İçduygu, Deniz Sert. Book Series: International Perspectives on Migration. Springer-Verlag. (International Perspectives on Migration), 159–206. [ilmumas]
- Jørgensen, Jens Norman 2008. Polylingual languaging around and among children and adolescents. – International Journal of Multilingualism vol. 5:3, 161–176
- Jørgensen, J. Normann; Karrebæk, Martha Sif; Madsen, Lian Malai; Møller, Janus Spindler 2011. Polylinguaging in superdiversity. – Diversities Journal Fall 2011, 32–54.
- Kärkkäinen, Henrik 2015. Facebook paljasti Suomi-lukuja. <http://www.itviikko.fi/uutiset/2015/04/15/facebook-paljasti-suomi-lukuja/20154707/7> (23.04.2016)
- Koreinik, Kadri; Praakli, Kristiina 2016. Emerging language political agency among Estonian native speakers in Finland. – Language Policy Beyond the State. Eds. Maarja Siiner, Kara D Brown, Kadri Koreinik. Springer Netherlands. (Language Policy ) [ilmumas].

- Lehtonen, Heini 2015. Tyyliellen: Nuorten kielelliset resurssit ja kielen sosiaalinen indeksisyys monietnisisessä Helsingissä. Väitöskirja. Helsinki: Helsingin yliopisto.
- Otsuji, Emi; Pennycook, Alastair 2010. Metrolingualism: fixity, fluidity and language in flux. – *International Journal of Multilingualism* 7:3, 240–254.
- Praakli, Kristiina 2009. Esimese põlvkonna Soome eestlaste kakskeelne keeikasutus ja koodikopeerimine. Tartu: Tartu Ülikooli Kirjastus.
- Praakli, Kristiina 2010. Eestlased ja eesti keel Soomes. – *Eestlased ja eesti keel välismaal*. Toim. Kristiina Praakli, Jüri Viikberg. Tallinn: Eesti Keele Sihtasutus, 455–501.
- Praakli, Kristiina 2014. On the bilingual language use of the Estonian-speakers in Finland. – *Sociolinguistic Studies* 8 (3). 383–409.
- Riionheimo, Helka 2007. Muutoksen monet juuret. Oman ja vieraan risteytyminen Viron inkerinsuomalaisten imperfektin muodostuksessa. – *Suomalaisen Kirjallisuuden Seuran Toimituksia* 1107. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Riionheimo, Helka 2009. Interference and attrition in inflectional morphology. A theoretical perspective. *Language contacts meet English dialects: Studies in honour of Markku Filppula*. Eds. Esa Penttilä, Heli Paulasto. Newcastle: Cambridge Scholars Publishing. 83–104.
- Riionheimo, Helka 2011. Complex morphologies in contact. The case of Ingrian Finns in Estonia. – *Language contact and language shift. Grammatical and sociolinguistic perspectives*. Ed. Kelechukwu Ihemere. Munich: LINCOM Publishers. 10–36.
- Riionheimo, Helka 2013a. Inkerinsuomen verbimorfologia kielikontaktissa ja attritiossa. – *Rapautuva kieli. Suomalaisen Kirjallisuuden Seuran Toimituksia* 1404. Toim. Kimmo Granqvist, Päivi Rainò. Helsinki: Suomalaisen Kirjallisuuden Seura. 37–74.
- Riionheimo, Helka; Frick, Maria 2014. The emergence of Finnish-Estonian bilingual constructions in two contact settings. – *Sociolinguistic Studies* 8(3) Equinox Publishing Ltd. 409–447.
- Zabrodsckaja, Anastassia 2006. Vene-eesti koodivahetuse funktsioonid Kohtla-Järve venekeelsete laste vestluses. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 2. Toim. Maria-Maren Sepper, Helle Metslang, Margit Langemets. Tallinn: Eesti Keele Sihtasutus 231–249.

Tilastokeskus 2016=Tilastokeskuksen PX-Web-tietokannat→Px-Web StatFin→Väestö→Väestörakenne. Kieli iän ja sukupuolen mukaan maakunnittain 1990–2015. <http://www.tilastokeskus.fi/tup/tilastotietokannat/index.html>. (15.08.2016)

Verschik, Anna 2008. Emerging bilingual Speech. From Monolingualism to Code-Copying. London, New York: Continuum.

## SUMMARY

# THE MULTIPLE FACES OF ESTONIAN-FINNISH CODE-SWITCHING SEEN FROM FACEBOOK CONVERSATIONS

The contemporary trans-border world has drastically changed our understanding of communication and socialization. Language environments have expanded into the virtual spheres and the major share of communication is performed on the Web. In terms of mobility and, in the first place, transnationalism, this means the substitution of traditional places of socialising (e.g. community evenings, cultural and club activity in native language, etc.), or, at the same time, participating in internet environments. Similarly, the main communication places of Estonian-speakers in Finland are the numerous virtual networks, of which there exist a couple dozen on Facebook, with the biggest comprising approx. 30,000 members. According to Statistics Finland, there are 48,087 people in Finland whose mother tongue is Estonian. In comparison, shortly after Estonia gained its independence this figure was 1,394; by the new millennium it had risen to 10,176, climbing further to 28,493 by 2009. The size of the Estonian-speaking population in Finland was impacted by Estonia's entry into the European Union in 2004 and the state of the Estonian labour market.

The present article concerns itself with code-switching functions in written communication in a Facebook group produced by Estonian-speakers in Finland. The data include postings of the group members during the period June 2015 to April 2016. The data are comprised of 421 texts. The group surveyed includes young adults, most of whom speak Estonian as their mother tongue, and many also speak Russian or Finnish. Although the background of the group studied is largely heterogeneous, they are, however, united



by their shared experiences in recent years and a common language and cultural background. Their initial contacts with the Finnish language mainly occurred in adulthood; therefore, their Finnish language skills are asymmetric.

Based on the results of this study on code-switching in written communication, one may make the following conclusions: The use of Finnish in conversations include functionality, references to the conversation's context, being speaker's identity marker and, furthermore, a marker of intragroup boundaries. In multi-language communication situations, code-switching acts as a device which carries, transfers, produces and reproduces meaning. At the same time, code-switching means speaking in many voices, whereby the different speakers' cultural-language experiences converge. In the case of the respondents one may say that their code-switching is an accepted mode of linguistic behaviour, and the use of code-switching is a norm, even in situations where the speaker(s) in question has/have only minimal knowledge of Finnish. On the other hand, there is a noticeable variation in the written communication of the informants, and the use of the Finnish language is one means for the group members to position themselves and demonstrate their status. It may also be viewed as an apparatus of power, as well as a means to distinguish themselves from the Estonian-speakers who reside on the other side of the Gulf of Finland).

# ESTONIAN LOTFITKA ROMANI AND ITS CONTACT LANGUAGES

**Anette Ross**

Tallinna ülikool

**Abstract.** The Estonian Lotfitka dialect is a Romani dialect whose speakers have migrated from Latvia to Estonia. This article provides an overview of the recent and current contact languages of the Estonian Lotfitka dialect – Latvian, Russian and Estonian – and draws attention to some of the contact-induced language changes. To provide a comprehensive insight into the intensity and scope of borrowing I have applied Thomason and Kaufman’s borrowing scale to categorize the contact languages. The relevant features behind the contact-induced changes that appear in Estonian Lotfitka dialect are listed.

**Keywords:** Romani linguistics, language contact, language change, Estonian Romani, Latvian Romani

## Introduction

This article will focus on the Lotfitka Romani dialect (also known as Latvian Romani) in Estonia and its language contact situation. The language’s speakers have migrated from Latvia to Estonia for couple of generations and refer to themselves as Lotfitka or Laloritka Roma.

Romani has not gained much attention as a minority language in Estonia and this article tries to shed light on the situation of Romani in Estonia. The emphasis is on its recent and current contact languages: Russian, Latvian and Estonian. The language contact situation is analyzed within the framework proposed by Thomason and Kaufman (1988). The contact-induced changes are listed and the languages are categorized according to Thomason and Kaufman’s borrowing scale. The borrowing scale has been applied to provide a

more easily graspable classification of the Lotfitka Romani contact situation.

The article is based on material analyzed in the scope of the author's Master's research on the the language, the focus of which is to describe the dialect in more detail and compare it to the Lotfitka spoken across the border in Latvia.

## I. Roma and Romani language in Estonia

The Romani language is an Indic language belonging to the Indo-Aryan branch of the Indo-European languages. The speakers migrated from the territory of India and reached Europe though the Balkans from where they migrated all over Europe.

According to the Estonian Population and Housing Census held in 2011 (Statistics Estonia), 456 Roma live in Estonia, of which 361 name Romani as their mother tongue. The data on Romani spoken as a second language is not publicly available, but would surely complement the data as some of the Romani speakers report other languages as their mother tongue. According to *Third report on Estonia* carried out by the European Commission against Racism and Intolerance, around 1,100 to 1,500 Roma live in Estonia (ECRI 2006: 35).

In the classification of the Romani dialects I follow that suggested by Matras (2002: 10), which is also followed by Tenser (2008) in his dissertation on the Northeastern Romani group. According to this classification the Romani dialects spoken in Estonia belong to the Northeastern sub-branch of the Northern group of Romani dialects. According to Tenser (2008: 12) dialects belonging to the Northeastern group are also spoken by Ruska Roma (also known as Xaladytka), Polska Roma (in northern areas of Poland), Litovska Roma and Lotfitka Roma. The dialects spoken in Estonia are Lotfitka (Latvian) and Xaladytka (Russian) Romani dialects.

Romani is a language that has been taught across generations without disrupt (Matras 2002: 191) and therefore classifies as a maintained language in Thomason and Kaufman's framework. Romani is

still common as a first communication language in Estonian Roma families.

Roma who have retained Romani as their mother tongue are commonly multilingual (Matras 2002: 191). Roma children in Estonia often learn Russian or Estonian in kindergarten or school as a second language. In a survey on educational issues which involved 88 Roma origin children (up to 17 years old) (Lutt et al. 2011), 28 were reported to speak Romani, Estonian and Russian. Even when attending schools where Estonian is the main language of instruction, Roma children still often learn Russian and use the language in every day communication. The reasons for their multilingualism, according to the Estonian Roma themselves, are family relations, the working environment, interactions with locals and involvement in small business (Ross 2013).

## 2. Research on Lotfitka Romani as part of the Northeastern dialect group

Out of Northeastern Romani dialects, the grammar of Russian Romani has been described by Ventzel (1980), that of Polish Romani by Matras (1999) and that of Lithuanian Romani by Tenser (2005). A short grammatical description of Latvian Lotfitka is part of the etymological dictionary of Latvian Romani (Mānušs et al 1997). In Estonia Ariste has collected data from Lotfitka speakers, and has published several articles on Lotfitka, e.g. Estonian loanwords in Lotfitka (Ariste 1983); loanwords in Lotfitka that give evidence of the migration route through Europe (Ariste 1958); Latvian verbal prefixes in Lotfitka (Ariste 1973), and on the Latvian noun derivation suffix *-uma* in Lotfitka (Ariste 1969).

A comparative study on Northeastern Romani dialects was conducted by Tenser (2008). In his dissertation Tenser (2008: 282) refers to Estonian and Latvian Romani as isolates in the Northeastern group. Some of the features differentiating Estonian and Latvian Romani from other Northeastern dialects are those shared with

Northwestern and Central dialects, e.g. masculine singular loan noun marker *-os* instead of *-o* as in other Northeastern dialects, and plural reflexive noun base *pen-* instead of *pes-*. Contraction of personal markers *ker-av-as* > *ker-âs* is only shared with Northwestern dialects. Interrogative ‘which’ *saj-* instead of *sav-* and contraction in ‘day’ *dyis* instead of *dyves* is shared only with the Sinti dialect (belonging to the Northwestern dialects). As unique features to Estonian and Latvian Romani, Tenser lists metathesis of *pǎ* > *ǎp* (and *kǎ* > *ǎk*), which are derived from aspirated consonants *ph* and *kh*; voicing of *ph-* in *phuè-* ‘ask’ as *buè-*; and loss of participles in *-ime(n)*. These features seem to be relevant only in Latvian Lotfitka and have not spread to Estonian Lotfitka.

### 3. Language interference in a situation of language maintenance

Thomason and Kaufman (1988: 37–39) divide the interference of languages into two basic types – borrowing situation in the case of language maintenance, and interference through shift. In the case of borrowing, foreign features are incorporated into the native language by the speakers. Borrowing is denoted as the transfer of features from one language to another in the areas of phonology, morphology, syntax and lexicon. Interference through shift, on the other hand, appears due to imperfect group learning. In that process the target language is learned imperfectly by new speakers. The features of this emerging variant are then adopted in the speech of native speakers and spread among all speakers of the language, thus forming a new *shifted* variant.

For the borrowing situation Thomason and Kaufman (1988: 73–95) propose a borrowing scale divided into five categories. These categories are based on two factors: the intensity of the language contact situation and the cultural pressure. This article uses these five categories to visualize the stage of borrowing from contact languages by Estonian Lotfitka Romani.

The first and lowest category in the scale describes casual contact that affects only the lexicon. The second and the third categories describe more intense contact in which slight structural borrowing is present. The changes in the second category might include borrowing function words such as conjunctions and adverbial particles, and in structure minor phonological, syntactic, and lexical semantic features. In the third category borrowings can include adpositions, and borrowed derivational affixes, phonemicization of previously allophonic alternations and syntactic changes. The fourth category is characterized by moderate structural borrowings under strong cultural pressure. The phonological changes can affect the native vocabulary; in the scope of syntax extensive word order changes can occur; concerning morphology, borrowed inflectional affixes and categories might be added to native vocabulary. In the fifth category, heavy structural borrowing happens under very strong cultural pressure. Thomason and Kaufman (1988: 74–76) describe these changes as ‘major structural features that cause significant typological disruption’.

#### 4. Data Collection

Data collection utilized the Romani Morpho-Syntactic (RMS) Questionnaire devised by Elđík and Matras (2001). The interviews were recorded in the informants’ homes over the course of one to a number of days. For the interviews in Estonia either Estonian or Russian were used as the language of elicitation. In Estonia altogether 11 linguistic RMS interviews were elicited, four of them with speakers of Xaladytka Romani, and 7 of them with speakers of Lotfitka Romani. In Latvia 14 interviews were elicited with Lotfitka Romani speakers.

The data on Estonian Romani dialects and the rest of the North-eastern Romani group was collected as part of the project *Finnish Romani and other northern dialects of Romani in the Baltic Sea area* lead by the University of Helsinki. The Estonian data was collected by Anton Tenser, Roman Lutt, Zalina Dabla and Anette Ross in

2013 and 2015. The Latvian data was collected by Anton Tenser and Dainis Krauklis in 2013, and transcribed by Dainis Krauklis, Anton Tenser and Anette Ross. The Estonian Lotfitka data has been transcribed or reviewed by the author of this article. The sociolinguistic background is based on the author's personal observations and interaction with Estonian Roma from 2013 to 2016, including work for her Bachelor's thesis at the University of Tartu on the language use of Estonian Roma (Ross 2013). The data collection included conversations with informants during interviews and events organized by the Roma community or Estonian state institutions concerning language and culture issues.

## 5. Contact languages

Matras (2002: 191) points out that Romani speakers have been multilingual for centuries and Romani has been in contact with other languages at least since the Early Romani. This has given Romani a layered structure of borrowings as the dialects have been in contact with different languages at different times. The borrowed lexicon is subject to being replaced with items from new contact languages, but the layer of stable lexical and grammatical borrowings from previous contact language will stay in the dialect. The previous contact languages usually continue to play a role in family interaction after the actual migration. The importance of the previous contact language might be increased, because the migration of Roma often involves migration of extended families and several generations. Matras (2002: 191–196) emphasizes that the contact languages' stratification profiles vary in numerous ways.

In order to classify the contact languages I use the distinction provided by Matras. Matras (1998: 300) proposes a distinction of three layers in order to stratify the grammatical borrowings. The layers are: older, recent and current contact language. The older L2 has had a considerable impact on the dialect, but is not spoken anymore. The recent L2 is spoken by the older generation and the

current L2 is spoken by all members of the community and plays a key role in communication with non-Romani people.

The older contact languages that have had an impact specifically on the whole Northeastern group are German and Polish (Tenser 2008: 221). Tenser lists Latvian as the recent L2 and Estonian and Russian as the current L2s for Estonian Romani (p. 222). This situation is plausible for the Estonian Lotfitka dialect as Latvian is still known to some extent by the Estonian Lotfitka speakers. Verification can also be found in the Estonian Census in 2000 (Statistics Estonia), which lists Latvian as a known foreign language for 120 of 542 Roma in Estonia.

The author finds that the Northeastern Romani dialects are not greatly affected typologically by the contact languages, since generally the contact languages and the conservative Romani share basic typological profiles. He adds that 'some of the contact-induced changes, however could be analyzed as forms of slight typological drift' (Tenser 2008: 236). Some of the more prominent contact-induced changes in Northeastern dialects are loss of articles, and use of aktionsart prefixes to modify verbs.

## 6. Russian Influence on Estonian Lotfitka Romani

Russian has had a stronger influence on the Estonian Lotfitka dialect when compared to Latvian and Estonian. Russian has been a contact language for Lotfitka both in Latvia and Estonia and we can see similar contact-induced changes in Latvian and Estonian Lotfitka. Another aspect strengthening the pressure of Russian is the close interaction with Xaladytka Roma. The strategy of replacing Latvian borrowings with Russian borrowings in communication with Xaladytka Roma was described by some of the Estonian informants.

Tenser (2008: 223) points out that the phonology of Northeastern Romani dialects is conditioned to a large degree by the contact languages with which Roma come into contact. The Russian language has affected speakers of Northeastern Romani dialects to the



extent that the Romani speech of some speakers show palatalization of consonants. This is certainly true for some of the Lotfitka speech, but as the currently dominant contact language and time of migration from Latvia varies among Estonian Lotfitka, the research on phonology would need a detailed speaker specific approach. Another borrowed feature that Tenser points out is the velar fricative /ɣ/ (ibid.) that is present in the speech of Estonian Lotfitka, e.g. *yaning* 'well', *yuèò* 'tall'. The sound exists in Ukrainian and southern dialects of Russian. Tenser also mentions velarization of aspirated phonemes, but the extent of this phenomenon needs to be investigated for Estonian Lotfitka.

In Latvian and Estonian Lotfitka there are shared contact-induced changes due to contact with Russian. In addition, there is also a layer of features in Estonian Lotfitka not shared with Latvian Lotfitka. The Estonian Roma population is approximately 10 times smaller than the Latvian Roma population, which is listed at 5,388 by the Central Statistical Bureau of Latvia (the number marks residents of Roma ethnicity). The relatively small Roma community in Estonia leads to stronger inter-group communication between Lotfitka and Xaladytka Roma, while Latvian Lotfitka Roma can continue with cultural practices in their own community.

In Latvian and Estonian Lotfitka in the category of indefinite pronouns the Russian specific marker *-mo* is combined with inherited relativizers forming the indefinites *so-ta* 'something' (Russian *что-то*), *kidi-ta* 'sometime' (Russian *когда-то*), and *kaj-ta* 'somewhere' (Russian *где-то*). In the free choice category only Estonian Lotfitka speakers have borrowed the Russian marker *-нибудь*, e.g. *kon-n'ibut* 'anyone', *so-n'ibut* 'anything', *kidi-n'ibut* 'anytime', *kaj-n'ibut* 'anywhere'. The free choice marker *-n'ibut* is very rare in Latvian Lotfitka, but is commonly used in other Northeastern Romani dialects influenced by Russian (Tenser 2008: 108). The following table (Table 1) shows the distribution of the indefinite pronouns in specific and free-choice category. The additional marker *vari(var/vaj)*- is added to provide a more reliable picture as the suffix

is also commonly used by speakers. The Romanian origin prefix *vari-* appears in a few examples in Estonian and Latvian Lotfitka with indefinite pronouns in the specific or free-choice categories.

**Table I.** Indefinite pronouns: specific and free-choice category

Category		Latvian Lotfitka	Estonian Lotfitka	Estonian Xaladytka
Specific	'something'	so-ta var-so-ta	so-ta vari-so	so-ta
	'sometime'	kidi-ta	kidi-ta	kagda-ta
	'somewhere'	kaj-ta var-kaj-ta	kaj-ta vari-kaj	kaj-ta
Free-choice	'anything'	so-na-so var(vaj)-so-ta	so-n'ibut'	so-n'ibut'
	'anytime'	kidi-na-kidi	kidi-n'ibut'	kagda-n'ibut' koli-n'ibut' kidi-n'ibut'
	'anywhere'	kaj-na-kaj	kaj-n'ibut'	kaj-n'ibut'

Another borrowed Russian marker in Estonian and Latvian Lotfitka is the diminutive adjectival suffix *-in'k-*; this is in fact found in all Northeastern Romani dialects except Polish Romani (ibid.: 66). Russian verbal prefixes are common with both borrowed verbs and inherited Romani verb stems, e.g. the Russian prefix *raz-* and Romani stem *phen-* 'to say' are combined as *ras-phenel* to express the meaning 'to tell (a story)' modeled after Russian *рассказать* 'to tell (a story)' < *сказать* 'to say'. Occasionally Russian verbs are left unintegrated in 3<sup>rd</sup> person singular and plural, and in the imperative form in Estonian Lotfitka. This is well attested in Xaladytka Romani (Tenser 2008: 121) in all persons and seems to appear in Estonian Lotfitka due to contact with Xaladytka Romani, not as a new strategy developed from within the dialect itself. The strategy is not attested for verbs borrowed from Latvian nor Estonian. The strategy to borrow verbs from contact languages retaining the conjugation of the contact language is apparent in more dialects (Elđík, Matras 2006: 135), e.g. the borrowings from Turkish into Romani dialects spoken in the Balkans retain the Turkic

conjugation (*ibid.*). The verb forms of 3<sup>rd</sup> person left as in the Russian conjugation is also supported by the tendency recognized by Elđik and Matras (2006: 102), namely that the 3<sup>rd</sup> person is most prone to borrowing and is most differentiated.

Concerning syntax, Estonian Lotfitka speakers sometimes omit the copula as in Russian. Absence of the copula in Estonian Lotfitka speech is probably not triggered only by Russian, but also by interaction with Estonian Xaladytka speakers that also tend to omit the copula. Another feature is the incorporation of Russian conditional/irrealis particle *õbi*, which has been borrowed into Xaladytka Romani (Tenser 2008: 143). The particle *õbi* is unsystematic, but used quite often in irrealis and conditional structures in Estonian Lotfitka. The particle is used mostly with remoteness marker *-as* added to perfective verb form, and more rarely added to verbs present tense as is common in Xaladytka. In Latvian Lotfitka the particle *õbi* is not used, and only the remoteness marker *-as* is added to perfective verb forms as in Early Romani, or to present tense personal marker as in Xaladytka.

The influence of Russian can be placed between level three and four in the borrowing scale, showing intense contact and strong cultural pressure. Level four is reached due to the phonological changes. Tenser (2008: 237) comments that phonology seems to be influenced more by contact languages in Northeastern Romani dialects than the morphology, and this contradicts Thomason and Kaufman's (1988) borrowing scale. Still, many distinctive features remain, such as aspirated consonants. The contradiction can be seen also in the influence of Latvian and Estonian on Lotfitka Romani.

## 7. Latvian Influence on Estonian Lotfitka Romani

Latvian is a recent, and to some speakers of Estonian Lotfitka also current contact language. In Latvia the Lotfitka dialect is under strong cultural pressure from Latvian and as a result is expected to develop in directions different from those expected for Estonian Lotfitka. Estonian Lotfitka is no longer affected by these changes,

because the cultural pressure from Latvian is low, and the interaction with Latvian Lotfitka is not sufficiently intense.

Due to Latvian influence a distinction between long and short vowels has appeared in the Lotfitka dialects (Tenser 2008: 223). It is maintained in Estonian Lotfitka as the new contact language, Estonian, makes this distinction as well. Latvian nouns and verbs are integrated into Lotfitka. Nouns are given feminine ending *-a*, e.g. Lat. *puīe* 'flower' becomes Lot. *puèèa*, or masculine ending *-os*, e.g. Lat. *vilks* 'wolf', Lot. *vilkos*; *-us*, e.g. Lat. *laiks* 'time', Lot. *lajkus*; *-is*, e.g. Lat. *bullis* 'bull' retaining its form in Romani as *bullis*. Verbs are adapted with loan verb adaptation marker *-in-*. This is followed by a personal marker, e.g. Lat. *rakstīt* 'write', Lot. *rakst-in-*, but occasionally the loan marker is followed by Greek-derived 3<sup>rd</sup> person singular and plural marker *-i*, e.g. Lat. *brauc* 'drives' becomes *brauc-in-i* 'he/she drives, they drive'.

Latvian influence on the lexicon includes adpositions such as *blakam*, *blakus*, *blaku* 'next to' and *preèu* 'opposite', which is according to Mânuðs (1997: 104) a contamination of Latvian *pret*, *pretī*, *pretim* and Polish *przeciw*, *przeciwko*. Estonian Lotfitka speakers have retained the Latvian superlative prefix *vis-* beside Russian adjective *сам-* 'most'. Some Latvian contrastive conjunctions are borrowed, e.g. *bet* 'but', *vaj - vaj* 'either or', *ne - ne* 'neither nor'. In the case of *vaj* 'or' the inherited Romani form and the Latvian conjunction are identical, so it is hard to say if the conjunction is preserved or borrowed. In the class of utterance modifiers, proposed by Matras (1998), are some Latvian borrowings, e.g. conjunctions, *pat* 'even', *īstes* 'really'.

In morphology, Latvian ambipositions such as *blakam* 'next to' are borrowed as prepositions and postpositions into Latvian Lotfitka, occurring in both positions, as is possible in Latvian. In Estonian Lotfitka they occur only as prepositions, which is similar to the behavior of the inherited Romani adpositions.

A sign of a stronger language contact situation is the presence of Latvian verbal prefixes. Many Latvian verbal prefixes are borrowed

into Estonian Lotfitka, e.g. *aiz-*, *ap-*, *at-*, *ie-*, *no-* and *pâr-*, e.g. *ajz-del* ‘to lend’ calqued from Latvian *aiz-ðemties*; and *pâr-dþal* ‘pass, overpass’ calqued from Latvian *pâr-iet*. Ariste (1973: 80) recorded three Latvian verbal prefixes: *uz-*, *no-* and *ie-*. Compared to Latvian Lotfitka, the frequency of Latvian verbal prefixes is currently quite low in Estonian Lotfitka and their use is not systematic. That said, some inherited verb stems are systematically used by Estonian Lotfitka speakers, e.g. Lat. verbal prefix *no-* in *nuo-peja pe* ‘to happen’. It is a combination of the Latvian aktionsart prefix *no-* on the base of Latvian verb *notikt* and Romani verb *peerel* ‘to become’ as an analogy with Latvian *tikt* ‘to become’. The borrowing of Latvian verbal prefixes is preconditioned by Slavic verbal prefixes that existed in the dialect before the contact with Latvian. The typological shift towards aktionsart prefixes has possibly happened already under the influence of Polish. Under the influence of Latvian the replacement of morphological material, i.e. replacement of the prefixes, has taken place.

Another morphological feature that is mentioned by Ariste as an influence from Latvian is the nominalization suffix *-uma* in *farduma* ‘jail’, *perðuma* ‘beginning’ (Ariste 1969: 179). Ariste does not believe that the suffix derives from the nominalization suffix *-imol-ima*, as it is also present in Lotfitka and the sound change from *-im* to *-um* has not happened in any other context (1969: 181). He believes that the suffix is formed along the lines of the Latvian *cietums* ‘jail’, GEN *cietuma*, as the Latvian basis *ciets* ‘hard’ complies with Romani *fardo* ‘hard’; and Lot. *perðuma* ‘beginning’, Lat. *sâkums*, GEN *sâkuma*. Tenser (2008: 46), on the other hand, considers the suffix *-ima* and *-oma/-uma* as variants of the Greek-derived suffix *-ima*, as the suffix is applied to lexicon of non-inherited origin in both cases (ibid.: 47). Also, the suffix *-oma* is present in Lithuanian Romani (ibid.: 46), e.g. *èemnoma* ‘darkness’, *radoma* ‘joy’ and therefore does not seem to be direct due to influence from Latvian.

The Latvian influence on Estonian Lotfitka dialect could be marked under category two and three in the borrowing scale, describing more intense contact and therefore some structural borrowing.

**Table 2.** Suffix *-ima/-uma*\*

	Lithuanian Romani	Latvian Lotfitka	Estonian Lotfitka	Xaladytka
'darkness'	čėmnoma	čėmn'uma	čėmnuma	(Estonian Xaladytka) t'omnyma t'omnuma
'light'	švjtālīma	švatluma	gajšuma svetloma	svētlīma
'joy'	radoma	raduma		radīma
'quiet, silence'	čixoma	čixuma		tixima

\*(The data on Lithuanian Romani and Xaladytka dialect (North Russian Romani) are from the ROMLEX database)

## 8. Estonian Influence on Lotfitka Romani

Estonian influence on Estonian Lotfitka Romani is mostly limited to the lexicon. Lexical borrowings are integrated into Romani. Estonian verbs *jalutama* 'to walk', *kasutama* 'to use' and *reisima* 'to travel' are all integrated with Romani loan-verb adaptation marker *-in-* as *jalut-in-*, *kazut-in-* and *reiz-in-*.

The nouns are adapted into Romani feminine or masculine class, e.g. Estonian nouns get a feminine ending *a-* as Estonian *sild* 'bridge' > *silta* and *ploom* 'plum' > *ploomā*, *pirn* 'pear' > *pirna*; masculine nouns get the endings *-os*, *-is*, e.g. Estonian *juust* 'cheese' > *juust-os*, *hunt* 'wolf' > *un't-is*. For integrating masculine loan nouns, use is rarely made of the markers *-as*, and *-us*, but see e.g. *maagus* 'stomach (internal organ)' from Estonian *magu*, and *täxtas* 'star' from Estonian *täht*. According to Tenser (2008: 47), Latvian Lotfitka has retained three masculine markers on loan nouns: *-os*, *-us* and *-is*.

Tenser (2008: 48) writes that in Latvian Lotfitka the masculine markers are distributed in accordance to the source language, *-os* is for Russian and *-is* for Latvian borrowings. In the current data Latvian masculine nouns in Latvian Lotfitka are dominantly integrated with *-os/-us* and rarely with *-is*, mostly only when the Latvian noun ends in *-is*, and so fully corresponds to Romani masculine loan marker *-is*, e.g. *trusis* 'rabbit', *kuņģis* 'stomach'. In Estonian Lotfitka

Estonian nouns tend to get either ending. The choice of the marker is connected to the Estonian stem vowel, as the genitive of *juust* ‘cheese’ is *juustu* > *juustos*; *kaev* ‘well’, GEN *kaevu* > *kaevos*; *loom* ‘animal’, GEN *looma* > *loomos*; *direktor* ‘director’, GEN *direktori* > *direktoris*, and genitive of *hunt* ‘wolf’ is *hundi* > *unt’is*. According to Elšik (2000: 19), one of the criteria for assigning gender and class to a loan in Romani is phonological shape of the noun in the source language, i.e. the final phonemes. For adaptation, the internal morphological structure of the given noun in Romani and in the source language are taken into account, and the stem of the source noun is adopted (Elšik 2000: 20). The pattern of adaptation of the nouns into Estonian Lotfitka is in accordance with the pattern provided by Elšik, i.e. the stem of the noun in the source language is the basis for adaptation.

Some Estonian adverbials are in use: *äkki*, *järsku* ‘suddenly’, *tavaliselt* ‘usually’ and *liiga* ‘too much’. Estonian vowels /x/, /æ/, /ø/ and /y/ are found in the loanwords *tähta* ‘star’, *sünnipääv* ‘birthday’, *küla* ‘village’ and also diphthongs that are not common for Romani as *lõuna* ‘lunch’. There is one semantic shift that is widespread in the Estonian Lotfitka dialect and is modeled after Estonian. The verb *dol-* ‘to receive, to get’ has taken on the meaning of ‘to be able to, can’ (Example 1) from Estonian *saama* ‘to receive, to get’ and ‘to be able to’. The more conservative forms to express the meaning ‘can’ and ‘cannot’ are inherited non-inflected *ašti* (Example 2) and *našti*, respectively, which are still in use, but being replaced by inflected *dol-* and *na dol-*, *na* being the common negation marker. In Latvian Lotfitka there are no examples of *dol-* attested with the meaning of ‘be able to’ and only the conservative inherited forms are present.

- (1) *Tu dolesa manca te jees?*  
 you can.2SG me.INST COMP come.2SG
- (2) *Tu ašti manca te jees?*  
 „Can you come with me?“

Estonian has had a relatively small influence on Estonian Lotfitka Romani dialect. This is due to Estonian being a contact language of Romani only in recent generations. In the borrowing scale of Thomason and Kaufman the Estonian impact on the Romani dialect could be estimated to belong in Category One, limiting the borrowing to lexical items only. However, the casual contact seems to be moving towards Category Two with changes being observed in cases of slightly more intense contact as relationships become more intimate and more Roma speak Estonian on a daily basis.

## 9. Xaladytka Romani influence on Estonian Lotfitka Romani

Lotfitka Romani and Xaladytka Romani both belong to the North-eastern group of Romani dialects. The dialect classification that this article is based on – Matras (2002) and Tenser (2008) – takes into account the genetic model and the geographical diffusion model. The genetic model divides the Romani dialects into branches after migrating from the Southern Balkans during the Early Romani period (roughly the Byzantine period) (Matras 2002: 215). According to the geographical diffusion model innovation is introduced in one location and then spreads gradually (*ibid.*: 265).

Lotfitka and Xaladytka in Estonia show mutual interference. At this stage we can attest variation in Estonian Lotfitka that is present due to contact with Xaladytka Roma. This article only describes Xaladytka influence on Estonian Lotfitka, and not vice versa. The variation is speaker specific, but the listed features appeared in the speech of more than one informant.

A change taking place in Xaladytka that is described by Tenser (2008: 67–68) is agreement between head noun and adjectives, numerals and demonstratives. In conservative Romani (Example 3) only the head noun would take the case marker and modifiers would take the oblique case marking. In Estonian Lotfitka, numerals normally do not take oblique case, unlike in other Northeastern dialect, in which numerals take the oblique marker *-e*, e.g. *trin* vs. *trin-e*



‘three’. In the Estonian Lotfitka data, some examples of case agreement (Example 4) of adjectives with head nouns was present due to Xaladytka influence. It has not yet been observed with demonstratives and numerals.

(3) *trin tern-e murš-en*  
 three young-OBL men-ACC.PL  
 ‘with three young men’

(4) *trin tern-en murš-en*  
 three young-ACC.PL man-ACC.PL

NE dialects calque the Russian and Polish way of constructing with reflexive enclitic the passive voice of the verbs, intransitivize verbs, and making reflexives from the transitive verbs, e.g. *garavel* ‘to hide (something)’ > *garavel pe* ‘to hide oneself’. In Xaladytka there exists the impersonal reflexive enclitic *pe(s)*. Latvian Lotfitka has personalized markers of reflexiveness on verbs: 1sg. *man*, 2sg. *tut*, and 1pl. *men* and 2pl. *tumen*; 3pl. is marked with reflexive pronoun *pe(s)* and 3pl. with *pen*. In Estonian Lotfitka verbs are often calqued into Romani from Russian, Polish or Latvian with the reflexive enclitic, but in Latvian Lotfitka the personalized enclitic is used instead of impersonalized *pe*, e.g. in Estonian Xaladytka ‘to meet’ *udykhel pe* is marked with clitic *pe* in all persons, e.g. *ame udykhasa(m) pe* ‘we will meet’, while in Latvian Lotfitka in 2pl. the enclitic would be *men*, e.g. *ame dikhasam men* ‘we will meet’.

In Estonian Lotfitka we see variation in all speakers’ samples – sometimes only the impersonal reflexive pronoun *pe(s)* is used and in some cases personal enclitics are used. The strategy of not integrating Russian verbs, seems to be influenced by Xaladytka as well.

Some Estonian Lotfitka speakers have replaced Lotfitka epistemic complementizer *si* (from *sir* ‘how’), modeled after Latvian, with Xaladytka complementizer *so* ‘what’, modeled after Russian. In these cases Estonian Lotfitka speakers vary between the forms of Latvian Lotfitka and Xaladytka or prefer only the Xaladytka feature as with the epistemic complementizer. The existence of both features

in the speech of Estonian Lotfitka, and the tendency to prefer the shared features with Xaladytka allows us to consider these features as recent influence on Estonian Lotfitka from the Xaladytka dialect.

## 10. Conclusion

This article draws attention to the contact situation of the Lotfitka dialect in Estonia and points out the degree to which the recent and current contact languages – Latvian, Russian and Estonian – have influenced Estonian Lotfitka, using Kaufman and Thomason's borrowing scale classification which describes intensities of contact and levels of interference. The classification was employed as a more comprehensible tool to observe the situation and relevant features of Estonian Lotfitka.

Russian has had a great influence on most of the dialects in the Northeastern group and has triggered contact-induced changes in Estonian Lotfitka the most when compared to Estonian and Latvian. Estonian Lotfitka exhibits borrowings from Russian of verbal prefixes, adjectival suffixes and indefinite pronouns, as well as heavy influence from Russian on phonology. These changes lead us to evaluate the Russian contact intensity and level of borrowings as being the Category 3 or 4.

Lotfitka has had intense contact with the Latvian language and some structural borrowing has taken place into Estonian Lotfitka. The layer of lexical borrowings is now narrowing and giving way to Russian and Estonian. In the borrowing framework the situation is between Categories 2 and 3.

Estonian has not influenced Estonian Lotfitka to the extent that Latvian and Russian have. There are semantic shifts and borrowing of lexicon. The lexical borrowings from Estonian into Romani also retain their phonetic structure and introduce Estonian vowels into Estonian Lotfitka. The language contact situation can be seen as casual contact moving towards slightly more intense contact, described by the borrowing scale as moving from Category 1 to

Category 2. As Latvian is losing its importance in the community, Estonian is replacing it as a strongest current L2.

Beside the three local contact languages, Xaladytka Romani spoken in Estonia has an effect on the Estonian Lotfitka dialect. There is a variation in the speech of Lotfitka speakers in Estonia that is triggered by contact with Xaladytka Roma.

For future research, samples of casual speech should be collected in order to evaluate the state of tendencies that are currently based solely on the translated questionnaires. There are currently no up-to-date material of narratives or dialogues that could complement the data presented here.

Although the samples were collected from different localities in Estonia, in the current research samples from the Estonian-Latvian border town of Valga/Valka are missing. The town has the biggest population of Roma (around 180 Roma inhabitants) and it is also known as a migration destination for Latvian Lotfitkas. Therefore, samples from the area would be important in order to investigate the interaction between Latvian and Estonian Roma and observe whether the innovations from Latvian Lotfitka are spreading. It would furthermore give a more rounded picture of the state of Estonian Lotfitka.

## REFERENCES

- Ariste, Paul 1958. Des Tziganes (Bohémiens) des pays Baltes. Omagiu lui Iorgu Iordan cu prilejul împlinirii a 70 de ani, 35–40. București: Editura Academiei.
- Ariste, Paul 1969. Ein lettisches Ableitungssuffix im Zigeunerischen. *Baltistica* 5, 2, 179–181. Vilnius: Mintis.
- Ariste, Paul 1973. Lettische Verbalpräfixe in einer Zigeunermundart. *Baltistica* 9, 1, 79–81. Vilnius: Mintis.
- Ariste, Paul 1983 (1985). Eesti laensõnu mustlaskeeles. *Emakeele seltsi aastaraamat*, kd 29, 25–29.
- Central Statistical Bureau of Latvia 2015. ISG07. Resident population by ethnicity at the beginning of the year <http://data.csb.gov.lv/> (01.05.2016)
- ECRI 2006. Third report on Estonia <http://www.coe.int/> (01.05.2016)

- Elšík, Viktor 2000. Romani nominal paradigms their structure, diversity, and development. Grammatical relations in Romani: the noun phrase. Edited by Elšík, Viktor; Yaron Matras. Amsterdam: Benjamins, 9–30.
- Elšík, Viktor; Yaron Matras 2001b. Romani Dialectological Questionnaire. University of Manchester, Department of Linguistics.
- Elšík, Viktor; Yaron Matras. 2006. Markedness and Language Change: The Romani Sample. Berlin: Mouton de Gruyter.
- Lutt, Roman; Jüri Ginter; Viktoria Neborjakina; Tatjana Neborjakina; Kadri Põldaru 2009. Romad Eesti haridussüsteemis. Uurimise aruanne. Tartu 2009–2011. [https://www.hm.ee/sites/default/files/romad\\_eesti\\_haridus-susteemis.pdf](https://www.hm.ee/sites/default/files/romad_eesti_haridus-susteemis.pdf) (01.05.2016)
- Mānušs, Leksa; Jānis Neilands; Kārlis Rudevičs 1997. Čigānu-latviešu-angļu etimoloģiskā vārdnīca un latviešu-čigānu vārdnīca. Rigi: Zvaigzne ABC.
- Matras, Yaron 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics*, Volume 36, Issue 2, 281–331.
- Matras, Yaron 1999. The speech of the Polska Roma. Some highlighted features and their implications for Romani dialectology. *Journal of the Gypsy Lore Society fifth series* 9, 1–28.
- Matras, Yaron 2002. *Romani: A Linguistic Introduction*. Cambridge, Cambridge University Press.
- Ross, Anette 2013. Eesti romade mitmekeelsus: keeleoskus, keelekasutus ja keelekasutuse mõjutajad. Tartu: Tartu University.
- Population and Housing Census 2011. Statistics Estonia <http://www.stat.ee/phc2011> (01.05.2016)
- Population and Housing Census 2000. Statistics Estonia <http://www.stat.ee/population-census-2000> (01.05.2016)
- ROMLEX Database = ROMLEX Lexical Database <http://romani.uni-graz.at/romlex/> (10.08.2016)
- Tenser, Anton 2005. *Lithuanian Romani*. Lincom Europa: Munich.
- Tenser, Anton 2008. *Northeastern Group of Romani Dialects*. Manchester: The University of Manchester.
- Thomason, Sarah Grey, Terrence Kaufman 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.
- Ventzel, Tatjana V. 1983. *The Gypsy Language*. Moscow: Nauka Publishing House.

## RESÜMEE

# EESTIS KÕNELDAVA ROMA KEELE LOTFITKA MURRE JA SELLE KONTAKTKEELED

Artikkel kirjeldab Eestis kõneldud roma keele Lotfitka murde olukorda, lähtudes kontaktkeeltest ja nende mõjul toimunud keelemuutustest. Lotfitka murre on Lätis kõneldud roma keele murre, mille kõnelejad on asunud Eestisse püsivalt elama ja nimetavad ennast Läti (Lotfitka) või Eesti (Laloritka) romadeks. Artikli eesmärk on pöörata tähelepanu roma keelele kui Eesti vähemuskeelele, mida on Eesti kontekstis väga vähe uuritud. Artikkel keskendub vene, läti ja eesti keelele, mis on oluliselt mõjutanud Lotfitka murret. Lisaks kontaktkeelele kirjeldatakse artiklis mõningaid Eestis kõneldava Xaladytka murde (Vene murre) mõjusid. Artikkel põhineb Helsingi Ülikooli roma keele dialektoloogia projekti raames kogutud 7 Eesti Lotfitka murde kõneleja ja 14 Läti Lotfikta murde kõneleja lingvistilistel tõlkeküsitlustel.

Kontaktkeelte mõju hindamiseks kasutatakse Thomasoni ja Kaufmani (1988) välja pakutud laenamise skaalat. Laenamisena mõistetakse keeleainese ülekandmist ühest keelest teise nii leksiikaalsel tasandil kui ka foneetilisel, morfoloogilisel ja süntaktilisel. Laenamise skaala on rakendatav loomulikult edasi antud keelte puhul. Eestis kõneldavat Lotfitka murret on järjepidevalt esimese keelena peres õpetatud ning seega kuulub murre laenamise tüübi alla.

Eestis kõneldavat Lotfitka murret on enim mõjutanud vene keel, mille mõju võib hinnata viietasemelise skaala piires foneetiliste joonte puhul neljandal ja muude tunnuste puhul kolmandal tasemel olevaks. Lisaks sõnavarale on laenatud mitmeid afikseid, näiteks verbiprefikseid ja omadussõnalisi sufikseid, mida liidetakse ka algupärasele sõnavarale. Süntaktilise muutusena jäetakse vene

keele eeskujul lausest välja koopula. Läti keele mõju võib hinnata tasemele kaks või kolm – laenatakse eessõnu, sidesõnu, verbiprefikseid ning läti keele mõjul eristatakse pikki ja lühikesi vokaale. Eesti keel on Eestis kõneldava Lotfitka murde kõige hilisem kontaktkeel ning lisaks sõnavara laenamisele on eesti keele eeskujul toimunud tähendusnihkeid. Eesti keelest laenatud sõnades säilitatakse vokaalid *õ*, *ä*, *ö* ja *ü*. Eesti keele mõju Lotfitka murdele on tasemel üks, sest peamiselt laenatakse sõnavara, mis ei ole põhisõnavara. Samas on eesti keele oskus Lotfitka romade seas levinud ning võib eeldada laenamise intensiivistumist.

Eesti Lotfitka romad käivad küllaltki tihedalt läbi Xaladytka romadega ning Lotfitka murdes on märgata varieeruvust, mida mõjutab roma keele Xaladytka murre. Murded on lähedas suguluses, kuuludes roma keele kirderühma.

# CODE-SWITCHING IN EMERGENT GRAMMARS: VERB MARKING IN BILINGUAL CHILDREN'S SPEECH<sup>1</sup>

Virve-Anneli Vihman  
University of Tartu

**Abstract.** This paper examines the code-switching of verbs in the speech of two children bilingual in Estonian and English (aged 3 to 7). Verbs typically have lower rates of code-switching than nouns, due to their central role in argument structure, lower semantic specificity, and greater morphological complexity. The data examined here show various types of morphological mixing, and include examples which violate the prediction from the literature that only finite verbs bear inflectional morphology from the other language, suggesting that children do not adhere to the same constraints as adults when code-switching.

**Keywords:** code-switching, acquisition, bilingual children, verb morphology

## 1. Introduction

The ways in which the cognitive, social and structural aspects of language interact are made particularly evident when two or more languages are at play. For more than one language to be used in a single social situation, the languages must also coexist in the speakers' minds, and the discourse often reflects this. When code-switching

---

<sup>1</sup> I gratefully acknowledge support from a Marie Curie IEF grant, funded by the European Union's Seventh Framework Programme (no. 623742), during the time this paper was written. I also wish to thank two anonymous reviewers, participants at the ISMBS 2015 (Chania), and the organisers and participants at 'The interaction between social factors and the occurrence of lexico-grammatical contact effects' at the Sociolinguistics Symposium 21 (Murcia), where I presented similar data, for their insightful discussion and comments. All remaining weaknesses are inevitably my own.

takes place within a clause, and more strikingly, across morphemes within a word, some of the (usually hidden) mechanisms underlying production are made visible. In first language acquisition research, this may be one route to evidence regarding children's knowledge and understanding of the structure of the languages they are acquiring.

In a prevalent view of code-switching, bilingual clauses are always structured according to one of the participating languages, meaning that the languages are clearly distinguishable and remain distinct in bilingual discourse. The language which imposes the basic structure of the clause is called the Matrix Language, and this approach has been best formalised and is best known as the Matrix Language Frame, or MLF, model (Myers-Scotton 2002, 2005; Myers-Scotton & Jake 2000). This has been further developed to incorporate evidence that morphemes are not all alike, as discussed below. One strength of this model is that it provides a principled framework for analysing code-switched utterances and for making predictions regarding code-switching. Myers-Scotton's approach focuses on code-switching within the clause, which is also the focus of this paper. She claims that „it is only in the bilingual clause that the grammars of both languages are in contact and [...] the basic hierarchical opposition [...] between the matrix language and the embedded language makes any sense“ (2005: 329).

An alternative view, which may also be construed as complementing and refining the MLF view, is that the two languages interact. Once two languages are used together in bilingual conversation, they may influence each other and they are both likely to show effects of mutual interaction, even when one is dominant or one is identified as the Matrix Language. As linguistic knowledge is not a steady state, but always dynamic, a bilingual language production model will allow interaction and mutual influence even in the production process (De Bot 2004, Vihman 2016). In addition, it has been shown that bilingual conversation is not always best analysed through a matrix language approach (Backus 2014, Johanson 2002, Verschik



2007), and that it is not always possible to identify a Matrix Language (Auer & Muhamedova 2005). The Matrix Language Frame model has also been criticised from a different perspective: e.g. MacSwan (2000, 2005) argues that some of the concepts are vaguely defined and overly rely on code-switching data rather than being independently motivated. We leave this open, and assume that the MLF framework is clear enough to be tested using the data we present.

This paper examines a dataset from two bilingual Estonian and English-speaking children, and focuses on utterances with code-switched verbs in order to explore predictions which have been made regarding morphology in code-switching. In particular, the study aims to investigate the following questions: Can children's code-switching data shed light on how much they generalise abstract knowledge in online production? Do children follow the same constraints proposed in the literature on code-switching more generally? What can the code-switching data reveal about the emergent knowledge of the bilingual child?

## 2. Background

Before turning to the data, this section introduces background relevant to the study. To begin with, section 2.1 discusses why it is important to investigate bilingual children's productions in addition to adults' discourse; following this, code-switching with verbs is discussed; an overview is given of relevant details of the morphosyntax of the two languages in this study in 2.3.

### 2.1 BILINGUAL CHILDREN'S SPEECH

Much of the research on childhood bilingualism has focussed on comparing groups of bilingual children with their monolingual peers, and asking whether we find quantitative or qualitative differences, and what factors affect any observed differences. Overall, findings suggest that bilingual acquisition proceeds along similar

developmental paths as monolingual acquisition, and the differences we find tend to be quantitative rather than qualitative, reflecting inevitable differences in amount of exposure to each language in comparison with monolinguals (see Unsworth 2013, Yip 2013). Additionally, much attention has been paid to the question of whether and to what degree the languages interact in bilingual children, with agreement that some interaction may occur, but only under certain conditions (see Hulk & Müller 2000, Müller & Hulk 2001). Interaction has been shown to take place more broadly (see Argyri & Sorace 2007, Pérez-Leroux et al. 2011), but the general consensus is that children separate their languages fairly early (Paradis & Genesee 1996). Even in the literature on adults, disagreement prevails as to what degree of interaction exists between languages in bilingual discourse (Backus & Verschik 2012, De Bot 2004, Johanson 2002). Young children in the process of acquiring their linguistic systems are still discovering what sorts of regularities to expect in language. They make predictions, discover co-occurrences and learn patterns in the languages they are exposed to. We may ask what mechanisms they use in online production to select lexical items from the appropriate language as well as how they structure utterances according to their emerging grammars.

Much recent research has demonstrated how implicit, statistical learning underlies first language acquisition, and operates on various levels of language (Boyle et al. 2013, Chang et al. 2012, Rebuschat & Williams 2012, Perruchet & Pacton 2006). Statistical learning, the „unconscious process of inducing structure and regularity from exposure to repeated exemplars“ (Boyle et al. 2013), has been applied to phonology, morphology and syntactic dependencies in first and second language acquisition. Typically, we think of first language learners as mostly operating via distributional, implicit learning, supplemented by a number of explicit learning processes. Generalisation over exemplars must take place, as learning of linear strings alone would mean that the productive output reflected only structures identical to what is heard in the input. However, the

attribution of pattern generalisation sometimes involves an inferential leap on the part of the analyst, since access to a child's input is limited, nor can we usually see the process underlying the output. This is related to the first research question: Does a particular utterance merely reflect knowledge of co-occurrences based directly on heard utterances, or does it derive from a more abstract level of knowledge, reflecting morphosyntactic generalisations?

Bilingual speech is a good locus for evidence of what knowledge underlies an utterance. With bilingual acquisition, the abstraction of grammatical and collocational regularities may be seen in innovative constructions on the part of the child. Constructions which the child has not heard in the ambient language may reveal something of the grammatical process underlying the production of an utterance. Code-switching is, of course, not limited to children, but the code-switching of young children, whose mental grammars are not yet fully formed, may provide insight into several questions of interest. First, does children's code-switching differ in quantity or quality from that of adults? The data reported on in this study come from the author's own diary notes, which neither provide information on frequency nor include a controlled study of the input speech. Nevertheless, we can assume for the purposes of this study that the intuitions are relatively accurate, regarding the types of code-switching which were present in the child-directed (the author's own and other familiar adults') speech and those which weren't. On the assumption, then, that the children in this study do indeed show evidence of combining their languages in different ways than adults do, we can ask these further research questions: Do the data reflect constraints proposed in the literature on code-switching; and what can these tell us about the emergent linguistic knowledge of the bilingual child?

## 2.2 VERBS IN CODE-SWITCHING

It has been noted in the code-switching literature that verbs show lower rates of code-switching than nouns (Bolonyai 2005, Cantone 2007:173, Myers-Scotton 2005: 330). This may be largely due to their central role in argument structure: because the verb forms both the semantic and syntactic core of the predicate, it can be seen to provide the core structure of a clause. Verbs also tend to carry more functional information and morphological complexity than other constituents, and to be less semantically specific, which may reduce the motivation to code-switch based on contextually specific lexical items (cf. Backus 2001, Isurin et al. 2009, Backus & Verschik 2012).

For examining the interplay between grammatical structure and lexical units, then, verbs are a useful category to explore. The languages included here both require the expression of certain information via morphological marking when a verb is utilised. Verbs are intimately connected not only to the argument structure of a clause, but also to the tense, aspect and mood of an utterance.

Additionally, clear predictions have been made in the Matrix Language Frame and 4M model of Myers-Scotton (Myers-Scotton 2002, 2005, Myers-Scotton & Jake 2000) regarding the morphosyntactic integration of code-switched verbs.

This model claims, first, that an asymmetry always exists between the participating languages within a clause. The Matrix Language is that which structures the bilingual clause, in that it supplies the morphosyntactic frame.<sup>2</sup> Crucially, the model predicts that only Matrix Language morphemes will indicate grammatical relations within mixed constituents.

The 4M model refines this to differentiate between four types of morphemes. First, content morphemes, lemmas with semantic

---

<sup>2</sup> A possible composite Matrix Language Frame has been proposed (e.g. Bolonyai 2002, Myers-Scotton 2002), but has not been fleshed out, nor have the implications been fully explored.

content associated with assigning or receiving thematic roles, are set apart from system morphemes. System morphemes are further subdivided into three types, early system morphemes and two late system morpheme types. Both the (thematic role-assigning) content morphemes and (non-assigning) early system morphemes (e.g. determiners and derivational affixes) are conceptually activated, and accessible in the mental lexicon. Late system morphemes, on the other hand, express grammatical information, are structurally assigned and hence accessed later in the production process; they are claimed to be available at the level of the formulator (Levelt 1989) and can therefore participate in code-switching. They mark relations between constituents rather than within a single constituent. These are further subdivided into bridge system morphemes and late outsider morphemes. Person and number in a noun phrase can be accessed within the noun phrase, and are therefore marked with early system morphemes. Person and number marking on a verb, however, is marked with late system morphemes, as it relies on information outside the verb phrase. Case-marking on a noun also derives from information outside the constituent; therefore case markers are also classed as late outsider system morphemes.

According to the 4M model, then, only finite verbs are predicted to bear inflectional morphology from the other language: non-finite verb forms are predicted to be accessed holistically and early, and hence ought not to participate in code-switching. Myers-Scotton & Jake (2000) note that all examples of content words from the embedded language with matrix language inflections<sup>3</sup> support the notion that lexical words with regular morphology are constructed online. We return to this notion after discussing the data.

---

<sup>3</sup> This is applied to outsider late system morphemes (Myers-Scotton 1997), but it is not always immediately clear how this plays out in languages other than the ones discussed. Bolonyai (2000), for instance, makes a case for Hungarian preverbs to be treated as both early and late morphemes, partly on the basis of code-switching data which indicate differential behavior.

### 2.3 ESTONIAN AND ENGLISH

The two languages in this study, one Finnic and one Germanic, have both congruences and dissimilarities which create the potential for structural conflict as well as innovative combinations of structures. English, as is well known, marks argument structure primarily through word order, with a very limited range of (pronominal) case-marking. Estonian, on the other hand, uses a complex system of morphological case-marking to signal grammatical relations, with more flexible word order. English is analytic, whereas Estonian is fusional-agglutinative (Erelt 2003), and uses a combination of post-positions, prepositions, and cases to signal the adverbial meanings usually encoded by prepositions in English.

The difference in morphological richness can be easily demonstrated by the verb paradigms. In English, a total of three overt morphemes are used in regular, finite verb conjugation, as well as frequent zero marking, and only one of the overt affixes signals person/number distinctions. In general, verbal marking signals tense or aspect, with present tense given null inflection in all but one cell, and an overt morpheme (*-ed*) marking past tense. The gerund (*-ing*) is a frequently used verb form which neutralises other grammatical information in the lexical verb, leaving person, tense and number marking to the auxiliary.

In Estonian, on the other hand, each cell in the regular verb paradigm for present and past tense is distinct (with one exception, the isomorphism of 2sg past and 3pl past), and each is marked by overt morphology. As shown in Table 1, the system is nearly agglutinative, with one affix marking past tense and a similar set of affixes marking person/number in both present and past tense: the first and second person morphemes are identical in present and past tense, whereas third person marking is unique in each cell. A total of eleven distinct verb endings are used for regular verbs in present and past tense. This is summarised in Table 1.

**Table I.** Regular verb paradigms in English and Estonian

			1 <sup>st</sup> person	2 <sup>nd</sup> person	3 <sup>rd</sup> person
English	Present	SINGULAR	-s		
		PLURAL			
	Past	SINGULAR	-ed		
		PLURAL			
Progressive		-ing			
Estonian	Present	SINGULAR	-n	-d	-b
		PLURAL	-me	-te	-vad
	Past	SINGULAR	-si-n	-si-d	-s- $\emptyset$
		PLURAL	-si-me	-si-te	-si-d

An additional complicating factor is that Estonian verb paradigms often involve stem changes. For instance, the verb *rääki-ma* ‘to talk’ has two stems (*räägi-* and *rääki-*), used in different parts of the paradigm. Present tense forms use the former (e.g. 1sg *räägi-n*, ‘I speak’), whereas the past tense is formed from the latter (e.g. 1sg *rääki-si-n*, ‘I spoke’). Participial forms of one verb may also derive from different stems in a paradigm, e.g. *katma* ‘to cover’ > passive participle: *on kae-tud* ‘is covered’ vs. active participle *on kat-nud* ‘has covered’). There is no room here to go into detail (instead see Blevins 2007, Erelt et al. 1995, Viks 1992), but the above will be relevant for some of the examples discussed in section 4.

In summary, it is crucial that the use of finite verbs in Estonian necessarily involves inflectional endings. When finite verbs are embedded in code-switched utterances framed by the other language, the speaker is forced to make a choice of how to mark the verb, according to the source or target language.

### 3. Data and Method

The data examined in this study derive from a diary kept by myself of my daughters' utterances. The children are bilingual sisters, aged 6;6-7;11<sup>4</sup> (M) and 2;10-4;3 (K) during the time of the study. The family was living in Estonia. The father, a native Estonian speaker, spoke Estonian with all family members; the mother, raised bilingually in the United States, spoke English with the children and Estonian with the father.

The family, thus, represents a one-parent, one-language model, yet did not strictly observe this model, as when the family setting naturally inclined one or the other parent to join in a discussion in the other language, or when other speakers were present. Estonian was the common parental language, as well as the main social language spoken by most of the family's interlocutors outside the home. The children were attending full-time day-care and school, which took place entirely in Estonian.

Some friends and relatives spoke English with the family. The children visited their grandmother and other relatives and friends at least once a year in English-speaking settings. Additionally, although television was not a constant fixture, it must be said that the abundance and popularity of English-language entertainment and educational media may have had an effect on the children's language skills and preferences. Although the children were born and raised in Estonia, they spent much of their free time (outside educational institutions) with their mother, thus ensuring sufficient English exposure and identity to lead to a preference for English even with each other. During the year when they attended the same day-care center ('pre-school'), they were reported to often speak English with each other on school premises, either as a sign of solidarity and

---

<sup>4</sup> This notation is commonly used in developmental literature to indicate age: it is to be read as x;y.z = x years, y months, z days. I will continue to use this notation following examples to note age of speaker.



belonging, or else to distinguish themselves from others. Nevertheless, their language dominance is not easy to assess, and having used no formal methods of assessment, I will refrain from saying anything further on the matter.

The diary data were gathered during everyday activities by myself (the children's mother), though at the time of the examples used in this study this did not reflect all mixed utterances, but rather those that were particularly noteworthy for any reason (e.g. they were not typical of adult code-switching), and those which I was able to record in my notes, often doing so in the course of a busy work and family schedule. The examples included here come from a corpus of 600 examples of mixed utterances, of which 85% come from K (2-4 years old). It may be worth noting that the high proportion of examples with Estonian insertions into English utterances does not necessarily say anything about actual frequency of the direction of insertions, but rather reflects the bias created by the observer's role (or Observer's Paradox, Labov 1972, Lanza 1997) – since I spoke English with the children, the utterances heard and noted most frequently came from an English discourse context.

The data were combed for examples with code-switched verbs. All examples with verb and predicate insertions were collected and analysed for type of code-switching and amount of inflectional morphology involved in the code-switching. The results are discussed in the next section.

#### 4. Results

As mentioned above, verb insertions necessarily involve choices regarding the morphology associated with the verb. The examples in the dataset were grouped according to what sort of verb morphology was used. In the following, I discuss the code-switched utterances, first finite and then non-finite verbs, according to the morphological make-up of the bilingual clause.

#### 4.1 FINITE VERBS

Numerous verbs in the data involved code-switching. Especially the younger child, K, used various strategies for integrating embedded verbs into clauses in the other language. These included three main combinations: null inflection, embedded language inflection, and matrix language inflection. Each of these are exemplified and discussed below.

##### 4.1.1. FINITE VERBS WITH NULL INFLECTION

Estonian verbs are often used in the data in English-language utterances without any morphological marking. Because many Estonian verbs undergo stem changes in their inflectional paradigms, it is pertinent to ask which stems are selected by the child in the context of insertions into another language. The data include stems which are used as uninflected bare stems in certain contexts in Estonian, such as those in (1–2). Estonian uses these stems in contexts with no affix: for instance, in negative clauses (where the negative particle blocks any person/number marking) and second person singular imperative constructions.

- (1) You *kleebi-ø* [stick] the wings on. (K, 3;10.2)  
 (2) When we do the *õppimise* [learning] thing...  
       when we *õpi-ø* [learn] then we don't go upstairs (K, 3;10.16)

The inserted stem which appears in examples (3–4), on the other hand, is never used alone in Estonian. The verb *jätma* 'to leave. TRANSITIVE' has alternating stems used in different parts of the inflectional paradigm, *jät-* and *jäta-*. The stem ending in a vowel is used with indicative personal endings and contexts without any inflection (2sg imperative, negative), whereas the shorter stem without the vowel is used in past tense forms and some participle formation. However, the CVC form of the shorter stem is more in accord with English phonology, which may be what prompted the forms used in (3–4).

- (3) *Jät-ø* [leave] some for me too! (K, 3;8.24)  
 (4) I *jät-ø* [leave/left] some for [M...] because she's hungry (K, 3;8.24)

The lack of inflection in many cases can be analysed as following the grammar of English, where typically no inflectional morphology is required (e.g. for present tense 2sg and 1pl in examples 1–2). However, example (4) is intriguing from this point of view, as this utterance denotes a past event and would require past-tense morphology in both English and Estonian. As English past tense verbs usually end in /d/ or /t/, and the lexeme in question corresponds to a one-syllable, irregular verb ending in /t/ in English, the stem *jät* is phonologically felicitous in that it is prosodically similar to the appropriate form, although it does not follow the morphological paradigm of either English or Estonian (though note the existence of similar-sounding, null-marked past-tense verbs in English such as 'set' and 'put').

Hence, examples (1–4) can all be analysed as lacking morphology. Yet, it is not entirely clear which language this should be related to: should it be interpreted as lacking morphology from the perspective of English, which requires null morphology in many of these examples, or from the perspective of Estonian, which would require overt endings. As all these examples have a clear English matrix frame, we would be justified in saying that null morphology is appropriate, yet the form selected for the embedded verb in (4) remains a puzzle. Does this form indicate a misanalysed Estonian stem, or does it reflect a subtle interaction of the morphophonological properties of the lexeme, drawing on knowledge from both languages? Bilingual adults in ordinary code-switching environments might use the form *jäta-* with no ending, but *jät-* is anomalous.

Examples (5–8) all have predicate verbs formed with an auxiliary and lexical verb. The lexical verb comes from language B (Estonian, in these instances) and is inserted with no inflection, appropriately for language A (English, which makes frequent use of auxiliaries). Hence, these are all cases of B-insertions which can be analysed as following A-morphology. In (8), we see another example like (3–4)

above, wherein the selected stem does not appear on its own as a bare stem in Estonian.

- (5) Then I **can't** *kasta-ø* [water.TRANS] the flowers when there isn't any rain to go in here. (K, 3;11.3)
- (6) I **don't** *mäleta-ø* [remember] what it was (K, 3;11.2)
- (7) Jakob says that the *soovikivi* [wishing-stone.NOM] **doesn't** *täida-ø* [grant wishes.NOM.PL] *soovid*. (M, 7;6.12)
- (8) These **will** *kadu-ø* [get-lost] in here. (K, 3;11.0)

Also importantly from the point of view of the Matrix Language Frame and the morphology used, two of these examples have direct objects. In (5), only the lexical verb is inserted, with the auxiliary 'can', the negative particle and the direct object all in English and following English matrix structure. Example (7), on the other hand, is a more complicated mix of two grammars. The matrix clause involves reported speech, which is most likely the reason this utterance includes several inserted lexemes: the subordinate clause takes its subject *soovikivi* 'wishing stone', lexical verb *täida* 'fill', and direct object *soov-id* 'wish-NOM.PL' from language B. Moreover, one might assume the phrase 'grant wishes' is taken wholesale from Estonian, but in Estonian, negative polarity requires an object case-marked with partitive case. This is an exceptionless and early-learned syntactic rule in Estonian (Argus 2009, Erelt et al. 1995), but this utterance is an example of two languages interacting in complex ways. Although the lexical verb, and indeed all the content words, are taken from Estonian, the core argument structure derives from English. English structure is marked by the definite article 'the', the negative contracted auxiliary 'doesn't', and the lack of partitive case-marking on the object.

## 4.1.2. FINITE VERBS WITH MORPHOLOGICAL INFLECTION

Finite verbs also appear in the data with inflection from both languages. In examples (9–11), the embedded language, B, contributes the lexical verb together with its inflectional morphology.

- (9) Jüri *kriimusta-s* [scratch-3SG.PAST] my hand when we were going outside. (K, 3;10.16)
- (10) When I was there, *Vanaema* [Granny] *korista-s* [clean-3SG.PAST] the *tänav* [street.NOM.SG]. (K, 3;9.21)
- (11) I knew daddy's letter but our teacher *ikka* [still] *õpeta-s* [teach-3SG.PAST] (**it to us**). (K, 3;9.21)

All these examples are third person singular and past tense, and hence would require morphology in either language. Importantly, example (10), like (5) and (7) above, has a direct object which reveals more of the syntactic structure of the clause. *Tänav* 'street' is in nominative case (the unmarked subject case), whereas in Estonian it would require overt, direct object case-marking (partitive or genitive). Additionally, it has an English definite article. Hence, this example contains system morphemes from two different languages, the past tense from Estonian immediately followed by a definite article and nominal case taken from English structure.

Myers-Scotton's 4M model predicts the usage of matrix-language tense and aspect inflection with embedded verbs: B-items with A-inflection. Examples (12–15) follow this pattern, but in the collected data are not distinguished from the examples in (9–11) by context, content or age. The examples are interspersed and show no detectable regularity of usage patterns.

- (12) Today we *võimle-d* [exercise-ENG.PAST] at preschool (K, 3;8.15)
- (13) Bazoo [=cat] *tagurda-d* [went-backward-ENG.PAST]. Like this. (M, 7;9.17)
- (14) When the big boy *karju-s* [yell-s] in the bathroom then my friends can't sleep. (K, 3;5.7)

- (15) *Issi kui me käisime arstis siis ma choose-i-si-n*  
 Daddy when we went doctor-LOC then I choose-PAST.1SG  
*selle sparkly konna.*  
 this.GEN sparkly frog.GEN  
 ‘Daddy when we went to the doctor’s then I **chose** this **sparkly**  
 frog.’ (K, 3;9.14)

Examples (12–13) both involve English past tense inflection on Estonian lexical verb stems. In (14), the verb has an ending which appears in Estonian on verbs, and in fact the form *karjus* means ‘yelled’. However, in this context, the subsequent clause disambiguates the *-s* ending as an English 3SG present tense inflection (rather than Estonian 3SG past): ‘When the big boy yells, then my friends can’t sleep’.

Finally, in (15), an Estonian matrix clause has an inserted English verb. The verb in English is irregular, and would produce a one-syllable past tense with no /d/ ending (‘chose’), but here it is given regular Estonian inflection, with a default theme vowel *-i-*, the past tense *-si* and first person singular *-n*. In this case, despite the B-inserted verb and adjective, the direct object is inflected fully according to Estonian object case-marking paradigms, with genitive case.

Example (14), though it can be straightforwardly analysed as an Estonian verb with English inflection, raises the issue of morphology seemingly ‘blended’ from the two languages, where it is not clear which is the source language. This might be better described, especially in the children’s code-switching data, as a means of combining inflectional resources from the available languages.

This leads to the question of analysis of examples such as those in (16–19), which were frequent for a period of nearly six months. In these, a clause typically displaying English matrix clause structure employs an inserted Estonian verb in a predicate in progressive aspect. However, the lexical verb is marked with a *-n* affix. This is simpler phonologically than *-ing*, as well as following the phonological structure of Estonian, which lacks /ŋ/. However, in Estonian, the *-n* inflection on verbs indicates first person singular. In (16), then, the *-n* could be drawing on the Estonian structure not

only phonologically, but also morphologically. It is clear that *-n* is an adapted form of the progressive ending, but it may also incorporate the frequent Estonian verbal *-n* morpheme, to produce a form which draws on grammatical resources and satisfies grammatical requirements of both languages. From examples (17–18), we can see that the *-n* ending is not limited to first person contexts.

- (16) I wasn't hitting you, I **was** *koputa-n* [knock-n] (K, 3;5.24)  
 (17) Why you're *kirjuta-n* [write-n], Mommy? (K, 3;5.25)  
 (18) a. K: This is a bad plate because this guy **is** *aja-n tema taga* [drive-n 3SG.NOM/GEN behind].  
       b. Mother: What is he doing?  
       c. K: He **is** *aja-ing tema taga* [driving 3SG.NOM/GEN behind = chasing him]. (K, 3;5.17)  
 (19) We **are** *mängi-n* [play-n] *peitust* [hide-and-seek.PAR] with the *karu* [bear] (K, 3;5.10)

In (18a), K comments on a picture of Mr. McGregor chasing Peter Rabbit with a rake, depicted on her dinner plate. The entire matrix clause and beginning of the subordinate clause are in English, until the lexical verb *taga ajama*, 'chase'. This is a challenging word for code-switching, a phrasal verb composed of a light verb *ajama* 'drive/make/push' and adverbial *taga* 'behind', the semantics of which is not compositional, but depends on the whole phrase. Perhaps because the direct object occurs inside the complex verb, the language shifts at this point, and the rest of the clause is in Estonian, but the verb has the same *-n* ending which seems to mark the progressive in combination with the English auxiliary, and the pronominal direct object is not in the expected partitive case, but rather in nominative / genitive (ambiguous with this pronoun). The construction seems so surprising that I ask K to repeat it, upon which she slows down and enunciates the progressive marker, resulting in a more difficult word to pronounce (with two successive non-diphthong vowels from *aja* + *-ing* as well as the /ŋ/ ending), while also demonstrating that she knows and can say the English progressive *-ing*.

Finally, the example in (19) involves the same pattern, this time in first person plural, with an English auxiliary and Estonian verb ‘play’ with the *-n* ending, and with a direct object following Estonian case-marking patterns.

Examples with the type of ‘blended’ morphology discussed above and exemplified in (14) and (16–19) were regularly present in K’s speech for approximately six months. These always involved Estonian verbs inserted into English utterances, with morphology which can be seen as drawing on both languages. A limited set of three inflections were used: *-d*, *-s*, and *-n*. These can all be mapped to English verb endings, but they (a) suit the Estonian verbs phonologically and (b) also occur in the Estonian verbal paradigm, though with different meanings, while (c) being selected according to similarity with English morphemes to map to grammatical functions required by English grammatical structure: *-d* (past), *-s* (third person singular present) and *-n* (progressive *-ing*).

#### 4.2 NON-FINITE VERBS

The above examples demonstrate that the child is constructing finite verb forms online, as well as producing flexible argument structure constructions around them. According to Myers-Scotton, the combining of regular inflectional morphemes from one language with content morphemes from another is „strong evidence that regular inflections are supported as individual elements in the mental lexicon“ (2005: 333). She claims that: „code switching also provides good evidence that Jackendoff’s semi-productive elements are based on single units in the mental lexicon; they are not constructed on line. The evidence is that Embedded Language nonfinite verb forms, especially for the participles, from different languages always appear as holistic units in code switching“ (Myers-Scotton, 2005: 333).

In this section, we examine code-switched participles to investigate whether this claim is well founded. Participles do participate in the code-switching data in question. In (20), K inserts a participle,



taken wholesale from Estonian, as predicted by Myers-Scotton, along with the participial morphology and the particle which comprises the phrasal verb ‘tied up’.

- (20) When your eyes are *kinni* [‘closed’] *seo-tud* [‘tied’] then you can’t see. And then you break piñatas and then you get candy!

(K, 3;4.27)

In (21–22), however, we see that participles are not necessarily retrieved as holistic units, which may cast doubt on the argument above. If participles are always single units in the lexicon, that implies that they cannot be formed through morphological patterning. Nevertheless, examples of children producing erroneous forms such as ‘eated’ or ‘putten’ are not uncommon in English, indicating that either the story of mental storage of participles is more complicated than suggested, or else that children store them differently, perhaps not having yet acquired the crucial distinction between holistic units and grammatical forms constructed online. We first add to this question attested examples of participial forms used as insertions in code-switched utterances with erroneous, Estonian-based forms in (21–22). The error in example (21) arises from selection of the incorrect stem, while the error in (22) involves an incorrect allomorph.

- (21) Now Katie’s necklace is \**tee-dud* [‘made’, pro: ‘*teh-tud*’]

(K, 3;6.2)

- (22) Now it’s even more \**mur-dud*. [‘broken’ pro: ‘*mur-tud*’]

(K, 3;11.8)

However, more problematic for Myers-Scotton’s argument is the fact that we also find examples of participles with code-switching across morpheme boundaries, a phenomenon which suggests that these forms are not simply stored or retrieved as holistic forms and that they may involve online construction. In (23), the Estonian verb *jätma* is given an English participial ending, resulting in *jät-en*, to express an English phrasal verb ‘left over’. In (24), an English verb is inserted into an Estonian matrix clause and combined with an

Estonian participial ending. The affix involves metathesis, presumably as a consequence of the repetition of 'i' in the lexical verb, but the use of a participle here is in accord with Estonian syntax: negative past is formed with the negative particle *ei* and an active past participle, typically ending in (Vowel)-nud.

(23) Kribu [=cat] ate a drumstick that was *jät-en* [leave-ENG.PRT]  
over (K, 3;6.5)

(24) *Ma jätsin pildi mida ma ei finish-onid*  
I leave.PAST.1SG picture what.PAR 1SG NEG finish-EST.PRT  
*lasteaeda*  
at-preschool  
'I left a picture that I didn't finish at preschool.' (K, 3;11.3)

These and other examples of code-switching in nonfinite verb forms across morpheme borders are clear violations of Myers-Scotton's prediction that participles, as holistic units, are embedded holistically in code-switching. They occur rarely enough that it would not seem plausible to claim the verbal stems are borrowings, for instance, but their recurrence in at least some children's productions merits attention in any framework for analysing code-switching data.

## 5. Summary and Conclusions

The examples discussed here are drawn from diary data from two children simultaneously bilingual in Estonian and English, with fairly balanced competence in both languages. As shown, verb insertions appear in the children's code-switching with various combinations of morphological marking. This includes embedded verbs with:

- a. zero marking in constructions which may be either well- or ill-formed in the matrix language of the utterance;
- b. morphology from the source language;
- c. morphology and function words from the target language;
- d. a blend, where both source and target languages constrain the forms used in mixing.

Returning to the research questions posed at the beginning, we claim that children's code-switching provides a fruitful vantage point for analysing their emerging knowledge of grammar and linguistic structure. The data examined here come from only two children and two languages, but suggests strongly that the proposed constraints are either not followed by children or not as generally applicable as has been suggested in the literature. This study looks at only a subset of data involving verbs, but it holds promise for future research on interaction between languages, emergent grammatical systems, and possible combinations in code-switching, drawing on grammatical resources from both languages.

Children, whose grammatical competence is less than fully formed, are in a continuous process of discovering regularities, making predictions, and analogising across forms, in addition to learning new concepts and vocabulary. Some code-switching reflects an imbalance in lexical knowledge between the languages, but the more interesting examples of how those embedded words are grammatically accommodated may shed light on the knowledge underlying children's productions. When they embed a verb from one language in an utterance which begins in a different language, they must immediately make choices regarding the use of grammatical resources and how they interact. Data from code-switching is akin to a naturally occurring novel-word experiment. The degree and type of abstraction of regularities can be clearly seen in some of these innovated, online constructions.

Examples from the data show that both finite and nonfinite forms can occur with morphemes drawn from either the source or target language. It is beyond the scope of this paper to assess whether these constitute evidence for a lack of separation in the emergent morphological inventory, but they raise questions.

The division of lexical items into function and content morphemes may not always be cross-linguistically comparable. In analysing a child's developing grammar, these categories must not be taken for granted, either. Likewise, the question of separation of lexicons

from each language must be revisited, as it is unclear whether bilingual children draw on two distinct lexical inventories, and how the distinctions between the languages emerge. Individual differences compounded with language experience may lead to greater effects of differential inhibitory control, executive function and verbal memory in bilingual children. This multiplied effect may also lead to greater differences among bilingual than monolingual children in the storage and retrieval of lexical items and grammatical structures.

Finally, violations of predicted constraints in children's code-switching data may mean that the predictions are incorrect, but it may also mean that children do not adhere to the same constraints on code-switching as adults do. This needs to be further explored, as much for analysing code-switching data as for what this implies about the developing cognitive systems, before we assess the merits of one or another set of constraints on code-switching. Regardless, it must be noted that these constraints have been criticised also from the perspective of adult code-switching (e.g. Backus 2014, Auer & Muhamedova 2005). The creative, online code switching of children affords a window onto the mutual influence of the languages in the mind, as well as the child's emergent knowledge of both linguistic systems.

#### REFERENCES

- Argus, Reili 2009. The early development of case and number in Estonian. In M. D. Voeikova & U. Stephany (Eds.) *Development of Nominal Inflection in First Language Acquisition: A Cross-Linguistic Perspective*, pp. 111–152. Berlin: Mouton De Gruyter.
- Argyri, Efrosyni; Sorace, Antonella 2007. Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition*, 10(1): 79–99.
- Auer, Peter; Muhamedova, Raihan 2005. 'Embedded language' and 'matrix language' insertional language mixing: Some problematic cases. *Italian Journal of Linguistics / Rivista di linguistica*, 17(1): 35–54 (special issue, ed. G. Berruto).

- Backus, Ad 2014. Towards a Usage-Based account of language change: Implications of contact linguistics for linguistic theory. In R. Nicolai (ed.) *Questioning Language Contact: Limits of Contact, Contact at its Limits*. Leiden: Brill, pp. 91–118.
- Backus, Ad; Verschik, Anna 2012. Copiability of (bound) morphology. In L. Johanson & M. Robbeets, *Copies versus Cognates in Bound Morphology*. Leiden: Brill, pp. 1–32.
- Bolonyai, Agnes 2000. „Elective affinities“: Language contact in the abstract lexicon and its structural consequences. *International Journal of Bilingualism*, 4(1): 81–106.
- Boyle, Whitney; Lindell, Annukka; Kidd, Evan 2013. Investigating the Role of Verbal Working Memory in Young Children’s Sentence Comprehension. *Language Learning*, 63(2): 211–242.
- Chang, Franklin; Janciauska, Marius; Fitz, Hartmut 2012. Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistic Compass*, 6(5): 259–278.
- De Bot, Kees 2004. The Multilingual Lexicon: Modelling Selection and Control. *International Journal of Multilingualism*, 1(1):17–32.
- Demirçay, Derya; Backus, Ad 2014. Bilingual constructions: Reassessing the typology of code-switching. *Dutch Journal of Applied Linguistics*, 3(1):30–44.
- Erelt, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1995. *Eesti keele grammatika*. [=Grammar of Estonian]. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- Erelt, Mati 2003. *Estonian Language*. (Linguistica Uralica, Supplementary Series 1) Tallinn: Eesti Akadeemia.
- Hulk, Aafke; Müller, Natascha 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, 3(3): 227–244.
- Johanson, Lars 2002. Contact-induced linguistic change in a code-copying framework. In: M.C. Jones & E. Esch (eds.) *Language change: The interplay of internal, external and extra-linguistic factors*. Berlin: Mouton de Gruyter, pp. 285–313.
- Levelt, Willem J. M. 1989. *Speaking, from Intention to Articulation*. Cambridge, MA: MIT Press.

- Müller, Natascha; Hulk, Aafke 2001. Crosslinguistic Influence in Bilingual Language Acquisition: Italian and French as Recipient Languages. *Bilingualism: Language and Cognition*, 4(1): 1–21.
- Myers-Scotton, Carol 2002. Contact linguistics, bilingual encounters and grammatical outcomes. OUP.
- Myers-Scotton, Carol 2005. Supporting a Differential Access Hypothesis. In J. Kroll & A. de Groot (eds). *Handbook of bilingualism, psycholinguistic approaches*, pp. 326–348. New York: OUP.
- Myers-Scotton, Carol; Jake, Janice L. 2000. Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics* 38: 1053–1100.
- Paradis, Johanne; Genesee, Fred 1996. Syntactic Acquisition in Bilingual Children. *Studies in Second Language Acquisition*, 18: 1–25.
- Pérez-Leroux, Ana Teresa; Cuza, Alejandro; Thomas, Danielle 2011. Clitic placement in Spanish / English bilingual children. *Bilingualism: Language and Cognition*, 14: 221–232.
- Perruchet, Pierre; Pacton, Sébastien 2006. Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233–238.
- Rebuschat, Patrick; Williams, John N. (eds.). 2012. *Statistical learning and language acquisition*. Berlin: Mouton de Gruyter.
- Unsworth, Sharon 2013. Current issues in multilingual first language acquisition. *Annual Review of Applied Linguistics*, 33: 21–50. <http://dx.doi.org/10.1017/S0267190513000044>
- Verschik, Anna 2007. Multiple language contact in Tallinn: Transfer B2 > A1 or B1 > A2? *The International Journal of Bilingual Education and Bilingualism*, 10 (1): 80–103.
- Vihman, Marilyn M. 2016. Prosodic structures and templates in bilingual phonological development. *Bilingualism: Language and Cognition*, 19: 69–88.
- Yip, Virginia 2013. Simultaneous language acquisition. In F. Grosjean and L. Ping (eds). *The Psycholinguistics of Bilingualism*, Wiley-Blackwell, pp. 119–136.

## RESÜMEE

### KOODIVAHETUS ARENEVAS GRAMMATIKAS: VERBI MARKEERIMINE KAKSKEELSETE LASTE KÕNES

Artiklis vaadeldakse kahe 3–7aastase eesti ja inglise keelt rääkiva kakskeelse lapse tegusõnade koodivahetust. Verbid osalevad tavaliselt nimisõnadest harvemini koodivahetuses ja seda eeldatavasti mitme teguri tõttu: keskne roll argumentstruktuuris, vähem spetsiifiline semantika ning morfoloogiline komplekssus. Vaadeldavad andmed hõlmavad erinevaid morfoloogilisi võtteid ning sisaldavad ka näiteid, mis räägivad vastu hüpoteesile, et vaid finiitsed verbid kannavad koodivahetuse puhul teise keele morfeeme. Need näited toetavad väidet, et laste koodivahetus ei allu samadele piirangutele nagu täiskasvanute koodivahetus.