

7 (2022)

Philologia Estonica

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

RESEARCH IN CHILDREN'S AND YOUTH MULTILINGUALISM

*Uurimusi laste ja noorte
mitmekeelsusest*

Toimetaja/Editor Anna Verschik

Tallinna Ülikooli Kirjastus
Tallinn 2022

Philologia Estonica Tallinnensis 7 (2022)
Research in children's and youth multilingualism
Uurimusi laste ja noorte mitmekeelsusest

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

Toimetuskolleegium / Advisory Board
Helle Metslang (Tartu, Tartu Ülikool), Margit Langemets (Tallinn, Eesti Keele
Instituut), Meelis Mihkla (Tallinn, Eesti Keele Instituut), Renate Pajusalu (Tartu,
Tartu Ülikool), Maria Voeikova (Viin/Sankt-Peterburg), Cornelius Hasselblatt
(Zuidhorn / Eesti Teaduste Akadeemia), Epp Annus (Tallinna Ülikool / Ohio
State University)

Keeletoimetajad: Rita Anita Forssten, Helin Kask, Eva-Maria Kinkar
Tõlkija: Geidi Kilp
Kaanefoto: Roman Nogin / Shutterstock

Väljaannet toetab grant EKKD33 „Eesti laste ja noorte
mitmekeelse suhtluse keelekogud ja korpused“

Autoriõigus: Tallinna Ülikooli Kirjastus ja autorid, 2022

ISSN 2504-6616 (trükis)
ISSN 2504-6624 (võrguväljaanne)

Tallinna Ülikooli Kirjastus
Narva mnt 25
10120 Tallinn
www.tlupress.com

Trükk: Grano Digital

SISUKORD

Anna Verschik

Sissejuhatus.....	7
<i>Introduction.....</i>	10

Elena Antonova-Ünlü, Çiğdem Sağın-Şimşek

Connectivity in Narratives of Turkish-English and Turkish-Russian Bilinguals.....	13
<i>Türgi-inglise ja türki-vene kakskeelsete narratiivide siduselementide kasutus.....</i>	37

Piret Baird, Reili Argus, Merilyn Meristo

Eesti keelt teise keelena omandavate laste eesti keele oskuse areng aasta jooksul.....	39
<i>Children's Estonian as a Second Language Development During a Year</i>	76

Piret Baird

Enabling tool: Estonian-English Code-Mixing of a 2-Year-Old with Balanced Input.....	80
<i>Võimendav tööriist: Eesti-inglise tasakaalus sisendiga kakskeelse 2-aastase koodivahetus.....</i>	102

Vlada Baranova, Kapitolina Fedorova

Accents of Russian in Performative Use: Ethnic Styles, Language Attitudes, and Identities of Young Non-native Speakers.....	103
<i>Vene keele aktsendid performatiivses kasutuses: noorte muulaste rahvuslikud stiilid, keelehoiakud ja identiteedid.....</i>	125

Inga Hilbig

Zwei Sprachen gleichzeitig? Nein, das schaff' ich nicht:

A Lithuanian-German Boy's Journey to Active

Bilingualism 126

Zwei Sprachen gleichzeitig? Nein, das schaff' ich nicht:

leedu-saksa poisi teekond aktiivse kakskeelsuseni 147

Oleksandr Kapranov

The English Fricative Consonant /z/ as a Challenge

to Norwegian L1 EFL Learners: An Error Analysis

of Phonemic Transcriptions 148

Inglise frikatiiv /z/ kui väljakutse norra emakeelega

inglise keele õppijatele: foneemiliste transkriptsioonide

veaanalüüs 185

Victoria V. Kazakovskaya

Acquisition of Russian nominal derivation

in monolingualism and bilingualism 186

Vene nimisõnatuletuse omandamine üks- ja kakskeelses

situatsioonis 214

Geidi Kilp

The Insertion of Person References Motivated by Pragmatic

Differences in Estonian-English-Japanese Facebook

Communication 215

Pragmaatilistest erinevustest ajendatud isikuviiidete

sisestamine eesti-inglise-jaapani Facebooki vestlustes 238

Eglė Krivickaitė-Leišienė, Ineta Dabašinskienė

The bilingual advantage: performing the non-word

repetition test 239

Kakskeelsuse eelis: väljamõeldud sõnade kordamise test 262

**Kristiina Praakli, Mari-Liis Korkus, Aive Mandel,
 Elisabeth Kaukonen, Annika Kängsepp, Triin Aasa,
 Kristel Alvere, Helen Eriksoo, Marion Mägi,
 Getri Tomson, Liina Lindström**

„Mis keeles ma räägin, <i>I don't know</i> “. Eesti sisuloojate inglise keele kasutusest YouTube'is	263
„Mis keeles ma räägin, <i>I don't know</i> “. <i>On the use of English by Estonian YouTubers</i>	290

**Virve-Anneli Vihman, Kristiina Praakli,
 Maarja-Liisa Pilvik, Mari-Liis Korkus**

Kas noored on inglise keelega <i>obsessed</i> ? Millest räägivad korpusandmed?	292
<i>Are Estonian teenagers obsessed with English? What do corpus data say?</i>	321

SISSEJUHATUS

Anna Verschik

Tallinna Ülikool

Artiklikogumik käsitleb laste ja noorte mitmekeelsust. Kuigi mitmekeelsusest üldiselt ning konkreetselt laste ja noorte mitmekeelsusest on juba ilmunud hulgaliselt kirjandust ja ilmub ka edaspidi, tundub, et mitmekeelsuse eri tahud on ammendamatu teema, mis on jätkuvalt oluline nii keeleteaduse teooria kui rakenduslike uurimuste jaoks. Nõnda juhtub osaliselt sellepärast, et kogu aeg tekib uusi sotsiolingvistikulisi olukordi ja uute keelte kombinatsioone (nagu näiteks türgi-vene varajane kakskeelsus); aga ka sellepärast, et juba tuntud mudeleid ja meetodeid rakendatakse uutes keelesituatsioonides, nagu näiteks hiljuti alustatud Eesti noorte keele uurinud. Samuti loovad uus meedia ja suhtlus interneti vahendusel uusi keelekasutuse võimalusi ja mustreid, mida on oluline arvestada mitmekeelsuse uurimisel.

Kogumiku artiklid keskenduvad niisugustele teemadele nagu varajane ja noorte mitmekeelsus, teise keele omandamine, perekonna keelepoliitika, mitmekeelne suhtlus YouTube'is ja TikToki ja sotsiaalmeedias, etnolektide kasutus. Metodoloogia poolest on nii eksperimentaalseid kui ka etnograafilisi kirjutisi.

Elena Antonova-Ünlü ja Çiğdem Sağın-Şimşek võrdlevad türgi-vene ja türgi-inglise varajaste kakskeelsete ning türgi ükskeelsete laste narratiiviloomet. Leiti, et temporaalsete konnektiivide kasutus on mitmekeelsetel kõnelejalatel teistsugune, mis kinnitab varasemate uuringute tulemusi. Üks põhjusi võib olla türgi keele kasutuse piiratud võimalus.

Piret Baird, Reili Argus ja Marilyn Meristo uurisid eesti keele oskustaset muu (enamasti vene) kodukeelega koolilastel. Võrreldi eesti keele omandamist kahes rühmas: ühe rühma lastel oli juba mõningane eesti keele oskus varasemast ajast ja teistel mitte.

Leiti, et kooliaasta lõpuks kahe rühma eesti keele oskus ei erinenud oluliselt.

Piret Baird vaatleb kaheaastase lapse inglise-eesti koodivahetust, kusjuures sisend mõlemas keeles on tasakaalus. Erinevalt tavapäraest perekonna keelepoliitika mudelitest („üks vanem, üks keel“ või „üks keskkond, üks keel“), kasutab pere „üks päev, üks keel“ mudelit, kuna vanemad valdavad teineteise keelt. Tasakaalus sisend avaldub ka lapse väljundis mõlemas keeles: on üsna palju koodivahetusega ütlusi, mis on ükskeelsetest ütlustest keerukamad.

Vlada Baranova ja Kapitolina Fedorova tutvustavad eri aktsentidega vene keelt, mida kasutavad eri etnolingvistikulise taustaga noored esinejad YouTube'is ja TikTakis. Aktsentide kasutus on suunatud ükskeelsete silmaklappide ja etniliste stereotüüpide vastu ning püüab kindlustada esinejate agentsust.

Inga Hilbig kirjeldab leedu-saksa kakskeelse poisi juhtumit ja leiab, et vastupidiselt laialt levinud arvamustele ei pruugi „üks vanem, üks keel“ põhimõtte olla piisav. Ta rõhutab lapse agentsuse olulisust ja näitab, et ka passiivset keelt saab hõlpsasti aktiveerida.

Oleksandr Kapranov käsitleb inglise frikatiivi /z/ omandamist Norra kõrgema kesktaseme inglise keele õppijate seas. Katsed näitasid, et konsonant on õppijatele problemaatiline ja et enamasti asendati /z/ /s/-iga, mis on oluline teave keelepädagoogika jaoks.

Victoria Kazakovskaya analüüsib vene käändsõnatuletust vene ükskeelsete ja simultaansete vene-saksa kakskeelsete laste hulgas. Uurimus kinnitab varasemaid tulemusi, et kakskeelsete areng kummaski keeles võib mõnevõrra maha jääda, kuid kumulatiivses keelearengus on nad ükskeelsetest ees.

Geidi Kilp keskendub pragmaatiliselt tingitud erinevustele isikuvideote kasutuses eesti-inglise-jaapani Facebooki-suhtluses. Keelte kombinatsioon on haruldane ja seetõttu on niisuguste kolmkeelsete keeakasutajate arv üsna piiratud. Kuigi osalejate jaapani keele oskus on erinev, tajuvad kõik erinevust kolme keele pragmaatiliste vahendite vahel ning kasutavad jaapani isikuvideoteid nagu *senpai* ja *sensei* ka eesti- ja ingliskeelsetes ütlustes.

Eglė Krivickaitė-Leišienė ja Ineta Dabašinskienė osutavad kakskeelsete laste edule väljamõeldud sõnade kordamise testis. Kaht eri keelekombinatsiooniga kakskeelsete laste rühma ja üht leedu ükskeelsete laste rühma võrreldi ning selgus, et kakskeelsete tulemused on paremad ilmselt sellepärast, et nad valdavad kaht erinevat fonoloogilist süsteemi.

Kristiina Praakli, Mari-Liis Korkus, Aive Mandel, Elisabeth Kaukonen, Annika Kängsepp, Triin Aasa, Kristel Algvere, Helen Eriksoo, Marion Mägi, Getri Tomson ja Liina Lindström uurisid inglise keele kasutust eesti noorte juutuuberite hulgas just noorte keele uurimise perspektiivist. Kuigi eesti noored juutuuberid kasutavad inglise keelt üsna sageli, ilmnesid kaheksa esineja vahel olulised individuaalsed erinevused.

Virve-Anneli Vihman, Kristiina Praakli, Maarja-Liis Pilvik ja Mari-Liis Korkus löid kaks eesti noortekeele korpust, nimelt *online*-keelekasutuse ja suulise keelekasutuse korpuse. Nad uurisid ingliskeelse sõnavara kasutuse mustreid. Selgus, et ingliskeelsete lekseemide arvu on võimatu ennustada vanuse või soo põhjal, samas on vanus ja sugu olulised tegurid kõige sagedamate inglise sõnade puhul.

Anna Verschik on üldkeeleteaduse professor, ta uurib mitmekeelsust ja keelekontakte.
anna.verschik@tlu.ee

INTRODUCTION

Anna Verschik

Tallinn University

The focus of the article collection is child and youth multilingualism. While the body of literature on multilingualism in general and child and youth multilingualism in particular is huge and steadily growing all the time, it appears that multilingualism and various aspects thereof are and always will be relevant both for linguistic theory and applied topics.

This happens partly because new sociolinguistic settings with new combinations of languages emerge (like Turkish-Russian early bilingualism) and partly because some already known models and methods are applied to new situations, like studies on Estonian youth language. Also new media and CMC create new opportunities and patterns of language use that are relevant in multilingualism research.

The studies in this volume address topics in early bilingualism, SLA, family language policy, multilingual communication on YouTube, Tik Tok and other social media platforms, and ethnolectal speech. The papers use different methodologies: some are experimental and some ethnographic in their nature.

Elena Antonova-Ünlü and Çiğdem Sağın-Şimşek compare aspects of Turkish narrative production among Turkish-Russian and Turkish-English early bilinguals on one hand and Turkish monolinguals on the other. They found that the use of temporal connectors in bilinguals differs from that of monolinguals, which is in line with the previous research on bilingual children who have Turkish as one of their languages. The reason may be the scarce use of Turkish in limited context only.

Piret Baird, Reili Argus and Merilyn Meristo investigated the development of Estonian proficiency in children with another

first language (mostly Russian). The proficiency in Estonian among two groups of learners was compared: children who had had some command in Estonian prior to the instruction and children without prior exposure to the language. It was discovered that at the end of the school year the command of Estonian did not differ much among the two groups.

Piret Baird discusses English-Estonian code-mixing in a two-year old child with balanced input in both languages. Differently from usual models (“one parent, one language” or “one environment, one language”), the model used in the family is “one day, one language”, since both parents are fluent in each other’s language. The balanced input reflected in the child’s proportional output in the two languages; many code-mixed utterances were produced, and these utterances were more complex than monolingual ones.

Vlada Baranova and Kapitolina Fedorova introduce a novel topic of the performative use of accented Russian in young TikTok and YouTube performers of a different ethnolinguistic origin. Performance of accents is aimed against monolingual bias and ethnic stereotypes and seeks to reclaim the performers’ agency.

In the contribution by **Inga Hilbig**, a case of Lithuanian-German bilingual boy is discussed, and it is suggested that, contrary to popular opinions, the OPOL policy may be insufficient. She emphasises the significance of a child’s agency and shows that passive languages may easily become activated.

Oleksandr Kapranov looks into the acquisition of the English fricative consonant /z/ among Norwegian higher intermediate learners of EFL. The experiments demonstrated that the consonant poses a challenge for the learners, as they mostly substituted /z/ with /s/, which has implications for language pedagogy.

Victoria Kazakovskaya analyses acquisition of Russian nominal derivation in Russian monolingual and Russian-German simultaneous bilingual children. The study confirms previous findings that bilinguals may lag in each of their languages behind their

monolingual peers but are superior in the cumulative language development.

The topic of the study by **Geidi Kilp** is the insertion of person references motivated by pragmatic differences in trilingual Estonian-English-Japanese Facebook communication. The combination of languages is rare and the number of such trilingual users is rather limited. The users' competence in Japanese may differ, but all of them perceive differences in pragmatic devices across the language and insert Japanese person references *senpai* and *sensei* in their Estonian and English utterances.

Eglė Krivickaitė-Leišienė and **Ineta Dabašinskienė** describe the bilingual advantage in non-word performance tests. Two groups of bilingual children with different languages and monolingual Lithuanian-speaking children were compared, and bilinguals demonstrated better results, possibly because of their command of two different phonological systems.

Kristiina Praakli, Mari-Liis Korkus, Aive Mandel, Elisabeth Kaukonen, Annika Kängsepp, Triin Aasa, Kristel Algvere, Helen Eriksoo, Marion Mägi, Getri Tomson, and Liina Lindström investigated the use of English among Estonian YouTubers from the perspective of youth language research. While English is widely used by Estonian YouTubers, the analysis of eight YouTubers exhibited significant individual differences.

Virve-Anneli Vihman, Kristiina Praakli, Maarja-Liis Pilvik, and Mari-Liis Korkus built two corpora of Estonian youth speech, those of spoken and online usage, and investigated patterns in the use of English lexical items. The amount of English items is not predicted by age or gender; however, age and gender do matter as far as the most frequently used English-language items are concerned.

Anna Verschik is professor of general linguistics at Tallinn University. Her field of research is language contacts and multilingualism. anna.verschik@tlu.ee

CONNECTIVITY IN NARRATIVES OF TURKISH-ENGLISH AND TURKISH-RUSSIAN BILINGUALS

Elena Antonova-Ünlü

Hacettepe University

Çiğdem Sağın-Şimşek

Middle East Technical University

Abstract. The present study draws on the narrative production of the Turkish-English and Turkish-Russian bilingual children in an attempt to examine whether the use of connectivity elements in the oral narratives of the bilingual children diverges from that of monolingual Turkish children. In particular, the study aimed to examine the use of temporal connectivity elements in the oral narratives of the Turkish-English and Turkish-Russian bilingual children in comparison to Turkish monolingual children focusing on the use of tense/aspect markers utilized to anchor narratives, temporal converbs used to link clauses in narratives, and also temporal connectors used to link clauses. The data were collected from two bilingual groups, Turkish-Russian (Group 1) and Turkish-English (Group 2), consisting of five children each and the control group consisting of seven monolingual Turkish children. The analysis of the data revealed that the Turkish-English and Turkish-Russian bilingual children performed differently than their Turkish monolingual counterparts in how consistently they used tense/aspect markers to anchor their narratives and in how they used converbial markers to indicate the sequentiality of the events in their narratives. The results are discussed in relation to prior research and the typological peculiarities of the languages.

Keywords: bilingual language acquisition, Russian-Turkish, Turkish-English, connectivity, narratives

I. Introduction

The present study draws on the narrative production of two groups of bilingual children (Turkish-English and Turkish-Russian) in an attempt to examine whether the use of connectivity elements in the oral narratives of the bilingual children diverges from that of monolingual Turkish children. Narrative abilities of children have often been studied to assess linguistic, cognitive, and pragmatic skills as they provide rich data regarding children's expressive language, including the knowledge of vocabulary, grammatical constructions, and story structure (Botting 2002; Iluz-Cohen, Walters 2012; Squires *et al.* 2014). As described by Labov (1972), a narrative is “a method of recapitulating past experience by matching a verbal sequence of clauses to the sequence of events which actually occurred”. Thus, in order to effectively narrate a coherent story, children need to link the sequence of events temporally and causally in their minds and verbalize the events by making use of relevant connectivity elements (Berman, Slobin 1994; Öger-Balaban, Aksu-Koç 2020). Temporality markers are basic connectivity elements in narratives as they indicate the flow of the timeline of narratives. Temporal connectivity is established through the appropriate use of linguistic devices, such as verbal temporal elements (tense/aspect markers) used to anchor a tense and to link clauses in narratives, and temporal connectors (time adverbials and other temporal connectives) used to link clauses.

The phenomenon of connectivity in the narrative abilities of mono-/bilingual children has been widely investigated in the context of language acquisition (Aarssen 2001; Berman, Slobin 1994; Bohnacker 2016; Montanari 2004; Uccelli, Páez 2007; Roch, Florit, Levorato 2016) for the following reasons. First, narratives allow researchers to examine multiple linguistic aspects in a single task, ranging from lexical and morpho-syntactic elements to discourse structure (Hickmann 2003). Second, narratives provide a baseline for literacy development (Dickinson, Tabors 2001), and exploring

narrative abilities of children allows scholars to reveal language development problems in children (Bishop, Donlan 2005; Norbury, Bishop 2002). Finally, peculiar to bilingual children, narratives allow eliciting phenomena that are unique to bilingual language use, such as code-switching and cross-linguistic influences (Iluz-Cohen, Walters 2012). Yet, the phenomenon of connectivity in the narrative abilities of bilinguals has received relatively less attention from a typological and comparative perspective.

Given the substantial role of narratives in bilingual children's language development (Chang 2004), this study, adopting a typological and comparative perspective, aims to examine whether the use of temporal connectivity elements in oral narratives of Turkish-English and Turkish-Russian bilingual children diverges from that of the monolingual Turkish control group with a focus on tense/aspect markers used to anchor the narrative and on temporal connectives used to link clauses in narratives, as well as on temporal connectors such as time adverbials used to link clauses. The language combinations of Turkish-English and Turkish-Russian were selected for the following reasons. First, while in Turkish and English narratives, consistent use of a tense/aspect marker is required to anchor a narrative, in Russian, tense shifts within a narrative are common. Second, all the three languages use language-specific means to link clauses due to their typological features, which may, in turn, result in the use of unique bilingual strategies to achieve connectivity in narratives.

The organization of the paper is as follows. First, we briefly describe the temporal connectivity markers in Turkish, Russian, and English. Then, we present previous studies on connectivity elements in Turkish narratives. Following the methodology and results, we discuss the findings.

2. Connectivity in Turkish

Turkish, belonging to the Turkic language family, is an agglutinating language in which verbs and nouns are richly inflected with suffixes. While the canonical order is SOV, Turkish allows flexibility depending on pragmatic constraints. Turkish does not have a formal article system and lacks grammatical marking for gender (Göksel, Kerslake 2005; Kornfilt 1997). Regarding its narrative structure, an important peculiarity of Turkish is described by Aksu-Koç as “one of the criteria for the well-formedness of a narrative is the choice of a consistent favored tense” (1994: 333) throughout the narrative (Akıncı 1999). Johanson (2007a, 2007b) suggests that aspectotemporal elements are realized depending on the discourse type in Turkish narratives. While *-DI*-based past narrative is described as the most differentiated discourse type, *-mİş*-based evidential (indirect) past narrative is used in traditional story-telling. In addition to past narrative markers, it is also possible to use *-(I)yor* and *-(I)r*-based narratives to describe events simultaneous to the speech event.

Expressing simultaneity and sequentiality of events in narratives is based on clause linkage. Turkish clause linkage relies predominantly on non-finite subordination and less on finite subordination, coordination, and use of temporal connectors. While in finite subordination the predicate may be verbal or nominal and marked in the same way as the predicate of a main clause, in non-finite subordination the predicate is verbal and marked by distinctive subordinating morphology (Göksel, Kerslake 2005; Kornfilt 1997; Kerslake 2007). In Turkish narratives, the converbial markers *-(y)IncA* (when, since, as), *-(y)ken* (while, when) and *-(y)Ip* (then) are used to establish temporal connectivity.

Turkish also uses adverbials such as *sonra* (later) and its various forms, *ondan sonra* (after that), *daha sonra* (later) and two-word combinations such as *o zaman* (that time), *o an* (that moment), which is a combination of a demonstrative/determiner and a noun (Karahan 2007; Özsoy 2021), to connect clauses as connectivity elements.

In this study, we will name such connectivity elements as temporal connectors.

3. Connectivity in Russian

Russian narratives do not require the consistent use of tenses, and native speakers of Russian commonly use tense shifts in their narratives (Bondarko 2005; Paducheva 2011). Bondarko (2005) and Paducheva (2011) distinguish several reasons for tense shifts in Russian, such as distinguishing between the foreground and background as well as between the topic and the focus, marking the consequence of actions in the narrative, and emotional-expressive actualization. Example 1 illustrates a typical tense shift taken from Rekemchuk (1962):

- (1) *Сплю (PRE) я сегодня ночью и приснилось (PAST) мне...*
Sleep (PRE) I today at night and dreamt (PAST) me...
"Today I was sleeping at night and saw in my dream..."

Russian clause linkage relies predominantly on finite subordination and coordination and less on non-finite subordination. Among non-finite subordination, converbs, which are also known as *deepričastie*, are used to establish temporal connectivity. Russian converbs have two forms, perfective, expressed by the morphemes *-в(вши)*, and imperfective, expressed by the morpheme *-я*. The perfective converbial form indicates the action expressed by the converbs that precedes the one expressed with the finite form in the main clause. While the imperfective form indicates that the two actions are simultaneous, Example 2 and Example 3 illustrate the use of perfective and imperfective converbs in Russian, respectively.

- (2) *Сделав домашнее задание, Ник пошел играть с друзьями.*
Having done his homework Nick went to play with his friends.

(3) *Ребенок шел по улице, **напевая** песню.*

A child was walking down the street singing a song.

Russian also uses adverbs such as *потом* (then), *затем* (then), *позже* (later) and two-word combinations such as *после этого* (after that), which is a combination of a preposition and a demonstrative pronoun, to connect clauses as connectivity elements.

4. Connectivity in English

Similar to Turkish, English requires a linear presentation of events and does not allow tense shifts within narratives (Kornfilt 1997). The English language does not have converbial forms but utilizes participles to fulfill a converbial function of marking adverbial subordination as in Example 4. The same forms are also used as participles or verbal nouns in English. As for the connectivity elements, adverbs *then*, *later* and two-word combinations such as *after that* are also used in English.

(4) *The child walked down the streets eating an apple.*

5. Research on the development of connectivity in Turkish

The development of connectivity in Turkish narratives has been examined in the monolingual and bilingual acquisition contexts. Research on acquisition of the converbials in Turkish shows that the converbial markers appear early in monolingual Turkish (Aksu-Koç 1994; Topbaş *et al.* 2012). In a recent and extensive study, Ögel-Balaban and Aksu-Koç (2020) examined the development of clause chains formed with converbial clauses. The study used narratives of 40 Turkish-speaking four- to eleven-year-olds and six adults elicited by a wordless picture book. The study demonstrated that there is a gradual increase by age in the variety of clauses combined, the length of the complex sentences, and their frequency of use.

Converbial clauses emerge as the earliest and most frequent type of clauses. Regarding the development of narrative organization, the study reported that children first establish aspectual-temporal continuity and then temporal-causal continuity in Turkish.

On the other hand, research on acquisition and use of temporal connectivity elements in Turkish as a heritage language demonstrates that in bilingual contexts, such as in German, Swedish, French, and Dutch contexts, temporal connectivity elements are used differently (Aarsen 2001; Bohnacker, Karakoç 2020; Boetschoten and Verhoeven 1986, Rehbein and Herkenrath 2015).

Rehbein and Karakoç (2004) reported that Turkish-German bilingual children in the German context use aspectotemporal elements in their narratives differently from their monolingual counterparts. The study concluded that the bilinguals shifted between aspectotemporal elements *-DI*, *-mİş* and *-(I)yor*, which was not observed in the Turkish monolingual data. Similarly, Karakoç (2007) studied connectivity by means of finite elements in Turkish-German bilingual children in Germany. The researcher reported that while all forms were used by Turkish monolingual children, the bilingual children used *-(I)yor* (present imperfective) and *-DI* (past perfective) forms in their narratives to maintain connectivity and refrained from using *-mİş* (perfective aspect/evidential modality) forms. In addition to aspectotemporal elements of connectivity the study highlighted a highly frequent use of temporal-deictic expressions, such as *o zaman* (at that time), *sonra* (than), *ondan sonra* (after that) by the bilingual children.

Based on the data obtained from Turkish-French children in France, Akıncı (1998) reported that children between the ages of 5 and 10, born to immigrant parents in France, revealed no clear and consistent “anchor tense”. The researcher, however, reported that the children at the age of 9–10 began anchoring either the present or past tense as the favoured one. In another study, based on the data gathered from first- and second-generation Turkish immigrants in France, Akıncı (2003: 296) reported that the majority of

the first-generation immigrants shifted tenses while the second-generation bilingual participants “used tense just as the monolingual high-educated participants do” after the age of 14 . The researcher presented social class attitude and the level of literacy as two important factors that had an impact on the development of tense usage of the bilinguals.

In another study, Schroeder (2016) examined the clause-combining strategies of Turkish-German bilinguals in a German context to interpret the dynamics of language shift. The study reported that the shift to using more finite clauses, clause initials, and semantic connectors in Turkish in Germany could be explained by two factors: first, the limited access to the structures of the formal register of Turkish that results in “generalization of structural elements of spoken Turkish”, and, second, to the “generalization of structures with a structural and functional correspondence in the contact language German” (2016: 97).

Akkuş (2019) investigated the converbial constructions in heritage Turkish in the Netherlands from a language contact perspective. Based on the data obtained from the first and second generations of Dutch-Turkish speakers, the study reported a gradual decrease in the frequency of converb use and unconventional usages of converbs in non-finite constructions of the second-generation speakers. The study suggested that the participants’ perception and production of the converbial constructions indicated a linguistic change regarding the frequency and pattern of use.

Turan *et al.* (2020) examined the perception and use of the converbs *-Ip* and *-Inca* in heritage Turkish in Germany. Based on the analysis of the data obtained through a grammaticality judgment task and a picture-story description task, the study revealed that the bilinguals’ perception of the grammatical constructions with *-Inca* and of the ungrammatical constructions with *-Ip* and *-Inca* differed significantly from that of the monolinguals, while the perception of the grammatical constructions with *-Ip* was reported to be similar. As for the production of the converbs, the bilingual participants

tended to use the converbs significantly less than the monolingual control group, which was reported as unconventional.

The studies above commonly suggest that converbs as connectivity elements emerge late in heritage Turkish and are used less frequently than in monolingual Turkish. Yet, based on the data obtained from 102 children between the ages of four and seven years old, Bohnacker and Karakoç (2020) reported that the case was a bit different in the Swedish context. The researchers stated that while temporal converbs were not very common at the ages of 4, 5 and 6, their frequency of use increased at age seven. The study concluded that extensive exposure to Turkish in the home environment might have promoted the children's active use of converbs and the development of their Turkish in general.

In a recent study, Özsoy (2021) examined heritage Turkish speakers' use of temporal connectors in Germany and the USA. The study utilized a large systematic corpus of semi-naturalistic narrations in the RUEG corpus (Wiese *et al.* 2020) and investigated the role of age, register, mode, and grammatical aspect in the use of temporal connectors. The study reported more frequent use of temporal connectors by Turkish heritage speakers in Germany and the USA compared to Turkish monolinguals. The researcher argued that because there is no significant difference in the use of grammatical aspect between the two Turkish varieties, the results cannot be explained in relation to language contact with German and English. Instead, the study argued that the difference is related to language use patterns as "they use [Turkish] mostly [to] communicate in informal, spoken settings, compared to monolinguals who use Turkish in written and formal settings too" (2021: 5).

6. Study

This study aims to answer the following research questions:

1. Do the narratives of the Turkish-English and Turkish-Russian bilingual children differ from that of monolingual Turkish children in terms of anchored tense markers, conversational forms and temporal connectors?
2. In case of the differences between the bilingual and monolingual participants, can cross-linguistic influence from L1 Russian and English be considered as an underlying factor accounting for the divergence?

7. Participants

The participants of the study were two bilingual groups, Turkish-Russian (Group 1) and Turkish-English (Group 2), consisting of five children each, and a control group consisting of seven monolingual Turkish children. The age of the participants in the Turkish-Russian group varied from 4,9 to 16,0 years old, and the age of the participants in the Turkish-English group varied from 5,0 to 6,0 years old. The Turkish-Russian bilinguals were born and raised in Russia in Turkish-Russian families where mothers are Russian and fathers are Turkish, while Turkish-English bilinguals were born and raised in England in families where both parents are Turkish native speakers. All the bilingual participants were recruited into the study during their summer vacation in Turkey. The participants in the control group were aged from 4,9 to 16,2. All of them were born and raised in Turkey in monolingual Turkish families. The parents in all the families were university graduates and reported their socioeconomic status as middle class.

8. Data Collection Tools

To collect the narrative data, a picture book, “Frog where are you?” (Mayer 1969) was used. The above-mentioned book has been often used in research focusing on the language development of monolingual and bilingual children (Aksu-Koç 1988; Berman 1999; Kupersmitt, Berman 2001, among others) because it allows researchers to collect data from participants of different age groups, including very young ones, and provides researchers with natural narrative data, which allows researchers to examine language devices occurring only in connected speech. The participants were requested to retell a story about a boy and his friend dog, who have lost their frog pet, using the pictures in the book. Both the researchers and one of the children’s parents were present during the data collection. The narratives of the children were recorded upon their parents’ consent and later transcribed by the researchers using the transcription software CLAN_CHILDES (MacWhinney 2000).

8.1. DATA ANALYSIS

8.1.1. RESULTS

8.1.1.1. TENSE ANCHORING

The ability to anchor tense in the narrative, that is, to consistently use a favoured tense, is considered as one of the criteria for the well-formedness of a narrative (Aksu-Koç 1994). Table 1 presents the percentage of tense consistency in the narratives for each of the participants in the bilingual and monolingual groups.

As Table 1 displays, from the age of approximately five to eleven, the bilingual participants do not anchor their narratives to one tense but they tend to shift between two (or three) tenses. Example 5 illustrates tense shifts from the narrative of the eight-year-old Turkish-English participant.

Table 1. Tense consistency in the narratives of the bilingual and monolingual participants

Participants	1	2	3	4	5
Turkish-Russian Group					
Age	4.9	6.0	8.3	11.3	16.0
Tense consistency (%)	51	60	63	79	100
Turkish-English Group					
Age	5.0	6.2	8.1	11.1	16.2
Tense consistency (%)	52	48	64	70	100
Monolingual Group					
Age	4.9	5.9	8.0	10.11	16.2
Tense consistency (%)	99.9	100	100	100	100

- (5) *Şimdi çocuk uyandı ve kurbağayı bulmaya çalışıyor çünkü şiseden çıktı.*

Now the child woke up and tried to find the frog because it had gone out.

In Example 5, the participant anchors the narrative with the past tense marker *-DI* and uses it in the first three utterances of the narrative. However, in the fourth utterance he uses the present imperfective marker *-(I)yor* and then shifts back to the past tense marker *-DI*. Such shifts take place throughout the whole narrative.

The analysis of the narratives produced by the bilingual participants at the age of 16 has revealed that they anchor their narratives through the use of a consistent tense. Example 6 illustrates the consistent use of tense from the narrative of the sixteen-year-old Turkish-Russian participant.

- (6) *Bir akşam çocuk ve köpek bir kurbağa yakaladılar. Sonra onu kavanoza koydular. Çocuk ve köpek gece uyurken kurbağa kavanozdan çıktı.*

One evening, a child and his dog caught a frog. They put it into the jar. When the child and the dog were sleeping, the frog went out of the jar.

In Example 6, the participant anchors his narrative with the past tense marker *-DI* and uses it through the whole narrative.

As for the monolingual control group, already at the age of five, the monolingual participants are able to use one tense consistently to anchor their narratives. Example 7 illustrates the consistent use of tense markers in the narrative of the monolingual participant aged 4.9.

(7) *Köpek aşağıya düşüyor. Sonra köpeğini alıyor çocuk ve ormana gidiyorlar.*

The dog is falling down. Then the child is taking the dog and they are going to the woods.

Example 7 illustrates how the monolingual child successfully anchors the narrative with the imperfective present tense *-(I)yor* and sticks to it throughout the narrative.

Further, the tense use in the narratives of both the bilingual and monolingual participants was examined. Table 2 presents the type and distribution of tenses used in the participants' narratives.

Table 2. *The distribution of tense markers in the narratives of the bilingual and monolingual participants*

Participant	1	2	3	4	5
Turkish-Russian Group					
Age	4.9	6.0	8.3	11.3	16.0
Tense marker use (%)					
<i>-DI</i>	51	48	63	21	100
<i>-mİş</i>	–	40	37	79	–
<i>-(I)yor</i>	49	2	–	–	–
Turkish-English Group					
Age	5.0	6.2	8.1	11.1	16.2
<i>-DI</i>	48	40	64	70	100
<i>-mİş</i>	–	–	56	–	–
<i>-(I)yor</i>	52	60	–	30	–
Monolingual Group					
Age	4.9	5.9	8.0	10.11	16.2
<i>-DI</i>	–	–	–	–	–
<i>-mİş</i>	–	99.8	100	100	100
<i>-(I)yor</i>	100	0.2	0	0	0

As evident from Table 2, the bilingual participants use past tense marker *-DI*, indirect evidentiality marker *-mİş* and present imperfective marker *-(I)yor*. Though all the three tense markers are used in the narratives, there seems to be a preference for the past tense marker *-DI* among the bilingual participants. This marker is also exclusively used throughout the narratives by the sixteen-year-old bilingual participants to anchor their narratives.

As for the monolingual participants, our data indicate that at the age of 4.9 they use the imperfective present tense *-(I)yor* to anchor the narrative, and at the age of 5.9 and later, they anchor their narratives with the indirect evidentiality marker *-mİş*. Example 8 illustrates the use of the indirect evidentiality marker *-mİş* to anchor the narratives by the monolingual participant aged 5.9.

- (8) *Bir çocuk varmıŝ ve onun da bir köpeęi varmıŝ. Bir de bir tane de kurbaęası varmıŝ. Kurbaęa kavanozdan çıkmıŝ. Sonra uyanınca, çocuk ve köpek kurbaęanın olmadıęını görmüşler ve onu aramaya karar vermışler.*

Once there was a child and there was a frog. The frog got out of the jar. After having woken up the child and the frog saw that there was no frog and they went to look for it.

8.1.1.2. USE OF TEMPORAL CONVERBIALS

In the following stage, we examined the use of temporal converbial forms used to link clauses and indicate the sequentiality of the events.

The data analysis revealed that neither Turkish-Russian nor Turkish-English bilingual participants use converbial forms before the age of eleven. Moreover, the bilingual participants do not use any other non-finite forms in their narrative before the age of eleven, and all their utterances are formed with the help of finite verbal forms, as demonstrated in Example 9, taken from the narrative of the eleven-year-old Turkish-English participant:

Table 3. *The use of converbial forms in the narratives of the bilingual and monolingual participants*

Participant	1	2	3	4	5
Turkish-Russian Group					
Age	4.9	6.0	8.3	11.3	16.0
Use of converbials (N)	0	0	0	1	2
Turkish-English Group					
Age	5.0	6.2	8.1	11.1	16.2
Use of converbials (N)	0	0	0	0	1
Monolingual Group					
Age	4.9	5.9	8.0	10.11	16.2
Use of converbials (N)	0	2	3	4	4

- (9) *Çocuk arıları gördü ve korktu. Arılar sinirlendi çünkü çünkü köpek onu rahatsız etti.*

The child saw bees and got scared. The bees got nervous because because the dog disturbed them.

The bilingual participants at the age of 16 use converbial forms. Example 10, taken from the narrative of the sixteen-year-old Turkish-Russian participant, is illustrative.

- (10) *Köpek düşüñce çocuk de aşağıya düştü ve köpeği tuttu.*

After the dog's falling down the child fell down as well and held the dog.

In Example 10, the bilingual participant uses the converbial *-(I)nce* (*düşüñce*) to link the events and indicate their sequence in the utterance.

As for the monolingual participants, the data analysis revealed that converbial forms appear in the narratives of the child aged 5.9, and further temporal converbials are available in the narratives of all the older monolingual participants. Example 11, taken from the narrative of the monolingual participant aged 5.9, illustrates the use of a converbial form.

- (11) *Sonra çocuk pencereden çıkıp bağırmış*
Then having got out of the window the child shouted.

As illustrated in Example 11, the monolingual children used the temporal converbial *-(I)p (çıkıp)* to indicate the sequential order of the events performed by the same agent.

8.1.1.3. TEMPORAL CONNECTORS

As for the use of temporal connectors by the participants, the data revealed that the temporal connector *sonra* and its variants *ondan sonra* and *daha sonra* are used by all the participants in their narratives. No other temporal connectors are found in the participants' narratives. Table 4 presents the use of temporal connectors as the number of the connectors per the number of clauses in the narrative.

Table 4. *The use of temporal connectors in the narratives of the bilingual and monolingual participants*

Participant	1	2	3	4	5
Turkish-Russian Group					
<i>Age</i>	4.9	6.0	8.3	11.3	16.0
Number of temporal connectors per number of clauses	21/41	10/40	9/35	9/38	6/35
Turkish-English Group					
<i>Age</i>	5.0	6.2	8.1	11.1	16.2
Number of temporal connectors per number of clauses	10/20	11/29	7/40	0/23	2/30
Monolingual Group					
<i>Age</i>	4.9	5.9	8.0	10.11	16.2
Number of temporal connectors per number of clauses	16/26	14/33	7/30	5/28	3/33

Discussion

The present study aimed to examine the use of temporal connectivity elements in the oral narratives of Turkish-English and Turkish-Russian bilingual children in comparison to Turkish monolingual children. The study focuses specifically on the use of tense/aspect markers utilized to anchor narratives, temporal converbs used to link clauses in narratives and also temporal connectors used to link clauses.

Our first research question concerned whether the Turkish-English and Turkish-Russian bilingual children diverge from the monolingual Turkish control group in their use of tense/aspect markers to anchor the narratives. Our data obtained from the bilingual participants demonstrated that the bilinguals between the ages of five and eleven have difficulties in anchoring a tense/aspect marker in their narratives, but rather they show a tendency to shift between two (or even three) tenses. Yet, the data also revealed that the bilingual participants who are at the age of 16 are able to anchor their narratives by consistently using one tense. When we further examined the tense/aspect markers favored by the bilingual participants, the results revealed that while the bilingual participants use three different markers (past tense marker *-DI*, indirect evidentiality marker *-mİş* and present imperfective marker *-(I)yor*), they seem to use the past tense marker *-DI* at higher frequencies than the other markers. The monolingual participants, on the other hand, are able to consistently use one tense to anchor their narratives. Our data demonstrate that at the age of 4.9 the monolingual participants use the imperfective present tense *-(I)yor* to anchor the narrative, and at the age of 5.9 and later, they anchor their narratives with the indirect evidentiality marker *-mİş*.

These results are consistent with Rehbein and Karakoç (2004) and Karakoç (2007), who reported differences between Turkish-German bilingual and Turkish monolingual children's ability to anchor a tense/aspect marker. Similar results were also reported by

Akıncı (1998; 2003) based on the data obtained from Turkish-French bilingual children. The bilingual children in the German and French contexts were not able to favour and anchor a tense, unlike their monolingual counterparts, and shifted tenses throughout their narratives. Nevertheless, the picture changes as the bilingual children grow older. Similar to our findings, the Turkish-French bilinguals (Akıncı 2003) were reported to be able to consistently use one tense in their narratives.

From a typological perspective, we expected a variation between the Turkish-English and Turkish-Russian bilinguals due to cross-linguistic influence. Since both Turkish and English narratives require anchoring a tense while in Russian narratives tense shifts are quite common, we expected the Turkish-English bilinguals to show more monolingual-like anchoring patterns than the Turkish-Russian bilinguals. Yet, our bilingual data of the two groups revealed similarities in their inability to anchor a tense/aspect marker. Thus, the deviation of the bilinguals from the monolinguals cannot be explained due to cross-linguistic influence and the typological properties of the languages. Since both the Turkish-English and Turkish-Russian bilinguals have different typological properties in their repertoires and have performed similarly, we hypothesise that the difficulty in tense anchoring between the ages of 4.9 and 11 might be explained in relation to tense anchoring being a marked ability in bilingual language acquisition, which may require prolonged exposure and use of the language. The fact that the 16-year-old bilingual participants were consistent in their use of tenses allows us to support this hypothesis. Although the scope and data of this study do not allow us to draw conclusions regarding the cognitive load of tense anchoring, as the monolinguals and the bilinguals above 16 can anchor their narratives, it is possible to speculate that tense anchoring might require an additional cognitive process when narrating events.

The second issue we investigated concerned the use of temporal converbials, which are utilised to indicate the sequentiality of

the events in narratives. The analysis of the bilingual data indicated that no converbial forms are utilised by the Turkish-Russian and Turkish-English bilinguals who are younger than 11. A common peculiarity observed in the bilingual data concerned the avoidance of using non-finite temporal converbs. All the clauses formed by the bilingual participants included finite verbal forms exclusively. Only the participants above 16 used non-finite converbs. The monolingual data, on the other hand, revealed the use of non-finite converbial forms in the narratives of the children aged 5.9 and above.

When we compare our findings regarding the use of temporal converbials with those presented in prior research, once again we see similarities between the performances of our Turkish-Russian and Turkish-English participants and those of the Dutch-Turkish (Akkuş 2019) and German-Turkish bilinguals (Turan *et al.* 2020). Even though the participants were adult bilinguals in Akkuş (2019) and Turan *et al.* (2020), the studies reported significantly less use of converbs by the bilinguals in comparison to the monolinguals.

When we examined how the events in narratives are connected to one another from a typological perspective, the bilingual data revealed that the bilingual participants refrain from combining clauses via subordination by means of temporal converbials. From a typological perspective, while Turkish clause linkage relies predominantly on non-finite subordination (on the use of converbials to indicate temporal connectivity) and coordination, Russian and English rely primarily on finite subordination and coordination and less on non-finite subordination. The fact that both the Turkish-English and Turkish-Russian bilingual participants showed similarity in their preference to use finite clauses, unlike the Turkish monolinguals, allows us to assume that the acquisition of temporal converbial subordination by bilinguals is a challenging phenomenon. Akin to the difficulties in tense anchoring, the use of temporal converbial markers seems to develop later in bilinguals than in monolinguals (Bohnacker and Karakoç 2020).

Finally, we examined whether the temporal connectors are used differently by the Turkish-English and Turkish-Russian bilingual children and the Turkish monolingual children. The results revealed that the temporal connector *sonra* and its variants *ondan sonra* and *daha sonra* are used by all the participants in their narratives with no difference. Thus, the similarity in the restricted use of converbial forms and overuse of temporal connectors by the Turkish-English and Turkish-Russian bilingual children in comparison to the monolingual control group cannot be explained via Russian and English influence, either.

These results are in line with the previous studies, especially those conducted in the German context based on Turkish-German bilingual data (Rehbein 2001; Karakoç 2007; Özsoy 2021). Since the bilingual participants had difficulties in using temporal converbials to indicate sequential relations between the events they narrated, they needed to use lexical temporal connectors to indicate the sequential relations. We believe the use of temporal connectors can be perceived as a communicative strategy applied by the bilinguals due to a linguistic gap in their repertoire, which is likely to occur due to the restricted use of Turkish, mainly in informal settings, in the dominant context of the other language.

Indeed, these results can also be explained in relation to length and rate of exposure to the heritage language, language use frequency, language use environment, educational level, reading habits, and/or perceptions of language prestige as well as age and typological proximity. Yet, since these factors, which we assume to probably have an important impact on the findings of the study, were not specifically investigated in this study, further research may look into their roles in narrative development of bilinguals.

REFERENCES

- Aarssen, Jeroen 2001. Development of temporal relations in narratives by Turkish–Dutch bilingual children. – Narrative development in a multilingual context. Eds. Ludo Verhoeven, Sven Strömqvist. John Benjamins Publishing Company, 209–233.
- Akıncı, Mehmet-Ali. 1999. Turkish spoken by bilingual children in France: Mother tongue or weak language? – Turkish Language in Diaspora. *TASG News* 48, 40–46.
- Akıncı, Mehmet-Ali 2003. France multilingue: richesse ou danger? Résultats d’une enquête dans les écoles élémentaires de Lyon. – *Ecarts d’identité* 102, 41–47.
- Akkuş, Mehmet 2019. A usage-based investigation of converbial constructions in heritage speakers’ Turkish living in the Netherlands. Unpublished doctoral thesis. Ankara. Middle East Technical University.
- Aksu-Koç, Ayhan 1988. The acquisition of aspect and modality: The case of past reference in Turkish. Cambridge: Cambridge University Press.
- Aksu-Koç, Ayhan 1994. Development of linguistic forms: Turkish. – Relating Events in Narrative: A Crosslinguistic Developmental Study. Eds. Ruth A. Berman, Dan Isaac Slobin. Lawrence Erlbaum Associates, 329–385.
- Berman Ruth A.; Slobin Dan Isaac (eds.) 1994 Relating Events in Narrative: A Crosslinguistic Developmental Study. Lawrence Erlbaum.
- Bishop, Dorothy; Donlan, Chris 2005. The role of syntax in encoding and recall of pictorial narratives: Evidence from specific language impairment. – *British Journal of Developmental Psychology* 23 (1), 25–46.
- Bohnacker, Ute; Karakoç, Birsal 2020. Subordination in children acquiring Turkish as a heritage language in Sweden. *Studies in Turkish as a Heritage Language*. John Benjamins Publishing Company.
- Bondarko, Aleksandr Vladimirovits 2005. Teorija morfoložičeskikh kategorij i aspektologičeskie issledovaniya [The theory of morphological categories and aspectological investigations]. *Jazyki Slavjanskikh Kul’tur*.
- Botting, Nicola 2002. Narrative as a tool for the assessment of linguistic and pragmatic impairments. – *ChildLanguage Teaching and Therapy* 18 (1), 1–21.
- Bohnacker, Ute 2016. Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. – *Applied Psycholinguistics* 37 (1), 19–48.

- Chang, Chien-Ju 2004. Telling stories of experiences: Narrative development of young Chinese children. – *Applied Psycholinguistics* 25, 83–104.
- Dickinson, David K.; Tabors, Patton O. 2001. Beginning literacy with language: Young children learning at home and school. Brookes Publishing.
- Göksel, Aslı; Kerslake, Celia 2004. Turkish: A comprehensive grammar. Routledge.
- Hickman, Maya 1991. The development of discourse cohesion: Some functional and cross-linguistic issues. – *Language bases... discourse bases: Some aspects of contemporary French-language psycholinguistics research*. Eds. Gilberte Piéaut-Le Bonniec, Marlene Dolitsky. Amsterdam: John Benjamins, 158–185.
- Iluz-Cohen, Peri; Walters, Joel 2012. Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. – *Bilingualism: Language and Cognition* 15 (1), 58–74.
- Johanson, Lars 2007. Aspectotemporal connectivity in Turkic Text construction, text subdivision, discourse types and taxis. – *Connectivity in grammar and discourse*. Eds. Jochen Rehbein, Christiane Hohenstein, Lukas Pietsch. John Benjamins Publishing Company, 187–198.
- Johanson, Lars 2007. Alternative subordination strategies in Turkish. – *Connectivity in Grammar and Discourse*. Eds. Jochen Rehbein, Christiane Hohenstein, Lukas Pietsch. John Benjamins Publishing Company, 231–258.
- Karahan, Leyla 2007. „Sonra, Önce“ kelimelerinin edat kategorisi içindeki durumu. – *Dil Araştırmaları* 1 (1), 39–48.
- Karakoç, Birsal 2007. Alternative subordination strategies in Turkish. – *Connectivity in Grammar and Discourse*. Eds. Jochen Rehbein, Christiane Hohenstein, Lukas Pietsch. John Benjamins Publishing Company, 199–227.
- Kerslake, Celia 2007. Alternative subordination strategies in Turkish. – *Connectivity in Grammar and Discourse*. Eds. Jochen Rehbein, Christiane Hohenstein, Lukas Pietsch. John Benjamins Publishing Company, 231–258.
- Kornfilt, Jaklin 1997. On the syntax and morphology of relative clauses in Turkish. – *Dilbilim Araştırmaları Dergisi* 8, 24–51.

- Kupersmitt, Judy R.; Berman, Ruth A. 2001. Linguistic features of Spanish-Hebrew children's narratives. – Narrative Development in a multilingual context. Eds. Ludo Verhoeven, Sven Strömquist. John Benjamins Publishing Company, 277–317.
- Labov, William 1972. Language in the inner city: Studies in the Black English vernacular. No. 3. University of Pennsylvania Press.
- MacWhinney, Brian 2000. The CHILDES Project: Tools for Analyzing Talk. 3rd edition. Lawrence Erlbaum Associates.
- Montanari, Simona 2004. The development of narrative competence in the L1 and L2 of Spanish-English bilingual children. – International Journal of Bilingualism 8, 449–497. <https://doi.org/10.1177/13670069040080040301>.
- Norbury, Courtenay Frazier; Bishop, Dorothy V. 2002. Inferential processing and story recall in children with communication problems: a comparison of specific language impairment, pragmatic language impairment and high-functioning autism. – International Journal of Language & Communication Disorders 37 (3), 227–251.
- Ögel-Balaban, Hale; Aksu-Koç, Ayhan 2020. Clause chaining and discourse continuity in Turkish children's narratives. – Frontiers in Psychology 11, 115. <https://doi.org/10.3389/fpsyg.2020.00115>.
- Özsoy, Onur 2021. Temporal connectors in relation to grammatical aspect in heritage Turkish in Germany and the United States. Unpublished Master's thesis. Humboldt University.
- Paducheva, Elena 2011. The linguistics of narrative: the case of Russian. Lambert Academic Publishing.
- Rehbein, Jochen; Herkenrath, Annette; Karakoc, Birsal 2009. Turkish in Germany – On contact-induced language change of an immigrant language in the multilingual landscape of Europe. – Multilingualism and universal principles of linguistic change. Special Issue of Sprachtypologie und Universalienforschung – Language Typology and Universals, vol 1. Eds. Gisella Ferraresi, Esther Rinke. Akademie Verlag, 6–43.
- Rehbein, Jochen; Herkenrath, Annette 2015. Converbs in monolinguals' and bilinguals' Turkish. Turkish in Germany – On contact-induced language change of an immigrant language in the multilingual landscape of Europe. – Ankara papers in Turkish and Turkic linguistics, Turcologica,

- vol 103. Eds. Deniz Zeyrek, Çiğdem Sağın Şimşek, Ufuk Ataş, Jochen Rehbein. Harrassowitz Verlag, 493–513.
- Rekemchuk, Aleksandr 1962. Molodo-zeleno. <http://www.rulit.net/books/molodo-zeleno-read87308-1.html> (26.10.2022).
- Roch, Maja; Florit, Elena; Levorato, Chiara 2016. Narrative competence of Italian–English bilingual children between 5 and 7 years. – *Applied Psycholinguistics* 37 (1), 49–67.
- Slobin, Dan I. 1988. The development of clause chaining in Turkish. Paper presentation. The 4th Conference on Turkish Linguistics, August 17–19, Ankara, Turkey.
- Schroeder, Christoph 2016. Clause combining in Turkish as a minority language in Germany. – *Exploring the Turkish linguistic landscape: Essays in honor of Eser Erguvanlı-Taylan*. Eds. Mine Güven, Didar Akar, Balkiz Öztürk, Meltem Keleş. John Benjamins Publishing Company, 81–102.
- Squire, Corinne; Andrews, Molly; Davis, Mark 2014. *What is narrative research?* Bloomsbury Publishing.
- Turan, Dilek; Antonova-Unlu, Elena; Sağın-Şimşek, Çiğdem; Akkuş, Mehmet 2020. Looking for contact-induced language change: Converbs in heritage Turkish. – *International Journal of Bilingualism* 24 (5–6), 1035–1048.
- Uccelli, Paola; Páez, Mariela M. 2007. Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. – *Language, Speech, and Hearing Services in Schools* 38, 225–236. [https://doi.org/10.1044/0161-1461\(2007/024\)](https://doi.org/10.1044/0161-1461(2007/024)).
- Verhoeven, Ludo 1989. Acquisition of clause linking in Turkish. – *Linguistics in the Netherlands*. Eds. Hans Bennis, Ans van Kemenade. De Gruyter Mouton, 153–162.
- Wiese, Heike 2020. Language Situations: A method for capturing variation within speakers' repertoires. – *Methods in Dialectology* 16, 105–117.

RESÜMEE

TÜRGI-INGLISE JA TÜRGI-VEENE KAKSKEELSETE NARRATIIVIDE SIDUSELEMENTIDE KASUTUS

Elena Antonova-Ünlü

Hacettepe University

Çiğdem Sağın-Şimşek

Middle East Technical University

Uurimus keskendub türki-inglise ja türki-vene kakskeelsete laste narratiivimoodustusele, püüdes välja selgitada, kas nende suuliste narratiivide siduselemendid erinevad ükskeelsete türki laste omadest. Eesmärgiks oli vaadelda türki-inglise ja türki-vene kakskeelsete laste ajasuhteid väljendavate konnektiivide kasutust suulistes narratiivides, võrreldes ükskeelsete türki lastega, keskendudes aja/aspekti markeritele, mida kasutatakse narratiivide ülesehitamiseks, sündmuste järjekorda näitavatele konverbimarkeritele, ja lisaks lauseid siduvatele ajasuhteid väljendavatele konnektiividele. Materjali koguti kahest kakskeelsest grupist, türki-vene (grupp 1) ja türki-inglise (grupp 2), mõlemas viis last, ning lisaks kontrollgrupp seitsme ükskeelse türki lapsega. Andmete analüüs näitas, et türki-inglise ja türki-vene kakskeelsed lapsed erinesid ükskeelsetest türki lastest selle poolest, kui järjekindlalt nad kasutasid narratiivi ülesehitamiseks aja/aspekti markereid ning narratiivi sündmuste järjekorda näitavaid konverbimarkereid. Tulemuste üle on arutletud varasemate uurimuste taustal ja tüpoloogiliste erisuste vaatenurgast.

Võttesõnad: simultaanne kahe keele omandamine, vene-türki, türki-inglise, siduselemendid, narratiivid

Elena Antonova-Ünlü is an Associate Professor at Hacettepe University, Department of Translation and Interpreting. Her main research interests include bi-multilingualism, bilingual first language acquisition, second and third language acquisition, heritage language development and heritage language reactivation.
elenaunlu@gmail.com

Çiğdem Sağın-Şimşek is a Professor at Middle East Technical University, Department of Foreign Language Education. Her domains of research include bi-multilingualism, language contact, second language acquisition, third language acquisition and Turkish Linguistics.
sagin@metu.edu.tr

EESTI KEELT TEISE KEELENA OMANDAVATE LASTE EESTI KEELE OSKUSE ARENG AASTA JOOKSUL

Piret Baird, Reili Argus, Marilyn Meristo

Tallinna Ülikool

Annotatsioon.¹ Eesti koolides on järjest enam erineva keeletasemega õpilasi. Samas on väga vähe empiirilisi andmeid selle kohta, kuidas eesti keele kui teise keele omandamine täpsemalt toimub. Siinse uuringu eesmärk oli teha kindlaks kahe erineva eesti keele oskusega grupi laste keeleoskus pärast üht aastat eesti keele õpet koolis. Ühte gruppi kuulusid lapsed, kes õppeaasta alguses eesti keelt üldse ei rääkinud, teise need, kelle keeleoskus oli vähene, kuid kes mingil määral eesti keelt siiski rääkisid (oskasid kasutada tervikuna omandatud eestikeelseid väljendeid ja sõnu, kuid ei olnud võimelised iseseisvalt eestikeelseid lauseid moodustama). Teine eesmärk oli katsetada autorite loodud hindamisvahendi efektiivsust. Hindamisvahendiga mõõdeti laste oskust mõista ja kasutada eesti keele grammatilisi kategooriaid ja konstruktsioone. Tulemustest selgus, et kahe erineva keeleoskuse tasemega grupi tulemused aasta lõpus ei erinenud, st et tugevama stardipositsiooniga laste tulemused ei olnud nõrgema stardipositsiooniga laste omadest üheski kategoorias paremad. Hinnatud 7–9-aastased õpilased mõistavad grammatilistest kategooriatest kõiki eesti keele eelA1-taseme kategooriaid ja vorme, kuid kasutada oskavad lapsed vaid ainsuse nimetava käände vormi ning kindla kõneviisi oleviku esimese ja kolmanda pöörde vorme. Fraasi- ja konstruktsioonitüüpide, ühendtepusõnade ning liitsõnade kasutamise oskuse kohta hindamisvahendiga kinnitust ei saanud, seega võib hindamisvahendit pidada ainult osaliselt efektiivseks.

¹ Uuringut on rahastatud HTM-i projektist „Professionaalne eestikeelne õpetaja mitmekeelses klassis“.

Võttesõnad: eesti keele kui teise keele omandamine, keelelise arengu hindamisvahend, kakskeelsus, eelA1 keeletase, grammatiliste kategooriate omandamine

I. Sissejuhatuseks

Eesti on praegu eestikeelsele õppele ülemineku protsessis, muude tegevuste hulgas on algatatud mitu pilootprojekti, mis peavad toetama eestikeelset õpet nii mitmekeelses lasteaias kui ka põhikoolis. Samas on ikka veel väga vähe empiirilisi andmeid selle kohta, kuidas eesti keele kui teise keele omandamine täpsemalt toimub: mis kategooriad mis järjekorras omandatakse, kas ja kuivõrd sõltub omandamine sellest, kas laps saab rohkem või vähem eestikeelset sisendit, mis vanusest ta eesti keelega üldse kokku on puutunud või missugune on tema stardipositsioon ehk keeleoskus koolis eesti keelt õppima asudes.

Eesti keele kontekstis leidub esialgseid andmeid küll õpetajate, laste ja vanemate hoiakute, Kohtla-Järve koolide venekeelsete laste eesti keele õppe kohta (Zabrodskaia 2004), eesti keele kui teise keele õppe üldise korralduse (Metslang *et al.* 2013) ning eesti-vene kakskeelsete väikelaste grammatiliste oskuste kohta (Hallap *et al.* 2014; Hallap, Padrik 2019), kuid puudu on uuringud, mis vaatleksid laste keelelist arengut dünaamiliselt, st kindla ajalõigu jooksul, ning keeleliste kategooriate ja vormide arenemise seisukohalt. Eesti Keele Instituudis on loodud laste eesti keele oskuse tasemekirjeldused (nt eelA1 keeletaseme kirjeldus), kuidas aga kirjelduses esitatud grammatilised kategooriad tegelikkuses lastel keele õppimise käigus arenevad, ei ole siiani vaadeldud. Samuti puudub laste grammatiliste oskuste uurimiseks ja hindamiseks väljatöötatud ja läbikatsetatud metoodika.

Teise keele omandamist käsitlevates uuringutes (Jia, Fuse 2007; Unsworth *et al.* 2014; De Wilde *et al.* 2021) on leitud, et keele omandamine on seda edukam, mida rohkem õppija keele sisendit saab (ka kumulatiivselt). Erineva keeletasemega ja erineval hulgal keelelist

sisendit saanud õppijad on mitmekeelse klassi igapäevane tegelikkus ka Eestis (vt nt Lauristin *et al.* 2011; Raud *et al.* 2011; Juurak 2020) ja paljude laste jaoks ongi traditsiooniline keeletund ainuke eesti keele sisendi saamise koht (Argus, Baird 2022). De Wilde jt (2021) uuring on kinnitanud, et tõhusa keeleomandamise jaoks vajavad algkooli-ealised õpilased traditsioonilisele keeletunnile lisaks kontekstuaalset kokkupuudet õpitava keelega, milleks sobib näiteks igapäevane vestlus sihtkeeles nii tunnis kui ka vahetunnis. Teadlik (*explicit*) ja juhuslik (*incidental*) keele omandamine toetavad teineteist (Muñoz 2006) ja laste puhul on leitud, et just juhuslik keele omandamine toimub kiiremini ja väiksema kognitiivse koormusega kui teadlik ehk õppetegevuse kaudu toimuv keeleõpe (Paradis 2004, 2009; Ellis 2005, 2009; Lichtman 2016). Juhuslik keeleomandamine on kontekstuaalset laadi, st et toimub siis, kui tehakse tegevusi, mis ei ole keeleõppega seotud, olgu selleks nt kehalise kasvatused või tööõpetuse tunnid. Watts-Taffe ja Truscott (2000) on leidnud, et keeleõppe edukuse jaoks on oluline ka võimalus kuulda ja näha teisi lapsi kõnelemas, mis samuti toetab juhuslikku keeleomandamist. Nad toovad näiteks vestlusringid, kus nii keeleõppijatel kui ka emakeelsetel kõnelejalatel on võimalik kasutada keelt tähendusrikkalt ja eesmärgipäraselt. Seega toimub keeleõpe edukamalt siis, kui lapsel on peale keeleõppetunni veel võimalusi eesti keeles kõneleda.

Õpitava keelega kokkupuute maht ja sagedus on sõnavara omandamisel määrava tähtsusega, sest ainult keeletunnis õpitav ei ole piisav, eriti olukorras, kus sihtkeelega puudub kooliväline kokkupuude, näiteks sotsiaalmeedias suhtlemine, sõpradega mängimine või muusika kuulamine (Ellis 2002; Puimège, Peters 2019). Juhusliku keeleomandamise (Hulstijn 2003; Ellis 2009) toetavat rolli keeleõppes ei tohiks alahinnata. Juhusliku keeleomandamise olukorrad on võimalikud ka koolis, kus õpetaja räägib õpilasega igapäevateemadel, pööramata tähelepanu keelekasutusele ja vigade parandamisele, või situatsioonid, kus õpetajad räägivad omavahel õpilaste kuulates sihtkeeles. Spontaansed õpetaja-õpilase vestlused toetavad ka juhuslikku grammatika omandamist (Ellis 2009; Shintani 2015).

Sihtkeele pidev kuulamine ja väljendite kordumine toetab lisaks sõnavarale grammatiliste vormide omandamist (van Zeeland, Schmitt 2013). Sõnavara omandamisel mängib kontekstuaalne keeleomandamine eriti suurt rolli verbide puhul, sest verbid vajavad rohkem situatiivset keskkonda (Puimège, Peters 2019). On oluline märkida, et varase ea sõnavara oskus korreleerub teiste keeleõppe aspektidega, näiteks grammatika oskusega (De Wilde *et al.* 2021).

Kuivõrd teise keele õppimist ja selle omandamise kiirust mõjutavad väga paljud faktorid (näiteks motivatsioon, hoiakud, võimekus, õpistrateegiad, kodune taust, suhtlemine teistega), peavad õpetajad lähtuma õpilaste erinevustest (vt Tomlinson 2001a; de Bot *et al.* 2006, 2007; Chamot 2012). Õppe diferentseerimine keeleõppes on õppijakeskne lähenemine, mis tähendab tegevuste raskuste valikut õpilaste vajadustest lähtuvalt, samas ei eeldata, et iga tegevus peab olema diferentseeritud, vaid lähtutakse põhimõttest, et see toetaks õpilast (Blaz 2006; Chamot 2012). Siinjuures on oluline, et diferentseerimine toimuks mõlemas suunas – keeleliselt edasijõudnum õpilane saaks areneda talle kohasel viisil, samal ajal kui madalamal keeletasemel olev õpilane õpib temale jõukohaste tegevuste kaudu (Tomlinson 2001b). Varasemad uuringud on näidanud, et ehkki diferentseerimise vajadust teadvustatakse ja seda ka tehakse, siis just edasijõudnumatele õpilastele napib neile kohaseid tegevusi (Gunnulfsen, Møller 2017; Brevik *et al.* 2018). Tihti nimetatakse neid õpilasi andekateks või võimekateks, tegelikkuses on see rühm palju heterogeensem ja hõlmab u 10–15% õpilastest (Gagné 2005; Renzulli 2005).

Eeltoodust tulenevalt on siinse uuringu eesmärgid:

1. lisada teadmisi selle kohta, mis järjekorras eesti keele kui teise keele grammatiliste vormide õppimine 7–9-aastastel lastel toimub;
- 1.1. vaadelda erineva keeleoskusega laste tulemusi võrdlevalt ehk teha kindlaks, kas ja kuidas erineb kahe erineva keeleoskusega laste grupi eesti keele oskus pärast üht aastat eesti keele õpet, kui õpetajate hinnangul eri keeleoskustasemega

õppijaid õpetati kõiki ühtviisi ehk õpet teadlikult diferentseerimata;

- 1.2. tuvastada juhusliku keeleõppe mõju keeleomandamisele ehk vaadelda seda, kuivõrd võib mõjutada laste keeleoskust see, kas nad viibivad eestikeelses keskkonnas ka väljaspool traditsioonilisi keeletunde;
2. katsetada projektis siiani kasutatud hindamisvahendi efektiivsust EKI loodud eelA1-taseme kategooriate loendis esitatud grammatiliste kategooriate mõistmise ja kasutamise oskuse mõõtmisel.

Et sõnavara hindamisvahendeid on teiste keelte eeskujul võimalik lihtsamini kohandada ja kasutada, on siinse uuringu materjali kogumiseks kasutatava hindamisvahendi keskmes just grammatiliste ja grammatilis-leksikaalsete (nt ühendverbid ja liitsõnad) kategooriate omandamiseks sobiva hindamisvahendi väljatöötamine. Eesti keele rikkast muutevormistikust ja grammatilisest iseloomust lähtuvalt ei ole võimalik teistes keeltes loodud hindamisvahendit üle võtta. Laste keeleoskuse hindamiseks kasutatakse teistes keeltes väga eri tüüpi hindamisvahendeid (nt MAIN-test, mis mõõdab jutustamisoskust, LITMUS-projekti loodud lausete järelekordamise test), grammatiliste oskuste mõõtmiseks kasutatavaid vahendeid ei ole aga kuigi palju. Kõige lähedasem siinses uuringus kasutatud hindamisvahendile on LITMUS-projektis loodud käändevormide kasutusoskust mõõtev test Case Contrasting², kuid et eesti keele käändesüsteem on tunduvalt laiem kui saksa keele oma, ei sobi see eesti keele tarbeks ja sobiv hindamisvahend tuleb nii või teisiti eesti keele jaoks luua ja läbi katsetada.

Uuringu eesmärkide täitmiseks otsitakse vastuseid järgmistele uurimisküsimustele: 1. Millised erinevused ilmnevad õpetajate hinnangul kahe erineva keeleoskusega laste grupi eesti keele grammatiliste vormide, konstruktsioonide, ühendverbide ja liitsõnade mõistmise ja kasutamise oskuses esimese õppeaasta lõpuks? 2. Mis

² Vt <https://www.bi-sli.org/other-tasks-from-the-action> (06.10.2022).

iseloomustab vaatlusaluste laste eesti keele grammatiliste kategooriate mõistmise ja kasutamise oskust ja kuidas vastavad EKI loodud eelA1-taseme kategooriad õpilaste omandatud keelele? 3. Kas võib leida seoseid laste grammatiliste oskuste ja eesti keelega kokkupuutumise kestuse ja kanali (ehk juhusliku keeleõppekeskkonna olemasolu või puudumise) vahel? 4. Kuidas aitab seni kasutusel olnud hindamisvahend mõõta grammatiliste kategooriate mõistmise ja kasutamise oskust?

2. Meetod ja valim

Aastal 2020 alustati Eestis pilootprojekti „Professionaalne eestikeelne õpetaja mitmekeelses klassis“. Projektiga liitunud koolides toetatakse mitmekeelsesse klassi ühe eestikeelse lisaõpetaja palkamist ning õpetajate erialaste oskuste arendamist. Projektikoolides alustasid lisaõpetajad tööd klassides, kus oli muukeelseid lapsi, ja projektiklasse on põhiliselt kolme tüüpi: eestikeelne klass, kus õpib ühes klassis mitme eri emakeelega lapsi; eestikeelne klass, kus õpib rohkem kui 10 vene emakeelega last; ning vene õppekeelega kool, kus õpe toimub osa ainetes eesti keeles ning kus klassis on peamiselt vene emakeelega lapsed.³ Õpilaste keeleoskust hinnatakse katva uuringuga kõigepealt projekti alguses (vt Argus, Baird 2022) ning seejärel iga kahe aasta tagant ning valikulise kvalitatiivse uurinuga kord aastas. Keeleoskuse uuringuks loodi hindamisvahend, mida iga järgneva uuringuga täiustatakse. Siinse artikli aluseks olev uurimismaterjal moodustab ühe osa projektiga liitunud koolide laste keeleoskuse hindamisest.

Valim. Programmi „Professionaalne eestikeelne õpetaja mitmekeelses klassis“ raames 2020. aastal eesti keelt õppima asunud 7–9-aastaste laste hulgast valis iga õpetaja hindamiseks kaks erineva nn stardipositsiooniga last. Laste valiku põhimõtteks oli leida kaks

³ Projekti kohta vt lähemalt <https://www.hm.ee/et/tegevused/eesti-keel-ja-voorkeel/eestikeelsed-opetajad-lasteaada-ja-kooli> (06.10.2022).

sellist last, kellest üks ei rääkinud projekti alguses (2020. a sügisel) veel üldse eesti keelt ehk kelle puhul oli õpetaja 2020. a sügisel hindamisel teinud hindamislehele vastava märkuse, ja teine laps, kes samal ajal juba vähesel määral rääkis eesti keelt ehk kuulus nende laste hulka, kelle kohta oli õpetaja 2020. aasta sügisel märkinud hindamislehele, et õpilane räägib vähesel määral eesti keelt (ei oska iseseisvalt eestikeelseid lauseid moodustada). Õpetajatel paluti valida eelpool kirjeldatud tunnustega õppija prototüüpsed esindajad. Kahe eri nn stardipositsiooniga õppija keeleliste oskuste hindamine annab võimaluse saada vastus küsimusele, kas ja mille poolest erineb nende laste keeleoskus pärast üht õppeaastat teadlikult diferentseerimata keeleõpet.

Keelelise arengu hindamise vahend jagati õpetajaile välja aprillis 2021 ja saadi täidetuna tagasi juuni alguseks 2021. Kokku on analüüsitud 35 lapse keelelist arengut (18 õpetajalt). Lastest 46% olid tüdrukud ja 49% poisid, seega on sooline jaotus valimis üsna võrdne. Kahe lapse puhul ei olnud õpetaja sugu märkinud. „Oskas natuke“ gruppi kuulus 6 tüdrukut ja 10 poissi, „Ei osanud üldse“ gruppi 10 tüdrukut ja 7 poissi. Lapsed olid hindamise ajaks osalenud projektis 6–10 kuud (keskmiselt 8,3 kuud). Laste vanus jäi 7–9 eluaasta piiridesse (keskmine 7,7 aastat). Valimisse kuuluvad lapsed õppisid 1) eestikeelses koolis, kus klassis oli mitu muukeelset õpilast), 2) eestikeelses koolis, kus klassis oli mitu muukeelset õpilast, kellest suurem osa on vene kodukeelega, 3) venekeelse kooli keelekümbelusklassis või venekeelses koolis, kus osa ainetest õpetatakse eesti keeles. Kõikidest vaatlusalustest oli kool ainukeseks eesti keelega kokkupuutumise kohaks 11 lapsel, suurem osa (7) neist lastest kuulus sellesse gruppi, kes eesti keelt projekti alguses veel ei rääkinud.

Enamik lapsi (77%) kuulub perre, kus mõlemad vanemad räägivad lapsega vene keelt. Ühe lapsega räägivad mõlemad vanemad ukraina keelt ning üks laps kasvab peres, kus mõlemad vanemad räägivad temaga poola keelt. Valimis oli ka simultaanseid kakskeelseid lapsi ehk lapsi, kes hakkasid kahe keelega kokku puutuma alates sünnist (termini kohta vt Yip 2013). Neid oli neli: kolm, kes rääkisid

kodus nii eesti kui ka vene keelt, üks laps rääkis vene ja armeenia keelt. Ülejäänud valimis olnud lapsed olid suksessiivsed kakskeelsed, mis tähendab, et nad hakkasid teist keelt kuulma hilisemas eas (termini kohta vt Yip 2013). Nendest lastest üks kõneles kodus poola, üks ukraina ja üks inglise keeles.

Ülejäänud lastest 27%-l algas kontakt eesti keelega siis, kui nad said 7-aastaseks ehk läksid kooli, 29% lastest algas see kontakt kolme aasta vanuselt (ilmselt seoses lasteaeda minekuga). Ülejäänud lapsed jaotusid kontakti algust silmas pidades juhuslikult (eesti keelega hakkasid need lapsed kokku puutuma teise ja kuuenda eluaasta vahel).

Peale selle, et kõikidel lastel oli eesti keelega kokkupuude lasteaias või on kokkupuude koolis, puutuvad uuringus osalejad kokku eesti keelega ka sõpradega mängides, õdede-vendade ja sugulaste kaudu (need lapsed on enamasti kakskeelsetest peredest pärit) ning TV, arvuti ja/või raamatute vahendusel. Vene emakeelega laste puhul on kõige tavalisem kokkupuude eesti keelega peale lasteaia või kooli kas arvuti, televisiooni või raamatute kaudu. Valimis oli 11 last, kelle jaoks oli õpetajate väitel lasteaed või kool ainuke koht, kus ta eesti keelega kokku puutub, kuivõrd elatakse piirkonnas, kus perel puudub eestikeelne suhtlusring. Täpsema ülevaate laste eesti keelega kokku puutumise võimalustest leiab tabelist 1.

Tabel 1. Kokkupuude eesti keelega (laste arv)

Lapse kodune keel	Koolis/ lasteaias	Sõpradega suheldes	Sugulastega suheldes	TV/arvuti/ raamatute vahendusel	Muul viisil
Vene (27)	27	3	2	15	1
Vene-eesti (3)	3	3	3	3	0
Muu keel(ekombinatsioon) (4)	4	3	2	4	0

Instrument. Laste keelelist arengut hindasid õpetajad selle projekti raames välja töötatud küsimustikuga. Ühtset laste keelelise arengu hindamise instrumenti ei ole Eestis veel ei eesti kui esimese ega ka

eesti kui teise keele omandamise jaoks olemas. Esimese kooliastme jaoks on olemas hindamisvahend lugemispädevuse (Toomela *et al.* 2020), kuid mitte keelelise arengu hindamiseks.

Kuigi võib oletada, et eri koolitüübid võivad eesti keele õppimist mõjutada, ei ole andmeid koolitüübiti siiski võimalik võrrelda, sest eri tüüpi koolid ei ole projektis võrdsel arvul esindatud ning üldistuste tegemiseks oleksid grupid liiga ebavõrdsed.

Lapse keelelise arengu hindamise vahend koosnes kõige üldisemal tasandil kolmest osast: metaandmetest, üldist suhtlusoskust puudutavatest küsimustest ja grammatiliste kategooriate kohta käivast osast. Metaandmete osa ehk koduse keelekeskkonna kohta käivad andmed võimaldavad võrrelda eri vanuses eesti keelega kokku puutuma hakanud laste keelelist arengut ning leida ka nn juhusliku sisendi mõju laste eesti keele kui teise keele arengule. Hindamisvahendi esimese osa aluseks on projekti esimesel, 2020. aasta sügisel kasutatud lapse suulist suhtlust puudutavad peamised küsimused, mille vastuste puhul kasutatud kolmepalline skaala asendati siinses uuringus analüüsi täpsuse huvides viiepallisega.

Sissejuhatavas metaandmete plokis küsiti laste koduse keelelise keskkonna kohta, nt vanemate kasutatavate keelte, lapse kokkupuutumise kohta eesti keelega jms (kokku kaheksa küsimust). Koduse keelekeskkonna andmed said õpetajad omakorda vanematelt, enamasti küsiti need andmed arenguestluse ajal või kui see ei olnud võimalik, kirjaliku suhtluse käigus⁴. Üldised suhtlusoskuse küsimused (kokku 17 küsimust) puudutasid toimetulekut igapäevasisuatsioonides, teksti ja lause tasandi oskusi, kirjeldusoskusi ning lisaks küsiti hääldusoskuse ja lapse kõnes sisalduva koodivahetuse määra kohta. Teise osa puhul oli tegemist täiesti uue grammatiliste oskuste hindamise vahendiga, mille aluseks oli EKI õpetaja tööriistadena loodud eelA1-taseme grammatiliste kategooriate loend⁵. Eel-A1

⁴ Koduse keelekeskkonna andmed küsiti teadlikult vanematelt, mitte lastelt, sest lastelt saadavad andmed ei pruugi olla korrektsed, nt ei ole kindel, kas lapsed mäletavad seda, kas nad hakkasid eesti keelega kokku puutuma enne kolmandat eluaastat või mitte.

⁵ Loend leitav aadressil <https://sonaveeb.ee/teacher-tools/#/home>.

keeletaset on kirjeldatud kui kõige olulisemat ja just lapse jaoks suhtlemiseks vajalikku keelevahendite kogumit. Eel-A1 taseme kirjelduse juures on toodud sõnavaraloend (u 1000 sõna) ning grammatiliste kategooriate ja konstruktsioonide loend, mida sel tasemel keeleõppija oskama peaks. Vormimoodustuse puhul on esitatud kõige varem omandatavad vormid, lausemoodustuse puhul kõige sagedasemad fraasi- ja konstruktsioonitüübid. Sõnamoodustuse juures on tegusõnade puhul öeldud, et eelA1-taseme keeleomandaja mõistab ja oskab kasutada kõige sagedasemaid ühendverbe ning nimisõnade puhul mõistab ja oskab ta kasutada tuttavatest osistest koosnevaid liitsõnu. Grammatiliste kategooriate puhul on tasemekirjelduses ka märgitud, mis kontekstis ja kui spontaanselt peaks keelekasutaja neid kategooriad kasutada oskama (nt „Oskab sagedasemaid eakohaseid tegusõnu kasutada käskiva kõneviisi ainsuse teise pöörde vormis (nt *tule*), et väljendada ühele kuulajale suunatud käsku või palvet“). Õpetajatel tuli hinnata laste grammatilisi oskusi (tegu sõna ja nimisõna vormide, fraasi- ja konstruktsioonitüüpide ning mõne ühendverbi ja liitsõna mõistmise ja kasutamise oskust). Kokku esitati grammatiliste kategooriate kohta lapsele mõistmisoskuse kindlakstegemiseks 12 väidet ning samade kategooriate kasutusoskuse määramiseks samuti 12 küsimust. Tegusõnavomide kohta esitati 11 väite-küsimuse paarikut. Nimisõna vormistiku (10 eri vormi), fraasi- ja konstruktsioonitüüpide (9 tüüpi) ja sõnamoodustuse (2 küsimust) puhul oli õpetajail abiks pilt (näide 1), mille kohta sai esitada küsimusi, et vastava grammatilise kategooria mõistmist või kasutamisoskust esile kutsuda ja kontrollida (vt näide 2). Vastused hinnati skaalal „õige“ vs. „vale“. Valede vastuste hulgas kodeeriti eraldi need vastused, mis olid sisult õiged, kuid kus laps oli kasutanud mitte ootuspärast vormi).

Küsimustiku esimeses osas kasutas õpetaja lapse suhtlusoskuse hindamiseks viiepallist Likerti skaalat (1 – *ei oska üldse*, 5 – *oskab väga hästi/veatult*). Kvantitatiivsete andmete analüüsimiseks kasutati statistikaprogrammi IBM SPSS Statistics 27. Andmete analüüsimisel kasutati valimi iseloomustamiseks ning respondentide

Näide 1. *Küsimuste esitamisel kasutatud pilt***Näide 2.** *Katseküsimus*

Vorm	Mõistmisoskus <i>Mina ütlen sulle midagi ja kui sinu meelest on see õige, siis ütle seda ja näita mulle seda ka pildi peal.</i>	Kasutusoskus Hindaja osutab isikule või objektile pildil ja küsib.
Osastav (keda/mida?)	Vanaema joob kohvi.	<i>Mida koer sööb?</i> Oodatav vastus: (Koer sööb) kooki/torti.

hinnangute võrdlemiseks (1. rühm, kes oskas enne projektiga liitumist mingil määral eesti keelt; 2. rühm lapsi, kes ei osanud eesti keelt projekti algul üldse) sagedustabeleid, keskmiste võrdluseid ja jaotust. Vastajate rühmade keskmiste tulemuste statistiliste erinevuste hindamiseks kasutati mitte-parametrilist testi (Mann-Whitney). Et hinnata eesti keelega kokkupuute aja ja liigi seost eesti

keele oskusega, jagati õpilased gruppidesse vastavalt sellele, kas neil oli kontakt eesti keelega ainult koolis või said nad ka koolivälisest sisendist. Seejärel loodi mõlema grupi jaoks neli eraldi koondskoori: üldine suhtlusoskus (9 hinnatavat komponenti), kirjeldamisoskus (6 hinnatavat komponenti), tegusõnade kasutamisoskus (11 hinnatavat komponenti) ning nimisõnade kasutamisoskus (11 hinnatavat komponenti). Üldise suhtlusoskuse ning kirjeldamisoskuse puhul kasutati õpetajate hinnangute keskmist ja grammatiliste oskuste puhul andsime õigete vastuste eest ühe punkti ning arvutasime selle põhjal välja kahe grupi keskmised tulemused. Seejärel kasutati kahe grupi statistiliste erinevuste hindamiseks Mann-Whitney U testi. Seos loeti statistiliselt oluliseks, kui p väärtus jäi alla 0.5. Küsimustiku kolmandas osas hindas õpetaja lapse keeleoskusi hinnangutega „õige“, „vale“ ja „sisult õige, kuid mitte ootuspärane“. Viimati mainitud hinnangu tingis vajadus aru saada, kas osalejad vastavad hindamisvahendi küsimustele nii, nagu koostajad eeldasid. Neid tulemusi analüüsiti χ^2 -testi abil, et selgitada välja kahe rühma vahelisi statistilisi erinevusi.

Meetodi sobivuse hindamiseks analüüsiti küsimustiku teise osa vastuseid kvalitatiivselt ehk nii, et vaadeldi ka nende vastuste hulka, mis ei sisaldanud ootuspärast vormi, kuid olid sisuliselt õiged. Samuti vaadati üle kõik valed vastused ning rühmitati need vastavalt ühistele tunnustele. Kindlate küsimuste puhul tekkinud sarnaste tunnustega vastused võimaldasid hinnata küsimuse eesmärgikohasust.

3. Laste keeleoskuse hindamise tulemused

Lähteülesanne oli õpetajal valida kaks last, kellest üks ei rääkinud projekti alguses eesti keelt ja teine, kes vähesel määral rääkis. Edasises keeleoskuse analüüsis ongi vaadeldud neid kaht gruppi eraldi: vastavalt grupid „Ei osanud üldse“ (18 last) ja „Oskas natuke“ (17 last). Nende laste hulka, kes eesti keelt veel ei rääkinud, arvati lapsed, kelle kohta oli õpetaja hindamislehele lapse

keelelise arengu kohta 2020. aasta sügisel ehk projekti alguses kirjutatud näiteks „ei rääkinud üldse eesti keelt“ või nt „on nn vaikivas perioodis“.

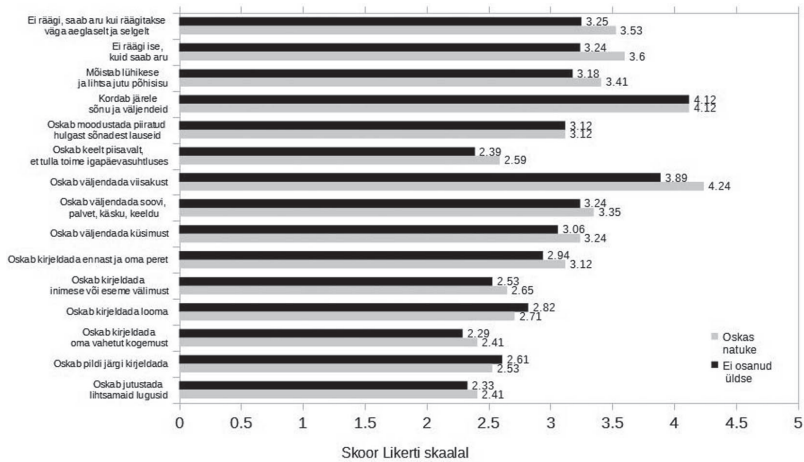
3.1. ÜLDINE SUHTLUSOSKUS JA SUULINE KÕNE

Suhtluseesmärkide täitmist vaadeldi kolmes osas. Esmalt pidi õpetaja hindama lapse üldist arusaamist eestikeelsest kõnест ning see osa oli jaotatud omakorda kolmeks ehk vaadeldi seda, kas räägitu mõistmisel laps: 1) ei räägi ise ja saab küsimusest või korraldusest aru ainult siis, kui see on öeldud väga aeglaselt ja selgelt; 2) ei räägi ise, kuid saab korraldustest ja küsimustest aru, nt suudab järgida lühikesi ja lihtsaid näpunäiteid tavapärasel viisil öelduna (nt *tule siia, pane riidesse*); 3) mõistab lühikese ja lihtsa jutu põhisisu.

Vahe kahe grupi keskmiste tulemuste vahel (vt joonis 1) on küll kõigi kolme küsitud oskuse puhul olemas, kuid mitteparameetrilise testi (Mann-Whitney) järgi ei ole erinevused statistiliselt olulised ($p > 0.5$) ja mõlema grupi laste keeleoskust iseloomustab ühtmoodi skaala keskmine väärtus ehk kolm. Statistiliselt mitteoluline tulemus muutub aga oluliseks uuringu konteksti valguses: keeleoskuse stardipositsioon oli osalejate seas erinev, kuid lõpp-positsioonis enam keelelist vahet pole.

Laste produktiivse keelekasutuse ehk kõnelemisoskuse üldisest hindamisest nähtub, et sarnaselt mõistmisoskusega paistavad ka kõnelemisel kahe grupi tulemused väga sarnased ja statistiliselt olulist erinevust ei ilmnenud (vt joonis 1). Sõnade-väljendite järelkordamisega paistavad hakkama saavat mõlema grupi lapsed, see oskus on tasemel, kus tehakse üksikuid vigu. Raskem ehk paljude vigadega on lastel lausete loomise oskus ning igapäevasuhtlusega toimetulekuks on keeleoskus veel napp (keskmine punktisumma üle kahe osutab sellele, et tehakse väga palju vigu).

Et vaadelda lapse oskust väljendada eri tüüpi suhtluseesmärke, pidid õpetajad hindama, kas laps oskab väljendada viisakust (tervitused, tänu jms), direktiivsust (soovi, palvet, käsku, keeldu) ja



Joonis 1. Laste üldine suhtlusoskus õpetajate hinnangute põhjal

küsimusi. Kuigi keskmised tulemused on grupil, kes oskas projekti alguses juba pisut eesti keelt rääkida, veidi paremad (vt joonis 1), ei ole erinevus siiski statistiliselt oluline. Viisakuse väljendamise oskus on mõlema grupi lastel olemas (nad teevad vähesel määral vigu) ja käskude-keeldude väljendamisega saadakse pisut paremini hakkama kui küsimuste moodustamisega, mõlema puhul on tulemus selline, et tehakse sageli vigu (vt joonis 1).

Arusaadavalt on tunduvalt vähem arenenud laste oskus kirjeldata ja jutustada, need on oskused, mis nõuavad oluliselt rikkamat keelevahendite pagasit kui vaadeldavatel lastel sel perioodil olemas. Peaaegu kõikide siinses osas hinnatud oskuste puhul on laste tulemus hinnatav kirjeldusega „oskab väga vähe / teeb väga palju vigu“ (vt joonis 1), ennast ja oma peret oskavad lapsed veidi paremini kirjeldada ning pisut paremad tulemused on lastel, kes projektiperioodi alguses natuke eesti keelt oskasid. Kahe grupi tulemuste vahe ei ole aga statistiliselt oluline.

Häädust pidid õpetajad hindama viiepallisel skaalal, alates sellest kui „häädus on eesti keelele omane, ei ole midagi, mis tähelepanu ärataks“ kuni selleni, et „häädusvead takistavad lapse kõne

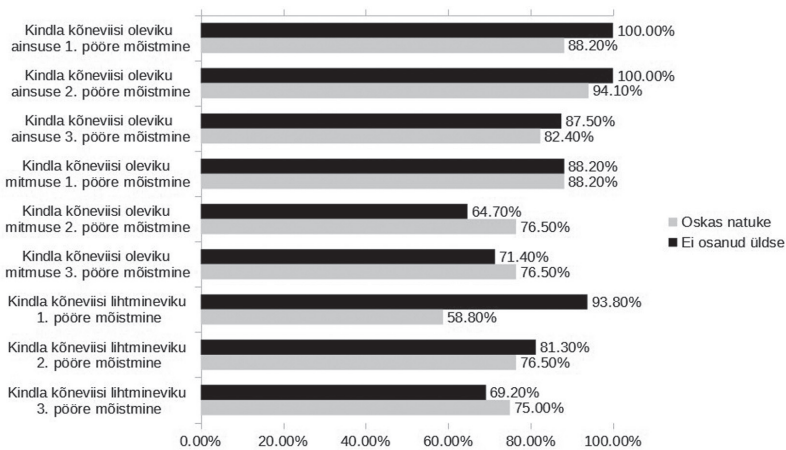
mõistmist“. Mõlema rühma laste hääldusoskus jäi aga väärtuse 3 („Häälduses on mitte eesti keelele omaseid tunnuseid palju, aga need ei takista eesti keeles suhtlemist“) ja 2 („Hääldusvigu on palju, kuid lapse kõne on siiski enamasti arusaadav“) vahele.

3.2. GRAMMATILISTE OSKUSTE OMANDAMINE

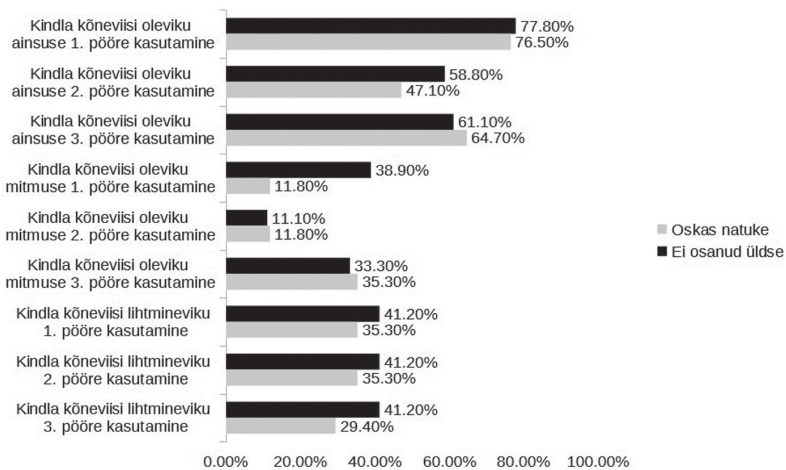
Laste grammatiliste oskuste hindamine oli jagatud nelja alaossa: eraldi vaadeldi tegu- ja nimisõna vorme, fraasitüüpe, öeldisega seotud konstruktsioone ja sõnamoodustust. Kõikide kategooriate puhul pidid õpetajad hindama nii keelendite mõistmise kui ka kasutuse oskust. Grammatiliste oskuste analüüs peaks andma vastuse, kas eelA1-taseme kategooriad on vaatlusalustel lastel omandatud või mitte. Kahe grupi statistiliste erinevuste hindamiseks kasutati Mann-Whitney U testi. Vastu ootusi tulemustes statistilist erinevust kahe rühma vahel ei esinenud.

Tegusõna vormistikust vaadeldi eelA1-taseme kirjelduses loetletud vormide, st kindla kõneviisi oleviku ja lihtmineviku vormide ning eitava kõne esimese ja kolmanda isiku vormide omandatust. Tegusõnavormide omandamise kohta on andmed esitatud joonistel 2 ja 3. Vastanute arv on tabelis küsimuseti varieeruv (mõni õpetaja oli jätnud mõne küsimuse hindamata).

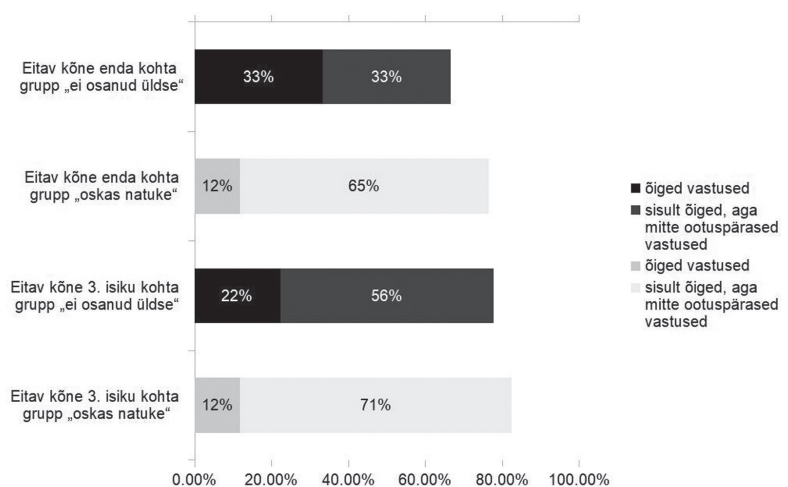
Olevikuvormidest mõistavad mõlema grupi lapsed kõiki küsitud ainsuse ja mitmuse verbivorme sisaldavaid lauseid. Ise kasutada oskavad (st ootuspäraste vastuste osakaal on üle 60%) aga ainult oleviku esimese ja kolmanda pöörde vorme (nt *mina teen, tema teeb*). Ootamatu tulemusena torkab silma, et teise pöörde vormide puhul on selle rühma, kes perioodi alguses eesti keelt veel ei rääkinud (vt joonis 2), tulemus parem kui tugevama stardipositsiooniga rühma oma, samas ei ole erinevus väikese valimi tõttu kuigi suur. Lihtmineviku verbivormide mõistmist kontrollivate küsimuste puhul võib öelda, et mõlemad rühmad mõistavad ainsuse 1., 2. ja 3. pöörde minevikuvorme (vt joonis 2). Kasutada ei oska aga neid vorme kumbki rühm (alla 42% õigeid vastuseid, vt joonis 3).



Joonis 2. Tegusõnavormide mõistmise oskus.
Ootuspäraste vastuste osakaal protsentides



Joonis 3. Tegusõnavormide kasutamise oskus.
Ootuspäraste vastuste osakaal protsentides

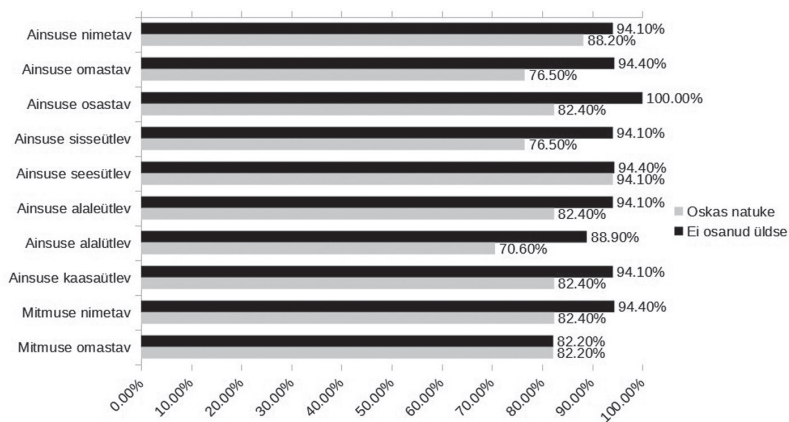


Joonis 4. Eitava kõne õigete vastuste ning sisult õigete, kuid vormilt mitte ootuspäraste vastuste osakaal protsentides

Eitava kõne 1. ja 3. pöörde vorme mõistavad mõlemad rühmad, kasutusoskuse kohta aga küsimustikust vastust ei saanud, sest ülekaalukalt (vt joonis 4) kasutasid lapsed sisult õiged, aga vormilt mitte ootuspäraseid vastuseid ehk vastasid enamasti ilma konkreetse verbivormita, ainult eituspartiklit kasutades (nt oodatud vastuse *ei oska* asemel *ei*).

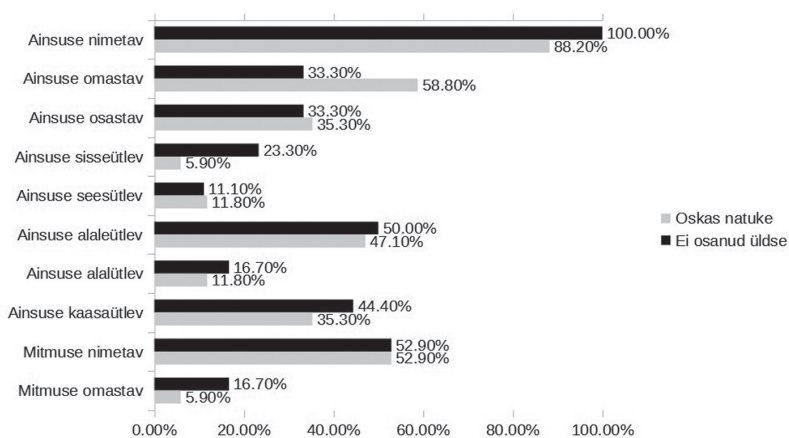
Nimisõna vormistikust pidid õpetajad hindama eelA1-taseme kirjelduses toodud vorme (sagedasemad ja varem omandatavad käändevormid). Jooniselt 5 on näha, et ainsuse käändevormidest mõistsid mõlema rühma lapsed kõiki küsitud käändevorme. Torkab aga silma, et kahe grupi võrdluses mõistavad need lapsed, kes projekti alguses eesti keelt ise veel ei rääkinud, kõiki vorme mõnevõrra paremini kui see grupp, kes projekti alguses juba oskas eesti keeles rääkida.

Nimisõnavormide kasutusoskus oli aga märkimisväärselt nõrgem kui mõistmine ja üle 60% ootuspäraseid vastuseid anti ainult ainsuse nimetava vormide kohta (vt joonis 6). Ainsuse omastava



Joonis 5. Nimisõnavormide mõistmise oskus.

Ootuspäraste vastuste osakaal protsentides



Joonis 6. Nimisõnavormide kasutamise oskus.

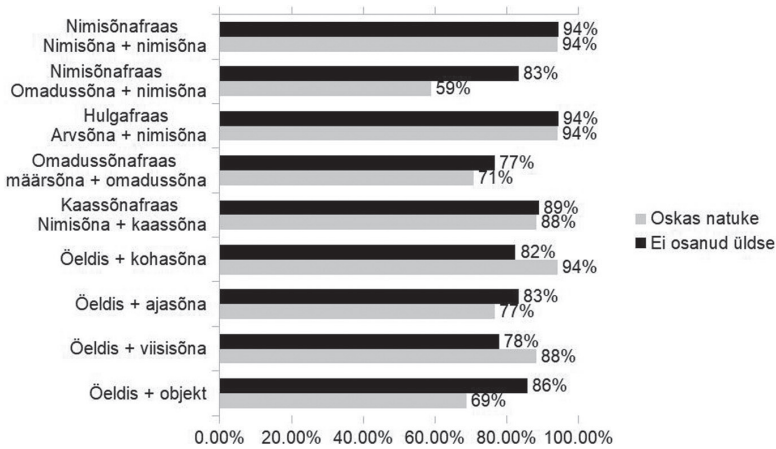
Ootuspäraste vastuste osakaal protsentides

ja osastava vormi puhul kasutati sageli näiteks nimetava käände vormi, sisseütleva puhul leidis valede vastuste hulgas nimetava käände vormi kõrval paljudel juhtudel ka vastamata jätmist, seesütleva, alaleütleva, alaltütleva käände vormi asemel kasutati samuti nimetava käände vormi, mitmuse nimetava puhul kõige sagedamini ainsuse nimetavat käännet, aga ka näiteks kvantorifraasi „kaks lind“, kus nimisõna on osastava asemel nimetavas. Torkab silma, et ainsuse omastava käände vormide kasutusoskus on tugevamal grupil mõnevõrra parem kui nõrgemal, kuid vahe ei ole statistiliselt oluline. Ainsuse omastava käände puhul (nt *poisi*) vastati tihti ainsuse nimetavat kasutades (nt *poiss*).

Mitmuse vormidest mõistavad mõlema rühma lapsed nimetava ja omastava vorme, kasutada ühtki mitmuse vormi ei osata (vt joonis 6). Mitmuse nimetava vormide kasutamise oskus on siiski parem kui omastava käände vormidel, seal oli õigeid vastuseid isegi alla 10% (rühmas, kes perioodi alguses eesti keelt juba natuke rääkis). Kokku kaheksal juhul (võrdselt õpilasi mõlemast rühmast) jättis laps küsimusele vastamata ja/või ei saanud küsimusest aru, ülejäänud juhtudel kasutati kas ainsuse või mitmuse nimetava vormi (oodatava vastuse *lindude* asemel *linnu* või lihtsalt *linnud*).

Fraasi- ja konstruktsioonitüüpidest vaadeldi nimisõna-, hulga-, kaassõnafaaside ning öeldist ja tema laiendeid sisaldavate konstruktsioonide mõistmise ja kasutamise oskust (vt joonis 7, 8 ja 9). Kõikides konstruktsioonides, mille mõistmist või kasutamist hinnati, oli kasutatud eelA1-taseme sõnavara.

Mõlema rühma lapsed mõistavad kaht tüüpi ehitusega nimisõnafraasi: nimisõna (omastavas) + nimisõna, nt *ema juuksed*; omadussõna (nimetavas) + nimisõna, nt *täpiline kann*. Tõendeid aga selle kohta, et lapsed neid fraase ka kasutada oskaksid, küsimustikuga kuigivõrd ei saanud. Nimisõnast ja omadussõnast koosneva fraasi puhul anti hulk sisult õigeid, aga vormilt mitte ootuspäraseid vastuseid, nt vastati ainult ühe komponendiga fraasist (*Kelle pilt on seinal? Lapse.*). Hulgafraasi puhul on kasutusoskuse kohta andmed samuti probleemsed (oodatava vastuse *kolm lille* asemel ütlesid

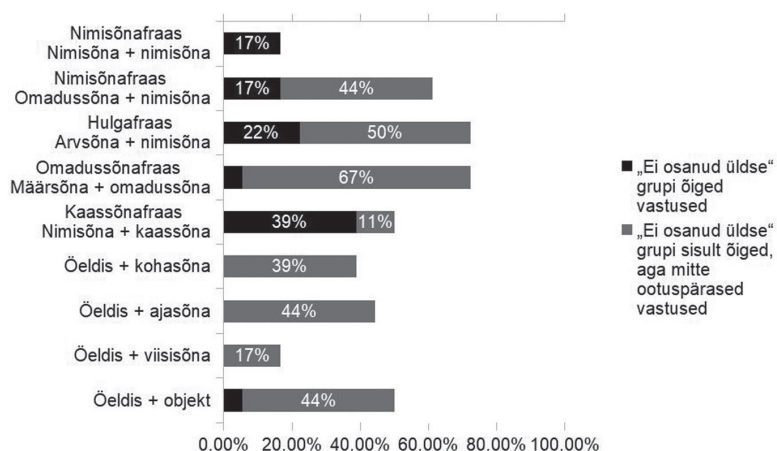


Joonis 7. Fraasi- ja konstruktsioonitüüpide mõistmise oskus.
Ootuspäraste vastuste osakaal protsentides

lapsed lihtsalt *kolm*), selline vastus on suhtluse seisukohast õige, kuid ei anna infot selle kohta, kas keelekasutaja oskab vastavat fraasi kasutada.

Omadussõnafraasi (nt *väga ilus*) mõistmisel andsid lapsed üle 70% õigeid vastuseid (vt joonis 7) ja seega võib öelda, et nad mõistavad määrsõnast ja omadussõnast koosnevat fraasi. Samasuguse fraasi kasutusoskuse kohta küsimustikust tõendust ei olnud võimalik saada, sest lapsed vastasid oodatud vastuse *väga rõõmus* asemel nii, et nimetasid ainult isiku, kes on väga rõõmus, nt *ema*).

Ka kaassõnafraasi (nimisõna + kaassõna, nt *tooli peal*) puhul on mõistmisel mõlema rühma tulemused üle 88%, kasutusoskuse tulemused jäävad aga alla 42%. Õeldisega seotud konstruktsioonidest (õeldis + kohasõna, nt *läheb õue*; õeldis + ajasõna, nt *tuleb kohe*; õeldis + viisisõna, nt *jookseb kiiresti*; õeldis + objekt, nt *tahab kooki*) on mõlemal rühmal omandatud nende mõistmine (vt joonis 7). Joonistel 8 ja 9 on esitatud mõlema grupi fraasitüüpide kasutusoskus protsentides ja vastavalt nende gruppide antud sisult õigete, vormilt mitte ootuspäraste vastuste hulk. Joonistelt on näha, et mõlemad

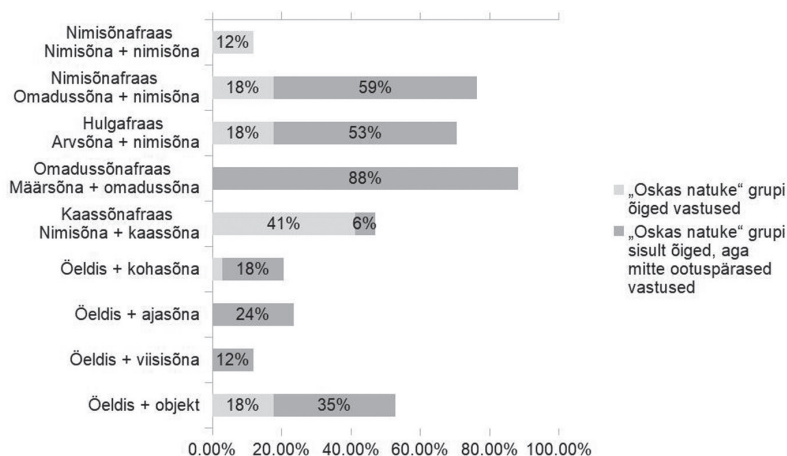


Joonis 8. Fraasi- ja konstruktsioonitüüpide kasutusoskus protsentides „Ei osanud üldse“ grupp

grupid andsid sisult õiged, vormilt mitte ootuspäraseid vastuseid samade fraasitüüpide ja konstruktsioonide kohta.

Eri fraasitüüpide ja konstruktsioonide kasutusoskuse kohta (joonis 8, 9) ei ole võimalik kasutatud õpetajate hinnangute põhjal vastust anda, sest sisult õigete ja vormilt mitte ootuspäraste vastuste hulk oli selleks liiga suur. Näiteks öeldise ja objekti ühendite puhul vastasid lapsed nii, et kasutasid ainult objekti (ehk oodatava vastuse *vaatab koera* asemel *koera*), mis on suhtluse seisukohast täiesti loomulik. Kui võrrelda kahe grupi puhul mitte ootuspäraste vastuste hulka, ei ilmne selget suundumust, nagu kasutaksid parema stardipositsiooniga lapsed rohkem selliseid keelendeid, mida hindamisvahendis ootuspäraseks ei loetud, kuid mis on suhtluse seisukohalt loomulikud (vrd joonis 7 ja 8). Eri konstruktsioonitüüpides on kord ühel, kord teisel rühmal rohkem mitte ootuspäraseid vastuseid.

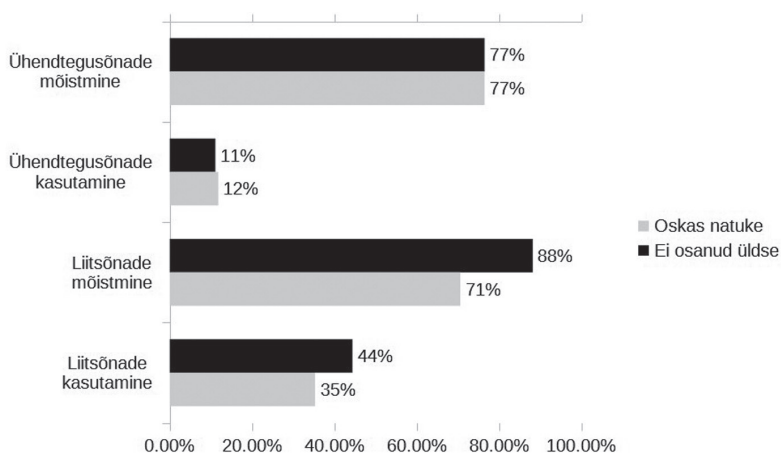
Sõnamoodustuse puhul vaadeldi ainult ühendtepusõnade ja liitsõnade omandamist. EelA1-taseme kirjelduse järgi peaksid lapsed mõistma ja oskama kasutada lihtsamaid ja tuttavaid ühendtepusõnu ja sagedastest sõnadest koosnevaid liitsõnu.



Joonis 9. Fraasi- ja konstruktsioonitüüpide kasutusoskus protsentides „Oskas natuke“ grupp

Hindamisküsimustikus kasutati liitsõnu, mille osised on esitatud eelA1-taseme sõnavaraloendis.

Nii lihtsamad ühendtegesõnad (nt *läheb katki*) kui ka liitsõnad (*lillevaas, raamaturiil*) on mõlemale rühmale mõistetavad (vt joonis 10). Kasutamise puhul ei olnud õigeid vastuseid piisavalt, et saaks öelda, et lapsed neid sõnu kasutada oskavad (vt joonis 10). Sageli vastasid lapsed ühendverbi asemel kas ainult partikliga (nt *läheb katki* asemel *katki*) või oli selge, et vastavat oodatud tege sõna lapse leksikonis ei ole (nt vastus *õhupall buh-buh*). Liitsõnade kasutuse küsimuses vastati sageli liitsõna asemel lihtsõnaga, nt *raamaturiil* asemel *riil* või *lillevaas* asemel *vaas*). Selliseid vastuseid võib jällegi suhtluse seisukohalt õigeaks pidada, kuid need ei võimalda hinnata oskust vastavaid sõnu kasutada. Huvitava tähelepanekuna võib mainida, et nende laste tulemused, kes perioodi alguses eesti keelt üldse ei osanud, on isegi paremad kui parema stardipositsiooniga laste omad (kuigi vahe ei ole statistiliselt oluline).

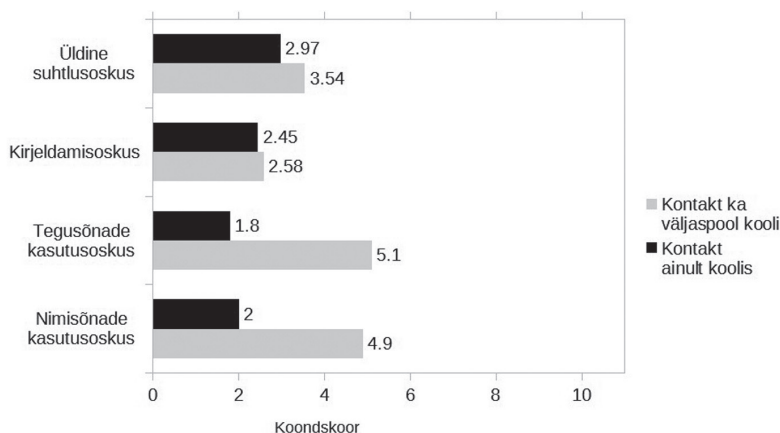


Joonis 10. Ühendverbide ja liitsõnade kasutuse ja mõistmise oskus. Ootuspäraste vastuste osakaal protsentides

3.3. EESTI KEELEGA KOKKUPUUTE AJA JA LIIGI SEOS EESTI KEELE OSKUSEGA

Valimis oli 11 last (7 last kuulus gruppi „ei osanud üldse“ ja 4 last „oskas natuke“), kelle puhul oli kool ainsaks eesti keelega kokku puutumise kohaks. Seetõttu on võimalus vaadelda, kas ja kui palju erines nende laste eesti keele oskus, kes puutusid keelega kokku ainult koolis, võrreldes nende lastega, kellel oli eesti keelega kokkupuude ka väljaspool kooli (sõpradega mängides, õdede-vendade/vanavanemate/sugulastega koos olles või TV/arvuti/raamatute kaudu).

Tulemustest ilmnes, et eesti keelega ainult koolis kokku puutuvate laste eesti keele oskus oli mõnevõrra nõrgem kui neil, kes saavad eesti keele sisendit ka väljastpoolt kooli. Üldiste suhtlusoskuste puhul olid nende laste tulemused keskmiselt 2,97 ja neil, kes ka väljastpoolt kooli eesti keele sisendit said, keskmiselt 3,54 ($p = 0.08$, Mann-Whitney U väärtus 88.5), kirjeldamisoskusele antud hinnangute keskmiste vahe ei olnud nii suur: 2,45 vs. 2,58 ($p = 0.29$, Mann-Whitney U väärtus 111.0), kuid kirjeldamisoskuse



Joonis 11. Eesti keelega kokkupuute seos eesti keele oskusega

tulemused olid kõikidel lastel madalamad kui üldise suhtlusoskuse omad (vt joonis 11). Tulemustest ilmnes, et grammatiliste oskuste puhul oli kahe grupi vaheline vahe suurem kui suhtlusoskuste puhul. Tegusõnade kasutusoskuse keskmine oli ainult koolis eesti keelega kokku puutuval rühmal 1,8 ja ka mujal kokkupuudet oma-val grupil 5,1 ($p = 0.002$, Mann-Whitney U väärtus 49.5). Nimisõnade kasutusoskuse keskmiste vahe oli sarnane: 2 vs. 4,9 ($p = 0.004$, Mann-Whitney U väärtus 36.0) ning mõlemad vahed olid statistiliselt olulised (vt joonis 11).

4. Meetodialalüüs ehk hindamisvahendi sobivus

Tulemuste põhjal võime öelda, et hindamisvahend sobis mõistmisoskuse hindamiseks väga hästi. Esiteks said lapsed hindamisvahendi puhul aru, mida nad tegema pidid; teiseks näitas küsimustele antud vastuste erinev hulk seda, et vahendi abil on võimalik parema keeleoskusega lapsi mitte nii hea oskusega lastest eristada. Ka ühe teemaploki sees (nt fraasitüübid ja konstruktsioonid) anti eri fraasivõi konstruktsioonitüüpidele eri hulgal ootuspäraseid vastuseid, mis tähendab, et vastused sõltusid küsitud keelendi keerukusest ja

omandatusest, mitte sellest, et sellist liiki küsimusele oleks üldiselt raske vastust anda.

Küll aga ilmnesid mõned hindamisvahendi kitsaskohad ja seda just grammatiliste kategooriate kasutusoskuse hindamisel. Selgus, et õpilased andsid osa küsimuste puhul palju sisuliselt ehk suhtlussituatsiooni arvestades loomulikke ja õigeid, kuid vormilt mitte ootuspäraseid vastuseid ning osa küsimuste puhul vastasid väga vähesed õpilased nii, nagu hindamisvahendi koostajad eeldasid. Fraasitüüpide ja konstruktsioonide kasutusoskuse puhul oli kaks küsimust, millele mitte ükski laps ootuspärast vastust ei andnud (tegemist oli aja- ja viisisõna sisaldavate konstruktsioonidega). Praegustele andmetele tuginedes jääb aga mõnel juhul selgusetuks, kas tegu oli meetodist tuleneva probleemiga või puuduski õpilastel hinnatav oskus. Kuna laste antud ootuspäraste vastuste hulk küsimusteti varieerus ja seda ka üht tüüpi küsimuste puhul (nt fraaside kasutusoskuse küsimuste hulgas tõusis esile kaassõnafraas, millele anti teistest oluliselt rohkem ootuspäraseid vastuseid), ei saa väita, et nõrgad tulemused olid põhjustatud alati just hindamismeetodist ja mitte õpilaste oskamatusest vastavat vormi kasutada.

Hindamisvahendi probleemsed kohad jagunesid vastuste järgi kolmeks: 1) kasutati ainult üht osa konstruktsioonist või fraasist (eriti verbivormide osas), 2) kasutati oodatava käändevormi asemel postpositsiooni, 3) muud, nt kasutatud pildiga seotud kitsaskohad.

4.1. VERBIVORMIDE, FRAASI- JA KONSTRUKTSIOONITÜÜPIDE KASUTUSOSKUSEGA SEOTUD PROBLEEMID

Eitava kõne ning pea kõikide uuritud fraasi- ja konstruktsioonitüüpide (v.a nimisõnafraas ehitusega nimisõna + nimisõna) puhul osutus probleemiks ainult ühe osa või ainult eituspartikli kasutamine. Eitava kõne puhul pidid lapsed esimese ja kolmanda pöörde vormide kasutamiseks vastama küsimustele *Kas sina näed pildil laeva?* ja *Kas sinu ema oskab lennata?*. Vastusteks oodati eitussõna ja põhiverbi, nt *ei näe*, *ei oska*. Kokku 35 lapsest vastas vastavalt

16 (ehk 44%) ja 21 last (ehk 58%) küsimustele lihtsalt eitussõnaga *ei*. Suhtluse seisukohast on selline vastus igati loomulik, kuid see ei võimaldanud hinnata laste oskust kasutada eitava kõne vorme. Lahenduseks võiks katsetada, kas tulemus oleks parem siis, kui anda lapsele üks lauseosa ette ja paluda tal lauset jätkata, nt (pilt kahe lapsega, üks sõidab jalgrattaga, teine ei sõida) *Mina ütlen sulle lause alguse ette ja sina lõpeta*: „Tüdruk sõidab jalgrattaga, aga poiss ei ... (sõida)“. Sellisel juhul võib aga muidugi tekkida ka olukord, kus laps vastab ikka jaatavas kõnes (mõne teise) verbivormiga.

Omadussõnast ja nimisõnast koosneva fraasi kasutamise oskuse hindamiseks küsiti õpilastelt pildi põhjal küsimus *Missugune tass on vanaema käes?* ning oodati vastuseks fraasi *täpiline/pruun/väike tass*. Pooled lapsed vastasid mõne värvi nimetamisega, kuid ilma nimisõna kasutamata (nt *punane, pruun*). Tegemist on jällegi suhtluse seisukohalt täiesti loomuliku vastusega, mis ometi ei võimalda hinnata seda, kuidas laps vastavat fraasi kasutada oskab. Omadussõnafraasi (ehitusega määrsõna + omadussõna, nt *väga ilus*) puhul andis sisuliselt õige, kuid vormilt mitte ootuspärase vastuse suisa 77% vastanud lastest. Näiteks vastasid lapsed *lapsed ja ema, vanaema, tüdruk ja koer*.

Kõikide tegusõnakonstruktsioonide puhul esines sama probleem: lapsed vastasid nii, et tegusõnavormi ei kasutanud ja sisult õigete, kuid vormilt mitte ootuspärase vastuste hulk ulatus 14 kuni 40 protsendini (vt tabel 2). Kõige sagedamini kasutasid lapsed ainult üht osa fraasist määrsõna ja omadussõna ühendites ning tegusõna ja objekti ühendites, nt küsimusele *Mitu lille on vaasis?* vastati enamasti vaid sõnaga *kolm*. Mõne konstruktsiooni puhul oli sisult õigete, aga mitte ootuspärase vormis vastuste hulk suurem kui vigaste vastuste hulk, nt omadussõnalise täiendiga nimisõnafraas, kuid on ka neid konstruktsioone, kus vigaste vastuste hulk on suurem, nt öeldist sisaldavad konstruktsioonid (vt tabel 2). Vigaste vastuste tüüpiliseks näiteks küsimuse *Millal ema koogiga tuppa tuleb?* puhul oli erineva käändelõpuga nimisõna, nt *tordi, tordile, koogiga* või küsimuse *Keda vanaema vaatab?* puhul nimetavas käändes *koer* või *vanaema vaatab koer*.

Tabel 2. Eri fraasitüüpide sisult õigete, kuid vormilt mitte ootuspäraste vastuste ning valede vastuste osakaal protsentides

Konstruktsioon	Sisult õigete, vormilt mitte ootuspäraste vastuste osakaal protsentides	Valede vastuste osakaal protsentides
Nimisõnafraas: omadussõna + nimisõna, nt <i>ilus pilt</i>	51,4	31,4
Hulgafraas: arvsõna + nimisõna, nt <i>kaks last</i>	51,4	28,6
Omadussõnafraas: määrsõna + omadussõna, nt <i>väga ilus</i>	77,1	20,0
Kaassõnafraas: nimisõna + kaassõna, nt <i>tooli peal</i>	8,6	51,4
Õeldise konstruktsioon: tegusõna + kohasõna, nt <i>läheb õue</i>	28,6	68,6
Õeldise konstruktsioon: tegusõna + ajasõna, nt <i>tuleb kohe</i>	38,2	61,8
Tegusõnafraas: tegusõna + viisisõna, nt <i>istub ilusti</i>	14,7	85,2
Tegusõnafraas: tegusõna + objekt, nt <i>tahab pliitsit</i>	40,0	48,6

4.2. POSTPOSITSIOONI KASUTAMINE KÄÄNDEVORMI ASEMELE

Ainsuse alalütleva käände puhul esitati lastele kaks küsimust, millega hinnata käändevormi kasutust kahes eri funktsioonis, koha ja omaja väljendajana: *Kus istub kass?* ja *Kellel on müts peas?*. Nendest kahest küsimusest osutus problemaatiliseks esimene. Kontekstis, kus alalütleva käände vorm oli kasutusel koha väljendamiseks, vastas 30% õpilastest postpositisiooni kasutades (*akna peal* või *akna juures*), mis on sellises situatsioonis ka täiesti loomulik keeleline käitumine. Kuigi suurem osa lastest (28) mõistis koha väljendamiseks kasutatud

alalütleva käände vormi, vastas sellele küsimusele ootuspäraselt ehk käänevormiga ainult viis last. Sisult õigeid, kuid vormilt mitte ootuspäraseid vastuseid anti 13 korral. Seega võib järeldada, et peale hindamisvahendist tuleneva probleemi võib olla tegemist ka sellega, et paljud lapsed lihtsalt ei osanud alalütleva käände vormi koha väljendamiseks kasutada. Siiski saaks hindamisvahendit nii muuta, et kaaluda läbi kõik eelA1-taseme sõnaloendis esitatud nimisõnad ja leida need, mille puhul on üldkeeles loomulikum kasutada alalütleva käände vormi ja mitte postpositsiooniga konstruktsiooni, ning kasutada siis pildil vastavat nimisõna.

Omajat väljendava alaleütleva käände vormiga küsimuse puhul oskasid pooled konstruktsiooni mõistnud lapsed seda ka kasutada (31-st 16). Tuleb aga mainida, et kaassõnafraasi (nimisõna + kaassõna, nt *tooli peal*) puhul ilmnes, kuigi väiksemas ulatuses, vastupidine probleem, kus kolm last vastas küsimusele *Kus vanaema istub?* ainult nimisõnaga (nt *toolil*) ja ilma kaassõna kasutamata. Hindamisvahendi parandamiseks tasuks ehk arvestada vastavate fraaside ja käänevormide puhul kasutusoskus positiivseks ka siis, kui laps on kasutanud kas käänevormi või fraasi sobivas funktsioonis mõnele teisele küsimusele vastamiseks. Samuti tasub proovida kasutada mõnd teist küsimust, kus tuleks kohasuhe selgemalt esile.

4.3. MUUD PROBLEEMID

Probleemseks osutus ka ainsuse seesütleva käände vormide kasutusoskuse hindamine. Selle küsimuse (*Kus on linnud?*) puhul tuleb kitsaskohaks pidada kasutatud pilti. Paljud lapsed (20%), ka need, kes valesti vastasid, ei pööranud tähelepanu pildil olnud eristusele „toas või õues“, vaid keskendusid sellele, et linnud olid puu okstel (küsimus oli *Kus on linnud?* ja ootuspäraseks vastuseks oleks olnud *õues*). Tundub, et laste jaoks oli pildi pealt raske hoomata, et osa tegevusest toimus aknast paistvas õues või et oodati just nimelt seda eristust. Kuigi vorm „õues“ valiti küsimusse teadlikult, sest tegemist

on sagedase ja varakult omandatava sõna(vormi)ga, tuleb kaaluda mõnd sellist seesütleva käände küsimust ja muuta pilti nii, et konkreetne kohasuhe ehk seesolu oleks selgemini nähtav.

Kaasaütleva käände vormi kohta käiva küsimuse vastuse puhul tuleb tõdeda, et lapsed olid pildi suhtes väga tähelepanelikud, sest küsimusele *Kellega vanaema kohvi joob?* vastas mitu last (20%), et *vanaema joob kohvi üksi*. Pildil on küll laual ka teine kohvitass ema jaoks, aga ema ei istunud veel laua ääres, vaid sammus laua poole, tort käes. Seega võib pilti tõesti tõlgendada ka nii, et vanaema joob kohvi üksi.

Valimi väiksus seadis uurimistulemuste tõlgendamisele samuti piirangud. Kuna õpetajad pidid valima kaks eri stardipositsiooniga n-ö prototüüpsemat last nende laste hulgast, kes kas ei rääkinud projekti alguses veel üldse eesti keelt, või kes rääkisid vähesel määral, võivad uurimistulemused peegeldada ainult nõrgema keeleoskusega laste omi ega pruugi kirjeldada kõikide keeleõppijate taset katvalt. Samas võis ka olla, et õpetajad valisid tahtmatult mõlema rühma seest siiski tugevama keeleoskusega või üldse akadeemiliselt edasijõudnumad lapsed. Seetõttu on oluline edaspidi laiendada hinnatavate õpilaste hulka just nimelt parema keeleoskusega õpilaste arvelt. Suurem valim annab võimaluse tulemusi ka mitmekülgsemalt analüüsida. Lisaks annavad siinsed tulemused võimaluse vaadelda edaspidi täpselt samade laste eesti keele oskuse arengut ehk kui sinne uuring fikseerib nende laste oskused mingil hetkel, saab edasise võrdluse toetada juba kindlamatele andmetele.

5. Arutlus ja kokkuvõte

Uuringus vaadeldud 35 lapsest vanuses 7–9 eluaastat oli suur enamik neid, kes kõnelesid vene keelt (77%), 11% lastest olid kakskeelsed (eesti-vene, vene-armeenia) ning ülejäänud laste koduseks keeleks oli mõni muu keel peale eesti või vene keele (nt poola või ukraina). Vaatlusalused lapsed moodustasid kaks gruppi: n-ö parema stardipositsiooniga lapsed (17 last), kes projekti alguses eesti

keelt mõnevõrra rääkisid, ja nõrgema stardipositsiooniga lapsed (18 last), kes 2020. a sügisel eesti keelt veel ei rääkinud. Selline jaotus andis võimaluse vaadelda, kuidas omandavad aasta jooksul eri tasemel keeleoskusega õpilased eesti keelt tingimustes, kus õpetajad õppetegevust (vähemalt teadlikult) ei diferentseerunud.

Vaadeldes õpetajate hinnanguid õpilaste üldistele suhtlusoskustele, paistab, et laste oskus väljendeid-sõnu järele korrata on tasemel, kus nad teevad üksikuid vigu. Samuti oskavad lapsed kasutada tavapäraseid viisakusväljendeid (nad teevad vähesel määral vigu). Käskude-keeldude väljendamisega saadakse pisut paremini hakkama kui küsimuste moodustamisega, mõlema puhul peegeldab tulemus aga oskust, mida iseloomustab skaala näitaja „tehakse sageli vigu“. Lausete loomise ja kirjeldus- ja jutustamisoskus on tasemel, kus nad teevad veel väga palju vigu ning õpetajad hindavad laste keeleoskust igapäevasuhtluses toime tulla napiks. Ühegi oskuse puhul kahe grupi tulemuste vahel olulist erinevust ei olnud.

Seega, õppeaasta jooksul on märkimisväärselt arenenud nende laste keeleoskus, kes aasta alguses eesti keelt üldse ei rääkinud, sama ei saa aga öelda nende laste kohta, kes juba rääkisid eesti keelt. Nende keelelised oskused ei erinenud üheski vaadeldud kategoorias oluliselt nõrgema stardipositsiooniga laste omadest. Seda näitab ilmekalt ka tulemustes kahe rühma vahel statistiliselt olulise erinevuse puudumine paljudes uuringu lõikudes. Kommentaarina olgu lisatud, et ka mitteootuspärase vastuste hulk ei olnud parema keeleoskusega alustanud laste puhul suurem, seega ei saa öelda, et tegemist oleks metoodilise probleemiga. Samuti kinnitas selline tulemus varasemates uuringutes mainitud nii alla- kui ülessuunalise diferentseerimise olulisust ehk seda, et õpetajal tuleb tähelepanu pöörata ka neile, kes üldisest tasemest vähem või rohkem keelt oskavad. Lisaks kinnitavad tulemused, et õppe vähene diferentseeritus võib mõjutada eelkõige n-õ paremal stardipositsioonil olevaid õppijaid (vt Subban 2006; Tomlinson 2001b). Saadud tulemused ehk kahe grupi tulemuste sarnasus osutab selgelt sellele, et õpetajad peavad pidevalt teadvustama eri keeletasemega õppijate erinevaid vajadusi,

kujundama sellele vastavalt tunni tegevused nii, et ka tugevama stardipositsiooniga laste keeleoskus areneks.

Grammatiliste kategooriate omandamise hindamisel selgus peamise tulemusena, et vaadeldud lapsed mõistavad kõiki eelA1-taseme kirjelduses esitatud grammatilisi kategooriaid ja vorme, fraasitüüpe, konstruktsioone ning sagedasemaid ühendverbe ja liitsõnu. Kuigi mõistmisoskus ongi keelelise arengu puhul kõnelemisoskusest tavaliselt ees ning kasutamise-/loomiskatsete tulemused võivadki olla kasutusoskuse katsete tulemustest paremad (vt Argus 2008: 39; Argus, Parm 2010: 37), on siinsetes tulemustes hindamisvahendi kahe poole tulemuste vahel erinevus väga suur. Lapsed mõistavad kõiki eelA1-kategooria grammatilisi struktuure (st et õigete vastuste hulk on üle 60%), kasutada oskavad nad aga ainult üksikuid vorme. Nimisõna kategooriatest oskavad õpilased selgelt kasutada ainult ainsuse nimetava käände vormi, tegusõnavormidest kindla kõneviisi oleviku esimese ja kolmanda pöörde vorme. Fraasitüüpide, ühendtegusõnade ning liitsõnade kasutamise oskuse kohta hindamisvahendiga kinnitust ei saanud. Samas tuleb silmas pidada ka seda, et katsemeetodi puhul ongi üldjuhul keeruline tagada just ootuspäraste vormide kasutamist ja madalam tulemus võib tuleneda osalt ka hindamisvahendist. Eriti ilmekalt näitasid seda fraasitüüpide ja konstruktsioonide osa vastused, kus sisult õiged ja vormilt valed ehk ebataielikud vastused tõstsid õigete vastuste osakaalu näiteks 17%-lt 40%-le.

Senised uurimistulemused on kinnitanud, et sõnavara oskus korreleerub teiste keeleõppe aspektidega, näiteks grammatika oskusega (De Wilde *et al.* 2021). Selle põhjal saame ka siinsete tulemuste juures oletada, et näiteks kui laps mõne tegevuse tähistamiseks (nii verbivormide ülesandes kui ka ühendverbide küsimuses) sobivat verbi ei olnud omandanud, ei saanud ka eeldada, et ta oskaks kasutada küsitud vormi. Seega näitab kasutatud küsimustik mitte ainult grammatika, vaid ka sõnavara omandatust. Eriti ilmekalt tuli see esile just ühendverbi „katki minema“ puhul, kus näiteks üks laps vastas oodatava verbi asemel grammatiliste tunnusteta imitatiiviga

„buh-buh“. Just sõnavara puhul on keeleõppes oluline kontekstuaalne ehk selline omandamine, kus sõnu õpitakse pigem suhtlusstuatsioonis kui keeletunnis, ning see mängib eriti suurt rolli verbide omandamisel, mis vajavad situatiivset keskkonda (Puimège, Peters 2019). Seega on oluline, et lapsel oleks traditsioonilise keeletunni kõrval veel kohti, kus sihtkeelega kokku puutuda ja sõnavara omandada. Sõnavara aga toetab omakorda grammatika omandamist.

Analüüsides eesti keelega koolivälist kontakti omavate laste tulemusi ainult koolis eesti keele sisendit saavate laste tulemustega, ilmnes, et ainult koolis eesti keelega kokku puutuvate laste eesti keele oskus oli mõnevõrra nõrgem kui neil, kes saavad sisendit ka väljaspool kooli (näiteks sõpradega mängides, trennis, sugulaste kaudu). Kui üldise suhtlusoskuse ja kirjeldamisoskuse puhul oli vahe väike ning polnud ka statistiliselt oluline, siis grammatiliste oskuste erinevuste vahe kahe grupi vahel oli märkimisväärssem: väljaspool kooli eesti keele sisendit saavate laste eesti keele grammatilised oskused olid paremad kui ainult koolis sisendit saavate laste oskused. Kooliväline kontakt eesti keelega tähendab ühtlasi ka suuremal hulgal sisendit, mis varasemate uuringute põhjal on olnud seotud edukama keeleomandamisega (Jia, Fuse 2007; Unsworth *et al.* 2014; De Wilde *et al.* 2021). Samuti on leitud, et juhuslik keele omandamine (tunniväline) toimub kiiremini kui teadlik ehk õppetegevuse kaudu toimuv keeleõpe (Paradis 2004, 2009; Ellis 2005, 2009; Lichtman 2016). Seda koolivälise kontakti olulisust ja juhusliku keeleõppe edukust näitavad ka siinse uuringu tulemused: väljaspool kooli keelega kokku puutunud laste keskmised tulemused, nii suhtlemiskui ka kirjeldamisoskus, tegusõnade ja nimisõnade kasutusoskus, olid paremad kui eesti keelega ainult koolis kokku puutunud lastel. Seega, eesti keelt teise keelena omandavate laste keeleõppe seiskohast on oluline leida viise, kuidas tagada neile eestikeelne sisend ka väljaspool koolikeskkonda.

Piirangud ja edasised ettepanekud. Võttes kokku kasutatud hindamisvahendi sobivusega seotu, võib öelda, et väga raske on otsustada, kas laste madalad tulemused tulenesid ainult hindamis-

vahendi probleemidest või oli tegemist ka sellega, et lapsed ei olnud vastavat vormi või konstruktsiooni omandanud. Kui nimi- ja tegusõnavormide puhul võib oletada, et lapsed ei olnud ootuspäraste vastuste andmiseks sobivaid vorme veel omandanud (andsid vales vormis vastuseid, nt ootuspärase sisseütleva või alalütleva asemel nimetavas käändes nimisõna), siis fraasi- ja konstruktsioonitüüpide puhul võib olla tegemist eelkõige hindamisvahendi probleemiga, sest vastati pigem ühe osaga fraasist ja mitte terve fraasiga. Samuti vastati eitava kõne verbivormide juures ainult eituspartikliga. Võimaliku lahendusena tuleks kaaluda lausete jätkamise ülesannet nii, et hindaja ütleb eituspartikli või lausest osa juba lapsele ette ja palub tal siis lauset jätkata.

Keerukam on pakkuda lahendust probleemile, mis ilmnis ainuse alalütleva käände puhul. Seal vastati postpositsiooniga küsimusele, kus oodati käändevormi. Kuigi postpositsiooniga vastus oli sisult õige (nt on ühtviisi sisult õiged *laua peal* kui ka *laual*), ei andnud see infot vormi kasutusoskuse kohta. Huvitav on see, et teises funktsioonis ehk omajat väljendavas konstruktsioonis oli alalütleva vormi puhul oli õigeid vastuseid rohkem. Ilmselt mängib rolli see, et omajat ei saa postpositsiooni sisaldava konstruktsiooniga väljendada.

Ühe võimaliku lahendusena tuleks leida mõni selline nimisõna, mille puhul on käändevormi kasutus postpositsiooniga konstruktsiooniga võrreldes tavapärasem, silmas tuleb muidugi pidada seda, et sõna oleks lapse jaoks tuttav ehk sage ja olemas eelA1-taseme sõnaloendis. Teiste hindamisvahendi probleemide puhul (seesütlev ja kaasaütlev kääne) puhul on võimalik pildi muutmisega suunata lapsi vastavaid vorme kasutama. Vorme või kategooriaid, mida eelA1-taseme kirjeldustes ei ole, laste vastustes ei leidunud ning seetõttu tuleb hindamisvahendit täiustada meetoodiliselt, mitte uusi grammatilisi struktuure lisades. Hindamisvahendi kasutamine parema keeleoskusega lastega, kes praegusest uuringust välja jäid, peaks samuti andma selgust, milliste kategooriate puhul on tegemist lihtsalt sellega, et laps seda omandanud ei ole.

Seega võib hindamisvahendi sobivuse kohta öelda kokkuvõtlikult, et grammatiliste struktuuride mõistmisoskuse hindamiseks väljatöötatud vahend sobib, kasutusoskuse hindamiseks tuleb aga seda oluliselt täiustada. Kindlasti tuleb kasuks ka see, kui katsealuste hulk on suurem, nii ei mõjutaks mõne üksiku lapse juhuslik vastus hindamistulemusi juhuslikus suunas. Samuti võimaldaks õpetajate tunnitegevuste kaardistus ja andmed selle kohta, kui võrd õpetaja eri keeletasemega lastele erinevat tegevust pakub ehk õpet diferentseerib, vaadelda eri stardipositsiooniga laste edasijõudmise seotust talle tunnis pakutavate tegevustega.

Artiklis esitatud eesti keelt teise keelena omandavate laste keeleoskuse tulemused on Eesti kontekstis esimesed omalaadsed ja annavad alles esmase vaate laste keeleoskusele ja selle hindamise võimalustele. Edasine süsteemne laste keeleoskuse hindamine, täiustatud hindamisvahendi kasutamine praegusest suurema valimiga võimaldaks saada terviklikuma ülevaate grammatiliste struktuuride omandamise järjekorra ja nende struktuuride ja üldse eesti keele oskust mõjutavate tegurite kohta, pakkudes toetuspunkte eesti keelt teise keelena õpetavatele õpetajatele.

KIRJANDUS

- Argus, Reili 2008. Psühholingvistiline katse eesti keele objekti käändevahelduse omandamise uurimise meetodina. – Emakeele Seltsi aastaraamat 54, 22–43.
- Argus, Reili; Parm, Sirli 2010. Eesti keele ajakategooria omandamisest – ajavormid ja ajasõnad. – Eesti Rakenduslingvistika Ühingu aastaraamat 6, 25–41.
- Argus, Reili; Baird, Piret 2022. Eesti keelt teise keelena omandavate 7–8-aastaste laste kodusest keelekeskkonnast ja eesti keele oskusest. – Eesti Rakenduslingvistika Ühingu aastaraamat 18, 75–95.
- Blaz, Deborah 2006. Differentiated instruction: A guide for foreign language teachers. Larchmont, NY: Eye on Education.
- Bot, Kees de; Lowie, Wander; Verspoor, Marjolijn 2006. Second language acquisition: An advanced resource book. New York, NY: Routledge.

- Bot, Kees de; Lowie, Wander; Verspoor, Marjolijn 2007. A Dynamic Systems Theory approach to second language acquisition. – *Bilingualism: Language and Cognition* 10 (1), 7–21.
- Brevik, Lisbeth M.; Gunnulfsen, Ann Elisabeth; Renzulli, Joseph S. 2018. Student teachers' practice and experience with differentiated instruction for students with higher learning potential. – *Teaching and Teacher Education* 71, 34–45.
- Chamot, Anna Uhl 2012. Differentiated instruction for language and learning strategies: Classroom applications. – *Perspectives on individual characteristics and foreign language education*. New York, NY: De Gruyter Mouton, 115–130.
- Ellis, Nick C. 2002. Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. – *Studies in Second Language Acquisition* 24 (2), 143–188. <https://doi.org/10.1017/S0272263102002024>.
- Ellis, Rod 2005. Measuring implicit and explicit knowledge of a second language: a psychometric study. – *Studies in Second Language Acquisition* 27 (2), 141–172. <https://doi.org/10.1017/S0272263105050096>.
- Ellis, Rod 2009. Implicit and explicit knowledge in second language learning, testing and teaching. (= *Second Language Acquisition* 42.) Clevedon: Multilingual Matters.
- Gagné, Francoys 2005. From gifts to talents: The DMGT as a developmental model. – *Conceptions of giftedness*. Eds. Robert J. Sternberg, Janet E. Davidson. Cambridge: Cambridge University Press, 98–120.
- Gunnulfsen, Ann Elisabeth; Møller, Jorunn 2017. National testing: Gains or strains? School leaders' responses to policy demands. – *Leadership and Policy in Schools* 16 (3), 455–474.
- Hallap, Merit; Padrik, Marika; Raudik, Signe 2014. Käänevormide kasutamise oskus eakohase arenguga vene-eesti kakskeelsetel ning spetsiifilise kõnearengu puudega ükskeelsetel lastel. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 10, 73–90.
- Hallap, Merit; Padrik, Marika 2019. Simultaanse kakskeelsusega 3- ja 4-aastaste laste grammatilised oskused eesti keeles. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 15, 35–52.
- Hulstijn, Jan H. 2003. Incidental and Intentional Learning. – *The handbook of second language acquisition* 19, 349–381.

- Jia, Gisela; Fuse, Akiko 2007. Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. – *Journal of Speech, Language, and Hearing Research* 50, 1280–1299.
- Juurak, Raivo 2020. Kui vene laps õpib eesti koolis. – *Õpetajate Leht*, 23. oktoober. [https://opleht.ee/2020/10/kui-vene-laps-opib-eesti-koolis/\(06.10.2022\)](https://opleht.ee/2020/10/kui-vene-laps-opib-eesti-koolis/(06.10.2022)).
- Lauristin, Marju; Kaal, Esta; Kirss, Laura; Kriger, Tanja; Masso, Anu; Nurmela, Kristi; Seppel, Külliki; Tammaru, Tiit; Uus, Maiu; Vihalemm, Peeter; Vihalemm, Triin 2011. Estonian Integration Monitoring 2011 (Summary). <http://pdc.ceu.hu/archive/00006517/> (06.10.2022).
- Lichtman, Karen 2016. Age and learning environment: Are children implicit second language learners? – *Journal of Child Language*, 43 (3), 707–730. <https://doi.org/10.1017/S0305000915000598>.
- Metslang, Helena; Kibar, Triin; Kitsnik, Mare; Koržel, Jevgenia; Krall, Ingrid; Zabrodskaja, Anastassia 2013. Kakskeelne õpe vene õppekeeleka koolis. Uuringu lõpparuanne. Tallinn: Tallinna Ülikool. https://www.hm.ee/sites/default/files/uuring_kakskeelne_ope_vene_oppekeeleka_koolis.pdf (06.10.2022).
- Muñoz, Carmen 2006. Age and the rate of foreign language learning (= *Second Language Acquisition* 19.) Clevedon: Multilingual Matters.
- Paradis, Michel 2004. A neurolinguistic theory of bilingualism. (= *Studies in Bilingualism* 18.) Amsterdam, Philadelphia: John Benjamins Publishing.
- Paradis, Michel 2009. Declarative and procedural determinants of second languages. (= *Studies in Bilingualism* 40.) Amsterdam: John Benjamins Publishing.
- Puimège, Eva; Peters, Elke 2019. Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. – *Language Learning* 69 (4), 943–977.
- Raud, Nina; Džalalova, Anna; Raik, Katri 2011. Plurilingualism In Teacher Education For Multicultural Schools And Kindergartens In Estonia. – *Problems of Education in the 21st Century* 32, 113.
- Renzulli, Joseph S. 2005. Applying gifted education pedagogy to total talent development for all students. – *Theory into Practice* 44 (2), 80–89.

- Shintani, Natsuko 2015. The incidental grammar acquisition in focus on form and focus on forms instruction for young beginner learners. – *TESOL Quarterly* 49 (1), 115–140.
- Subban, Pearl 2006. Differentiated instruction: A research basis. – *International Education Journal* 7 (7), 935–947.
- Tomlinson, Carol Ann 2001a. What is differentiated instruction. – *Education Leadership* 58 (1).
- Tomlinson, Carol Ann 2001b. Standards and the art of teaching: Crafting high-quality classrooms. – *NASSP Bulletin* 85 (622), 38–47.
- Toomela, Aaro; Mädamürk, Kaja; Soodla, Piret; Härma, Eliis (toim.) 2020. Arvutipõhised hindamisvahendid lugemis- ja matemaatikapädevuse hindamiseks põhikooli I ja II kooliastmes. Juhendid testide läbiviimiseks ja tulemuste interpreteerimiseks. https://www.innove.ee/wp-content/uploads/2020/05/Juhend_o%CC%83petajatele_kujundatud.pdf (06.10.2022).
- Unsworth, Sharon; Argyri, Froso; Cornips, Leonie; Hulk, Aafke; Sorace, Antonella; Tsimpli, Ianthi 2014. On the role of age of onset and input in early child bilingualism in Greek and Dutch. – *Applied Psycholinguistics* 35 (4), 765–805.
- Watts-Taffe, Susan; Truscott, Diane M. 2000. Focus on research. Using what we know about language and literacy development for ESL students in the mainstream classroom. – *Language Arts* 77 (3), 258–265.
- Wilde, Vanessa De; Brysbaert, Marc; Eyckmans, June 2021. Formal versus informal L2 learning: how do individual differences and word-related variables influence French and English L2 vocabulary learning in Dutch-speaking children? – *Studies in Second Language Acquisition* 44 (1), 1–25.
- Yip, Virginia 2013. Simultaneous language acquisition. – *The psycholinguistics of bilingualism*. Malden: Wiley-Blackwell, 119–136.
- Zabrodskaia, Anastassia 2004. Eesti keelt õppiv Kohtla-Järve vene koolide muukeelne laps. – *Keel ja Kirjandus* 4, 841–853.
- Zeeland, Hilde van; Schmitt, Norbert 2013. Incidental vocabulary acquisition through L2 listening: A dimensions approach. – *System* 41 (3), 609–624.

SUMMARY

CHILDREN'S ESTONIAN AS A SECOND LANGUAGE DEVELOPMENT DURING A YEAR

Piret Baird, Reili Argus, Merilyn Meristo

Tallinn University

Estonia is currently in the process of transitioning the education system to Estonian as the sole medium of instruction. Several pilot projects with the aim to support Estonian language learning in day nurseries and upper secondary schools have been started. However, there is not much empirical data on how Estonian as a second language acquisition takes place: which categories are acquired in which order, what kind of role input and age of onset play in the acquisition process. There is no research that would cover learning Estonian as a second language during a certain time period in a child's development from the point of view of language categories and forms. Also, no methodology has been tested for studying and assessing the grammar skills of children.

In SLA studies (De Wilde *et al.* 2021; Jia, Fuse, 2007; Unsworth *et al.* 2014) it has been found that the more a student receives input (also cumulatively) the more successful he/she will be. De Wilde *et al.* (2021) have found that in addition to traditional language classes primary school children also need contextual contact with the language. Language learning is more successful if the child has other opportunities besides the language class to speak the target language. The amount and frequency of contact with the language are important in vocabulary learning (N. Ellis 2002; Puimège, Peters 2019). Hearing the L2 often also supports the acquisition of grammar forms (van Zeeland, Schmitt 2013). Contextual language learning is of great importance, especially in the acquisition of verbs because they need a situative environment (Puimège, Peters 2019).

Differentiating the instruction means choosing the activities based on the needs of the students (Blaz 2006; Chamot 2012). Earlier studies have shown that although the need to differentiate is acknowledged and it is

implemented, the more advanced students do not always get appropriate activities (Brevik *et al.* 2018; Gunnulfsen, Møller 2017).

The purpose of the current study is to see how the skills in Estonian differ between two groups of learners after one year of learning when students at different language levels were all taught the same without differentiating instructions. The second goal was to test the effectiveness of the assessment tool.

Each teacher participating in the programme “Professional Estonian language teacher in a multilingual classroom” chose two 7–9-year-old students who were at different levels: one student who spoke no Estonian at the beginning of the school year and one who spoke a little Estonian. The teachers evaluated the language development of 35 children using a questionnaire which was created specifically for that purpose.

The questionnaire was composed of three parts: metadata, questions on general communication skills and a part about grammatical categories. The metadata included questions about the home language environment and contact with Estonian. The general communication skills part consisted of questions about managing in everyday language situations, sentence and text level skills, as well as the ability to describe things. These were evaluated on a 5-point Likert scale. The Estonian Language Institute's grammatical category list for the pre-A1 level was utilized for evaluating the grammatical skills. In this part the teachers had to evaluate children's abilities to understand and use certain grammatical forms (verb and noun forms, phrase types and word formation). A picture was used to ask questions to check for understanding and for the ability to use grammatical categories. When analysing data, the children were divided into two groups: those who spoke no Estonian at the beginning of the school year (18 children) and those who spoke some Estonian at the beginning of the school year (17 children).

Most of the children in the study spoke Russian at home (27 out of 35). There were also 4 bilingual children (3 Estonian-Russian, 1 Russian-Armenian) and a Polish-speaking child, a Ukrainian-speaking child and an English-speaking child. 27% of the children started having contact with Estonian upon entering primary school at age 7. And except for the bilingual children, the rest had initial contact during the daycare years

(between ages two and six). For 11 children school was the only place where they had contact with the Estonian language.

When analyzing students' general communication skills, the data showed that they were able to repeat words and expressions with only a few mistakes. The children were also able to use polite expressions without making many mistakes. They were also able to express commands somewhat better than they were able to ask questions, but both groups made mistakes often. The sentence formation, description and storytelling abilities of both groups were at a level where many mistakes were made and the teachers evaluated their abilities to handle everyday communications situations as meager. The differences in all the skills between the groups were not statistically significant.

As we can see from the results, the skills of the children who did not speak Estonian at the beginning of the school year had improved, while the same cannot be said about the group who was more advanced at the beginning of the school year. Their language skills were not much better in any of the categories than the skills of the children with a weaker starting position. This result confirmed the importance of differentiating instruction, as mentioned in earlier studies, and supported the view that this influences mostly the students who are at a better starting position (see Subban 2006; Tomlinson 2001b).

In the acquisition of the grammatical categories studied, the results showed that the children understood all the grammatical categories and forms, phrasal types and most common phrasal verbs and compound words at the pre-A1 level. However, they were only able to use a few of the forms. From the noun categories, the students can only use the singular nominative form. In regard to the verb forms, children know how to use first person and third person present indicative forms. We could not determine the usage skills for phrasal types, phrasal verbs and compound words. In some cases, it was difficult to determine if the results were low in some categories because the children had not acquired a certain form or construction or because there were problems with the assessment tool.

The assessment tool was suitable for evaluating the understanding of grammatical categories, but in some cases problems were detected in evaluating the usage of certain grammatical categories. Often the children

gave answers which were appropriate responses, but did not contain the form that was being evaluated. For example, many answers contained only one part of the phrase being evaluated. Hence, while the assessment tool worked for evaluating understanding of grammatical structures, it needs to be improved for evaluating their usage.

Keywords: Estonian as a second language acquisition, an evaluation tool for language development, bilingualism, pre-A1 language level

Reili Argus on Tallinna ülikooli eesti keele professor, on tegelenud peamiselt eesti keele kui esimese keele omandamise uurimisega, aga vaadelnud ka eesti keele kui teise keele omandamist.
reili.argus@tlu.ee

Piret Baird on Tallinna Ülikoolis doktorant ja uurib peamiselt varajast kakskeelsust.
pbaird@tlu.ee

Merilyn Meristo on Tallinna Ülikooli prantsuse keele dotsent ja uurib nii võõrkeelte õpetamist kui ka keeleõppe motivatsiooni, samuti eesti keele kui teise keele omandamist.
merilyn.meristo@tlu.ee

ENABLING TOOL: ESTONIAN-ENGLISH CODE-MIXING OF A 2-YEAR-OLD WITH BALANCED INPUT

Piret Baird

Tallinn University

Abstract.¹ In recent years several studies have focused on bilingual children's code-mixing in light of usage-based theory (Gaskins *et al.* 2019a; Quick *et al.* 2020; Yow *et al.* 2018). However, most studies on bilingual children so far have focused on families that employ the one-parent-one-language or minority language at home strategies, in which cases children often receive significantly more input in one language. The current case study focused on a 2-year-old (2;4–2;10) Estonian-English bilingual whose language input was more balanced between her two languages. The results showed that the child's balanced input was reflected in the output proportions of her two languages and in her mean length of utterance scores. The child produced many code-mixed utterances, which also had the highest mean length of utterance score and were more complex than monolingual utterances.

Keywords: bilingualism, code-mixing, MLU, usage-based, balanced input, Estonian, English

1. Introduction

It is well known that bilingual children code-mix. Code-mixing in this current article is defined as “the mixing of elements of two languages together in one utterance” (Paradis *et al.* 2000: 245). For

¹ **The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:** This research was funded by the Estonian Ministry of Education and Research grant “Data and corpora of Estonian children and youth multilingual communication” [grant number EKKD33].

decades, researchers in different fields have studied code-mixing in adults and children. Most studies of code-mixing have adopted a formalist view (see Bernardini, Schlyter 2004; Cantone 2007; Gawlitzek-Maiwald, Tracy 1996; Genesee 1989; MacSwan 2000; Myers-Scotton 1997; Poplack 1980; Quay 1995), but a detailed discussion of them and their development over time is out of the scope of this article.

Recent studies have started to investigate code-mixing from a usage-based perspective. Researchers studying code-mixing in bilingual children have attempted to cover various language pairs, looked at switch placement in bilingual combinations, studied the relationship between code-switching and linguistic competency, and partially schematic constructions (Gaskins *et al.* 2019a; Quick *et al.* 2020; Quick *et al.* 2018c; Yow *et al.* 2018). Most of the participants in these studies have come from families where the one-parent-one-language (OPOL) strategy is used or where the minority language is spoken at home (ML@H). These strategies usually result in a situation where language input for the children is fairly unbalanced². If the family resides in a country where one of the parents' languages is spoken, and especially in a situation where the primary caregiver also speaks the societal language, it often results in little input from the non-societal language. This, in turn, can result in bilingual children having a dominant language or varying development speeds in different languages.

The current study aims to fill a gap in the field by studying code-mixing in a child whose language input is more balanced and not separated by person or place. This allows us to see the interplay of two languages in an acquisition situation that to the best of author's knowledge is not yet covered in the literature. The research questions of this article are the following: 1) Does the child distinguish

² For example, Gaskins *et al.* (2019b) report 75% vs 25% input between two languages for two of their participants. However, it should be acknowledged that it is a complicated matter as various factors influence language balance for a bilingual child.

her languages? 2) How does the more balanced input influence the child's output proportions? 3) How much code-mixing takes place, and are there differences between languages? 4) What are the mean length of utterance scores (MLU) for monolingual and code-mixed utterances, and do they reflect the language input pattern?

2. Usage-based theory and its approach to code-mixing

The last two decades have seen a rise in studies involving usage-based theory. The main claims of this theory are that language emerges from usage events, children use innate cognitive skills (like intention reading, pattern finding, generalisation, analogy) to acquire language and they learn their language piece by piece (Bybee 2010; Tomasello 2003). This is in contrast to the generative approach, which claims that language learners do not receive enough input to fully learn a language and there is an innate Universal Grammar in place for language learning (see Valian 2014 for discussion on this topic). Universal Grammar also argues that the categories and principles of the core syntax do not have to be learned as people are born with them (Behrens 2006).

However, according to usage-based theory, language use is item-based, meaning it is organised around concrete, particular phrases, like *Could you please...*, *How-ya-doin?* These expressions are stored and produced as single units (Tomasello 2000; 2003). Language learning stems from usage (input and output) and takes place on a continuum with various levels of schematicity. First, children acquire fixed chunks (sometimes also called frozen phrases in the literature), which can be either single words, for example *cat*, or multiword expressions, for example *What's this?*. These chunks are unanalysed wholes, which later on in the acquisition process will be segmented and children will acquire their language piece by piece. Second, from fixed chunks they move on to frames with an open slot (also called slot-and-frame patterns), for example, *What's X?*, where the X can be replaced with other words or expressions (Tomasello

2003). The existence of fixed chunks and frames with an open slot in children's speech has been shown in literature. For example, Lieven *et al.* (1997) studied 11 children (1;8–2;8) and found that 60% of their recorded spontaneous speech was composed of the child's first 25 lexically based patterns, like the above-mentioned example *What's X?* and 31% were fixed chunks. This shows that children's early language use contains an abundance of reusing a limited number of patterns with different slot fillers. Third, from those fixed chunks and slot and frame patterns children move on to more abstract constructions (for example *NP aux neg Verb*). Each move along the continuum allows children to be more productive with their language (Ambridge, Lieven 2011).

Children not only themselves produce a great number of chunks and slot and frame patterns in their speech, but also a significant proportion of their input contains these types of multi-word units. Cameron-Faulkner *et al.* (2003) conducted a detailed analysis of the speech of 12 English speaking children between the age of 2 and 3 years and their mothers. They showed that half of the utterances by the mothers were characterised by 52 item-based frames. They also found that 45% of the utterances the mothers said to their children began with one of just 17 words (and this excluded communicators like *hello*, *goodbye*, *thank you*, which if included would have increased the percentage even more). Other studies have also found that the frequency of frames in child directed speech (CDS) is connected to acquisition. Stoll *et al.* (2009) studied monolingual Russian, German and English two-year-old children and likewise found a considerable amount of lexical repetitiveness at the beginnings of utterances in CDS. The repetitiveness of CDS means that it is easier for children to detect patterns and to extract linguistic knowledge that facilitates their early language acquisition (Schmid 2017).

The repetitiveness of speech is connected to entrenchment. Entrenchment is a set of cognitive processes that takes place in the minds of speakers. These processes are mainly memory consolidation, chunking, and automatisisation. A wide range of variables

influence entrenchment, but frequency and repetition in context are most prominent (Schmid 2017). According to usage-based theory, constructions become entrenched when they are used repeatedly. Every use of a construction, whether in comprehension or production, strengthens it (Dabrowska 2014). Once a construction is entrenched, it is activated more quickly and the activation itself requires less effort. Hence, those constructions are more likely to be repeated and used. Schmid (2017) calls this a feedback loop in which frequency is both the cause and the effect of entrenchment.

It is not only the repetition that leads to the abstraction of information. The mind recognises similarities and differences, forms categories and generalises from them by comparing the information that is already stored with new units. This way schemas are formed (Behrens 2006). Langacker (1987: 492) defines a schema as a “semantic, phonological, or symbolic structure that, relative to another representation of the same entity, is characterised with lesser specificity and detail.” These generalisations allow the formation of patterns at different levels of abstraction, which were mentioned before as part of a continuum from fixed chunks to abstract constructions. Frame-and-slot patterns have an important role on the continuum of schematicity as the open slots are where the productivity develops and grows as the child inserts new words or phrases into the slots.

Usage-based theory is especially interesting in terms of bilingual children whose input includes more than one language. How do their two languages interplay as the child produces speech? One such phenomenon of language interplay is code-mixing. Code-mixing is prevalent in bilingual children’s speech. Different studies report varying rates of code-mixing: 4–9% (Poeste *et al.* 2019), 7–10% (Quick *et al.* 2018a) and 1–10% (Allen *et al.* 2002). Various researchers have also attempted to examine the reasons why young children code-mix. One suggestion has been that children code-mix because of an imbalance between their languages (Bernardini, Schlyter 2004). When children speak in their weaker language they use code-mixing to fill their lexical and syntactic gaps (Gawlitsek-Maiwald,

Tracy 1996). However, studies like Cantone and Müller (2005) and Jorschik *et al.* (2011) have shown that gap filling can be excluded as the only reason for code-mixing as children had translation equivalents available in most occasions. Hence, there must be other reasons why young children code-mix.

According to usage-based theory one possible explanation for code-mixing could lie in partially schematic units. As mentioned above, a study by Lieven (1997) found that 60% of young children's speech is composed of the first 25 lexically based patterns (partially schematic units), like *What's X?* or *There's an X*. It seems like these partially schematic utterances provide a way for a child to produce longer and more complex utterances, and thereby, to be more communicative in expressing themselves. At first children's constructions are lexically fixed, whereafter some slot and frame patterns start to develop. The slot and frame patterns allow children to be more productive, as they already have a pattern in use and they learn to insert some other piece of their existing language into that open slot. Lieven *et al.* (2009) found in their study of four English speaking children that with increasing language experience the material inserted into the slots also became more complex. Moreover, Quick *et al.* (2018b) studied code-mixing of a German-English-Spanish trilingual child (1;10–3;1) and looked at the degrees of lexical specificity of his utterances. They found that slot and frame patterns were very important in his code-mixing as those utterances often formed a slot and frame pattern where the slot was filled with material from the other language (a partially schematic construction *ich x it* 'I x it' existed where the slot was filled with either German or English elements like *ich zip it* 'I zip it' or *ich spielen it* 'I play it'). They related their findings to entrenchment and activation issues of multi-word units or patterns.

Some other recent studies have looked at language interplay and acquisition regarding code-mixing. Quick *et al.* (2018a) studied three German-English bilinguals ages 2;3–3;11 and found that MLU for each child followed their input patterns and language preferences.

However, code-mixed utterances were the ones with the highest MLU, and they were also syntactically more complex. Also, Gaskins *et al.* (2019b) and Quick *et al.* (2018b) report that MLU follows input patterns and language preference. Quick *et al.* (2020) involved in their research several language pairs to see if the same findings about the length of MLU hold true. Their study of German-English, English-Polish, Finnish-English and French-Russian bilingual children found that children's input patterns reflected their MLU scores and their language use. Also, the children's code-mixed utterances had a higher MLU score than their monolingual utterances and were also syntactically more complex. However, though covering different language pairs, most of these participants received significantly more input in one of their two languages (the exception being Lily in Quick 2018a), leaving the question of whether these findings hold true in a more balanced input situation.

But why have some studies found that compared to monolingual utterances code-mixed utterances are longer and more complex? Quick *et al.* (2018a) suggest that entrenchment plays a role in it. The more entrenched a particular unit is, the easier it is for the speaker to activate it. If a particular structure has low entrenchment in Language A, it can result in being uttered as a fragment in monolingual utterances. But if the child has higher entrenchment for an equivalent structure in Language B, then he/she can use that instead thereby forming a code-mixed utterance and being able to form a full sentence. This also makes code-mixed utterances have a higher MLU and be more complex. One can therefore say that bilingual children employ all of their language resources and code-mixing helps them to communicate better.

Above-mentioned studies by Quick *et al.* (2018a 2020) and Gaskins *et al.* (2019b) have shown that MLU tends to follow the input quantity and code-mixed utterances are the longest and more complex than monolingual utterances. Do the same findings hold true in a situation where the input is more balanced between the two languages involved? In this paper the language proportions

(monolingual Estonian, monolingual English, code-mixed utterances) of a simultaneous English-Estonian child along with her MLUs and utterance complexity are investigated to see if her language proportions and MLU reflect her input patterns. This study also aims to add to the small amount of literature on Estonian-English early bilingualism that is currently available.

3. Methodology

3.1. PARTICIPANT AND DATA

The participant of this study was a simultaneous English-Estonian bilingual child. The mother is Estonian-speaking and the father English-speaking, but both parents speak the other's language well. The family resides in Estonia, but the child has not attended day-care, and therefore, most of her input up to the end of recording sessions had come from her immediate family. The family uses a language policy where Estonian is spoken on Mondays, Wednesdays and Fridays by the entire family and English is spoken on Tuesdays, Thursdays, Saturdays and Sundays. This has been a consistent family language policy since the birth of the first child. Taking into account sleep, naptime, media usage and visits/phone calls from grandparents the child's input during an average week is fairly balanced between Estonian and English.

The recording sessions took place at home during play and snack times. Most of the recording sessions took place with the mother present, but sometimes also with the father or older siblings. The parents did not use code-mixing in their speech, but the older siblings did use it occasionally (5% of the speech of the 5-year-old brother and 4% of the speech of the 7-year-old sister were comprised of code-mixed utterances³). The recordings were done between the ages of 2;4 and 2;10. The recordings were usually done weekly, but

³ For the 5-year-old there was 10 h 45 min of data and for the 7-year-old there was 6 h 59 min of data.

sometimes due to time constraints there was a longer gap between sessions. On average, each month had 4–6 h of recordings (all together 35 h). There were more recordings done on days when the family spoke in Estonian, but each month had at least one English session. 6,853 utterances were included in the analysis. The data were recorded and transcribed by the author using the CLAN program and the CHAT format (MacWhinney 2018).

3.2. ANALYSIS

All utterances were coded as Estonian monolingual, English monolingual or code-mixed. Unintelligible utterances were left out of the analysis. Also, utterances where it was not possible to determine the language (*yep*, *mhmh*, *okei/okay*) were left out of the analysis. While *yep* instead of *jah* ('yes') or *yes/yeah* was fairly frequent in the recordings, other ambiguous utterances were quite rare. Language proportions were calculated.

MLUs, in words, were calculated separately for Estonian, English and bilingual utterances. Quick *et al.* (2018a) was followed for the rationale of calculating MLU separately for monolingual and code-mixed utterances. Also, to see changes over time, MLU was calculated for 3 periods: for data from October to December, January to February and March to April.

Utterances were also coded for syntactic completeness and complexity. This was done separately for monolingual and code-mixed utterances. Three different groups were created for coding: sentences, phrases and fragments (Table 1). An utterance was assigned into the sentence category for Estonian utterances when it included a subject and a verb or only a verb when the verb ending indicated the subject as well. For example, *istun emmega* ('I am sitting with mommy') does not include a subject, but the verb ending *-n* indicates that the subject is *I*. An utterance was coded as a sentence in English if it included a subject and a verb. In the case of a code-mixed utterance both languages were taken into consideration. For example,

an utterance *jumpis siin* ('He was jumping here') was coded as a sentence because though the verb is in English and would require a subject, the Estonian case ending indicates the subject clearly. Utterances were assigned to the phrase category when they formed a group of words which acted together as a grammatical unit, but did not form a sentence. For example, *mommy pliiatsid* ('mommy's pencils') belonged to the phrase category. Utterances which did not fit into the sentence and phrase category were assigned to the fragment category. Examples of fragments were *seal* ('there'), *no* ('ei') and *emme two* ('mommy two'). For the complexity analysis one recording session from each month was included.

Table 1. *Examples of complexity analysis*

Level	Examples
Sentence	<i>Emme doggy smellib me</i> 'Mommy, doggy is smelling me' <i>Emme where is mommy sitting?</i> 'Mommy, where is mommy sitting?' <i>I vaatan instructionit.</i> 'I am looking at the instruction' <i>Look emme!</i> 'Look mommy!' <i>Istun emmega.</i> 'I am sitting with mommy'
Phrase	<i>Mommy pliiatsid</i> 'Mommy's pencils' <i>Veel üks car</i> 'One more car' <i>Something to süüa</i> 'Something to eat' <i>Väga sunny</i> 'Very sunny'
Fragment	<i>Pole</i> 'not' <i>Emme two</i> 'Mommy two' <i>No</i> <i>This</i> <i>Bye</i>

4. Results

4.1. LANGUAGE PROPORTIONS

The data shows that the child adheres to the family language policy by speaking mostly in the language of the particular day. However, the data also reveals that the subject uses a lot of code-mixing. The percentage of code-mixed utterances (40–42% depending on the language day, see Table 2) is higher than is reported in most other studies for simultaneous bilingual children of similar age (Allen *et al.* 2002; Bernardini, Schlyter 2004; Quick *et al.* 2020).

Table 2. Language proportions by speaking day

	Estonian days	English days
Code-mixed utterances	42%	40%
Estonian utterances	44%	15%
English utterances	14%	45%

The data shows that, regardless of which language the family is speaking that day, the percentages of code-mixed utterances and monolingual utterances are strikingly similar. 44–45% of the utterances were monolingual in the language that the family spoke that day and 14–15% were monolingual in the other language. Based on the data, it is clear that the more balanced input between two languages is reflected in the results as the child produces about the same percentage of utterances in both languages.

4.2. MLU SCORES

First, mean MLUs (measured in words) across all data will be reported. Thereafter, the changes in MLU scores over time will be shown. The mean MLU across all data for monolingual English utterances was slightly higher (2.28) than for Estonian utterances (2.12), but the difference is rather small. The difference could also be

due to differences in the morphology of the given languages, as Estonian uses more case endings while English uses more pre- and post-positions. However, MLU for all code-mixed utterances is higher (3.88) than for monolingual utterances in either language.

Looking at MLU scores longitudinally reveals a similar picture (Figure 1). For the first few months of recording sessions, the Estonian MLU was a bit higher (2.23) than the MLU for monolingual English utterances (2.04), and the code-mixed utterances had the highest MLU (3.62). For the next two months, the MLU-s had increased, except for Estonian monolingual utterances, where it stayed about the same (2.17). For the last two months of the recording sessions, the MLU for English monolingual utterances was higher (2.91) than the MLU for monolingual Estonian utterances (2.36). However, code-mixed utterances had the highest MLU again (4.30).

Hence the data reveals that the MLU for code-mixed utterances was higher (approximately 1.5–2 words) than for monolingual utterances throughout the recording period.

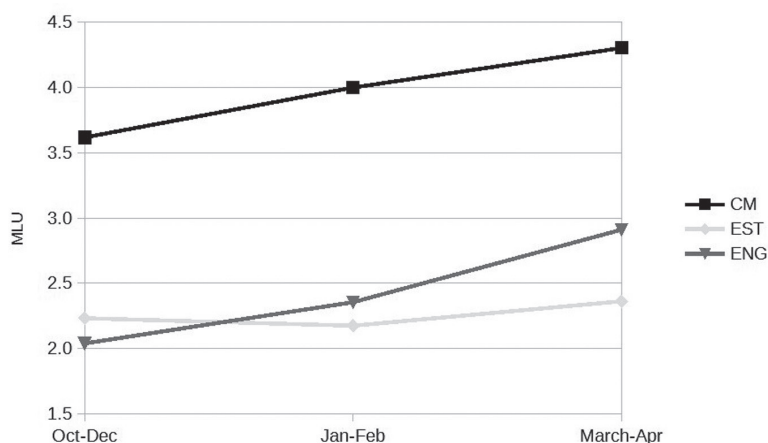


Figure 1. MLU scores developmentally

4.3. COMPLEXITY ANALYSIS

All utterances were categorised into sentences, phrases and fragments. The complexity analysis revealed that the code-mixed utterances were more complex than monolingual utterances (Figure 2). 78% of code-mixed utterances belonged to the sentence category, while 39% of Estonian and 35% of English monolingual utterances could be included in that category.

The input balance is also evident in the results of the complexity analysis. Monolingual utterances had proportionally similar percentages in each category. 39% of monolingual Estonian utterances were in the phrase category and 22% were fragments. The numbers for monolingual English utterances were respectively 38% and 27%.

To summarise the findings, it can be argued that the input balance is reflected in language proportions, in MLUs and in utterance complexity. The data also showed that the child uses a high proportion of code-mixed utterances, which have the highest MLU, and are more complex than monolingual utterances. In the next section, a usage-based explanation is proposed for the results.

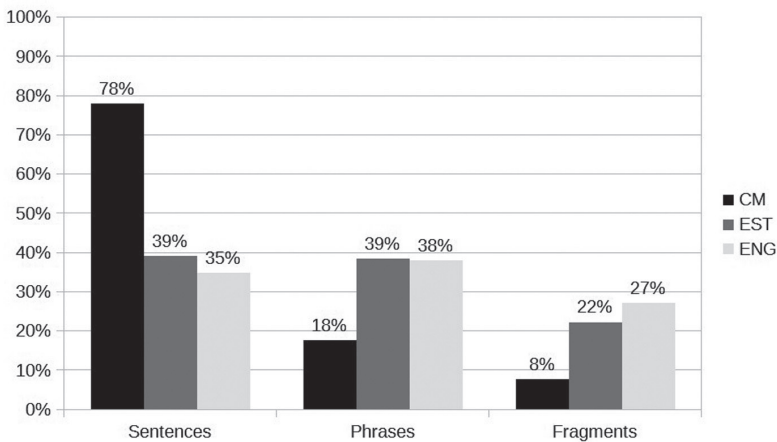


Figure 2. *Results of the complexity analysis*

5. Discussion

In this study the language development of an Estonian-English bilingual child (2;4–2;10) was analysed by measuring her language proportions along with code-mixed utterances, calculating her MLUs for monolingual and code-mixed utterances and conducting a construction type analysis. The data showed that the child distinguished her two languages and used a high proportion of code-mixed utterances when speaking in both of her respective languages. Code-mixed utterances also had a higher MLU and were more complex than monolingual utterances.

In line with previous studies (Gaskins *et al.* 2019b; Quick *et al.* 2018a 2020), the results of this study found that the MLU of a bilingual child reflected her input pattern. In above-mentioned previous studies the input of a bilingual child had been unbalanced. These studies involved children who had a dominant language (with the exception of one child, Lily, in Quick 2018a) and the analyses demonstrated that the more a child received input in one language the more he or she also produced output in that language, and this, in turn, according to the results, was also evident in the MLU score. The data from this study revealed that when the input from both languages is balanced then the output proportions and MLU scores are also more balanced. Though this is only a case study and more data with similar input pattern is required to further support these findings, these results further support the argument that MLU scores in the respective languages mirror the input pattern.

The MLU was the highest for code-mixed utterances, which were also more numerous in the speech of the participant than has been reported in other studies (Allen *et al.* 2002; Poeste *et al.* 2019; Quick *et al.* 2018a, 2020) with same-aged children. It could be suggested that this was the case because her input was not separated by place or speaker, as is the case with OPOL and ML@H strategies, which have been prevalent in most studies reporting on code-mixing. Dedicating specific weekdays to speak specific languages

creates a supporting environment for code-mixing; this is despite the fact that the input itself encouraged language separation as caregivers themselves did not use code-mixing in their speech. With both parents and siblings speaking both languages to the participant and the lack of other constant prevalent monolingual environmental factors⁴ (like daycare), an environment may have been created, which did not cue her to monolingual mode where she would have felt that only monolingual speech was accepted. The lack of monolingual cues and being understood when using code-mixed utterances could, therefore, be a factor contributing to the high rate of code-mixed utterances.

Another suggestion for the high rate of code-mixing is that it could be developmental at this stage of language acquisition and the supportive environment simply increases the use of it. Code-mixing being a developmental phenomenon has been suggested by Gaskins *et al.* (2019a) and Gaskins *et al.* (2021). Code-mixing is suggested to be more prevalent during this age of language development due to the lack of complete mastery of vocabulary and grammar. Genesee *et al.* (1995) point out that at times most (if not all) children code-mix at least until age three. Moreover, Yow *et al.* (2018), for example, point out that the code-mixing rate has developmental shifts and different stages have different reasons for engaging in it. According to Yow *et al.* (2018), younger children tend to code-mix to fill lexical gaps while older children do it for sociocultural or pragmatic purposes. As this data covers only a six-month long period, it is not possible to look at the developmental shifts in the rate of code-mixing, though the present data does lend support to previous research presenting

⁴ It should be noted that the participant was 1;9 when COVID-19 pandemic reached her region and subsequent lockdowns and recommendations for reducing social gatherings were in place before, during and after the recordings took place; hence, the child spent more time in her only bilingual environment as other social gatherings were often restricted or not recommended by the government. Subsequently the child spent majority of her time with her bilingual family and spent less time in only monolingual environments (e.g., grandparents) as would have been typical in normal circumstances.

the presence of code-mixing in the speech of 2-year-old bilingual children. It would be extremely beneficial to extend the time period of longitudinal studies of bilingual children to cover more than the year or two they usually are. This would allow the researchers to better determine developmental shifts in code-mixing while keeping some factors constant (as much as it is possible in the ever fluid factors surrounding bilingualism).

Several authors have suggested code-mixing to take place due to lexical gap filling. Cantone and Müller (2005) claim that filling a lexical gap and an uneven development of languages are sometimes the reasons for code-mixing, but that those are not the only reasons to code-mix for bilingual children. In their study of four 2-year-old bilinguals, Cantone and Müller (2005) found that translation equivalents were present for most code-mixes in the same recording; hence suggesting that though sometimes lexical gap filling is the reason to code-mix, it is not the sole reason behind it. In the current study, MLU and construction analysis indicated that the languages of the participant were evenly developed, indicating that the uneven development of languages was not the driving force behind code-mixing. Though it was not one of the research questions of this article, the high amount of code-mixed utterances raised the question about the existence of translation equivalents in the speech of the child. A review of the data indicated that the child had translation equivalents available in her speech, but she opted for code-mixing⁵. For example, in the transcript of a recording from the same day the following utterances can be found:

Mommy, mul ei ole bite in my mouth. 'Mommy, I do not have a bite in my mouth.'

Emme, mina võtsin see bite. 'Mommy, I took that bite.'

⁵ The analysis of translation equivalents was not overarching as it would have been out of the scope of this article. However, enough data were analysed to see the existence of translation equivalents.

So we can see that the word for *mommy* is used in both languages in close proximity to one another. Hence, also the findings of this study lend support to the claim that there are other reasons beyond lexical gap filling for using code-mixing.

Several studies (e.g., Quick *et al.* 2020; Yow *et al.* 2018) have suggested that code-mixing is a tool that bilingual children use to enhance their communicative competence. This argument of code-mixing being an **enabling tool** is supported by usage-based theory, which claims that children learn their language piece by piece. It has been shown that multi-word units (fixed chunks and slot-and-frame patterns) form an integral part of children's language acquisition, and this holds true for input and output (Quick *et al.* 2019). Such multi-word units could play an important role in a child's code-mixing. A study done by Quick *et al.* (2018c) with a German-English bilingual child found code-mixing to be very formulaic and that it contains many partially schematic utterances. Using lexically fixed items and partially schematic utterances in speech production requires less effort, as they are easier to store and easier to retrieve. If a given slot and frame pattern is entrenched in one language, it is more likely to be activated and used, whereafter the open slot can be filled with an item from the other language. This activation of units could contribute to code-mixed utterances having higher MLUs, as when children use both of their languages they are able to use a wider array of syntactic and lexical elements available to them due to using two languages. They can then pick the constructions that are more entrenched, easier to activate and do so regardless of the language, producing then code-mixed utterances. However, to evaluate entrenchment, a denser dataset would be needed than what was available for this study. A dense dataset involving children with different input patterns, including a more balanced bilingual child, would possibly shed more light on the role of entrenchment in code-mixing.

In the same regard, to understand better why some children code-mix more than others it would be necessary to study the

environmental factors surrounding the participating children (family language policies/strategies, bilingual/monolingual environments surrounding the child, variety of different recording environments, etc.). Future research should also tap into the developmental shifts in code-mixing rates and see if entrenchment is an important variable only in younger age groups of bilinguals and what kind of a role it has for older children.

6. Conclusion

In this paper the language proportions, MLU and utterance complexities of a 2-year-old (2;4–2;10) Estonian-English balanced bilingual child were examined. Unlike in other studies of bilingual children, the family of the participant separated the two languages by days of the week, rather than by speaker or place, as the common OPOL and ML@H strategies do. This, in turn, meant a more balanced input in both languages. The analysis showed that the balanced input was reflected in the language output proportions, as the child adhered to the family language policy by producing more speech in the respective language of the day. Interestingly, however, the participant produced a high proportion of code-mixed utterances (approximately 40%). The code-mixed utterances also had a higher MLU score and utterance complexity analysis revealed that the code-mixed utterances included more sentences, while monolingual utterances had more phrases and fragments. Code-mixing has an enabling effect on the child's speech allowing her to form longer and more complex utterances. Code-mixing is a tool enhancing communicative competence because it allows the child to use pieces from both languages to communicate a thought. Also, for the child in this study, the surrounding environment was supportive of code-mixing, as the input was not separated by speaker or place. According to usage-based theory, children learn language piece by piece, in chunks and slot and frame patterns. As some chunks and patterns become more entrenched, some neurological paths are activated

more easily. This could possibly be one explanation for code-mixing, though further studies need to be done to further evaluate this claim.

REFERENCES

- Allen, Shanley; Genesee, Fred; Fish, Sarah; Crago, Martha B. 2002. Patterns of code mixing in English-Inuktitut bilinguals. – *Proceedings of the 37th annual meeting of the Chicago Linguistic Society* 2, 171–188.
- Ambridge, Ben; Lieven, Elena 2011. *Child language acquisition: contrasting theoretical approaches*. Cambridge University Press.
- Behrens, Heike 2006. The input-output relationship in first language acquisition. – *Language and Cognitive Processes* 21, 2–24. <https://doi.org/10.1080/01690960400001721>.
- Bernardini, Petra; Schlyter, Suzanne 2004. Growing syntactic structure and code-mixing in the weaker language: The Ivy Hypothesis. – *Bilingualism: Language and Cognition* 7, 49–69. <https://doi.org/10.1017/S1366728904001270>.
- Bybee, Joan 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Cameron-Faulkner, Thea; Lieven, Elena; Tomasello, Michael 2003. A construction based analysis of child directed speech. – *Cognitive Science* 27, 843–873. https://doi.org/10.1207/s15516709cog2706_2.
- Cantone, Katja Francesca; Müller, Natascha 2005. Codeswitching at the interface of language-specific lexicons and the computational system. – *International Journal of Bilingualism* 9, 205–225. <https://doi.org/10.1177/13670069050090020501>.
- Cantone, Katja Francesca 2007. *Code-switching in bilingual children*. Dordrecht: Springer
- Dabrowska, Ewa 2014. Recycling utterances: A speaker's guide to sentence processing. – *Cognitive Linguistics* 25, 617–653. <https://doi.org/10.1515/cog-2014-0057>.
- Gaskins, Dorota; Backus, Ad; Quick, Antje Endesfelder 2019a. Slot-and-Frame Schemas in the Language of a Polish- and English-Speaking Child: The Impact of Usage Patterns on the Switch Placement. – *Languages* 4, 8. <https://doi.org/10.3390/languages4010008>.

- Gaskins, Dorota; Frick, Maria; Palola, Elina; Endesfelder Quick, Antje 2019b. Towards a usage-based model of early code-switching: Evidence from three language pairs. – *Applied Linguistics Review* 12, 2, 179–206. <https://doi.org/10.1515/applirev-2019-0030>.
- Gaskins, Dorota; Bailleul, Oksana; Werner, Anne; Endesfelder Quick, Antje 2021. A Crosslinguistic Study of Child Code-Switching within the Noun Phrase: A Usage-Based Perspective. – *Languages* 6, 29. <https://doi.org/10.3390/languages6010029>.
- Gawlitzek-Maiwald, Ira; Tracy, Rosemarie 1996. Bilingual bootstrapping. – *Linguistics* 34, 901–926. <https://doi.org/10.1515/ling.1996.34.5.901>.
- Genesee, Fred 1989. Early bilingual development: One language or two? – *Journal of Child Language* 16, 161–179. doi:10.1017/S0305000900013490
- Genesee, Fred; Nicoladis, Elena; Paradis, Johanne 1995. Language differentiation in early bilingual development. – *Journal of child language* 22, 611–631. <https://doi.org/10.1017/S0305000900009971>.
- Jorschick, Liane; Quick, Endesfelder Antje; Gläßer, Dana; Lieven, Elena; Tomasello, Michael 2011. German-English speaking children's mixed NPs with 'correct' gender. – *Bilingualism: Language and Cognition* 14, 2, 173–183. <https://doi.org/10.1017/S1366728910000131>.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Vol 1. Stanford University Press.
- Lieven, Elena; Pine, Julian M.; Baldwin, Gillian 1997. Lexically-based learning and early grammatical development. – *Journal of Child Language* 24, 187–219. <https://doi.org/10.1017/S0305000996002930>.
- Lieven, Elena; Salomo, Dorothé; Tomasello, Michael 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. – *Cognitive Linguistics* 20. <https://doi.org/10.1515/COGL.2009.022>.
- MacSwan, Jeff 2000. The architecture of the bilingual language faculty: evidence from intrasentential code switching. – *Bilingualism: language and cognition* 3, 1, 37–54.
- MacWhinney, Brian 2018. *CLAN Manual*. <https://doi.org/10.21415/T5G10R>.
- Myers-Scotton, Carol 1997. *Dueling Languages: Grammatical Structure in Code Switching*. Oxford: Clarendon Press.
- Paradis, Johanne; Nicoladis, Elena; Genesee, Fred 2000. Early emergence of structural constraints on code-mixing: evidence from French-English

- bilingual children. – *Bilingualism: Language and Cognition* 3, 245–261. <https://doi.org/10.1017/S1366728900000365>.
- Poeste, Meike; Müller, Natascha; Arnaus, Gil L. 2019. Code-mixing and language dominance: bilingual, trilingual and multilingual children compared. – *International Journal of Multilingualism* 16, 459–491. <https://doi.org/10.1080/14790718.2019.1569017>.
- Poplack, S. 1980. Sometimes I'll start a sentence in Spanish y termino en espanol': toward a typology of code-switching. *Linguistics* 18, 581–618. <https://doi.org/10.1515/ling.1980.18.7-8.581>.
- Quay, Suzanne 1995. The bilingual lexicon: Implications for studies of language choice – *Journal of Child Language* 22, 369–87.
- Quick, Antje Endesfelder; Lieven, Elena; Backus, Ad; Tomasello, Michael 2018a. Constructively combining languages: The use of code-mixing in German-English bilingual child language acquisition. – *Linguistic Approaches to Bilingualism* 8, 393–409. <https://doi.org/10.1075/lab.17008.qui>.
- Quick, Antje Endesfelder; Lieven, Elena; Carpenter, Malinda; Tomasello, Michael 2018b. Identifying partially schematic units in the code-mixing of an English and German speaking child. – *Linguistic Approaches to Bilingualism* 8, 477–501. <https://doi.org/10.1075/lab.15049.qui>.
- Quick, Antje Endesfelder; Backus, Ad; Lieven, Elena 2018c. Partially schematic constructions as engines of development: Evidence from German-English bilingual acquisition. – *Cognitive contact linguistics: placing usage, meaning and mind at the core of contact-induced variation and change*. Ed. Eline Zenner, Ad Backus, Esme Winter-Froemel. De Gruyter Mouton, 279–304. <https://doi.org/10.1515/9783110619430-010>.
- Quick, Antje Endesfelder; Hartmann, Stefan; Backus, Ad; Lieven, Elena 2019. Entrenchment and productivity: The role of input in the code-mixing of a German-English bilingual child. – *Applied Linguistics Review* 12, 225–247. <https://doi.org/10.1515/applirev-2019-0027>.
- Quick, Antje Endesfelder; Gaskins, Dorota; Bailleul, Oksana; Frick, Maria; Palola, Elina 2020. A gateway to complexity: A cross-linguistic comparison of child bilingual speech. – *International Journal of Bilingualism* 25, 3, 800–811. <https://doi.org/10.1177/1367006920956721>.

- Schmid, Hans-Jörg 2017. A framework for understanding entrenchment and its psychological foundations. – Entrenchment and the psychology of language learning. How we reorganize and adapt linguistic knowledge. Ed. Hans-Jörg Schmid, Boston, MA: De Gruyter, 9–39.
- Stoll, Sabine; Abbot-Smith, Kirsten; Lieven, Elena 2009. Lexically Restricted Utterances in Russian, German, and English Child-Directed Speech. – *Cognitive Science* 33, 75–103. <https://doi.org/10.1111/j.1551-6709.2008.01004.x>.
- Tomasello, Michael 2000. First steps toward a usage-based theory of language acquisition. – *Cognitive Linguistics* 11, 1/2, 61–82.
- Tomasello, Michael 2003. *Constructing a language: a usage-based theory of language acquisition*. Harvard University Press.
- Valian, Virginia 2014. Arguing about innateness. – *Journal of Child Language* 41, 78–92. <https://doi.org/10.1017/S0305000914000336>.
- Yow, W. Quin; Tan, Jessica SH; Flynn, Suzanne 2018. Code-switching as a marker of linguistic competence in bilingual children. – *Bilingualism: Language and Cognition* 21, 1075–1090.

RESÜMEE

VÕIMENDAV TÖÖRIIST: EESTI-INGLISE TASAKAALUS SISENDIGA KAKSKEELSE 2-AASTASE KOODIVAHETUS

Piret Baird

Tallinna Ülikool

Viimastel aastatel on järjest enam uuritud kakskeelsete laste koodivahetust kasutuspõhise teooria valguses (Gaskins jt. 2019; Quick jt. 2020; Yow jt. 2018). Kuid suurem osa uuringutest on siiani keskendunud perekondadele, kes kasutavad keelte eristamiseks üks vanem, üks keel või vähemuskeel kodus meetodeid. Nende puhul on tihti tegemist situatsioonidega, kus laps saab ühes keeles märkimisväärselt rohkem sisendit kui teises. Käesoleva artikli juhtumianalüüs keskendub 2-aastasele (2;4–2;10) eesti-inglise kakskeelsele, kes saab sisendit mõlemas keeles üsna võrdselt, sest peres on keelekasutus jagatud nädalapäevade järgi.

Uuringu tulemustest ilmnes, et lapse tasakaalus sisend väljendus ka tema kahe keele väljundi proportsioonides ja väljendi keskmise pikkuse (VKK) skooris. Eesti ja inglise keelsete lausungite VKK oli uuringuperioodi algul sarnane (2.23 ja 2.04) ja uuringuperioodi lõpus oli inglise keelsete lausungite VKK mõnevõrra kõrgem (2.91 ja 2.36). Lapse kõnes esines rohkelt (ligikaudu 40%) koodivahetusega lausungeid, mille VKK oli kõrgeim (alguses 3.62 ning lõpus 4.30) ja need olid ka komplekssemad kui ükskeelsed lausungid.

Võtmesõnad: kakskeelsus, koodivahetus, VKK, kasutuspõhine, eesti, inglise

Piret Baird's research interest is early bilingualism.
pbaird@tlu.ee

ACCENTS OF RUSSIAN IN PERFORMATIVE USE: ETHNIC STYLES, LANGUAGE ATTITUDES, AND IDENTITIES OF YOUNG NON-NATIVE SPEAKERS

Vlada Baranova

Das Nordost-Institut – IKGN

Kapitolina Fedorova

Tallinn University

Abstract. The paper deals with the performative use of different ethnic accents by young multilingual speakers of Russian. It aims at revealing the ways ethnic accents are used to challenge existing monolingual bias and reclaim agency opposing ethnic and linguistic prejudices. The study is based on an analysis of videos representing and discussing different non-native accents of Russian, created both by professional comedians and by amateurs and published on YouTube and TikTok. Metadata and comments on the videos expressing attitudes towards the performance of accents and linguistic stereotypes were also included in the analysis to reveal typical audience reactions.

Keywords: youth multilingualism, language ideology, ethnic identity, Russian, migration, comedy, vlogs

I. Introduction

As a result of mass migration to Russia from the former Soviet republics during the 1990s–2010s, linguistic diversity of the country has been on the rise, with languages other than Russian becoming more visible. Unlike in the Soviet Union in the 1960s–1980s and even in the early 1990s in Russia, when newcomers from, e.g. Uzbekistan or Georgia could speak Russian fluently after many years of studying

it in schools, nowadays, most migrants have no or very limited knowledge of Russian upon arrival. Internal migration to populous Russian cities from Russia's national republics with a high level of multilingualism is also significant; therefore, there is an increasing number of non-native speakers of Russian, especially in Moscow and St. Petersburg. However, despite being *de facto* multilingual, Russian cities tend to keep a 'monolingual façade' (Baranova, Fedorova 2019: 26): Russian totally dominates urban linguistic landscapes, and using languages other than Russian is usually not welcome in public spaces (Baranova, Fedorova 2019, 2020). Moreover, non-native accents and perceived imperfections in Russian are also treated with suspicion and provoke ethnic and linguistic prejudices.

In our previous research focused on Russian native speakers' linguistic prejudices and language attitudes (Baranova, Fedorova 2020), we discovered many cases of overtly expressed intolerance towards both other languages and accented Russian. There is also evidence that people often do not distinguish between different 'non-European' accents of Russian, merging, e.g. Armenian, Georgian, Tajik or even Chinese ones into an overgeneralised category of 'Eastern' (= 'Asian') accent (Panova 2012). Traditionally, in Russian ethnic jokes only a few accents, or ethnic styles of speaking (Ukrainians, Jews, Estonians, Georgians, and Chukchi) were represented (Shmeleva, Shmelev 2002, ch. 3). This oversimplified discriminatory discourse is still powerful among the Russian-speaking majority, and ethnic stereotypes and derogatory depictions of non-native speakers are widely represented in popular culture. However, with the development of the internet and social media, access to representation became easier for everyone, including non-native speakers of Russian, both newcomers and locals, i.e. the native citizens of national republics and the second generation of migrants. Nowadays, as we will try to show, they can become popular not only despite their accents but also due to them, by using them creatively in the web-space.

This paper aims to analyse approaches to linguistic diversity from the perspective of young people living in Russia and speaking with different accents. It emphasises the empowerment of speakers and their resistance to linguistic prejudices of the Russian-speaking majority. It is based on the study of data from professional videos (sketch and stand-up shows) and non-professional vlogs created by young non-native speakers of Russian who aim at performative use of different accents and ethnic styles of speaking. It should be noted that the field is heterogeneous: the Russian stand-up scene consists of actors who prefer to engage in TV-shows and the independent network of cooperation among co-performances (Gavrilov 2022). Both the speech of video bloggers and audience comments are analysed to reveal different types of discourses on multilingualism.

The paper is organised as follows. The next section discusses theoretical approaches towards non-native accents and their use as a part of creative practices and cultural industry; it also introduces the data and research methodology. Section 3 deals with stand-up comedy and comedic sketches employing ethnic jokes and non-native accents. Section 4 analyses videos created by non-native speakers of Russian as a form of their representation of everyday linguistic practices and distinctive features of different ethnic styles of speaking. The concluding section discusses the research results.

2. Theoretical background and research methodology

The ethnical way of speaking is often referred to as an ‘ethnolect’. Since Labov’s seminal works on AAVE (Labov 1972) and studies on Yiddish (Verschik 2007), or the speech of Turkish speakers in Germany (Wiese 2009), the term has been used to describe different linguistic features and more or less stable codes of non-native or bilingual speakers. More recent studies have emphasised the inadequate explanatory power of the term ‘ethnolect’, ethnicisation of language (Jürgen 2017) and an ‘outer’ view on the data, due to the fact that ‘ethnolects are perhaps as much produced by external

observers <...> as they are by speaking ones' (Lo 2020, p. 79). Here we prefer to avoid the term 'ethnolect' and focus on phonological transfer from various languages to Russian among non-native speakers. Kern (2011, p. 5) analyses new linguistic practices among adolescents from the second generation of migrants and calls them 'ethnic styles of speaking', emphasising the variability and linguistic resources that are available to the speakers. Benor (2010) also highlights the idea of multiple resources that make speech more or less 'ethnic' and can be used consciously to express one's identity. Thus, the issues of speakers' intentionality and identity are highly relevant for the studies in the field of ethnic styles.

As Rampton (1995) shows, groups may use new linguistic variants. Hybrid language forms or pronouncing words with a deliberate and exaggerated accent can be aimed at resisting ethnic prejudice, emphasising one's origin. This strategy can be seen in adolescents speaking English-based creoles. The sequence of adaptation includes the first generation of migrants who used Creole as the only resource available to them in their daily communication with native English speakers and the second generation who can speak English fluently and uses the pronunciation with an accent as a creative practice, as a marker of their identity, or for the sake of the elder generation. In another context, adolescents deliberately use the 'wrong' version of the article indicating the gender, reproducing the 'ethnic' version of Dutch (Cornips 2008, p. 119). Linguistic varieties associated with certain ethnic styles, therefore, can be used for claiming one's identity, for expressing solidarity with a group or for contrasting with some other groups (Blackledge, Creese 2015).

The representation of ethnic styles of speaking and accents in the media, for example, in TV comedy shows and stand-up performances, takes different forms; another important issue is the distinction between internal and external usage of some linguistic features to perform accents. By external usage we mean employment of those features by speakers normally not using them in their everyday speech and not belonging to the speech community associated

with them. For example, in Germany, Kanakcomedy ('Kanake' is a derogatory, racist term for Turkish immigrants) is a stereotyped parody of Kiezdeutsch, created by people without Turkish background; however, some artists of Turkish origin, like Fatih Akin, tend to reproduce a more natural way of speaking; the use of the variety in hip-hop turned it into a literary medium (Loentz 2006, p. 39). Linguistic stereotypes and derogatory terms, therefore, can be reclaimed and used creatively, which provides an opportunity for empowerment for linguistic minorities. A performance gives comedians an opportunity to go beyond conventions and existing social relations (Da Silva 2015: 206).

Young speakers identifying themselves ethnically and / or linguistically are becoming more visible in the comic scene than before. Digitalisation of communication and the opportunities provided by social media can further promote diversification of accents and styles among young speakers. Analysing Israeli humorous emails, Boxman-Shabtai and Shifman (2015) state that moving onto online ethnic humour transforms it significantly, especially in terms of locality. The growing influence and popularity of live (streaming) video from popular platforms such as YouTube and TikTok decrease barriers to inclusion in the comic industry. At the same time, interaction with the audience becomes easier as anyone can watch a video, post a comment and even interact with a performer. For researchers, this means the emergence of a new kind of data on youth multilingualism and language attitudes.

For the purpose of this study, we have chosen videos representing and discussing different non-native accents of Russian, created both by professional comedians (14 videos) and by amateurs (22 videos) and published on YouTube and TikTok. These videos were analysed to reveal particular linguistic features ascribed to different ethnic accents, ethnic and linguistic categories used by the speakers, and performative strategies and techniques employed by them. The second source of data was comments on the YouTube videos published by the viewers. Regarding videos that had prompted

a considerable number of comments, only the first 300 most recent and popular ones were collected, whereas all comments of a video were considered if they were fewer than 300. From these texts, we selected those expressing attitudes towards the performance of accents and linguistic stereotypes and used them to reveal typical audience reactions.

3. The performative use of non-native accents by professional comedians

The depiction of non-native accents for comic purposes was a rather marginal phenomenon in Soviet comedy. In popular films, there were only a few characters from the Caucasus and Central Asia speaking with certain phonetic and grammatical features. Among the most popular of them were Vasilii Alibabaeovich from some unspecified Central Asian republic (Alexandr Seryi's film *Gentlemen of fortune*) and Valiko from Georgia and Ruben from Armenia (Georgii Daneliia's film *Mimino*). Exaggerated accents, ungrammatical phrases and literal translations from other languages were aimed at creating a comic effect and exploiting the image of the Other (see, e.g. Andronikashvili 2015, p. 543). Some of those heavily accented lines became catch phrases known to most Soviet and even many post-Soviet Russian speakers. The famous Soviet comedian Arkadii Raikin also employed an unspecified 'Caucasian' or 'Central Asian' accent in some of his performances.

The same tendency to create stereotypical images of several ethnic groups by employing overgeneralized non-native accents was found in official comedian shows on Russian television throughout the 1990s and 2000s. Aimed at native Russian speakers, shows like *KBH (Клуб веселых и находчивых 'Club of the Funny and Inventive')* and *Comedy Club* would not normally try to represent accents realistically and make a distinction between different types of accents. The most obvious example is the *Hausa Russia 'Our Russia'* show (on air in 2006–2011), which depicted certain stereotypical

characters. Labour migrants from Tajikistan were represented by the comic figures ‘Rafshan’ and ‘Dzhamshud’, performed, respectively, by Mikhail Galustyan (Armenian) and Valerii Magdjash (Moldovan). When speaking ‘Tajik’, they pronounced gibberish words and syllables which are neither Tajik nor any other language, but which resemble the way Russian speakers usually imitate the speech of foreigners ‘from the East’, for example, using partial reduplication: *ахалай-махалай, иишгермек-бишгермек* (cf. typical ‘Turkic’ pattern of reduplication with changing the first consonant [Stolz 2008]). When speaking in Russian, the actors reproduce various stereotypes of ‘non-European’ accents which have very little in common with the speech produced by native Tajik speakers when speaking Russian. In particular, they add vowels after final consonants, thus turning closed syllables into open ones: *нащальника* instead of *начальник* ‘boss’. This feature is typical for native speakers of languages such as Chinese, where there is only a limited number of syllable final consonants, and which therefore is widespread in the speech of native speakers of Mandarin who have immigrated into Russia, have no formal tuition in Russian, and whose Sino-Russian idiolect features mainly Russian vocabulary (Frajzyngier *et al.* 2021); the same is true for a replacement of affricates by sibilants. Tajik speakers, on the contrary, have no reason to pronounce Russian words this way. Moreover, Rafshan’s and Dzhamshud’s speech lack the opposition between voiced and voiceless consonants: *каварил* instead of *говорил* ‘said’, which is, again, not typical for Tajik speakers. The show was very popular in Russia, and the names of Rafshan and Dzhamshud are still used by many people, even journalists, as a derogatory term to refer to labour migrants from Central Asia. Thus, for example, a news article informing about a proposed decree demanding migrants to sign a ‘loyalty agreement’ when entering Russia, is titled *Равшан и Джамишут пообещают больше не грабить и не насильничать?* ‘So Rafshan and Dzhamshud will promise not to rob and rape anymore?’ Unsurprisingly, the show was heavily criticised as racist and xenophobic by people from

Tajikistan and other countries as well as in the national republics within Russia. There were even official protests and attempts to stop the screening of the show and its consecutive film *Яйца судьбы* 'The balls of fate' by Tajik officials,¹ and rumours still circulate online that the actors had been threatened or even physically abused by some representatives of the Tajik diaspora.

Further development of stand-up comedy in Russia through various reality shows and competitions, alongside with the fast growth of social media, first of all YouTube, resulted in the emergence of new comedians who have begun to integrate their origins and accents into their performance personae. In most cases, they can speak Russian with just a slight accent, or even without it (as they sometimes are native speakers of Russian or early bilinguals), but they 'turn on' an accent as a performative tool. Rasul Chabdarov, a Balkar stand-up comedian from Nalchik, turns his accent into a weapon to attack xenophobic Russian speakers. In the following example, he describes a conversation with a taxi driver who complains about non-Russian drivers and their poor driving skills: *Они у себя в аулах на баранах ездят!* 'They ride sheep in their auls [a word of Turkic origin referring to villages in the Caucasus]!' The character replies with a strong accent and in an extremely high-pitched voice, typical for a stereotypical 'Caucasian': *На баранах?!* 'Sheep?' and then addresses the audience in a less accented tone: *Всю жизнь же на ишаках было, блядь!* 'For their entire life, they were donkeys, for fuck's sake!'. Indeed, popular xenophobic phrases about people from the Caucasus and Central Asia usually mention donkeys as their means of transportation. By addressing the issues of racism and xenophobia, Chabdarov 'turns the table' and transforms himself and his fellow 'non-Russians' from an object of mockery into a mocking subject, thus reclaiming his agentivity. The exaggerated accent he suddenly employs signals this turning point as if he were saying 'at least get your racist stereotypes right'. In his numerous jokes about the way

¹ <https://www.interfax.ru/russia/128901> (May 31, 2022).

Russians see other ethnicities, he reclaims his accent and style of speaking the same way he reclaims the image of the ‘savage from the aul’ by telling risqué jokes about sex with sheep. When talking about Russian speakers depicting a ‘Caucasian accent’, he states:

Вы не умеете, вы все время кого-то одного и того же грузина пародируете, че за расизм? Мы же блядь не азиаты!

You can’t do it, all the time you make parodies of the same one Georgian guy, that’s racism, ye? We are not fucking Asians!

The final racist slur can be either an unintentional expression of the speaker’s feeling of superiority towards Asian people, or something more subtle – a conscious attempt to keep the mask and further problematize racism. This second interpretation is supported by the context. First, Chabdarov, apart from performing under his own name, published a video of himself performing as a native Russian speaker named Alexandr Lurje.² In this video, staged at a club, the depicted audience is part of the show. The performer spent almost 30 minutes telling banalities, racist and homophobic jokes and complaining about the downfall of the Russian people to the applause of his listeners; after that he was shot to death by a ‘non-Russian’ and taken from the scene. This symbolic killing of the speaker during his performance reframes it and turns the mocking subject into an object of mockery. Second, when attacking the racism of Russians while performing as his ‘ethnic’ personality, he opposes to them all ‘non-Russians’ or ‘Easterners’, including not only people from the Caucasus but also, e.g. Tajiks. Moreover, he put in the same group both Russian citizens (Balkars, Chechens, Dagestanis) and foreigners (Georgians and Armenians), e.g. when he talks about what Caucasians would do in Hollywood—the Armenian would be Tony Stark (‘Iron Man’) and the Dagestani would be Spider-Man). He repeatedly opposes ‘us’ and ‘you’ (i.e. native Russian speakers):

² <https://www.youtube.com/watch?v=uh2q890GCrQ> (May 31, 2022).

Вы на нас срываете все негативные эмоции

‘You vent all your negative emotions on us’

Мы начинаем перед вами понтоваться нашими обычаями

‘We begin to show off our customs in front of you’

Вам нравится экзотика, а мы вам даем эту экзотику, еще и утрируем

‘You like exotic, and we give you this exotic, and we also exaggerate.’

On the other hand, Chabdarov demonstrates different accents and ‘teaches’ his, mostly native Russian-speaking, audience to depict different accents: Armenian, Chechen, Dagestani, etc.³ Thus, he simultaneously deconstructs the stereotype of ‘Easterners’, a homogeneous group of non-Europeans speaking Russian with one and the same ‘funny accent’ and contributes, at the same time, to it by stressing the opposition of ‘us’ (non-Russians from the Caucasus and Central Asia) and ‘you’ (native Russian speakers) thus uniting with other non-native speakers in this common identity imposed on them by Russians.

Similar dubious statuses and controversial strategies of ethnic representation are typical for other comedians working in this ‘ethnic field’. In 2014, in his interview for the Kyrgyz news-portal Limon.KG,⁴ Akim Karasaev who had previously appeared on the Russian TV show ‘Comedy Battle’ as ‘an oriental man in Kazakh khalat’ stated:

Я не говорил, что буду показывать именно Кыргызстан, я был представителем из Бишкека, из Кыргызстана, который показывает собирательный образ восточного человека. Моей целью было вообще познакомить россиян с востоком, чтобы не думали, что мы гастарбайтеры. У многих из нас

³ <https://www.youtube.com/watch?v=mAuBR75Q9Nk> (May 31, 2022).

⁴ <https://limon.kg/ru/news:63136> (May 31, 2022).

есть акцент, который веселит русских, так почему было не сыграть на нем? <...>. Акцент, возможно, смахивает на узбекский, может, на таджикский. <...> Нас очень часто путают с бурятами, калмыками, корейцами, таджиками. Разные национальности, но для россиян мы все одинаковые. Я часто рассказываю, чем они отличаются между собой, строю на этом свои шутки.

'I never said it was Kyrgyzstan that I aimed to show, I was a representative of Bishkek, from Kyrgyzstan who showed a collective image of an oriental person. My goal was to acquaint Russians with the East in general, so that they would not think of us as gastarbeiters [guest workers]. Many of us have an accent which makes Russians laugh, why couldn't I play it? ... The accent resembled probably Uzbek, or maybe Tajik. ... We are very often confused with the Buryats, Kalmyks, Koreans, Tajiks. Different nationalities, but for Russians we are all the same. I often tell how they differ from each other, I build my jokes on this'.

Comments on YouTube videos of ethnic stand-up reveal the same controversy. Some of the people who can identify with the depicted 'Easterners' accuse comedians of maintaining negative and/or false stereotypes. Thus, in his video,⁵ Evgenii Chebatkov, a native Russian speaker but a citizen of Kazakhstan, depicts 'his Kazakh friend Dosik in Canada', and some commentators have expressed offence at his 'exaggerated accent' and unflattering depiction of the Kazakhs:

С акцентом переигрывает, я такой не встречал
'He overacts the accent, I have never seen such a thing'

Бесит он как будто все казахи так разговаривают какими то калхозниками выставляет казахов
'He annoys me, as if all Kazakhs speak like that, he portrays the Kazakhs as some collective farmers'.

⁵ <https://www.youtube.com/watch?v=zvOWYEjMHDo> (May 31, 2022).

However, other commentators state that Chebatkov, first, portrays the Kazakhs very accurately and, second, has a right to do so as he is from Kazakhstan himself:

Он в точности копирует қазақша акцент

‘He copies Kazakh [the word is written in the Kazakh language] accent precisely’

Потому что Чебатков казах, про своих он может шутить

‘Because Chebatkov is Kazakh, he can make jokes about his people’.

Most interestingly, there are numerous comments from the Kazakhs living abroad praising the comedian for his true representation of ‘the sweet Kazakh speech’ evoking nostalgic memories:

Так передать казахский акцент, как будто на родину вернулся

‘He conveys the Kazakh accent so well, as if I’ve returned to my homeland’

Спасибо. Чуть взгрустнул за границей, давно не слышал родной казахской речи лол

‘Thank you. Became a bit sad abroad, haven’t heard my native Kazakh speech for a long time lol’.

It is evident that there is a request for representation among non-native Russian speakers residing in Russia and consuming its cultural products. A younger generation savvy both in consumption and production of video-content starts filling this gap in their non-professional video performances.

4. Grassroots level of performative use of non-native accents

This section deals with the internal representation of ethnic accents among the youth, mostly with a migration background. Young people of non-Russian ethnic origin spread short YouTube and TikTok

videos imitating the accents of different ethnic groups. Some videos become viral, and even many years after being uploaded they are still watched and generate thousands of comments. In terms of their online existence, therefore, there is no evident line of demarcation between popular stand-up comedy and those non-professional videos. Moreover, their authors, sometimes, after achieving success of their vlogs, start a professional career, as happened, for example, to Anisa Murtaeva, the author of the popular video *Типа диалекты, ну* (2012).⁶ At the same time, the video clips created for followers have some common features that will be discussed here.

In comparison with the ‘official’ comedy, where the imitation of accents is usually integrated in a more general comic framework focused on social and ethnic stereotypes, the sketches created by non-professionals pay more attention to the peculiarities of speech although stereotypes also play an important role. Instead of distorting words and exaggerating ‘wrong sounds’, which was typical for the *Hawua Russia* actors, the main strategies for imitating accents in videos created by non-native Russian speakers are pitch and intonation, as well as speech tempo, word order and specific discourse markers. More or less, even if exaggerated, they reflect real linguistic features of Russian as spoken by different non-native speakers. Moreover, different speakers use a similar way of describing a particular style of speaking. For example, they pronounce a bilabial [w] instead of Russian labiodental [v] and [y] instead of unstressed [a] in ‘Chechen Russian’. This suggests that this imitation of speech is based on an actual communicative experience and listening to different pronunciations and distinguishes this internal approach to accents from an external approach (merging all the accents together) typical for Russian speakers and/or aimed at a Russian, or even a foreign audience. For example, in the video *The most remarkable accents of the Russian language*⁷ created in 2021 by the vlogger *The Alex [Russian Teacher]*

⁶ <https://www.youtube.com/watch?v=sU0MZVrFQSM> (May 31, 2022).

⁷ <https://www.youtube.com/watch?v=5ji1Tnu2oO0> (May 31, 2022).

the speaker presents and explains in English, Ukrainian, Georgian (illustrated by Arkadii Rajkin's video from the Soviet times), Armenian, Chechen (illustrated by the speech of the notorious Chechnya leader Ramzan Kadyrov) and also 'northern' and 'southern' accents of Russian mixing up actual linguistic features with stereotypes, such as the use of a discourse marker *дон* typical for Kadyrov and, consequently, ascribed to all Chechen speakers. When asked about Central Asian accents in the comments, he replies:

It's pretty hard to think of many certain pronunciation features, I think they also change the stress or add one more stress in words, also they sometimes "omit" the "p" sounds (like "касавица" instead of "красавица") and in an extreme case they change "ч" to "ц" ("нащальник" instead of "начальник"). Usually their Russian is very poor and super-hard to understand, unlike that of most Caucasians :)

Evidently, in his comment he describes less a real Tajik or Uzbek accent, but more a *Hausa Russia* version of it.

'Internal' portraying of accents is, therefore, closer to reality and less affected by language stereotypes than the external one. However, ethnic stereotypes are employed widely and are quite often very similar to those represented in official TV shows. Caucasians can be depicted as criminals and Central Asians as socially inferior and not very bright people. Mentioning a particular ethnic label serves as a reference to the stereotype, as in the extract below from the video *Памирец говорит на русском с разными акцентами*.⁸

– Давай азербайджанский начинай тогда
 – Брааат! Первый мафия, первый мафия это азербайджанцы
 'Come on, start with Azeri then'
 'Brother! The first [=most important] mafia, the first mafia are the Azeris'

⁸ <https://www.youtube.com/watch?v=SNmUUrCnwH8&t=7s> (May 31, 2022).

It should also be noted that the set of portrayed vernaculars varies among speakers. Two points can be emphasised here. First, the selection of a represented variety reflects the vloggers' backgrounds—their area of living, birthplace and other biographical details. There is a common set of accents typically represented in non-professional videos; it includes both minority languages of the Caucasus and foreign migrants' vernaculars. However, a particular speaker can choose only some specific varieties. For example, in the video *Акцент разных наций Кавказа* (@luna69166, Tik-Tok, 2020) the speaker shows 'Dagestani', 'Ossetian', 'Ingush or Chechen, whatever', 'Kabardian' and 'Balkar'. This list, therefore, follows the administrative structure of the North Caucasus, but it marks some varieties as more or less familiar: the speaker is not sure about the differences between Ingush and Chechen but distinguishes Kabardian and Balkar very explicitly. Meanwhile, the speaker in the video quoted above (*Памирец говорит на русском с разными акцентами*) portrays Central Asians including some regional accents, e.g. Pamir as different from Tajik accent (he was himself born in Pamir) but provides a less detailed picture of Caucasian varieties ('Azeri', 'Chechen', 'Dargwa', and 'all other Dagestani accents'). Many speakers mix regional and linguistic groups, in particular, in the case of Dagestan. The Republic of Dagestan is highly multilingual; at the same time, the local Russian has many common areal features typical for speakers with different native languages (Daniel *et al.* 2010). The videos, therefore, highlight groups that are important for the speaker; they represent different categorizations and provide more or less detailed descriptions of each group. The name for a specific accent shows a process of 'making language' (Krämer *et al.* 2022) or 'language labelling' (Pennycook, Otsuji 2015).

The second point concerns the interaction between speakers from different minority groups. An important feature of such videos is their dialogical nature. In opposition to an imitation of dialogue in stand-up comedy (where contributions by the audience can be very limited and designed and manipulated by the comedian),

home-made stand-up videos are usually staged as a natural interaction between two or more speakers, one of whom plays a role of an imitator of accents. The audience (which can be both language- and gender-mixed company) asks this person to show a specific accent or ask questions (as in the examples above).

In other videos, the focus is not on just one speaker but on a representation of naturally occurring dialogues, as, for example, in a video recorded in a school class during a break where a student tries to imitate his friend's accent and employs other ethnic styles of speaking in the process of language play.⁹ Such videos represent a process of negotiation on language and ethnic identity; solidarity is expressed through demonstrating some knowledge about languages or accents of each other. Metalinguistic teasing ('as you [name of a particular group] speak') becomes a popular way of negotiating identity among Russian youth with different heritage languages and so should be treated as an important social practice. In the video *кавказские диалекты...смех до слез*, a group of young people are at a restaurant, and the speaker, provoked by his friends to start depicting different Caucasian accents, teases other speakers:

*Ты кто? Вика? Вика, я представляю себе голубоглазая блондинка, а не какая-нибудь носатая армянка*¹⁰

'Who are you? Vika? Vika, I imagine a blonde with blue eyes [the camera shows the girl] but not an Armenian girl with a long nose'.

Mocking other participants, the speaker also uses derogatory terms and characteristics describing himself. He replies as if quoting racist talk of Russian speakers:

А! Вы про этих что ли? Обезьяны, которые разговаривать умеют?

'Oh, so you speak about them? Those monkeys who can speak?'

⁹ The video was accessed March 15, 2019; later it was deleted from the platform and is not accessible.

¹⁰ <https://www.youtube.com/watch?v=3VT8pKSCBaw> (May 31, 2022).

There are other different derogatory terms for unification, i.e. *мигранты*, *нерусские*, *черные* ‘migrants’, ‘non-Russians’, ‘black’ that appear in the videos. In making the video, the young people in it use such terms to empower themselves thanks to their sense of irony and non-acceptance of racism. In the comments to this video, people fight about those words: some take them at face value and accuse the vlogger of racism, while others defend them as ironic and antiracist. The main argument for them, again, is the speaker’s ethnicity; he is described as Georgian or Caucasian and so ‘has a right’ to mock other Caucasians, and his words regarding them should be treated as sarcasm:

Это же сарказм, люди! Он же сам кавказец, по нему же видно. Это просто “ирония”

‘It is sarcasm, people! He is a Caucasian himself, you can tell by looking at him. This is simply “irony”’

So, ethnic stereotypes are used as a reverse tool to attack the racism of native Russian speakers, the same manner as in Rasul Chabdarov’s stand-up performances analysed above. At the same time, the reclaiming of derogatory terms is not fully understandable without the context.

Numerous approving comments to the non-professional videos appraise their linguistic correctness and stress the fact that commentators can relate to them as representing their own accents and experiences:

Я сам осетин. :) Осетина чётко сделал, и Чеченца. Bravo! :)

‘I am Ossetian myself. :) You did Ossetian just fine, and Chechen as well. Bravo!’

At the same time, if the speakers in the video are perceived by their non-native Russian audience as Russians and, therefore, outsiders, they can be heavily criticised for their poor performance of different accents (regardless of their actual correctness), as, e.g. in the case of the video *АКЦЕНТЫ, что это и как их пародировать?*¹¹

¹¹ https://www.youtube.com/watch?v=VhE1_5_PrFw (May 31, 2022).

Хрень! Вообще не похожие акценты

‘Bullshit! The accents are not similar at all’

Я ожидал большего. В итоге стандартный русский который пытается делать акценты

‘I expected more. As a result, it’s a typical Russian who tries to do accents’

In such cases, there is a tendency for commentators with ethnically marked nicknames (e.g. *Магомед Магомедов* or *Любовь Хартунян*) to be more critical than others.

On the other hand, comments to all accent-related videos, regardless of their authorship, look particularly interesting in terms of multilingualism: usually, there are some phrases in languages other than Russian, e.g. in Kazakh or Uzbek, and some commentators try to imitate different accents in a written form:

Ээ малаладес настроения паднял (instead of *Молодец, настроение поднял*)

‘Good boy, you improved my mood’

Ты дон тэк сэ мнэй нэ рэзгуэрыай дон (instead of *Ты так со мной не разговаривай*)

‘You shouldn’t speak to me like that’.

In this sense, the videos provoke discussions on language-related questions, stimulate linguistic creativity and make the Russian web-space more multilingual and diverse.

5. Conclusions

When comparing our two sets of data, there is an evident distinction between mainstream TV stand-up comedy represented by the shows like *KBH* or *Наши Russia*, and the imitations of accents by minority speakers. In the first case, stereotypes of broken ‘non-Russian’ speech are employed, and the artists do not claim reliable representation or authenticity. Overall, they do not contribute to the awareness of the non-native accents produced by migrants or linguistic minorities of

the Russian Federation. On the other hand, non-professional videos pay more attention to actual linguistic features found in the speech of non-native Russian speakers. They represent both the linguistic experience of a performer and the communication of people with different native languages using Russian as a *lingua franca*. At the same time, they also employ stereotypes, exaggerate some features and add comic content to amuse their audience. The same is true for official stand-up artists of ethnic origin who can turn exaggerated accents into a tool to confront the viewers' prejudices while still making them laugh in the process.

The reaction of the audience, however, is less focused on the differences between more and less realistic ways of accents representations by professional stand-up artists and by minority speakers themselves. Instead, the comments evaluate the 'authenticity' of speakers based on information about their ethnic origin. The insider position of artists presupposes that they share some common knowledge with their audience and thus have the moral right to mock the accents. According to the majority of the audience, their performances are appraised as realistic representations of the given accents. However, videos created by ethnically Russian comedians or vloggers, even when employing the same linguistic features, can be perceived as a parody and racism. Interestingly, some commentators have remarked (sometimes with displeasure and even dismay) that there has been an increasing number of comments expressing offence in recent years. Indeed, old comments to the videos uploaded around ten years ago tend to be less critical than the recent ones. This may attest to the growing awareness of minority language speakers of their right to challenge monolingual ideology and ethnic prejudices of the Russian-speaking majority.

In terms of linguistic performance, unofficial videos create a colourful picture of communication between minorities. Mixed ethnic communities with a migration background begin reflecting on linguistic strategies and share their stereotypes about language with an accent. It represents the process of categorization of a

minority group by other people with minority background. The discussion of different accents and making a choice of which accent is worth being represented add them to the map of known idioms and groups. The context of the informal videos reflects the communication of migrant youth from different ethnic minorities. They mock and tease each other but identify themselves as members of the same group opposed to 'Russians'; there are even some highly derogatory terms that minority youth reclaim for self-identification. The new solidarity of different linguistic minorities is a sign of empowerment and agency aiming to overcome the racism of the Russian society. Asking for representation and providing it in a 'non-serious', comic form, they simultaneously take a step towards claiming 'a (socio)linguistic citizenship' (Stroud 2015; Rampton *et al.* 2018), to be a part of the soundscape of a diverse multilingual space.

ACKNOWLEDGMENTS

We are grateful to Anna Verschik and two anonymous reviewers for their useful comments and help.

REFERENCES

- Andronikashvili, Zaal 2015. Glaubensbrüder oder Reich des Bösen: Russlandbilder der georgischen Literatur. – *Osteuropa* 65, 7/10, 531–548.
- Baranova, Vlada; Fedorova, Kapitolina 2019. 'Invisible minorities' and 'hidden diversity' in Saint-Petersburg's linguistic landscape. – *Language & Communication* 68, 17–27.
- Baranova, Vlada; Fedorova, Kapitolina 2020. Overcoming aggressive monolingualism: prejudices and linguistic diversity in Russian megalopolises. – *Open Linguistics* 6, 672–689.
- Benor, Sarah B. 2010. Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity. – *Journal of Sociolinguistics* 14, 159–183.
- Blackledge, Adrian; Creese, Angela 2015. Emblems of identities in four European urban settings. – *Language, Youth and Identity in the 21st Century: Linguistic Practices across Urban Spaces*. Eds. Jacomine Nortier, Bente A. Svendsen. Cambridge: Cambridge University Press, 167–182.

- Boxman-Shabtai, Lillian; Shifman, Limor 2015. When ethnic humor goes digital. – *New Media & Society* 17 (4), 520–39.
- Daniel, Mikhail; Knyazev, Sergei V.; Dobrushina, Nina 2010. Highlander's Russian: case study in bilingualism and language interference in Central Daghestan. – *Instrumentarium of Linguistics: Sociolinguistic Approach to Non-Standard Russian*. Helsinki: Slavica Helsingiensia, 65–93.
- Da Silva, Emanuel 2015. Humor (re)positioning ethnolinguistic ideologies: 'You tink is funny?' – *Language in Society* 44 (2), 187–212.
- Frajzyngier, Zygmunt; Gurian, Natalia; Karpenko, Sergei 2021. *Language Formation by Adults. The Case of Sino-Russian Idiolects*. Leiden; Boston: Brill.
- Gavrilov, Veniamin 2022. *Mainstream and Alternative Comedians: The Structure of Russian Stand-up Comedy Field*. Bachelor thesis. Ms. St. Petersburg, HSE.
- Jürgen, Jaspers 2017. The colour of Dutch: Some limits and opportunities of identifying Dutch ethnolects. – *Dutch Journal of Applied Linguistics* 6 (2), 231–246.
- Cornips, Leonie 2008. Loosing grammatical gender in Dutch. The result of bilingual acquisition and/or an act of identity. – *International Journal of Bilingualism* 12 (1–2), 105–124.
- Krämer, Philipp; Vogl, Ulrike; Kolehmainen, Leena 2022. What is "language making"? – *International Journal of the Sociology of Language* 274, 1–27.
- Labov, William 1972. *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press.
- Lo, Adrienne 2020. Race, language, and representations. – *International Journal of the Sociology of Language* 263, 77–83.
- Loentz, Elizabeth 2006. Yiddish, Kanak Sprak, Klezmer, and HipHop: Ethnolect, minority culture, multiculturalism, and stereotype in Germany. – *Shofar: An Interdisciplinary Journal of Jewish Studies* 25, 33–62.
- Ранова, Екатерина 2012. Отношение к языку: речь ученика и восприятие его учителем [Language attitudes: students' speech and its perception by teachers]. – *Язык, общество и школа [Language, society and school]*. Ed. Kapitolina Fedorova. Moscow, Novoe Literaturnoe Obozrenie.

- Pennycook, Alastair; Otsuji, Emi 2016. Lingoing, language labels and metro-lingual practices. – *Applied Linguistics Review* 7 (3), 259–277.
- Rampton, Ben 1995. *Crossing: Language and ethnicity among adolescents*. London, New York: Longman.
- Rampton, Ben; Cooke, Mel; Holmes, Sam 2018. Sociolinguistic citizenship. – *Journal of Social Science Education* 17 (4), 68–83.
- Shmeleva, Elena; Shmelev, Aleksej 2002. Русский анекдот. Текст и речевой жанр [Russian joke. Text and speech genre]. Moscow: Yazyki slavyan-skoj kul'tury.
- Stolz, Thomas 2008. Total reduplication vs. echo-word formation in language contact situations. – *Language Contact and Contact Languages*. Eds. Peter Siemund, Noemi Kintana. Amsterdam, Philadelphia: John Benjamins, 107–132.
- Stroud, Christopher 2015. Linguistic citizenship as Utopia. – *Multilingual Margins* 2 (2), 20–37.
- Verschik, Anna 2007. Jewish Russian and the field of ethnolect study. – *Language in Society* 36 (2), 213–232.
- Wiese, Heike 2009. Grammatical innovation in multiethnic urban Europe: New linguistic practices among adolescents. – *Lingua* 119 (5), 782–806.

RESÜMEE

VEENE KEELE AKTSENDID PERFORMATIIVSES KASUTUSES: NOORTE MUULASTE RAHVUSLIKUD STIILID, KEELEHOIAKUD JA IDENTITEEDID

Vlada Baranova

Das Nordost-Institut – IKGN

Kapitolina Fedorova

Tallinna Ülikool

Uurimus tegeleb noorte mitmekeelsete vene keele rääkijate erinevate etniliste aktsentide performatiivse kasutusega. Eesmärgiks on näidata, kuidas etnilisi aktsente kasutatakse ükskeelsete silmaklapide vaidlustamiseks ja agentsuse tagasi saamiseks, et võidelda rahvuslike ja keeleliste eelarvamuste vastu. Uurimus põhineb TikTokis ja Youtube'is avaldatud videote analüüsil, mille autorid on nii professionaalsed koomikud kui ka amatöörid ja mis esitlevad ja arutlevad muulaste erinevate aktsentide üle vene keeles. Et näidata publiku tüüpilisi reaktsioone, sisaldas analüüs ka videote metaandmeid ja kommentaare, mis väljendavad suhtumist aktsentide esitlusse ja keelelistesse stereotüüpidesse.

Võtmesõnad: noorte mitmekeelsus, keeleideoloogia, rahvuslik identiteet, vene keel, migratsioon, komöödia, vlogid

Vlada Baranova's research interests are in the fields of multilingualism, language revitalization and language documentation, language activism and how networks efforts of individual actors can interact with policy makers and official language planning. vladakharada@gmail.com

Kapitolina Fedorova's research interests are sociolinguistics, language contact, multilingualism, language ideology, border and migration studies, interethnic communication, and linguistic landscape studies. kapitolina.fedorova@tlu.ee

ZWEI SPRACHEN GLEICHZEITIG? NEIN, DAS SCHAFF' ICH NICHT: A LITHUANIAN- GERMAN BOY'S JOURNEY TO ACTIVE BILINGUALISM

Inga Hilbig

Vilnius University

Abstract.¹ Whereas many children in bilingual settings do not speak the minority language, very little is known about receptive bilingualism from the onset of speech and about such bilinguals activating their dormant language. Drawing on longitudinal ethnographic data, this paper reports on a case study of a receptive simultaneously bilingual Lithuanian-German boy who later started speaking both of his languages. Parents can do much for their children's bilingualism, but the child's agency is very important as well. The latter is much determined by the macro-socialisation factors, primarily by the communicative motivation of the child to use the minority language outside the bilingual home. Next to confirming possible insufficiency of the OPOL model, the paper demonstrates how quickly passive languages can be activated and highlights the importance of continuity of input and the value of receptive bilingualism.

Keywords: early simultaneous bilingualism, receptive bilingualism, minority language, OPOL model, Lithuanian, German

I. Introduction

Children's bilingualism in families is usually desired by the parents. However, unlike those of the majority language, the minority language competences, especially the productive ones, are not always

¹ This article was partly funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952366.

attained (De Houwer 2017, 2020). In fact, we know that approximately one in four children in bilingual settings who hear the minority language at home do not speak it at school age (De Houwer 2020). It means that this is not something exceptional. De Houwer's (2007) large scale study on language patterns in families in bilingual settings has revealed that in the homes where both parents use only the minority language the success rates to raise their children to speak both languages are the highest, whereas in the households where both parents speak the majority language and, in addition, one of the parents speaks the minority language they are the lowest. Thus, it can be argued that interethnic families where one of the parents is the majority language speaker and the other one is the minority language speaker using both family languages at home, also with the children, face more difficulties in raising the children to speak the minority language (Hilbig 2019). The OPOL (One Person One Language) model was held as ideal from the very first scholarly accounts on early bilingual development and is still widely deemed by the broad public as very effective and, moreover, very natural in bilingual families. However, by now it is clear that this principle is much more limited compared to the ML@H (Minority Language at Home) model, which, however, oftentimes is hard to implement due to the lack or non-existence of minority language skills in majority language parents.

Whereas problems with the minority language as children get older and the majority language dominates are well documented (see Lanza 1998; Barron-Hauwaert 2004; Slavkov 2015, among others), very little is known about children being receptive bilinguals from the very onset of speech. Moreover, there is so far only anecdotal evidence of such silent bilinguals activating their dormant language. In order to fill this gap, my sociolinguistic study demonstrates a rarely covered case of a child opting for the majority language from the very start of his bilingual first language acquisition and him later turning actively bilingual.

2. The subject of the study and his family

Pranas (P) was born in 2010 in Vilnius, Lithuania. P's father (F) is a German, with standard German (DE) as his native language. F is a medical doctor by education and occupation. P's mother (M) is a native standard Lithuanian (LT) speaker and is a linguist working at a university². P has a younger sibling Jonas (J), born two years later, in 2012. The first three years of P's life (September 2010–July 2013) were spent in Lithuania. Then the family moved to a village in Southern Germany for two years (July 2013–August 2015). After that, they returned to Lithuania.

P's parents understand and speak each other's languages although F is much more competent and fluent in LT than M is in DE. The main language of the couple's communication is LT. However, seeing that P would not speak DE, at some point F switched to DE when talking to M in P's presence. And after some time in Germany, seeing that P's LT was not active anymore, F started to use LT with M again in order to increase the minority language exposure. As for their communication with the child, M and F consistently adhered to the OPOL model. The main care-giver in the first five years of P's life was M, who was on a maternity leave with her both sons. P was attending day care in LT medium language from the age of 2. After moving to Germany, being 3 years old, he was enrolled at a local kindergarten where the local Swabian dialect was spoken. Until the age of 5, P used to spend only a few hours per day in the kindergarten in both countries, whereas the rest of his time was spent with M and J. P's F used to be much at home in P's first two years of life as well, since he was still studying. P has two sets of grandparents, living in their countries of origin, Lithuania and Germany. The Lithuanian grandparents do not speak any DE, and vice versa.

P was speaking almost exclusively the societal language of the country he was living in at the time, despite receiving plenty of input

² P's M is the author of this article.

in both of his native languages at home. When the family lived in Lithuania, it was the LT that was active. After moving to Germany, a quick and complete switch to DE took place. A breakthrough towards active bilingualism was finally noted only after the family settled back in Lithuania again, P being almost 5 years old.

3. Data collection, research methods, and research questions

This case study adopts a sociolinguistic approach to early bilingual development. It is based on ethnographic data: longitudinal observations documented in M's diary entries and audio-recordings. Field notes were kept roughly 2–3 times per week in 2010–2016 (until P turned 6). They include general observations on P's bilingual first language acquisition, its milestones, the most interesting cases of mixed utterances, P's thoughts and anecdotes that reveal his stance towards his two native languages in different periods of his early life. The diary notes also include relevant contextual remarks and explanations, notes on the changes in family situation, as well as M's comments, emerging questions and possible interpretations of P's linguistic behaviour. In 2012–2016, also naturalistic audio recordings (10–30 min in length each) were being made two to four times a week. M made sure to record an approximately equal number of P's one-on-one interactions with her, with F, and with both M and F. Later, M listened to some parts of the recorded material, making additional notes and writing some excerpts of conversations down. A selection of 10 hours has been fully transcribed and used as a complementary data source. The collected conversations were analysed using ethnographic and interactional sociolinguistics approaches.

4. Aim of the study and research questions

This paper continues a long tradition of linguistic biographies of simultaneous bilingual children (Afshar 1998; Dewaele 2017; Leopold 1970; Saunders 1988, among others). The article aims to present a sociolinguistic case study of early simultaneous receptive bilingualism from the very onset of speech acquisition. It follows how the child's active LT became passive and his passive DE was activated at the cost of his LT, and describes the breakthrough towards active bilingualism.

The research questions are the following: What has led to the situation in which P was speaking only the majority language (LT while in Lithuania and DE while in Germany), despite being exposed to two languages at home? Which factors were the most significant? What do manifestations of P's metalinguistic awareness reveal about his language attitudes and choices and the role of the child's agency in becoming or not becoming actively bilingual? What has played a crucial role in finally breaking the receptive bilingualism pattern?

The study is intended to emphasise the importance of the pragmatic communicative need for children to speak the minority language and to demonstrate difficulties in creating this need in inter-ethnic bilingual households. It also aims to show how quick the processes of passivation and activation of languages in young bilinguals can be, provided there are good children's receptive skills and favourable extralinguistic circumstances. As for practical implications, the intention is to advocate the value of receptive bilingualism among parents, educators and society at large, encouraging minority language speakers to use their native languages with their offspring even if the latter speak the majority language.

5. P's bilingual first language acquisition history

5.1. THE VERY BEGINNING: P IN LITHUANIA (0–2;10³)

LT clearly dominated P's speech during the one-word and early multi-word stages already (they started at 1;2 and 1;7, respectively). LT was the base language with rare DE insertions that were becoming more and more grammatical over time. The mixed language was easy for M to notice and write down, as there were relatively few instances of it. In 10 transcribed hours (2;4–2;10), only 5% of the output was DE. Unsurprisingly, P was mixing more when interacting with DE-speaking F. The child had a minimal amount of active DE vocabulary and phrases. Receptive bilingualism does not necessarily mean zero production, and, in line with findings by Nakamura (2018, 2019), this scarce minority language output was mainly of re-active origin, meaning that it consisted mostly of non-original, imitated or routine words and phrases in reaction to what F was saying. P was not using DE independently and had not moved beyond the two-word utterances level in it. Rare instances of P's mixed language aside, P and F were constantly engaging in what is called dual-lingual discourse (see Nakamura 2018; Saville-Troike 1987), meaning that they were systematically using different language codes in their conversations. P's minimal productive DE was in sharp contrast with his fast-developing speaking skills in LT. LT as P's stronger language also manifested itself in his LT accent when speaking DE. Sometimes P faced difficulties pronouncing DE words, although his articulation in LT was usually very clear. It is important to note that no language mixing outside the home was reported by kindergarten teachers or the Lithuanian grandparents when P had no bilingual parents at his side. Thus, the sensitivity to the context and interlocutors, typical for even very young emergent bilinguals (see Lanza 1998), was clearly there although P was sometimes mixing language "labels", calling LT DE and DE LT.

³ Child's years;months

P's reactions to M trying to experiment a bit and speak some DE to him were always negative. P used to "argue" with his parents, how one or another thing should be called. Sometimes the boy insisted on using DE words, but more often he preferred the LT ones. P also demonstrated a clear favouring of the majority language by offering quite a few metalinguistic comments on his language preferences in the analysed period.

For example, in the conversation excerpt No. 1, M remarks that she is frying eggs. P notices that *fried eggs* can also be said in a "different", F's way. At this point, F joins the conversation reminding P of the DE word. However, the child expresses his preference for the LT version of it by stating how *he* personally wants to call the dish, repeating the LT word twice and emphasising it the second time:

- (1) M: *Kepu kiaušiniene dabar.*
 P: *Tėtis sako kitaip...*
 F: *Ich sage „Spiegelei“.*
 P: *Kiaušiniene. Aš sakau KIAUŠINIENĖ!*
 M: *I'm frying eggs now. (LT)*
 P: *Dad says it differently... (LT)*
 F: *I say "Spiegelei" (DE)*
 P: *Fried eggs. I say FRIED EGGS (LT)!*
 2;6, M's diary note

In excerpt No. 2, even more explicit negative orientation towards DE and inclination towards LT is expressed. It is obvious that by asking F why he calls a peach "differently", and not "persikas" (LT), P not only genuinely asks a question, but in fact also confronts F. By making a judgement, P also implies that F should rather use the LT word as well:

- (2) P: *Aš sakau „persikas“. Kodėl tu sakai kitaip?*
 F: *Ich sage „Pfirsich“.*
 P: *Negražu taip sakyti!*

P: *I say "peach". Why do you say it in a different way?* (LT)

F: *I say "Pfirsich"* (DE).

P: *It's not nice to say it this way!* (LT)

2;8, M's diary note

It should be noted that the DE words are not articulated by P in either of those two conversations, and it is not clear whether he remembered them and could have produced them if willing. However, P proved to be capable of activating some of his DE when absolutely needed, for instance, in communication breakdowns, making use of his bilingual skills, like here:

(3) M: *Apie ką nori knygelę paskaityti?*

P: *Apie tenuką.*

M: *Apie ką tokį?!*

P: *Apie tenuką!* (vietoje „traukinuko“)

M: *Ką? Nesuprantu (susierzinusi). Ar negali man kaip nors kitaip pasakyti?*

P: *Apie ZUG* (vok. „traukinys“)

M: *What do you want me to read for you today?* (LT)

P: *About tenukas* (an invented word for train that M is not able to grasp). (LT)

M: *About what?!* (LT)

P: *About tenukas!* (LT)

M: *What? I don't understand* (irritated). *Can you say that in another way?* (LT)

P: *About ZUG!* ("train" in DE) (LT)

2;8, recorded conversation

On the other hand, P was always refusing to translate from LT to DE or repeat the DE words that parents were trying to "put in his mouth" if he sensed that M and F were just probing him in an attempt to activate his DE. If at all, P was solely translating from his weaker DE to LT, but not in reverse. In the episode No. 4, we see how P refuses to translate a phrase into DE by declaring not knowing DE altogether:

(4) M: *Pranuk, gal gali kitaip pasakyti?*

P: *Aš nemoku vokiškai!*

M: *Pranas, can you say it in another way? (LT)*

P: *I don't know any German! (LT)*

2;6, recorded conversation

P's strong determination to stick to the majority LT even while receiving abundant DE input is especially vivid in the following example. Here F is reading a classic German children's story about a train called Henriette (already mentioned in the conversation in example No. 3). The train is rolling down the fields and everybody is happy to see her. The book was particularly loved by P and was being read to him repeatedly, both in DE by F and in loose LT translation by M. This time, F stops for a second to encourage P to complete the sentence in which rabbits greet Henriette:

(5) F: *...Gleich danach hört man ein Rufen aus dem Brommelbeeren-schlag. Vierzehn Hasen rufen heiter: „Henriette, ...*

P: *... Labas!*“ (vietoje *Guten Tag*)

F: *... And then comes a shout from a blackberry bush. Fourteen rabbits shout cheerfully: “Henriette, ... (DE)*

P: *... Hi!*“ (LT) (instead of *Guten Tag*)

2;10, M's diary note

One may have anticipated that an intense flow of DE being listened to for several minutes could have broken P's LT-only pattern at least for a moment. Especially because the required phrase was so simple and absolutely easy to recall, and the rhyme and familiarity of the story were additional facilitators. But that did not happen.

To support the minority DE at home and to increase its input for P, F started talking DE to M as well. However, this change in inter-couple communication did not have any noticeable effect on P. Perhaps, it was even counterproductive and only strengthened the dual-lingual P and F communication pattern.

In his first three years, P was getting oriented in his bilingual environment, sorting out his native languages, their vocabularies

and grammar. He was successfully learning to speak in LT, whereas his DE seemed to be developing at a much slower rate. In spite of some signs of his productive DE emerging, P's clear predisposition for the majority LT from the very beginning was revealed both implicitly, by his actual linguistic practices, and explicitly, by his expressed metalinguistic comments.

The family was regularly visiting P's grandparents in Germany. Trips to the minority language country are known to be very effective for strengthening and/or activating the child's minority language and are highly recommended for bilingual families (see Barron-Hauwaert 2004; Bourgogne 2013; Slavkov 2015; Uribe de Kellett 2002). However, the biannual two-to-four week long stays in Germany did little in P's case. P was speaking LT also with his paternal grandparents, relying on M and F to interpret for him when it was absolutely needed. Thus, even when spending time with monolingual DE speakers P seemed to lack incentives and perhaps also capability to put his DE in active use. This is because, unlike children who turn into passive bilinguals at some point later in their childhood (cf. Slavkov 2015; Uribe de Kellett 2002), he had never really spoken DE before (for similar cases, see Nakamura 2018).

P's bilingual development was clearly not harmonious (for the definition of the concept, see De Houwer 2015). P's parents, especially F, felt discouraged and upset by the situation, as they were not sure if their son would ever learn and/or want to speak DE with F and his family. M and F believed that using OPOL consistently, providing rich and abundant input, that is, interacting much, reading to the child, and fostering his connections with the minority language country and family members leads to active bilingualism by default, and so they expected it to happen in their family as well. However, the results were unsatisfying. One of the reasons for the family to temporarily move to Germany after F was offered a job there was to place P into a monolingual DE environment for a longer period of time so that he would get a natural motivation to activate his DE.

5.2. THE BIG CHANGE: P IN GERMANY (2;10–4;11)

In Germany, P's DE became the language of the wider social environment. Just as expected, after settling in, it did not take too long for him to start actively using his dormant DE for the first time in his life. He became very open to DE and eager to master it. At home, he was mixing in more and more DE words and phrases into his LT. In three weeks' time P was already able to construct his first multiword utterances and more complex sentences in DE. His active DE vocabulary was rapidly growing and he started to initiate and hold conversations in DE. According to the kindergarten teachers' testimonies and from what the parents witnessed themselves, P's sufficient fluency in DE was reached in a mere month. In a diary comment from that time, M noticed that *P is making progress in DE not just within days, but within hours*.

On the other hand, the activation of DE was at the cost of P's LT. The latter was now restricted solely to the family context. The steadily and rapidly increasing quantity of DE utterances outnumbered LT ones to such extent that after two months there were only three to four LT insertions in P's DE per day when interacting with M. P was speaking only DE with F and J. P's productive LT grammar, vocabulary and native-like LT accent seemed to be affected by attrition. In cases of miscommunication, P was happy to define and translate DE words for M when he was able to. But just as before, P refused to translate into his weaker language, which was LT now, when asked without any apparent real-life reason. Also, he would not repeat LT words or phrases for language maintenance purposes only and, if at all, provided only routine and imitated short responses in LT for M. However, he neither insisted that M spoke DE to him nor ever demonstrated any comprehension difficulties and still enjoyed listening to books being read in LT and hearing LT songs. Obviously, P's receptive LT skills were intact.

The situation with the performance and balance of P's native languages in Germany was a mirror view of the former situation in

Lithuania. In the table below, I have listed and compared the most important sociolinguistic and extralinguistic factors that likely have influenced and could explain P's language attitudes and practices in both countries and the shift of his active and passive languages.

Table 1. *The main factors that could have influenced P's linguistic attitudes and behaviour in Lithuania and Germany and his active and passive languages swapping places*

Stable parameters	P (0–2;10 years) in Lithuania Active language – LT	P (2;10–4;11 years) in Germany Active language – DE
Language spoken to P by F	DE	DE
Language spoken to P by M	LT	LT
Daily time spent at home	≈ 8 hours per day	≈ 7 hours per day
Daily time spent with M and J	≈ 8 hours per day	≈ 7 hours per day
Daily time spent with F (mostly all family being together)	≈ 5 hours per day	≈ 3 hours per day
Daily time spent in majority language child-care	≈ 4 hours per day	≈ 5 hours per day
Language spoken by M to F (in P's presence)	LT	LT
Contacts with minority language peers	None	None
P's meetings with minority language relatives (in minority language/third country)	2 times, 1 week long	3 times, 1 week long
Changed parameters	P (0–3 years) in Lithuania Active language – LT	P (3–5 years) in Germany Active language – DE
Kindergarten and community language	LT	DE
Languages spoken by F to M (in P's presence)	LT, later DE	DE, later LT
Daily reading time for P	≈ 15 min in LT + ≈ 15 min in DE	LT and DE every second day ≈ 20 min
Visits to the minority language country	6 times, 1–4 weeks long each	1 time, 1 week long each

As it is shown, many parameters in the home domain were the same or only slightly different in both countries, with a sole exception of the fact that LT in Germany was getting less support than DE in Lithuania in terms of reading time and the number of visits to the minority language country. However, what was going on in the home environment in both countries seemed to have little influence on P's active usage of the minority language. Obviously, P simply saw no point using his weaker language with his bilingual parents. The major determinant for one language to be active and the other passive and them swapping places was by all means external, as it was the issue of a community language. It was the monolingual DE environment outside home that created a real need and steered P to start using his former passive DE. This macro factor has overwhelmed both the stable and changed micro factors at home. However, P's LT was not kept alongside and pushed away from active usage as socially and communicatively redundant, which meant that the boy still remained a passive bilingual.

As for the community language, P's encounters with monolingual German peers in the playground and especially in the kindergarten were the most significant forces. One could tell that the presence of not only adults, but also other children who spoke DE and understood no LT made an immense impression on him right away. Since P especially loved to be out of home and play with other children, it was suddenly very important for him to speak their language. It is worthy to add that although hearing only standard DE at home, P picked up much of the local Swabian variety and partially adopted the local accent. This is a proof of a strong linguistic peer group influence on even very young children.

During the most dynamic first weeks in Germany, P was completely concentrated on adapting and fitting into the changed linguistic and socio-cultural environment as quickly as possible. Although he liked it in the kindergarten very much and had no problems there whatsoever, the seemingly smooth and speedy reorientation was obviously strenuous. P was very tired in the evenings

and had frequent temper tantrums. Seemingly, P lacked capacity to activate one language while keeping the other active and to alternate between them and was not willing to develop those skills either because it was not necessary for him to speak both languages heard in his environment. To save effort and energy, he chose to be pragmatic and use solely the relevant DE with everyone, also with his bilingual family members. One episode, documented during the first month in Germany, exemplifies P's determination very well:

- (6) *Tonight, Pranas wanted to share something exciting about what happened in the kindergarten with F. While speaking, he couldn't retrieve one needed DE word. Instead of simply substituting it with a LT one, P got very frustrated, burst into tears and needed to get help to finish his sentence and get consoled.*

2;11, M's diary note

P enjoyed making fun of LT and LT words, e.g. by distorting them, which could be interpreted as a manifestation of his lack of care for his other native language. On the other hand, a half a year before the family returned to Lithuania, there was a short conversation going on between P and M, the main part of which is presented in example No. 7. Being aware that they are going to go back to Vilnius, P expressed his inability to speak LT with some sorrow and anxiety in his voice, also revealing a low self-confidence as an actively bilingual person:

- (7) P: *...weil LT sprechen kann ich nicht, gel?*
 M: *Kai grįšim į Lietuvą, vėl galėsi. Ir mes tikimės taip pat, kad su tėčiu ir toliau DE kalbėsi...*

P: *Zwei Sprachen gleichzeitig? Nein, das schaff' ich nicht...*

P: *...because I can't speak LT, can I? (DE)*

M: *When we are back to Lithuania, you will be able to again. And we hope very much that you will continue speaking DE with your dad... (LT)*

P: *Two languages at the same time? No, I won't manage... (DE)*

4;4, M's diary entry

This vivid example makes it even more obvious that in fact it was not because of some principled rejection of the LT language that P was not speaking it. Rather, it was because of P's lacking bilingual performance skills, despite sufficient receptive knowledge in both languages, and him wanting to act economically. No motivation to use the minority LT has led to its passivation and diminished productive skills, whereas poor performance skills on their part made it impossible to activate LT promptly even if willing.

The situation with bilingualism at home was not harmonious, just like it was not before in Lithuania. The fact that M and her boys⁴ spoke different languages felt strange and alienating to her. Still, since the family had a plan to return to Lithuania, parents were confident that the reactivation of LT was guaranteed, so they both mostly rejoiced over P's and J's developing DE. However, for a very short period of time, when M and F doubted and were seriously considering an option to stay in Germany permanently, M became really desperate, facing the prospect that P and J might not ever be able to speak her language.

5.3. THE BREAKTHROUGH AND BEYOND:

P BACK IN LITHUANIA (4;11–6;0)

Based on their previous experience, the parents were unsure what is going to happen to their sons' DE when they come back to Lithuania. M and F's concern was that according to the same pattern, P and also J will now cease speaking DE.

In the very first days back in Vilnius, P noted with astonishment: *They all speak LT!* It was relevant and socially needed to master LT again, especially for peer-group communication, which was the

⁴ J was following his brother's developmental path very closely in most regards. Only that his situation was exactly opposite in terms of active and passive languages because he was learning to speak in Germany. Despite spending much of time at home with M, he was receiving much of DE input both from P and F and in the *creche* which he started to attend 4 hours per day 3 times per week from the age of 1 year. From the very first words and utterances, he spoke almost exclusively in DE, also with M.

strongest stimulus for a change again. M's documented P's first LT sentence after two year's break was put together when trying to catch the attention of a neighbour boy to show him a new toy car. But apart from that, it was actually impossible for M to observe the activation process of P's LT because for one month he was speaking exclusively DE at home. The revival of the productive LT was taking place in the kindergarten. The teacher reported a breakthrough in P's Lithuanian in approximately 1,5 months. Since M was back to her full-time job, the boys were spending most of their days in the kindergarten. Being in the same group meant that they could help each other with the language, when needed, and that made the transitional period easier.

P and J had not stopped using DE with F. Moreover, P continued speaking DE also with M for nearly four months. DE was also kept as the dominant inter-sibling communication language for as long as nearly one year. Evidently, DE was positively associated with many nice sociocultural and personal experiences from the time in Germany (e.g. child-friendly countryside environment, interesting activities in the kindergarten, nice friends, impressing family trips) and was held as something valuable and worth maintaining. P was a role model for J and the one who kept an eye on the "right" language code. The older brother used to exercise some language control over the younger one (e.g. *Jonas, we speak DE!*). P was especially keen and determined to protect DE-only linguistic territory at home. For instance, one morning, M noticed that P talks in LT to her for a few minutes already:

(8) M: *Pranai, tu su manim lietuviškai kalbi!* (džiaugsmingai nuste-busi)

P: *Nein!* (tvirtai)

(P iš karto persijungia atgal į DE)

M: *Pranas, you speak Lithuanian to me!* (surprised, joyfully) (LT)

P: *No!* (firmly) (DE)

(P instantly switches back to DE)

5;5, M's diary entry

P now had no difficulties to switch between languages and was doing that frequently with M and especially with J. With the time, LT was dominating more and more in those interactions. As for F, P was clearly avoiding switching with him and was making attempts to stick to DE, at least at the discourse level. Unlike before, P was interested in and willing to translate when requested and on his own initiative in both directions, ask questions, compare the languages, discuss their differences and talk about his bilingualism quite often. This clear rise in his metalinguistic bilingual awareness had obviously to do with him reaching a new developmental stage.

By his sixth birthday and in more than one year after coming back to Lithuania, P's bilingualism could be called established, stabilised and balanced. In any case, the most vulnerable and dynamic period of P's earliest bilingual life seemed to be over. The child was majority language dominant, but also a competent speaker in DE according to his age⁵. P's DE was being further developed through books, audio-recordings, cartoons, films, educational TV programs and visits to Germany, etc. The parents thought that the success in maintaining DE had also much to do with P's close and warm relationships with F and the paternal grandparents.

6. Discussion and conclusions

From the onset of speech, P was speaking only the societal language of the country he was living in at the time (Lithuania or Germany). Although exposed to and stimulated by both languages at home, he would systematically reject the minority language and stick to the majority one. His active bilingualism emerged only after the family changed countries of residence twice, P already being 5 years old. Two things were crucial in this: P's experience living in both linguistic communities, especially his socialisation with monolingual peers,

⁵ A subjective evaluation of the German grandmother, an elementary school teacher in retirement.

as well as him getting more cognitively mature, which has enabled and enhanced his bilingual performance and switching skills.

The study confirms possible insufficiency of the idealised, as De Houwer (2009) puts it, OPOL model to produce active bilingual children if they have no pragmatic incentives to speak the minority language. OPOL can upset the language balance too much because the minority language parents often are the only source of input and the only conversational partners in children's day-to-day life. But even more importantly, children may not feel any necessity to put effort in practising their weaker minority language with a parent whom they know to understand and even speak their stronger majority language. As my evidence suggests, creating the pragmatic communicative need for the minority language in bilingual households can be very challenging, if not impossible sometimes, without the support of powerful macro-socialisation forces. Temporary relocation to the minority language country, where children have no other choice but to reawaken their dormant language in the local community, is optimal, but it is of course seldom feasible. On the other hand, shorter-term full immersions, e.g. summer camps or stays with monolingual minority language relatives could also be very useful, at least with some children.

P's parents had positive attitudes to bilingualism and were providing steady high quantity and quality input for their son. However, in P's very earliest years M and F were not aware of the specific recommended discourse strategies (see Lanza 1998) that have been proven to foster the usage of the minority language within the same conversation. Those include parents feigning lack of comprehension, asking for clarifications, repetitions and translations if the child responds in the inappropriate language, or at least recasting what the child says in the other language, etc. M and F were applying some of those strategies by intuition, but only sporadically, and they probably started doing that too late. The dual-lingual communication pattern was by then well-established and thus difficult to reverse. P was not being socialised to answer in the language he was addressed in from the very beginning and overtly resisted being nudged into the

minority language later. On the other hand, it is questionable if those strategies could have really been effective in P's case (see also Slavkov 2015), considering that the majority language dominated already in the one-word stage and also because of the boy's strong character. Apart from parental engagement and impact beliefs, that is, the conviction that parents can directly affect their children's language practices (see De Houwer 1999), the role of the child's agency is also very important and can not be underestimated. Based on what was presented, it must be admitted that success in early bilingualism in children does not lie entirely in parents' hands. With their own linguistic attitudes, preferences and agenda children also have a very significant role to play in the family language policy and can steer its course opting for receptive bilingualism or even monolingualism.

On a positive note, P's case clearly demonstrates how quickly the child's passive language can be revived, provided there are favourable extralinguistic circumstances and solid comprehension skills (see also Dahl *et al.* 2010; Slavkov 2015; Uribe de Kellett 2002). In this paper, I want to highlight the importance of continuity of input of the dormant language. P's bilingual first language acquisition clearly proceeded further in the passive bilingualism phases as well although the non-existent output left parents in doubt. The value of receptive bilingualism and the importance of parents' continued use of their native minority language with their children even if the latter systematically reply in the majority language needs to be more promoted among parents and educators. In the words of Slavkov (2015, p. 730), receptive bilingualism can be viewed as a valuable asset worth maintaining rather than a lost cause.

Receptive bilingualism is also bilingualism. Being able to comprehend another language is already very much and very valuable. Moreover, receptive skills also involve a potential for active bilingualism later in life. These ideas could be encouraging for parents willing that their children possess all family languages and striving for that. In case of a failure to socialise offspring into speaking the minority language with their parents, appreciation of the receptive

bilingualism (next to awareness that it is not something exceptional and, moreover, may be just a temporary phase) can possibly help parents to cope with emotional struggles (see Hilbig 2020) and assist them in generating the necessary energy and resources so that they can confidently continue to escort their children in their early bilingualism journey.

REFERENCES

- Afshar, Karin 1998. *Zweisprachigkeit oder Zweitsprachigkeit?: zur Entwicklung einer schwachen Sprache in der deutsch-persischen Familienkommunikation*. Münster: Waxman.
- Barron-Hauwaert, Suzanne 2004. *Language Strategies for Bilingual Families: The One-Parent-One-Language Approach*. Bristol, Buffalo, Toronto: Multilingual Matters.
- Bourgogne, Annika 2013. *Be Bilingual: Practical Ideas for Multilingual Families*. Amazon Digital Services LLC.
- Dahl, Tove Irene; Rice, Curt; Steffensen Marie; Amundsen, Liudmila 2010. Is it language relearning or language reacquisition? Hints from a young boy's code-switching during his journey back to his native language. – *International Journal of Bilingualism* 14, 4, 490–510. <https://doi.org/10.1177%2F1367006910371024>.
- De Houwer, Annick 1999. Environmental factors in early bilingual development: the role of parental beliefs and attitudes. – *Bilingualism and Migration*. Eds. Guus Extra, Ludo Verhoeven. New York: Mouton de Gruyter, 75–96.
- De Houwer, Annick 2007. Parental language input patterns and children's bilingual use. – *Applied Psycholinguistics* 28, 411–424. <https://doi.org/10.1017/S0142716407070221>.
- De Houwer, Annick 2009. *Bilingual First Language Acquisition*. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781847691507>.
- De Houwer, Annick 2015. Harmonious bilingual development: young families' well-being in language contact situations. – *International Journal of Bilingualism* 19, 2, 169–184. <https://doi.org/10.1177%2F1367006913489202>.
- De Houwer, Annick 2017. Minority language parenting in Europe and children's well-being. – *Handbook of Positive Development of Minor-*

- ity Children and Youth. Eds. Natasha J. Cabrera, Birgit Leyendecker. Springer, 231–246. <https://doi.org/10.1007/978-3-319-43645-6>.
- De Houwer, Annick 2020. Why so many children who hear two languages speak just a single language? – *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 25, 1, 7–26.
- Dewaele, Jean-Marc 2017. The Stories of our Multilingual Children. Livia: A Multilingual Journey. – *Raising Multilingual Children*. Eds. Jean-Marc Dewaele, Julia Festman, Gregory J. Poarch. Multilingual Matters, 5–33.
- Hilbig, Inga 2019. Dvikalbystė Oslo lietuvių šeimose. – *Emigrantai. Kalba ir tapatybė II: keturi sociolingvistiniai portretai*. Ed. Meilutė Ramonienė. Vilnius: Vilniaus universiteto leidykla, 183–241.
- Hilbig, Inga 2020. Nedarnioji dvikalbystė mišrioje emigrantų šeimose. – *Taikomoji kalbotyra* 14, 1–20. <https://doi.org/10.15388/Taikalbot.2020.14.1>.
- Lanza, Elizabeth 1998. Raising children bilingually in Norway. – *International Journal of the Sociology of Language* 133, 73–88.
- Leopold, Werner 1939–1949/1970. *Speech Development of a Bilingual Child: A Linguist's Record*. Evanston, IL: Northwestern Press.
- Nakamura, Janice 2018. Parents' use of discourse strategies in dual-lingual interactions with receptive bilingual children. – *Cross-linguistic Research in Monolingual and Bilingual Speech*. Ed. Elena Babatsouli. Chania: ISMBS, 181–200.
- Nakamura, Janice 2019. Receptive bilingual children's use of language in interaction. – *Studies in Language Sciences: Journal of Japanese Society for Language Sciences* 18, 46–66.
- Saville-Troike, Muriel 1987. Dilingual discourse: The negotiation of meaning without a common code. – *Linguistics* 25, 81–106.
- Saunders, George 1988. *Bilingual Children from Birth to Teens*. Clevedon: Multilingual Matters.
- Slavkov, Nikolay 2015. Language attrition and reactivation in the context of bilingual first language acquisition. – *International Journal of Bilingual Education and Bilingualism* 18, 6, 715–734. <https://doi.org/10.1080/13670050.2014.941785>.
- Uribe de Kellett, Angela 2002. The recovery of a first language: a case study of an English/Spanish bilingual child. – *International Journal of Bilingual Education and Bilingualism* 5, 3, 162–181.

RESÜMEE

ZWEI SPRACHEN GLEICHZEITIG? NEIN, DAS SCHAFF' ICH NICHT: LEEDU-SAKSA POISI TEEKOND AKTIIVSE KAKSKEELSUSENI

Inga Hilbig

Vilniuse Ülikool

Kuna paljud lapsed, kes on mitmekeelsetes keskkondades, ei räägi vähemuskeelt, on väga vähe teada retseptiivsest mitmekeelsusest kõne algstaadiumis ja selliste kakskeelsete tukkuva keele aktiveerumisest. Kasutades etnograafilist pikiuuringu andmestikku, käsitleb antud uurimus retseptiivset simultaanset kakskeelset leedu-saksa poissi, kes hakkas hiljem mõlemat keelt kõnelema. Vanemad saavad küll kakskeelse lapse jaoks palju ära teha, kuid lapse enda agentsus on ka oluline. See agentsus on mõjutatud makrosotsialiseerumise faktoritest, eelkõige lapse kommunikatiivsest motivatsioonist vähemuskeelt koduväliselt kasutada. Lisaks ÜVÜK (üks vanem, üks keel) mudeli võimalike puudujääkide tuvastamisele näitab antud uurimus, kuidas passiivsed keeled võivad kiiresti aktiveeruda, ja rõhutab sisendi järjepidevuse olulisust ning retseptiivse mitmekeelsuse väärtust.

Võtmesõnad: varajane kakskeelsus, retseptiivne kakskeelsus, vähemuskeeled, ÜVÜK mudel, leedu, saksa

Inga Hilbig is an assistant professor at Vilnius University. Her research interests comprise sociolinguistics of early bilingualism and family language policy.
inga.hilbig@flf.vu.lt

THE ENGLISH FRICATIVE CONSONANT /Z/ AS A CHALLENGE TO NORWEGIAN L1 EFL LEARNERS: AN ERROR ANALYSIS OF PHONEMIC TRANSCRIPTIONS

Oleksandr Kapranov
NLA University College Oslo

Abstract. English fricatives, such as /z/, are thought to pose substantial challenges to the students of English as a Foreign Language (EFL) whose first languages (L1s) are characterised by phonetic repertoires that are dissimilar to that of English as far as fricatives are concerned (Kallio *et al.* 2021; Kanokpermpoon 2007). The absence of the fricative /z/ in the phonetic inventory of Norwegian is reported to impact negatively on the Norwegian L1 EFL learners' speech production in English (Rugesæter 2014). The study that is further presented in the article aims to analyse potential challenges associated with the English fricative consonant /z/ experienced by a group of Norwegian L1 EFL students (hereafter "participants") on the B2 level of proficiency in English according to the Common European Framework of Reference for Languages (The Council of Europe 2011). To that end, the participants were requested to execute a series of phonemic transcriptions in the International Phonetic Alphabet (IPA): one transcription at the end of the autumn semester (Task 1) and another one at the end of the spring semester (Task 2). The phonemic transcriptions in the study were regarded as a diagnostic tool (Fouz-González, Mompean 2021; Lintunen 2005) that provided an index of the participants' familiarity with /z/ and, indirectly, reflected their use of /z/. The error analysis of the participants' transcriptions revealed that the majority of them made mistakes by substituting /z/ for /s/. Considering that the substitution of /z/ for /s/ was common in Task 1 and persisted in Task 2, it was concluded that the participants, who were on the upper-intermediate level of English proficiency, found /z/ challenging. The linguo-didactic implications of the findings are discussed in the article.

Keywords: English as a Foreign Language (EFL), the English fricative consonant /z/, /s/ – /z/ contrast, upper-intermediate EFL learners, Norwegian L1

1. INTRODUCTION

Successful foreign language (FL) learners should master pronunciation in an FL and foster their awareness of individual sounds of the FL they study (Bjelaković, Čubrović 2021; Huensch, Thompson 2017; Verschik 2017). However, an FL learner's success in mastering pronunciation may be hampered by a host of variables (Szyszka 2017; Waniek-Klimczak, Klimczak 2005), such as individual sounds (also referred to as “segmentals”) that pose significant problems to the learners of English as a Foreign Language (EFL) whose first languages differ significantly from English (Bjekić, Čubrović 2021; Kallio *et al.* 2021; Kanokpermpoon 2007; Roa *et al.* 2021; Waniek-Klimczak 2011; Zhang *et al.* 2021). For instance, English fricative consonants, e.g. /z/, are reported to be associated with substantial challenges to the EFL learners whose first languages (L1s) are not typologically related to English, such as Chinese, Thai, etc. (Kanokpermpoon 2007; Strange 1992). However, EFL learners whose L1s are genetically close to English also experience production and perception difficulties that involve English fricatives (Lersveen 2018; Rugesæter 2014). In particular, the English fricative consonant /z/ is found to be difficult to perceive and pronounce by Norwegian L1 and Swedish L1 EFL learners (Flege, Hillenbrand 1986; Haugen 1967; Lersveen 2018; McAllister *et al.* 2002; Rugesæter 2014). Whilst Norwegian and Swedish as Germanic languages are typologically very similar to English (Bech, Walkden 2016; Kapranov 2014), the absence of /z/ both in Norwegian and Swedish is argued to impact negatively on the Norwegian L1 and Swedish L1 EFL learners' speech production in English (Flege, Hillenbrand 1986; Lersveen 2018; McAllister 2007; Rugesæter 2014). In this regard, the literature indicates that a common /z/-related mistake made by Norwegian L1 EFL learners

involves the substitution of /z/ for its closest Norwegian equivalent, i.e. the fortis fricative /s/ (Rugesæter 2014). The substitution of /z/ for the Norwegian /s/ is argued to constitute a typical feature found in the speech by Norwegian L1 immigrant population in the USA (Moen 1988), as well as by young Norwegian L1 EFL learners on the beginner's level of proficiency in English (Lersveen 2018; Nilsen 1989; Rugesæter 2014).

Building upon the previous literature (Flege, Hillenbrand 1986; McAllister *et al.* 2002; Nilsen 1989; Rugesæter 2014), this article presents a study that seeks to shed light upon whether or not the English fricative /z/ would represent a challenge to a group of Norwegian L1 EFL students (hereafter “participants”) on the B2 level of proficiency in English according to the Common European Framework of Reference for Languages, or CEFR (The Council of Europe 2011). Whilst the prior studies pay attention to /z/-related mistakes in the perception and production of English speech either by Norwegian L1 EFL learners on the beginner's level of proficiency or Norwegian immigrants in the English-speaking countries (Moen 1988; Nilsen 1989; Rugesæter 2014; van Dommelen, Hazan 2010), the novelty of the present study involves the research focus on the group of participants who are university EFL students on the upper-intermediate level of proficiency in English. It should be emphasised that the B2 level of English proficiency according to CEFR (The Council of Europe 2011) is routinely overlooked in the literature on EFL pronunciation (Rugesæter 2014). In addition, there seems to be a paucity of published research that analyses Norwegian L1 EFL learners' /z/-related errors through the lenses of phonemic transcription in the International Phonetic Alphabet (IPA). Seeking to bridge the existing gap, the aim of the present study is to identify and analyse possible /z/-related errors in the participants' phonemic transcriptions in the IPA by means of addressing the following research questions (RQs):

1. Would the participants make any mistakes associated with /z/ in the phonemic transcriptions in the IPA?

2. Would the participants' possible /z/-related mistakes in the phonemic transcriptions in the IPA decrease, increase and/or remain stable in the course of two semesters of study?

Prior to discussing the RQs, this article proceeds as follows. First, the theoretical framework of the study is provided in Section 2. Thereafter, in Section 3, a review of the previous literature is outlined. Section 4 discusses how phonemic transcription in the IPA is employed in a variety of EFL contexts. In Section 5, the status of the English language in Norway is explained. Additionally, Section 5 summarises the teaching and learning of English in Norwegian contexts. That is followed by the description of the present study, inclusive of the participants, methodology, tasks, results and their discussion in Section 6. Finally, the article concludes with the summary of the major findings and their linguo-didactic implications in Section 7.

2. Theoretical Background

The theoretical background of the present study is based upon the Speech Learning Model (SLM). In accordance with the SLM, the learner's L1 plays a prominent role in the acquisition of the phonetic system of one's foreign language (FL), second language (L2), or EFL, for that matter (Flege 2005). The SLM postulates that

... the phonetic systems used in the production and perception of vowels and consonants remain adaptive over the life span, and those phonetic systems reorganize in response to sounds encountered in an L2 through the addition of new phonetic categories, or through the modification of the old ones. (Flege 1995: 233)

It should be emphasised that, according to the SLM, the mechanisms of learning one's L1 sound system last over the life span and can be successfully applied to the learning of an L2 and/or an FL (Flege 1995). In the process of L2 learning, the sounds that are phonetically similar to the learners' L1 are assimilated into the L2 and/

or FL categories (Chan 2012; Evans, Alshangiti 2018; Flege 2009). Specifically, the recently revised SLM (SLM-r) model suggests that L2 learners map the sounds they encounter in L2 words onto their L1 phonetic categories by means of the so-called cross-language mappings that take place subconsciously and automatically (Flege *et al.* 2021: 85). It follows from the SLM, as well as from the SLM-r, that the process of cross-language mappings is not straightforward, given that it is exacerbated by the learners' maturing categories in their L1 that can block the formation of new categories of the FL sounds (Best, Tyler 2007; Rojczyk 2010; Munro, Bohn 2007). Specifically, it is argued that the maturation of the L1 categories can lead to potential problems associated with the perception of phonetic contrasts in the L2 and/or FL (Flege, Hillenbrand 1986: 508).

In addition to the compromised perception of the L2/FL contrasts (for instance, the /s/ – /z/ contrast in English), the SLM points to the substitution of FL-specific phonemes that are absent in the learners' L1 by the closest L1 phoneme as a typical strategy used by an FL learner (Evans, Alshangiti 2018). From the vantage point of the SLM, the substitution implies that the FL learner either construes a mental association of the unfamiliar FL phoneme with the L1 phoneme or fails to establish the connection between them (Chan 2012; Flege 1995). It is inferred from the SLM that the novel FL sound could be linked by the FL learner to the closest L1 sound or sounds (Flege 1995).

In light of the above-mentioned factors, the SLM suggests that the learners' L1 exerts phonetic and phonological influences on the FL sound system (Amengual 2021; Flege 2009). In the SLM, the cross-linguistic influence that is associated with the learners' L1 is deemed to be a cause of the FL speech production with the so-called “foreign” accent that involves a range of pronunciation errors on the part of an EFL learner (Bjekić, Čubrović 2021; Marković 2020; Waniek-Klimczak 2008; Waniek-Klimczak *et al.* 2015). Additionally, the SLM assumes that the cross-linguistic phonetic and phonological influences are one of the reasons of “hearing with the

accent” (Amengual 2021), i.e. the compromised phonemic ability of the FL learners to perceive and process the FL-specific segmentals. It is inferred from the SLM that a compromised perception ability (in other words, hearing with the accent) is likely to be concomitant with a similarly compromised speaking ability, i.e. speaking with the accent. This contention is explained in the framework of the SLM-r by positing that segmental production and perception in the learner’s FL co-evolve owing to a bi-directional connection that is thought to exist between them (Flege *et al.* 2021; Flege, Bohn 2021).

In line with the SLM, it is assumed that EFL speakers whose L1s lack /z/ might experience challenges with its production and perception due to a variety of reasons (Flege, Hillenbrand 1986). Following the SLM-r, several variables could be involved in the compromised /z/ perception and production, for instance, EFL learners’ individual characteristics, the amount of EFL exposure, and inter-subject variability, to name just a few (Flege *et al.* 2021; Flege, Bohn 2021). The following section of the article provides a review of the prior literature that seeks to establish variables that could be the cause of EFL learners’ errors associated with /z/.

3. Literature Review

There is a cornucopia of previous publications that focus on EFL learners’ and speakers’ problems with /z/ (Broersma 2010; Bryła-Cruz 2021; Demirezen 2016; Flege, Hillenbrand 1986; Lersveen 2018; McAllister 2007; Rugæseter 2014; Roa *et al.* 2021; Zhang *et al.* 2021). The prior research literature focuses on i) the perception of the /s/ – /z/ contrast by EFL speakers and their English L1 controls (Broersma 2010; Flege, Hillenbrand 1986), ii) the perception of the /s/ – /z/ contrast by EFL speakers without the reference to the English L1 controls (McAllister 2007; Rugæseter 2014; Roa *et al.* 2021), iii) EFL speakers’ perception and production of English fricatives inclusive of /z/ (Demirezen 2016; Lersveen 2018; Zhang *et al.* 2021),

and iv) gender differences in the production and perception of /z/ (Bryła-Cruz 2021).

The perception of the /s/ – /z/ contrast by Swedish L1 and Finnish L1 EFL speakers on the one hand and the English L1 controls on the other hand is investigated by Flege and Hillenbrand (1986). They have found that whilst English L1 speakers rely on phonological cues, such as the duration of fricatives, in order to identify /z/, Swedish and Finnish participants, whose L1s do not possess a /s/-/z/ contrast, show no significant effect of fricative duration (Flege, Hillenbrand 1986: 513). Additionally, the Swedish L1 and Finnish L1 participants' phonological awareness and phonetic sensitivity to fricative duration as a cue to the English /s/ – /z/ contrast do not correlate with their exposure to the English language in the English-speaking countries (Flege, Hillebrand 1986: 514). Similar to Flege and Hillenbrand (1986), Broersma's (2010) attention is centred on the durational cues for final fricative discrimination in English by Dutch L1 EFL speakers, who are contrasted with a group of English L1 controls. Broersma (2010) observes that, unlike the English L1 controls, the Dutch EFL speakers do not rely on the durational cues in the perception of the English final fricative contrasts. She concludes that a durational cue for the L1 fricative contrast is insufficient for successful perception of the /s/ – /z/ contrast (Broersma 2010).

Whilst research design in Broersma (2010), as well as in Flege and Hillenbrand (1986), involves the English L1 controls, there are several studies (McAllister 2007; Rugæseter 2014; Roa *et al.* 2021) that focus on the production of /s/ – /z/ contrast by proficient EFL speakers without comparing them to the English L1 speakers. For instance, McAllister (2007) indicates that English voiced fricatives are poorly acquired by Swedish L1 advanced EFL speakers, who fail to produce /z/. Likewise, Rugæseter (2014) has established that less than five per cent of the Norwegian L1 participants in his study produce the /s/ – /z/ contrast systematically in a reading-aloud task in English. He notes that the majority of the Norwegian L1 participants

substitute /z/ for /s/ consistently in their speech production in English. Analogously to Norwegian and Swedish, there is no /z/ in Spanish. Subsequently, Spanish L1 EFL speakers do not exhibit phonological awareness of the /s/ – /z/ contrast and fail to produce /z/ in a variety of positions in the word, particularly in the word-final position (Roa *et al.* 2021).

Whereas the study by Roa *et al.* (2021) focuses exclusively on the /s/ – /z/ contrast, Demirezen (2016), Lersveen (2018), Zhang *et al.* (2021) analyse EFL speakers' perception and production of /z/ and other English fricatives. Specifically, Demirezen (2016) investigates Turkish L1 EFL students' problems with the production and perception of /z/, /θ/, and /ð/. He has found that /z/ does not pose a substantial problem to the Turkish L1 EFL students due to the presence of a similar phoneme in Turkish. In like manner, Zhang *et al.* (2021) explore the perception and production of /z/ and other English fricatives, such as /f/, /v/, /θ/, /ð/, /s/, /ʃ/, and /ʒ/, by Chinese L1 EFL cohorts on the beginner's and intermediate levels of proficiency. They posit that /z/ is problematic for Chinese L1 EFL learners, who typically substitute it for /s/ and/or /ts/ due to the absence of /z/ in the Chinese phonetic inventory. By means of examining the perception and production of the unvoiced-voiced pairs of alveolar stops /d/-/t/, alveolar fricatives /s/-/z/, postalveolar fricatives /ʒ/-/ʃ/ and affricates /dʒ/-/tʃ/, Lersveen (2018) concludes that both the perception and production of the /s/-/z/ contrast by Norwegian L1 EFL speakers is compromised due to the L1 input.

Gender differences in the production and perception of English consonants, inclusive of /z/, are explored by Bryła-Cruz (2021). She suggests that the /s/ – /z/ contrast is difficult to Polish L1 EFL speakers in the word-final position. Specifically, it is difficult for 20% of female and 37.5% male participants in her study. Bryła-Cruz (2021) argues that whilst there is a Polish counterpart of /z/, Polish L1 EFL speakers' problems with the perception of /z/ are associated with insufficient attention to vowel duration as a temporal parameter in the /s/ – /z/ contrast (Bryła-Cruz 2021: 130).

It follows from the current literature that research studies focus, primarily, on EFL learners' perception and production of /z/, which are analysed, mainly, through the lenses of the /s/ – /z/ contrast. The literature, however, does not seem to utilise the IPA transcriptions as a diagnostic means of investigating EFL learners' errors associated with /z/. Further, in Section 4, there is an outline of prior studies that employ phonemic transcription in the IPA in a variety of EFL contexts.

4. Phonemic Transcription in the IPA in EFL Contexts

The literature in applied linguistics and EFL studies suggests that phonetic alphabets, for instance, the IPA, are invaluable tools in EFL settings (Messerklinger 2009: 27) that play an important role in the teaching and learning of English pronunciation (Allegra 2018: 1). The IPA in EFL contexts is problematised in the literature as a form of metalanguage that enables EFL instructors, as well as students, to visualise, represent, and communicate about the peculiarities of the English phonological system by means of phonetic symbols (Lintunen 2005; Mompean 2017; Mompean, Fouz-González 2021; Sordelli *et al.* 2022; Trinh *et al.* 2022). In the current research studies, the IPA transcription is regarded “an umbrella term that is used to refer to several types of transcription” (Lintunen 2004: 27), such as i) phonetic transcription (also known as narrow transcription), which is employed to represent nuanced phonological differences and ii) phonemic transcription, or broad transcription, which is used to separate one phoneme from another without delving into phonological details (Lintunen 2013; Marshall 2020; Mompean 2015; Mompean, Fouz-González 2021; Sordelli *et al.* 2022; Trinh *et al.* 2022).

In a number of prior studies, the phonemic (broad) transcription in the IPA is employed as a diagnostic tool in i) assessing EFL learners' awareness of the English sounds, ii) perceptual training associated with the sound categories, iii) facilitating a general insight into the phonetic system of English and its main varieties (Atkielski 2005;

Cox, Fletcher 2017; Fouz-González, Mompean 2021; Lintunen 2013; Marshall 2020; Mompean 2015; Mompean, Fouz-González 2021; Sordelli *et al.* 2022; Trinh *et al.* 2022). The literature seems to share a contention that the application of the IPA transcription, mainly in its phonemic variant, has a beneficial effect on the teaching and learning process in an EFL classroom (Atkielski 2005; Mompean 2005; Mompean, Fouz-González 2021; Sordelli *et al.* 2022; Trinh *et al.* 2022). For instance, Mompean (2005) argues that

The use of phonetic symbols in foreign language teaching and learning is potentially very advantageous. Provided that the values of phonetic symbols are known and that the foreign language learner can produce and discriminate the sounds symbols stand for, these advantages include, among other things, increased awareness of L2 sound features, “visualisation” of such intangible entities as sounds, increased learner autonomy when checking pronunciation in dictionaries, etc. (p. 1).

As far as the “visualisation” (Mompean 2005: 1) of the English sounds in the IPA is concerned, it is argued to play a diagnostic role in identifying and understanding EFL learners’ pronunciation errors. The logic behind this argument is that when the learners record and transcribe their own speech, or, alternatively, EFL instructors do so, the static and visual IPA symbols enable the visualisation of the pronunciation errors and facilitate their correction (Atkielski 2005: 1). In relation to the pronunciation errors, Komar (2017: 162) posits that phonemic transcriptions are reflective of EFL students’ pronunciation errors that eventuate in their actual speech. Consequently, phonemic transcription could be seen as a diagnostic tool that is indicative of EFL learners’ actual performance (Atkielski 2005; Lintunen 2005; Messerklinger 2009; Mompean, Fouz-González 2021; Sordelli *et al.* 2022; Trinh *et al.* 2022).

It should be observed, however, that whilst the use of phonemic transcription is considered “very advantageous” (Mompean 2005: 1) in a variety of EFL contexts (Komar 2017; Lintunen 2013), it is not

commonly employed in the EFL teaching and learning at primary and secondary schools in Norway (Rugesæter 2012). Further, the article proceeds to the description of the Norwegian EFL contexts and the status of English in Norway.

5. The Status of English in Norway and an Outline of the EFL Teaching and Learning in Norwegian Contexts

The status of English in Norway is characterised by the notions of both prestige and necessity, given that English is widely used for educational, professional, and recreational purposes (Brevik, Hellekjær 2018; Kapranov 2019). Generally, Norwegians are considered to be highly proficient users of English, especially in terms of oral and conversational skills (Vold 2022). The high level of English proficiency by Norwegians is based upon several variables, such as the daily presence of the English language on Norwegian TV, extensive travel to the English-speaking countries, and the necessity to use English in order to communicate with foreign workers, refugees, and foreign tourists in Norway (Sunde, Kristoffersen 2018). In particular, English is “omnipresent in Norway’s written and audio-visual media and popular culture, and Norwegian youth often immerse themselves in leisure activities involving rich English input” (Sunde, Kristoffersen 2018: 280). The omnipresence of the English language in Norway has facilitated a view of the current status of English as hybrid (Hellekjær 2007; Rindal, Piercy 2013; Simensen 2005; Vold 2022). Specifically, it is argued that English in Norway is regarded as an L2 rather than an FL (Rindal, Piercy 2013: 212). However, it should be noted that English does not have the official status of an L2 there (Hellekjær 2007; Rindal, Piercy 2013). Its hybrid status is reflective of the current socio-linguistics context, where English is associated with an important part of everyday life (Eide 2021).

It could be argued that the status of English in Norway is commensurate with the context of the Norwegian EFL teaching and learning process. Due to the aforementioned hybrid status of English,

some EFL professionals suggest that English is taught in Norway less as an FL and more as an L1 (Simensen 2005). For instance, Norwegian L1, as well as English, are taught at primary school starting from the age of six (Brevik, Hellekjær 2018).

English is a compulsory subject at primary and lower secondary school in Norway. According to the curriculum requirements that are set by the Norwegian Ministry of Education, there are 138 teaching hours in Years 1–4 and 228 hours in Years 5–7 at primary school, whereas in Years 8–10 at lower secondary school there are 222 hours (Udir 2022a). From Year 1 onwards, the school subject of English involves the focus on such obligatory components as i) oral skills, ii) writing skills, iii) reading skills, and iv) digital skills (Udir 2022b). Given that English pronunciation as a part of oral skills is pivotal to the present study, let us note that the Norwegian Ministry of Education posits that “oral skills in English are to create meaning through listening, speaking and conversation” (Udir 2022b).

It follows from the description of the basic oral skills provided by the Norwegian Ministry of Education (Udir 2022b) that not much specific attention is paid to English pronunciation. However, by the end of secondary school a Norwegian EFL learner is expected “to use key patterns for pronunciation in communication” (Udir 2022c). Whilst there is a reference to English pronunciation in the competence aims, the Norwegian Ministry of Education does not describe what the key patterns of pronunciation are. In this regard, the prior literature indicates that “English pronunciation does not seem to play a central role in the development of communicative competences of a Norwegian L1 EFL learner” (Kapranov 2020: 73).

Taking into account the current context of EFL teaching and learning in Norway, it could be reasonable to assume that Norwegian EFL learners on the upper-intermediate level of proficiency might experience challenges with the pronunciation of the English sounds that have no analogues in their L1, Norwegian. Given that /z/ is absent in the phonological inventory of the Norwegian language, it remains to be elucidated whether or not the English fricative

consonant /z/ poses challenges to the group of participants, who are Norwegian L1 EFL learners on the B2 level of proficiency in English. That is further explored in the study, which is presented in Section 6 of the article.

6. The Present Study and Its Assumptions

From the vantage point of applied linguistics, the present study aimed at contributing to the existing body of knowledge about the acquisition of the English fricative /z/ by a cohort of EFL learners whose L1s did not have the equivalent fricative sound in their phonological repertoires (Flege, Hillenbrand 1986; McAllister 2007; Rug  seter 2014; Roa *et al.* 2021; Zhang *et al.* 2021). Specifically, the study focused on the group of participants on the upper-intermediate B2 level of EFL proficiency whose L1 was Norwegian. From a theoretical perspective, the study was informed by the tenets of the SLM and SLM-r. First of all, the study took into consideration the role of segmental production and perception of the FL sounds that were theorised to co-evolve and influence each other in the process of the FL category formation, which, according to the SLM-r, could take place regardless of the age of first exposure to an FL (Flege, Bohn 2021: 42). In addition, the study took into account the SLM-r tenet, which pointed to the non-linearity and inter-subject variability of an FL learner's phonetic performance (Flege, Bohn 2021). Importantly, however, the study factored in the SLM principle of the FL phonetic category formation that involved the FL learner's awareness of cross-language phonetic differences and ensuing establishment of perceptual links between L1 and FL sounds (Flege, Bohn 2021).

Yet, another theoretical and methodological consideration that was central in the study involved the prior research (Atkielski 2005; Lintunen 2005; Marshall 2020; Mompean 2005; Mompean, Fouz-Gonz  lez 2021; Sordelli *et al.* 2022; Trinh *et al.* 2022) which established that EFL students' transcriptions in the IPA were reflective of their pronunciation skills. In particular, the study factored in that

EFL students, “who were the best transcribers were also the ones whose pronunciation developed the most” (Lintunen 2005: 5). In line with the prior literature, phonemic transcriptions in the IPA were treated in the study as a diagnostic tool that allowed the identification of /z/-related mistakes made by the participants.

In light of the aforementioned theoretical and methodological backgrounds, it was assumed in the study that an EFL learner on the B2 level of proficiency according to CEFR (The Council of Europe 2011) would be aware of the English fricative consonant /z/ and, as posited by the SLM-r (Flege, Bohn 2021: 43), would be able to establish a perceived phonetic dissimilarity between /z/ and its closest Norwegian equivalent, the fortis fricative /s/. In other words, Assumption 1 was based upon the contention that the participants would not make any /z/-related mistakes in a set of phonemic transcriptions, in particular, they would not substitute /z/ for its Norwegian equivalent /s/.

Concurrently with Assumption 1, however, Assumption 2 was considered in the study. Assumption 2 rested on the participants’ possible lack of awareness of the English fricative consonant /z/ that could stem from i) the absence of /z/ and the /s/-/z/ contrast in the phonological system of Norwegian, the participants’ L1 (Flege, Hillenbrand 1986; Haugen 1967; Nilsen, Rugesæter 2015; Rugesæter 2014) and ii) insufficient attention to English pronunciation in the Norwegian EFL contexts (Bøhn, Hansen 2017; Kapranov 2020). All that, subsequently, would map onto the participants’ errors associated with /z/ in phonemic transcriptions in the IPA, where, according to the SLM (Flege 1995; Flege *et al.* 2021), the participants, potentially, would substitute /z/ for its Norwegian equivalent /s/. In line with the aforementioned Assumption 1 and Assumption 2, two RQs were formulated (see the introductory part of the article). Based upon the RQs, the specific aim of the study was to identify, quantify and classify /z/-related errors in the participants’ phonemic transcriptions in the IPA.

6.1. THE STUDY CONTEXT

The study was contextualised in the university course in English phonetics offered to pre-service EFL teachers at a university in Norway. The course consisted of two semesters of study (the autumn semester and the spring semester), which was organised around the topics in the course book *English Phonetics for Teachers* (Nilsen, Rugesæter 2015). The lectures and seminars in the course of English phonetics involved the topics from the course book by Nilsen and Rugesæter (2015) that are summarised in Table 1 below.

Table 1. *The Topics in the Course in English Phonetics*

N	Lecture/Seminar Topics	Semester
1	Chapter 1. Introduction Chapter 2. Sound Foundation (manner and force of articulation, phoneme and allophone, the syllable)	Autumn semester
2	Chapter 3. Consonants (stops, fricative, nasals, approximants)	Autumn semester
3	Chapter 4. Vowels (monophthongs, the weak vowels, diphthongs)	Autumn semester
4	Chapter 5. Stress, Rhythm, and Sounds in Company (word stress, sentence stress, weak forms, assimilation and elision)	Autumn semester
5	Chapter 6. Intonation (pitch, tones, the five tones in English)	Spring semester
6	Chapter 7. Teaching pronunciation (the teaching and learning of pronunciation, language practice)	Spring semester
7	Varieties of Spoken English (accent and dialect, variation in Britain, variation in the US, Australian English, pidgin and creoles, English as an international language)	Spring semester

As far as the teaching and learning content associated with /z/ was concerned, it was emphasised in the course book that

The importance of teaching the difference between /s/ and /z/ cannot be stressed too strongly, because a great many pairs of words are distinguished solely by the use of these two fricatives. Teachers will do their pupils a disservice if they do not insist on the correct pronunciation of the “inflectional -s”. (Nilsen, Rugesæter 2015: 46)

In addition to the theoretical and methodological considerations, the course book offered pronunciation practice (see Excerpt 1) that involved segmentals and important contrasts, such as the /s/ – /z/ contrast, which the participants were requested to listen to, practise and analyse. An example of pronunciation practice in the course book is provided in Excerpt (1) below.

(1) **Pronunciation practice 3.3**

- (a) soup, psalm, course, psychology, dismiss, cement;
- (b) zoo, amaze, raisin, please, misery, examine, possess;
- (c) lice – lies; face – phase; niece – knees [...] (Nilsen, Rugesæter 2015: 46)

In addition to pronunciation practice, the course in English phonetics had a strong focus on transcription exercises that involved individual words and sentences to be transcribed in phonemic transcription in the IPA, as illustrated by Excerpt 2.

(2) **Exercise 3.4. Transcribe the following words:**

Sparks, sits, smiles, busy, bus, matches, saves, chips, pears, conceal, goose, horse, mixes, amazes, glass [...]

Transcribe the following sentences:

- (a) Sarah’s husband was disturbed by his Swiss cousin’s singing.
- (b) Suzy received an offer as assistant manager.
- (c) We discovered all the cows grazing near the fancy swimming pool. [...] (Nilsen, Rugesæter 2015: 46)

It should be noted that the participants were introduced to the IPA at the beginning of the autumn semester, when it was used in order to familiarise them with the consonant sounds and a number of contrasts, such as /w/-/v/, /s/-/z/, that were deemed to be of critical

importance to Norwegian L1 EFL learners. Additionally, it should be observed that the participants worked with the phonemic transcription in the IPA on a routinely basis during the two semesters of study either by transcribing individual words, such as *sparks*, *sits*, *smiles*, etc. in Exercise 3.4. (see Excerpt 2) or transcribing the whole sentences, as in examples (a) – (c) in Excerpt 2.

The course in English phonetics was aimed, primarily, at pre-service EFL teachers, who were expected to be at the B2 level of proficiency in English. Given that the participants in the study were on the B2 level according to CEFR (The Council of Europe 2011), it would be pertinent to specify its competences associated with pronunciation. Below, Table 2 summarises the competencies in terms of the overall phonological control, sound articulation, and prosodic features that are expected to be mastered by an EFL learner on the B2 level of proficiency.

Table 2. *Descriptors of Pronunciation-Related Linguistic Competencies on the B2 Level according to CEFR (The Council of Europe 2011)*

N	CEFR Descriptor Scheme	Descriptor
1	Overall Phonological Control	Can generally use appropriate intonation, place stress correctly and articulate individual sounds clearly; accent tends to be influenced by the other language(s) they speak, but has little or no effect on intelligibility.
2	Sound Articulation	Can articulate a high proportion of the sounds in the target language clearly in extended stretches of production; is intelligible throughout, despite a few systematic mispronunciations.
3	Prosodic Features	Can employ prosodic features (e.g. stress, intonation, rhythm) to support the message they intend to convey, though with some influence from the other languages they speak.

6.2. PARTICIPANTS

The study involved 16 participants (11 females and 5 males, mean age = 22.3 y.o., standard deviation = 3.7), who were on the B2 level of EFL proficiency according to CEFR (The Council of Europe 2011). The participants' proficiency level was documented by their secondary school leaving certificates that stated that they had passed their English exams on the B2 level. All participants were enrolled in an EFL programme for pre-service EFL teachers at a university in Norway.

The participants' L1 was Norwegian and English was an FL to all of them. There were neither early balanced nor early sequential English/Norwegian bilinguals among the participants. None of the participants reported any knowledge of a third language. The participants' formal exposure to English started at the age of six at primary school. In addition, all participants informed the author of the article of their short stays in the English-speaking countries, predominantly, in the United Kingdom (the UK) and the United States of America, either as tourists or students at the Norwegian Centre in York (the UK). The mean duration of their stays in the English-speaking countries was two weeks per participant.

The participants were requested to sign a consent form that allows the author of the present article to process, analyse and publish their written data for scientific purposes. To ensure confidentiality, the participants' real names were coded. The following coding scheme was used in the study, e.g. P as in "participant" and the number (P1, P2, ... P16).

6.3. METHODOLOGY AND PROCEDURE

The study involved the following methodological considerations. In line with the prior literature (Atkielski 2005; Lintunen 2005; Marshall 2020; Mompean 2005; Mompean, Fouz-González 2021; Sordelli *et al.* 2022; Trinh *et al.* 2022), phonemic transcription in the

IPA was regarded as a diagnostic means of error identification as far as the participants' /z/-related mistakes were concerned. In light of the methodological approach adopted by Lintunen (2005), Mompean (2017), and Mompean and Fouz-González (2021), /z/-related mistakes made by the participants in the phonemic transcriptions were assumed to be indicative of the lack of awareness of /z/ and, consequently, its correct use in their speech production in English.

Two tasks, Task 1 and Task 2, that involved phonemic transcriptions in the IPA were used in the study. Task 1 was executed by the participants by the end of the autumn semester after they had been explicitly taught how to transcribe phonemically in the IPA. Task 2 was offered to the participants at the end of the spring semester. It should be noted that the participants had explicit instruction and practice in phonemic transcription in the IPA both in the autumn and in the spring semesters. In Task 1, as well as in Task 2, the participants were instructed to transcribe phonemically one short written text per task. Short texts for Task 1 and Task 2, respectively, were film plot synopses that were adapted from the web-site www.imdb.com (the Internet Movie Database, or IMDb). The reason for choosing the IMDb's plot synopses was accounted for in the prior literature (Kapranov 2019) that pointed to the successful application of feature films synopses to the execution of phonemic transcription tasks due to the generic and easily understandable summaries about popular feature films that, as a rule, were devoid of specialised vocabulary and aimed at the public at large. Whilst the short texts that were used in Tasks 1–2 originated from the IMDb's website, they were adapted by the author of the article in such a manner that each text contained words with the fricative consonant /z/. Specifically, there were six words that contained /z/ (two words with word-initial /z/, two words with word-medial /z/, and two words with word-final /z/) per each text in Task 1 and Task 2, respectively. It was ensured that no words that contained /z/ in Task 1 were repeated in Task 2. In addition, it should be observed that all six words with /z/ per task pertained to the frequently used lexical items that would be

typically encountered in stylistically neutral texts (for instance, *is*, *position*, *zoo*). The descriptive statistics of the tasks are summarised in Table 3.

Table 3. *The Descriptive Statistics of Task 1 and Task 2*

N	Descriptive Statistics	Task 1	Task 2
1	Total number of short texts to be transcribed per task	1	1
2	Total number of words per text	40	65
3	Total number of sentences per text	4	4
4	Total number of words that contained /z/ per text	6	6

The participants executed Tasks 1 and 2, respectively, at home and sent their phonemic transcriptions to the author of the article, who analysed them manually to identify /z/-related mistakes. Once the participants’ /z/-related errors were identified, they were entered into the Statistical Package for Social Sciences, or SPSS (IBM 2011) in order to compute means and standard deviations of the errors per group. It should be specified that in the discussion of the results the terms “/z/-related mistake/mistakes” and /z/-related error/errors” are used interchangeably.

6.4. RESULTS AND DISCUSSION

The results of the data analysis revealed that the majority of participants made /z/-related mistakes, specifically 75% of all participants in Task 1 and 87.5% of them in Task 2. Whilst none of the participants transcribed /z/ as an omission (i.e. no symbol instead of /z/), they, nevertheless, substituted /z/ for /s/ both in Task 1 and Task 2. No other types of substitution, for instance, /ʃ/ instead of /z/, was found in Tasks 1–2.

These findings are further discussed in the article through the prism of the RQs in the study. To reiterate, RQ 1 aims at establishing

whether or not the participants make any /z/-related mistakes in the phonemic transcriptions in the IPA, whereas RQ 2 seeks to discover whether or not /z/-related mistakes decrease, increase, or remain stable in Task 1 and Task 2.

6.4.1. THE DISCUSSION OF RQ 1

As previously mentioned, the majority of the participants made /z/-related mistakes in phonemic transcriptions in Tasks 1 – 2. The only type of mistakes involves the substitution of /z/ for /s/. This finding supports Assumption 2 in the study, which factors in the absence of /z/ and the /s/-/z/ contrast in the participants' L1 (Flege, Hillenbrand 1986; Haugen 1967; Nilsen, Rugesæter 2015; Rugesæter 2014) on the one hand and insufficient attention to pronunciation in the Norwegian EFL contexts (Bøhn, Hansen 2017; Kapranov 2020) on the other hand. Arguably, the participants' lack of awareness of /z/ maps onto the /z/-related errors in Task 1 (the total number (N) of /z/-related mistakes = 39) and in Task 2 (N of /z/-related mistakes = 27).

Given that there are 16 participants in the study and six occurrences of /z/ per Task, the total number of /z/-related mistakes, potentially, could be 96 in each task per group (16 participants multiplied by six errors = 96). It should be borne in mind that each task involves the maximum of six occurrences of /z/ that are represented by two occurrences in the word-initial position, two in the word-medial and two in the word-final positions. The comparison between the highest possible number of /z/-related mistakes (N = 96) and the actual number of errors associated with /z/ per group in Tasks 1–2 is emblematised by Figure 1 below.

Against the hypothetical number of /z/-related mistakes, i.e. 96 in each task per group, the total number of the actual /z/-related errors in Task 1 (i.e. 39) does not seem to be substantial. However, if we analyse the number of /z/-related errors per participant in each of the tasks, the error analysis reveals that only four participants out of 16 (i.e. 25%) have no mistakes associated with /z/ in Task 1. At

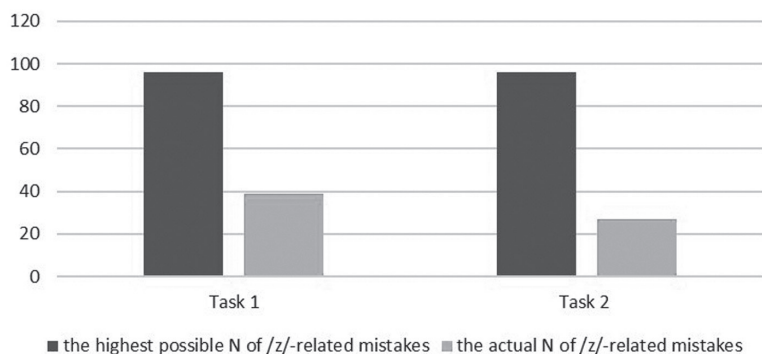


Figure 1. *The Total Number of Actual /z/-Related Mistakes per Group Compared to the Total Number of Potential /z/-Related Mistakes per Group in Tasks 1–2*

the same time, however, five participants (31.25%) stand out of the group by making the highest number of mistakes that involve the correct transcription of /z/. In particular, three participants (18.75%) have made the maximal number of /z/-related mistakes ($N = 6$) in Task 1, and two participants (12.5%) have 4 /z/-related errors each, whilst mean (M) mistake in total per group is 3.25, standard deviation (SD) = 1.83, in the same task. To reiterate, all /z/-related mistakes both in Task 1 and Task 2 involve the substitution of /z/ for /s/, whereas other types of mistakes associated with /z/, such as the omission of /z/ from the transcription or its substitution by other fricatives (e.g. /ʃ/) have not been identified in the error analysis. The distribution of /z/-related errors per participant in Task 1 is illustrated by Figure 2.

It is evident from Figure 2 that between the two opposite extremes of the participants with no /z/-related mistakes (25%) and the participants with the highest number of errors (31.25%) associated with /z/ there is a subgroup of participants with a lower number of mistakes; specifically four participants (25%) have made two mistakes each in Task 1. As shown in Figure 2, their mistakes are associated, predominantly, with substituting /z/ for /s/ in the word-final

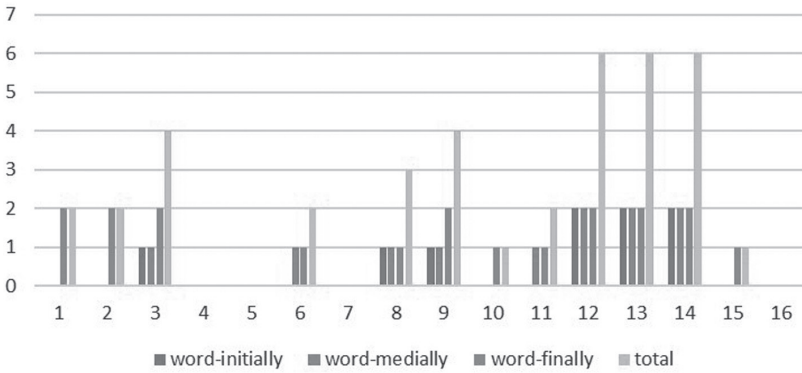


Figure 2. *The Distribution of /z/-Related Errors per Participant in Task 1*

position, represented in Task 1 by the words *lives* and *sisters*. Compared to Task 1, the distribution of the /z/-related errors seems to be similar as far as the mistakes in the transcription of /z/ in the word-final positions are concerned. This finding is graphically represented by Figure 3.

It is seen in Figure 3 that none of the participants has made the maximum number of /z/-related mistakes in Task 2, i.e. $N = 6$. The highest number of mistakes associated with /z/ and made by one participant (6.25%) is $N = 4$. That is followed by three participants (18.75%) with three /z/-related mistakes each, which are associated, mainly, with /z/ in the word-final position (see Figure 3). Otherwise, there is a substantial number of participants (37.5%) who have made only one mistake associated with /z/ in Task 2.

In terms of the relationship of the errors associated with /z/ and gender differences, the results of the error analysis in Task 1 indicate that three female participants and one male participant make no such mistakes, whereas the rest of them substitute /z/ for /s/ irrespective of gender. In Task 2, there are only two participants who have no /z/-related mistakes. These participants are female. Arguably, this finding provides indirect support to the previous literature

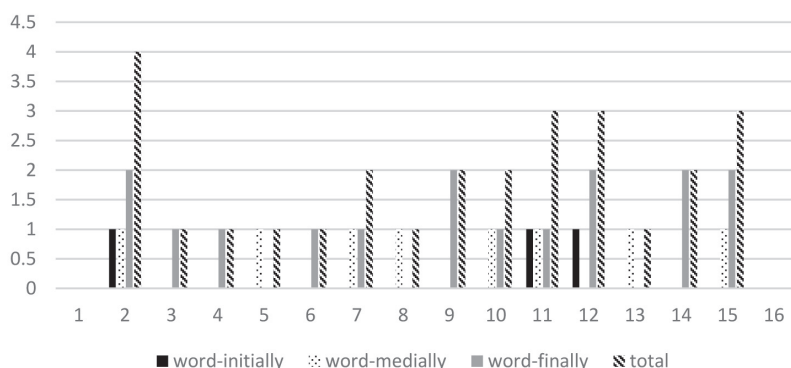


Figure 3. *The Distribution of /z/-Related Errors per Participant in Task 2*

(Bryła-Cruz 2021) which reports that female EFL learners make less /z/-related mistakes than their male counterparts. It should be mentioned, however, that the present study is not gender balanced with five males and 11 females. Hence, the gender-related findings in the study should be treated with caution.

Summarising RQ 1, it could be posited that whilst there is a decrease in the number of /z/-related mistakes from Task 1 to Task 2, the majority of participants, nevertheless, made them abundantly in the tasks. From a broad theoretical perspective, these findings corroborate the prior research (Haugen 1967; Lersveen 2018; Moen 1988; Nilsen 1989; Rugesæter 2014) that points to the substitution of /z/ for /s/ by Norwegian L1 EFL speakers. Presumably, the substitution of /z/ for /s/ by the participants can be accounted by the SLM and SLM-r (Flege 1995; Flege *et al.* 2021; Flege, Bohn 2021), which posit that EFL learners, and FL speakers generally, use the closest L1 phoneme instead of the FL-specific phoneme that is absent in their L1. Given that all /z/-related mistakes in the study involve only one type of errors, i.e. the substitution of /z/ for /s/, it could be argued that this mistake is not only recurrent in the tasks, but also typical to the group of participants. In terms of the typicality and consistency

of the type of /z/-related mistakes in the tasks, the present findings are in line with the SLM, which suggests that substitutions of a novel FL phoneme by the closest phoneme in the FL learner’s L1 constitutes a typical strategy (Evans, Alshangiti 2018).

6.4.2. THE DISCUSSION OF RQ 2

As previously indicated in the article, RQ 2 is concerned with a possible decrease or increase in /z/-related mistakes in Task 1 and Task 2. The error analysis shows that there is a decrease in the total number of /z/-related mistakes from Task 1 (N = 39) to Task 2 (N = 27). In addition, the error analysis indicates that the decrease involves the occurrence of /z/-related errors in all three positions of /z/ in the words that are used in the tasks, namely /z/ in the word-initial, word-medial, and word-final positions. These findings are presented in Table 4, below, in the form of means (M) and standard deviations (SD) per group in each task.

Table 4. /z/-Related Mistakes Made by the Participants in the IPA Transcription Tasks per Group

N	Types of the /z/-Related Mistakes	M and SD of /z/-Related Mistakes in the IPA Task 1	M and SD of /z/-Related Mistakes in the IPA Task 2
1	/s/ instead of /z/ word-initially	M 1.5 SD 0.5	M 1.0 SD 0.0
2	/s/ instead of /z/ word-medially	M 1.37 SD 0.48	M 1.0 SD 0.0
3	/s/ instead of /z/ word-finally	M 1.58 SD 0.49	M 1.45 SD 0.47

It follows from Table 4 that whilst the decrease in the errors from Task 1 to Task 2 is observed in the data, it does not seem to be substantial.

The observation is further supported by the statistical analysis. In particular, the application of the paired samples t-tests in SPSS (IBM 2011) to the data that are summarised in Table 4 has yielded no statistically significant results. Specifically, the difference between means in Task 1 and Task 2 is not significant at $p < .05$ as far as the word-initial errors are concerned, e.g. $t(15) = 1.65145$, $p = .054539$. Similarly, the word-medial errors (e.g. $t(15) = 0.79241$, $p = .217173$), as well as the word-final errors (e.g. $t(15) = 0.64253$, $p = .262705$) are not significant at $p < .05$. In other words, despite the observed decrease in the total number of /z/-related errors, the absence of the statistically significant differences in the distribution of /z/-related errors in Task 1 and Task 2 is indicative of the persistent nature of the /z/-related mistakes that the participants make even after they have had two semesters of training in the IPA transcription.

Whereas the total number of /z/-related mistakes per group decreases from Task 1 to Task 2, data analysis reveals that the number of participants who make them actually increases in Task 1 compared with Task 2. This finding is illustrated by Figure 4 below, where the total number of errors associated with /z/ per individual participant is plotted against the tasks.

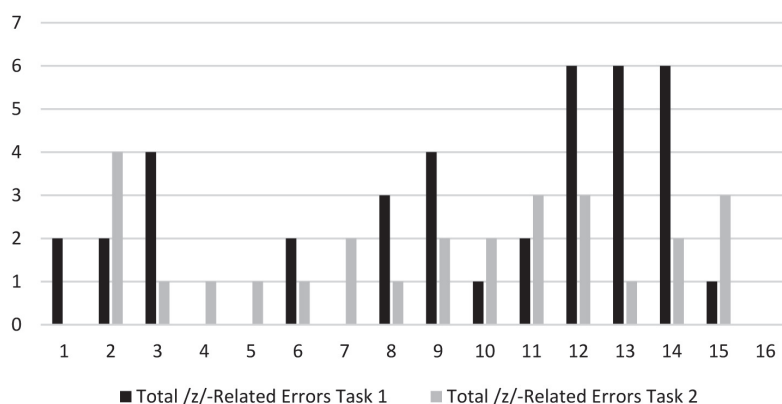


Figure 4. *The Comparison of Total /z/-Related Errors per Individual Participant in Tasks 1–2*

It follows from Figure 4 that the increase in the individual participants who make /z/-related mistakes in Task 2 (14 participants) in contrast with Task 1 (12 participants) involves Participants 4 and 5, whose phonemic transcriptions in Task 1 are error-free as far as the transcription of /z/ is concerned. It may seem paradoxical that, concurrently with the decrease in the mean /z/-related errors, Task 2 is characterised by the increase in the participants who continue to make /z/-related mistakes. Put differently, we observe the participants' non-linear performance in the Tasks 1–2. It could be assumed that the participants' non-linear or, perhaps, unstable performance in the tasks is indicative of their lack of awareness of /z/, its production and perception. In this regard, these findings support the literature (Bryła-Cruz 2021; Flege, Hillenbrand 1986; McAllister 2007; Roa *et al.* 2021; Zhang *et al.* 2021) which indicates that advanced EFL learners whose L1s do not possess /z/ in their phonological inventories fail to exhibit awareness of /z/. To reiterate, the present study employs phonemic transcriptions in the IPA as a diagnostic tool to establish whether or not the participants know how to transcribe /z/ correctly. Consequently, errors in the transcription signal about the participants' insufficient awareness of /z/. Given that the participants invariably transcribe /z/ by substituting it for /s/, it could be suggested that the participants' category formation for /z/ is compromised. In line with the SLM-r (Flege *et al.* 2021; Flege, Bohn 2021), the substitution of /z/ for /s/ in the present study illustrates the contention that if an FL phonetic category is not properly formed, a composite L1-L2 phonetic category may be developed on the basis of the closest L1 phoneme (Flege, Bohn 2021: 42).

Arguably, the persistence of the substitution of /z/ for /s/ in Task 2 cannot be explained by the participants' individual differences, since this mistake is made by the majority of them. Another variable that should be factored out involves the participants' sojourns abroad in English-speaking countries. In this regard, the author of the article concurs with Flege and Hillenbrand (1986), Lersveen (2018) and McAllister (2007), who point to the absence of positive

gains associated with the FL exposure in English-speaking countries as far as the acquisition of the /s/-/z/ contrast is concerned. Furthermore, in light of substantial everyday exposure to English in Norway (Brevik, Hellekjær 2018; Kapranov 2019; Sunde, Kristoffersen 2018), a stay abroad as a means of getting exposure to English appears less relevant in the Norwegian EFL contexts (Lersveen 2018; Vold 2022).

Presumably, the main variable that seems to be involved in the participants' /z/-related mistakes could be associated with phonetic factors. Following the SLM-r (Flege *et al.* 2021; Flege, Bohn 2021), phonetic factors are deemed to be the primary force behind the formation or non-formation of a new phonetic category for an FL sound (Flege, Bohn 2021: 42). In particular, the SLM-r posits that the phonetic factors related to the FL category formation involve the “degree of perceived phonetic dissimilarity from the closest L1 sound, and the precision with which the closest L1 category is specified when L2 learning begins” (Flege, Bohn 2021).

In conclusion to the discussion of RQ 2, it appears possible to consider the following. Notwithstanding that the number of /z/-related errors decreases in Task 2 in contrast to Task 1 (see the means summarised in Table 4), the number of participants who make /z/-related mistakes increases in Task 2. This finding suggests that the variables of continuous learning, sojourns abroad and other forms of L2 exposure as posited in the SLM-r (Flege, Bohn 2021) are not sufficient to override the phonetic factors at hand, namely the absence of /z/ and the /s/-/z/ contrast in the participants' L1, Norwegian.

7. Conclusions and Linguo-Didactic Implications

The study focuses on the English fricative consonant /z/ as a challenge to the group of EFL university students on the B2 level of proficiency. Given that scholarly attention to the upper-intermediate cohorts of EFL learners on the B2 level of proficiency is underrepresented in the literature (Raeisi-Vanani, Baleghizadeh 2022), the

study might offer a novel insight into /z/ as a challenge to the study participants whose L1 does not include this fricative consonant in its phonological inventory. Specifically, the study demonstrates that on the B2 level of EFL proficiency there is insufficient awareness of /z/, which is evident from multiple /z/-related mistakes in phonemic transcriptions in Task 1 and Task 2.

It should be emphasised that the study has demonstrated how phonemic transcription can be successfully employed as a diagnostic tool in identifying the participants' /z/-related errors. Specifically, the error analysis of the participants' phonemic transcriptions has revealed that they make the typical and persistent mistake of substituting /z/ for /s/ in Tasks 1–2. The error analysis of the participants' transcriptions indicates that whilst the substitution of /z/ for /s/ in the word-initial and word-medial positions declines in Task 2, it still persists in the word-final positions in this task. It could be concluded that the present findings are indicative of the participants' compromised phonological awareness of /z/.

Whilst the present findings provide an addition to the prior literature (Flege, Hillenbrand 1986; Haugen 1967; Lersveen 2018; Rugesæter 2014) and offer novel avenues to explore, the study involves several shortcomings that should be remedied in the subsequent research. Specifically, the study would benefit from the recordings of the participants' spontaneous and semi-prepared speech in English. The recordings should be analysed in conjunction with the participants' phonemic transcriptions in order to arrive at a broader picture associated with their use of English fricatives, inclusive of /z/ in their actual speech. Additionally, the study would benefit from the participants' reflections concerning their awareness of /z/, the /s/ – /z/ contrast in English and the possible reasons that might compromise their perception and production of /z/.

Arguably, the present study is relevant not only to EFL students whose L1 is Norwegian, but also to other cohorts of EFL learners whose L1s lack /z/ and the /s/-/z/ contrast, for instance, Finnish, Swedish, and Thai. The findings in the study are indicative of

the following linguo-didactic suggestions that could be applied to a variety of EFL teaching and learning contexts. First, given that /z/-related errors are persistent on the B2 level of EFL proficiency, it appears reasonable to incorporate pronunciation instruction on this level of proficiency (Metruk 2017: 15). Second, EFL learners on the B2 level of proficiency whose L1 backgrounds lack /z/ should pay specific attention to /z/ and the /s/ – /z/ contrast in the English language. Third, EFL learners on the B2 level of proficiency should be encouraged to use the IPA transcription as a (self)-diagnostic tool in assessing one's potential problems associated with English pronunciation. Fourth, EFL students whose L1s do not have /z/ in their phonological inventories should be taught the English fricatives explicitly in a systematic manner.

REFERENCES

- Allegra, Mary 2018. Role of phonetic alphabets for teaching pronunciation. – The TESOL Encyclopedia of English Language Teaching. Ed. John I. Lontas. London: Wiley Blackwell, 1–7. <https://doi.org/10.1002/9781118784235.eelt0867>.
- Amengual, Mark 2021. Exploring cross-linguistic effects and phonetic interactions in the context of bilingualism: Introducing the special issue. – *Languages* 6, 1, 1–7. <https://doi.org/10.3390/languages6010054>.
- Atkielski, Anthony 2005. Phonetic transcription can be a useful tool for teaching or correcting pronunciation in the ESL/EFL classroom. – *Using Phonetic Transcription in Class*, 1–12. https://www.academia.edu/9750926/Phonetics_Using_Phonetic_Transcription_in_Class.
- Bech, Kristin; Walkden, George 2016. English is (still) a West Germanic language. – *Nordic Journal of Linguistics* 39, 1, 65–100. <https://doi.org/10.1017/S0332586515000219>.
- Best, Catherine T.; Tyler, Michael 2007. Nonnative and second-language speech perception: Commonalities and complementarities. – *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Eds. Ocke-Schwen Bohn, Murray J. Munro. Amsterdam: John Benjamins Publishing, 13–34.

- Bjekić, Ana; Čubrović, Biljana 2021. Acoustic characteristics of American English monophthongs in Serbian EFL Speakers – A case study. – *Philologia* 19, 1, 65–84. <https://doi.org/10.18485/philologia.2021.19.19.6>.
- Bjelaković, Andrej; Biljana Čubrović. 2021. Attitudes regarding the realisation of the lot vowel in advanced Serbian EFL speakers. – *Belgrade English Language and Literature Studies* 13, 1, 11–28. <https://doi.org/10.18485/bells.2021.13.1>.
- Bøhn, Henrik; Hansen, Thomas 2017. Assessing pronunciation in an EFL context: Teachers' orientations towards nativeness and intelligibility. – *Language Assessment Quarterly* 14, 1, 54–68. <https://doi.org/10.1080/15434303.2016.1256407>.
- Brevik, Lisbeth M.; Hellekjær, Glenn Ole 2018. Outliers: Upper secondary school students who read better in the L2 than in L1. – *International Journal of Educational Research* 89, 80–91. <https://doi.org/10.1016/j.ijer.2017.10.001>.
- Broersma, Mirjam 2010. Perception of final fricative voicing: Native and non-native listeners' use of vowel duration. – *The Journal of the Acoustical Society of America* 127, 3, 1636–1644. <https://doi.org/10.1121/1.3292996>.
- Bryła-Cruz, Agnieszka 2021. The gender factor in the perception of English segments by non-native speakers. – *Studies in Second Language Learning and Teaching* 11, 1, 103–131. <http://doi.org/10.14746/ssllt.2021.11.1.5>.
- Chan, Alice 2012. Cantonese English as a second language learners' perceived relations between “similar” L1 and L2 speech sounds: A test of the speech learning model. – *The Modern Language Journal* 96, 1, 1–19. <https://doi.org/10.1111/j.1540-4781.2012.01291.x>.
- Cox, Felicity; Fletcher, Janet 2017. – *Australian English Pronunciation and Transcription*. Cambridge: Cambridge University Press.
- Demirezen, Mehmet 2016. Perceptual identification and perception of sibilants of English language by Turkish English majors. – *Procedia-Social and Behavioral Sciences* 232, 750–758. <https://doi.org/10.1016/j.sbspro.2016.10.102>.
- Eide, Emily 2021. Meir engelsk i språket: – Det er nesten umogeleg å ikkje bruke engelsk. – NRK, November 11. <https://www.nrk.no/kultur/engelsk-blir-ein-storre-del-av-daglegtalen-til-nordmenn-1.15726515> (24.10.2022).

- Evans, Bronwen G.; Alshangiti, Wafaa 2018. The perception and production of British English vowels and consonants by Arabic learners of English. – *Journal of Phonetics* 68, 15–31. <https://doi.org/10.1016/j.wocn.2018.01.002>.
- Flege, James Emil 1995. Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* 92, 233–277.
- Flege, James Emil 2005. Origin and development of the Speech Learning Model. – Talk presented at the 1st ASA Workshop on L2 Speech Learning, Simon-Fraser University, Vancouver, BC. April 14–15. http://www.jimflege.com/files/SLMvancouver_updated.pdf.
- Flege, James Emil 2009. Give input a chance! – *Input Matters*. Eds. Thorsten Piske, Martha Young-Scholten. Clevedon: Multilingual Matters, 175–190.
- Flege, James Emil; Aoyama, Katsura; Bohn, Ocke-Schwen 2021. The revised speech learning model (SLM-r). – *Second Language Speech Learning: Theoretical and Empirical Progress*. Ed. Ratree Wayland. Cambridge: Cambridge University Press, 84–118.
- Flege, James Emil; Bohn, Ocke-Schwen 2021. The revised Speech Learning Model (SLM-r). – *Second language speech learning: Theoretical and empirical progress*. Ed. Ratree Wayland. Cambridge: Cambridge University Press, 3–83. <https://doi.org/10.1017/9781108886901.002>.
- Flege, James Emil; Hillenbrand, James 1986. Differential use of temporal cues to the /s-/ /z/ contrast by native and non-native speakers of English. – *The Journal of the Acoustical Society of America* 79, 2, 508–517. <https://doi.org/10.1121/1.393538>.
- Fouz-González, Jonás; Jose A. Mompean 2021. Phonetic symbols vs keywords in perceptual training: the learners' views. – *ELT Journal* 75, 4, 460–470. <https://doi.org/10.1093/elt/ccab037>.
- Haugen, Einar 1967. The Prosodic System of Norwegian. – *Scandinavian Studies* 39, 1, 47–51.
- Hellekjær, Glenn Ole 2007. Fremmedspråk i norsk næringsliv – engelsk er ikke nok! – *Fokus på språk*. Eds. Gerard Doetjes, Beate Trandem. Halden: Fremmedspråksenteret, 6–80.

- Huensch, Amanda; Thompson, Amy S. 2017. Contextualizing attitudes toward pronunciation: Foreign language learners in the United States. – *Foreign Language Annals* 50, 2, 410–432. <https://doi.org/10.1111/flan.12259>.
- IBM 2011. IBM SPSS statistics for Windows, version 20.0. – New York: IBM Corp.
- Kallio, Heini; Suni, Antti; Šimko, Juraj 2021. Fluency-related temporal features and syllable prominence as prosodic proficiency predictors for learners of English with different language backgrounds. – *Language and Speech*, 1–27. <https://doi.org/10.1177/00238309211040175>.
- Kanokpermpoon, Monthon 2007. Thai and English consonantal sounds: A problem or a potential for EFL learning. – *ABAC Journal* 27, 1, 57–66.
- Kapranov, Oleksandr 2020. Intermediate EFL Students' Self-Assessment of Phonetically Difficult Words in English. – *Ostrava Journal of English Philology* 12, 2, 69–91. <https://doi.org/10.15452/OJoEP.2020.12.0011>.
- Kapranov, Oleksandr 2019. Self-Assessment of the sounds of the English language that pre-service EFL teachers consider problematic to pronounce. – *ELOPE: English Language Overseas Perspectives and Enquiries* 16, 1, 77–99. <https://doi.org/10.4312/elope.16.1.77-99>.
- Kapranov, Oleksandr 2014. Syntactic performance in online written discourse by an English/Swedish bilingual with Asperger's syndrome: A Case Study. – *East European Journal of Psycholinguistics* 1, 1, 74–80.
- Komar, Smiljana 2017. The relationship between the perception and production of four General British vowels by Slovene university students of English. – *Linguistica* 57, 1, 161–170. <https://doi.org/10.4312/linguistica.57.1.161-170>.
- Lersveen, Liv-Randi 2018. The perception and production of nonnative English consonants in native Norwegian speakers. MS thesis, NTNU. <https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2574550/Masteroppgave,%20Liv-Randi%20Lersveen.pdf?sequence=1> (24.10.2022).
- Lintunen, Pekka 2013. The Effect of Phonetic Knowledge on Evaluated Pronunciation Problems. – *Proceedings of the Phonetics Teaching and Learning Conference*. Eds. Joanna Przedlacka, John A. Maidment, and Michael Ashby. London: University College London Press, 55–58.
- Lintunen, Pekka 2005. Phonemic transcription and its effect on learning. – *Proceedings of the Phonetics Teaching and Learning Conference 2005*.

- Ed. John A. Maidment. London: London: University College London Press, 1–4.
- Lintunen, Pekka 2004. Pronunciation and Phonemic Transcription: A Study of Advanced Finnish Learners of English. Turku: University of Turku.
- Marković, Maja 2020. Phonetic feature interpretation in second language phonology: The acquisition of the final obstruent voicing contrast. – *Belgrade English Language and Literature Studies* 12, 1, 27–49. <https://doi.org/10.18485/bells.2020.12.2>.
- Marshall, Todd Alden 2020. Seven secrets to improving pronunciation. – *Approaches to English for Specific and Academic Purposes: Perspectives on Teaching and Assessing in Tertiary and Adult Education*. Eds. Michal Ennis, Jemma Prior. Bolzano: Bolzano University Press, 193–121.
- McAllister, Robert 2007. Strategies for realization of L2-categories: English /s/ – /z/. – *Language Experience in Second Language Speech Learning*: In Honor of James Emil Flege. Eds. Ocke-Schwen Bohn, Murray J. Munro. Amsterdam: John Benjamins Publishing Company, 153–166.
- McAllister, Robert; Flege, James E.; Piske, Thorsten 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. – *Journal of Phonetics* 30, 2, 229–258.
- Messerklinger, Josef 2009. Teaching English with the IPA. – *CELE Journal* 19, 3, 27–33.
- Metruk, Rastislav 2017. Pronunciation of English dental fricatives by Slovak university EFL students. – *International Journal of English Linguistics* 7, 3, 11–16. <https://doi.org/10.5539/ijel.v7n3p11>.
- Moen, Per 1988. The English pronunciation of Norwegian-Americans in four Midwestern states. – *American Studies in Scandinavia* 20, 2, 105–121.
- Mompeán, José A. 2017. Doing phonetic transcription in a modern language degree. 1 – *Estudios de filología inglesa: Homenaje al profesor Rafael Monroy*. Murcia: EditUM, 479–505.
- Mompean, José A. 2015. Phonetic notation in foreign language teaching and learning: Potential advantages and learners' views. – *Research in Language* 13, 3, 292–314. <https://doi.org/10.1515/rela-2015-0026>.
- Mompean, José A. 2005. Taking advantage of phonetic symbols in the foreign language classroom. – *Proceedings of the phonetics teaching and learning conference*. London, UK, 27–30 July 2005. Ed. John A. Maidment. London: University College London, 1–4.

- Mompean, Jose A.; Fouz-González, Jonás 2021. Phonetic symbols in contemporary pronunciation instruction. – *RELC Journal* 52, 1, 155–168. <https://doi.org/10.1177/0033688220943431>.
- Munro, Murray J.; Bohn, Ocke-Schwen 2007. The study of second language speech. – *Language Experience in Second Language Speech Learning*. – In Honor of James Emil Flege. Eds. Ocke-Schwen Bohn, Murray J. Munro. Amsterdam: John Benjamins Publishing Company, 3–11.
- Nilsen, Thor S. 1989. Do You/Ond'taend? An Error Analysis of Norwegian Students' Pronunciation of English. – <https://eric.ed.gov/?id=ED321585>.
- Nilsen, Thor S.; Rugesæter, Kåre Per Nitter 2015. *English Phonetics for Teachers*. Bergen: Fagbokforlaget.
- Raeisi-Vanani, Amin; Baleghizadeh, Sasan 2022. The contributory role of grammar vs. vocabulary in L2 reading: An SEM approach. – *Foreign Language Annals*. <https://doi.org/10.1111/flan.12606>.
- Rindal, Ulrikke; Piercy, Caroline 2013. Being 'neutral'? English pronunciation among Norwegian learners. – *World Englishes* 32, 2, 211–229. <https://doi.org/10.1111/weng.12020>.
- Roa, Leonardo Contreras; Mairano, Paolo; Bouzon, Caroline; Capliez, Marc 2021. The acquisition of /s/ –/z/ in a phonemic vs neutralised context: comparing French L1, Italian L1 and Spanish L1 learners of L2 English. – XVII AISV Conference, Feb 2021, Zurich (online). Switzerland. <https://hal.archives-ouvertes.fr/hal-03343042/document> (24.10.2022).
- Rojczyk, Arkadiusz 2010. Temporal and spectral parameters in perception of the voicing contrast in English and Polish. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Rugesæter, Kåre Per Nitter 2014. Difficult contrasts: an analysis of phonemic distinctions in the English of young Norwegian learners seen against the backdrop of incidental foreign language learning. – *Acta Didactica Norge. Tidsskrift for fagdidaktisk forsknings- og utviklingsarbeid i Norge* 8, 1, 1–20.
- Rugesæter, Kåre Per Nitter 2012. Uttaleundervisning i engelsk–nye tider, nytt fokus. – *Communicare* 2, 56–60.
- Simensen, Aud Margit 2005. Ja, engelsk er noe vi møter som barn og det er ikke lenger et fremmedspråk. – *Sprogforum-tidsskrift for sprog-og kulturpædagogik* 35, 57–61.

- Sordelli, M. Laura, Susana Chiatti, Sofía Romanelli 2022. "EFL teachers don't stop to teach pronunciation": an interpretation of learner beliefs on pronunciation learning at an English teacher preparation programme. – *Entramados: Educación y Sociedad* 9, 11, 214–231.
- Strange, Winifred 1992. Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. – *Speech Perception, Production and Linguistic Structure*. Eds. Yohichi Tohkura, Eric Vatikiotis-Bateson, Yoshinori Sagisaka. Tokyo: Ohmsha IOS Press, 197–219.
- Sunde, Anne Mette; Kristoffersen, Martin 2018. Effects of English L2 on Norwegian L1. – *Nordic Journal of Linguistics* 41, 3, 275–307. <https://doi.org/10.1017/S0332586518000070>.
- Szyska, Magdalena 2017. Pronunciation Learning Strategies and Language Anxiety. In *Search of an Interplay*. Cham: Springer. <https://doi.org/10.1007/978-3-319-50642-5>.
- The Council of Europe 2011. Common European Framework of Reference for Languages: Learning, Teaching, Assessment, 2011. https://www.coe.int/en/web/language-policy/home?e1_en.asp.
- Trinh, Quoc Lap; Nguyen, Thi Dieu Lien; Le, Thanh Thao 2022. Using explicit instruction of the International Phonetic Alphabet system in English as a Foreign Language adult classes. – *European Journal of Educational Research* 11, 2, 749–761. <https://doi.org/10.12973/eu-jer.11.2.749>.
- Udir 2022a. Timetall. <https://www.udir.no/lk20/eng01-04/timetall> (24.10.2022).
- Udir 2022b. Grunnleggende ferdigheter. <https://www.udir.no/lk20/eng01-04/om-faget/grunnleggende-ferdigheter> (24.10.2022).
- Udir 2022c. Kompetansemål og vurdering. <https://www.udir.no/lk20/eng01-04/kompetansemaal-og-vurdering/kv4> (24.10.2022).
- Van Dommelen, Wim A.; Hazan, Valeri 2010. Perception of English consonants in noise by native and Norwegian listeners. – *Speech Communication* 52, 11–12, 968–979. <https://doi.org/10.1016/j.specom.2010.05.001>.
- Verschik, Anna 2017. Language Contact, Language Awareness, and Multilingualism. – *Language Awareness and Multilingualism, Encyclopedia of Language and Education*, 3rd ed. Eds. Jasone Cenoz, Durk Gorter, Stephen May. Cham: Springer, 2–13. https://doi.org/10.1007/978-3-319-02325-0_21-1.

- Vold, Eva Thue 2022. Learner spoken output and teacher response in second versus foreign language classrooms. – *Language Teaching Research*, 1–35. <https://doi.org/10.1177/13621688211068610>.
- Waniek-Klimczak, Ewa 2011. Aspiration in Polish: A sound change in progress? – *New Perspectives in Language, Discourse and Translation Studies*. Eds. Mirosław Pawlak, Jakub Bielak. Berlin: Springer, 3–11.
- Waniek-Klimczak, Ewa 2008. Variability in Accents of English: A Non-Native Speaker Perspective. – *Issues in Accents of English*. Ed. Ewa Waniek-Klimczak Newcastle upon Tyne: Cambridge Scholars Publishing, vii–viii.
- Waniek-Klimczak, Ewa; Klimczak, Karol 2005. Target in speech development: learners' views. – *English Pronunciation Models: A Changing Scene*. Eds. Katarzyna Dziubalska-Kołaczyk, Joanna Przedlacka. Bern: Peter Lang, 229–249.
- Waniek-Klimczak, Ewa; Rojczyk, Arkadiusz; Porzuczek, Andrzej 2015. 'Polglish' in Polish eyes: What English studies majors think about their pronunciation in English. – *Teaching and Researching the Pronunciation of English*. Eds. Ewa Waniek-Klimczak, Mirosław Pawlak. Cham: Springer, 23–34. https://doi.org/10.1007/978-3-319-11092-9_2.
- Zhang, Buyi; Zhang, Jiaqi; Lee, Sook-Hyang 2021. Perception and production of English fricatives by Chinese learners of English: Error patterns and perception-production relationship. – *Phonetics and Speech Sciences* 13, 1, 25–36. <https://doi.org/10.13064/KSSS.2021.13.1.025>.

INGLISE FRIKATIIV /Z/ KUI VÄLJAKUTSE NORRA EMAKEELEGA INGLISE KEELE ÕPPIJATELE: FONEEMILISTE TRANSKRIPTSIOONIDE VEAANALÜÜS

Oleksandr Kapranov
NLA University College Oslo

Inglise keele frikatiivid, näiteks /z/, põhjustavad arvatavalt raskusi inglise keele kui võõrkeele õppijatele, kelle esimesed keeled (K1) on frikatiivide poolest inglise keelest teistsuguse foneetilise repertuaariga (Kallio, Suni, Šimko 2021; Kanokpermpoon 2007). Kuna norra keele foneetilises süsteemis ei ole frikatiivi /z/, mõjutab selle puudumine uurimuste kohaselt negatiivselt norra emakeelega inglise keele õppijate kõneproduktiooni (Rugesæter 2014). Selle uurimuse eesmärk on analüüsida inglise frikatiivi /z/ võimalikke raskuseid, mida kogevad norra emakeelega inglise keele õppijad (edaspidi *osalejad*) B2 tasemel CEFR-i järgi (The Council of Europe 2011). Selleks paluti osalejatel teha seeria foneemilisi transkriptsioone rahvusvahelises foneetilisel tähestikus (IPA), üks transkriptsioon sügissemestri lõpus (ülesanne 1), teine kevadsemestri lõpus (ülesanne 2). Foneemilisi transkriptsioone kasutati siin diagnostiliselt (Lintunen 2005; Fouz-González, Mompean 2021), et määrata nende /z/ tundmist ja kaudselt nende /z/ kasutust. Osalejate transkriptsioonide veaanalüüs näitas, et enamik neist tegi /z/ asendamisel /s/ häälikuga vigu. Arvestades, et /z/ asendamine /s/ häälikuga oli ülesandes 1 tüüpiline ning esines ka ülesandes 2, võib öelda, et kõrgema kesktaseme õppijate jaoks oli /z/ keeruline. Keelelis-didaktiliste implikatsioonide üle arutleti artiklis.

Võtmesõnad: inglise keel võõrkeelena, inglise frikatiivide /z/, /s/ – /z/ kontrast, kõrgema kesktaseme inglise kui võõrkeele õppijad, norra keel K1

Oleksandr Kapranov's research interests involve academic writing, cognitive linguistics and psycholinguistics.
oleksandr.kapranov@nla.no

ACQUISITION OF RUSSIAN NOMINAL DERIVATION IN MONOLINGUALISM AND BILINGUALISM

Victoria V. Kazakovskaya

Institute for Linguistic Studies, Russian Academy of Sciences

Abstract. The case study focuses on the acquisition of Russian derivational morphology in terms of nouns by monolingual (Russian) and simultaneous bilingual (Russian-German) children of early age. The results are based on analysis of representative natural longitudinal recordings transcribed and stored in CHAT format using the CHILDES system. The first patterns and methods of nominal word-formation along with the morphemes used by children are revealed. The properties of word-formation that indicate the productive use of the nominal derivatives, such as the presence of *simplex-derivative* pairs, chains and word families, as well as occasionalisms are noted. The similarities and differences in the acquisition of nominal derivatives, including their semantic domains, in mono- and bilingual situations are discussed.

Keywords: nouns, word-formation, derivatives, simplexes, compounding, Russian, German

1. Introduction

The article highlights early development of nominal derivation in Russian. The results of a comparative analysis of this process in monolingual (Russian) and bilingual (Russian and German) situations are discussed. Despite the fact that the features of bilingual (and more broadly, multilingual) development of children are currently being intensively studied all over the world (e.g. Bayram *et al.* 2018; De Houwer, Ortega 2019 among others), the grammatical aspects and, especially, the derivational morphology have still been

insufficiently investigated. The early natural simultaneous “one parent – one language” bilingualism also remains almost undescribed (De Houwer 1990; Döpke 1992, 2000). In addition, in recent years researchers into the bi- and multilingual language acquisition have increasingly mentioned the need to take into account the individual peculiarities in each case and therefore appealed for collection of new data of child speech (CS).

Russian-German bilingualism, the one we are focusing on, is of particular interest as it combines the grammatical systems of, first, a morphologically rich language and a morphologically poor one¹ and second, a language with a predominantly affixal system of word-formation (typical for Slavic languages) and a language with a predominantly compound system of word-formation (typical for a number of Germanic languages). Thus, the paper touches upon the problem of interference, i.e. the influence of how the typological features of one language learned by a child are reflected in the acquisition of this fragment of the language system in another one. Simultaneous bilingualism provides some opportunities for such observations. Simultaneous bilingual children, having “a remarkable ability to differentiate their two languages from early in development”, show “signs of cross-linguistic influence, or processing their two languages in ways that show influence from the other language” (Nicoladis 2018: 81, see also De Houwer 1990; Döpke 1992, 2000).

The last question involves the use of recent results obtained by an analysis of the acquisition of affixation and compounding by monolingual children in typologically different languages (e.g. Dressler *et al.* 2017; Mattes *et al.* 2021, see also Argus, Kazakovskaya 2013 and

¹ The concept of morphological richness has been developed in Dressler (1999 among his others). Its importance for language acquisition was demonstrated in, e.g. Xanthos and Gillis (2010); Xanthos *et al.* (2011); Savickienė and Dressler (2007), Dressler *et al.* (2017, 2022). Based on the data of inflectional morphology, compounding and such derivatives as diminutives, it is shown that “a greater richness of the morphological structure facilitates acquisition” (Dressler *et al.* 2022).

Kazakovskaya, Argus 2018 for the mostly agglutinating Estonian and the inflectional Russian). Based on them we can assume the presence of a greater number of compounds in the Russian speech production of a bilingual child acquiring German than in the speech of a Russian-speaking monolingual child, as well as other possible manifestations of interference. Until recently, the Russian-German type of child bilingualism was considered to have been poorly studied when compared to other bilingual pairs (for example, those with Spanish or English). However, this gap is gradually being closed (e.g. Gagarina *et al.* 2017, 2018; Mak *et al.* 2019; Protassova 2007; Stadtmiller *et al.* 2022; Tribushinina *et al.* 2017).

When learning word-formation models, an important but not the only (see, e.g. Swan 2004; Mattes 2022) indicator of their productivity is considered to be the presence of an occasionalism in CS, built on a certain model (e.g. Clark, Berman 1984)². As regards the Russian language, this was demonstrated in a number of investigations, starting with the pioneering studies of A. N. Gvozdev (1949/1961), see also (Ceitlin 1989/2009; Kharčenko, Ozerova 1999; Jurjeva 2006 among others). As regards German, productivity problems are discussed in Dressler (2007), Mattes (2018). Meanwhile, with rare exceptions (Gagarina, Reichel 2013; Argus, Kazakovskaya 2013; Kazakovskaya, Voeikova 2021) the results obtained in terms of Russian word-formation are based on the analysis of disparate diary observations of parents and are not compared in a so-called bilingual aspect. The involvement of longitudinal data of spontaneous speech of a monolingual and a bilingual child, which our study is based on, will make it possible to receive a more complete and more reliable overall picture of the development of the word-formation component of system-language competence.

Thus, the main purpose of our study is to detect similarities and differences in the acquisition of Russian nominal derivation by

² On the terminological distinction between neologisms and occasionalisms see, e.g. Mattiello (2017: 23–26).

mono- and bilingual children. Achieving this goal involves solving the following specific issues:

- identification of the proportion of nominal derivatives and their stems;
- identification of the leading type of word-formation processes;
- detection of frequent word-formation morphemes, patterns and models;
- determining the order of development of semantic domains of nominal derivatives; and
- establishing the frequency of occasionalistic derivatives.

In addition to finding similar and distinctive features of nominal word-formation in mono- and bilingualism, two research questions on related problems are raised. Do non-derived nouns (i.e. simplexes) always precede nominal derivatives in the early CS? Do word-formation morphemes begin to be used primarily with simplexes and only after that with derivatives and compounds? The solution is related to the verification of the hypothesis according to which the development of derivation is carried out in accordance with a building-block model of complexity (Zurek 1990; Dziubalska-Kołaczyk 2014), see also Argus and Kazakovskaya (2018) based on Estonian and Russian L1.

In the next part of the article, the language data under observation, including how it was collected and analysed, will be described (Section 2). After that the results obtained will be presented, providing an answer to each of the questions posed and grouped as similarities (Section 3.1) and differences (Section 3.2) in the acquisition of Russian nominal derivation in mono- and bilingualism. Finally, the main conclusions and the prospects for research will be formulated (Section 4).

2. Data and method

Longitudinal recordings of the spontaneous speech of the Russian-speaking boy Filipp (Russia, St. Petersburg) and the girl Anna (Germany, Berlin), simultaneously acquiring Russian and German languages, serve as material for observing the development of nominal word-formation. In both instances, the recordings of adult/caregiver (mother) speech interaction were conducted several times a month in a natural setting—usually at home while playing, bathing, or eating, but on the street or outside the city as well. Spontaneous speech was recorded on a dictaphone, and then all the data of both the children and their caregivers were transcribed and stored in CHAT format using the CHILDES system (MacWhinney 2000)³.

The data for approximately the same observation period is included in the analysis of each of the corpora (see Table 1). In general, the size of the analysed data amounted to 45 hours of recordings containing 122,697 tokens.

Table 1. Data analysed

Subject (language/s)	Age of sub- jects	Length of obser- vation (months)	Length of recordings (hours)	Child speech (tokens)	Child- directed speech (tokens)	All tokens
Filipp (Russian)	1;5–2;8	16	28	16,468	40,253	56,721
Anna (Russian, German)	3;0–4;2	15	17	21,455	44,512	65,976
Total		31	45	37,923	84,765	122,697

Anna has been learning to speak two languages from birth—German, in which her father and her German relatives speak to her, and

³ We sincerely thank T. Pranova, M. Voeikova and N. Gagarina for the data provided.

Russian, which she uses for communication with her mother. Before 2;6 the influence of the German input was more abundant, which made German the dominant language. However, after 2;6 Russian has begun to prevail in the girl's speech environment. This type of bilingualism can be characterised as "asymmetric and simultaneous" (Gagarina, Reichel 2013: 197).

The period beginning from 3;0 (i.e. a few months after the start of the strengthening of the Russian input) was chosen for the comparative analysis. The mean length of the girl's utterances at that time was about two words (MLU: 1.8). By the end of the observations it had increased to three words (MLU: 2.9). It is important that during the first recording (3;0) her Russian speech production consisted mainly of repetitions after the mother, along with affirmative (*yes-*) or negative (*no-*) reactions to the mother's utterances. This suggests that Anna already understood well the phrases addressed to her in Russian and had a receptive vocabulary. At the same time, she often found it difficult to express her thoughts, to select a suitable word and to choose the appropriate language that causes both the code switching and code mixing (e.g. Lanza 2001).

At the beginning of the observation period, at the age of 1;5 Filipp's MLU was the same, 1.8. A similar MLU (being one of the indicators of so-called language age) of both subjects permits comparison, despite their different biological age. However, by the end of the observations, this index exceeded that of found in Anna's speech and approached five words (MLU: 4.8). Also Filipp's strategy in language acquisition could be defined as repetitive (e.g. Voeikova 2015), which also gave an additional basis for comparing these children.

Concluding the data review, one should note that certain features of Anna's speech development—viz. her grammatical errors and the use of spatial prepositions—were described in Gagarina and Reichel (2013) and Jakovleva (2016). The speech portrait of Filipp has been studied to a greater extent. In particular, the acquisition of his verbs is described in Gagarina (2008), adjectives in Voeikova (2015), and pronouns in Krasnoščekova (2016). The development of

dialogical (i.e. conversational) skills as well as epistemic modality is reflected in Kazakovskaya (2019, 2020). Some facets in the acquisition of word-formation–affixation and compounding—are shown in Protassova and Voeikova (2007), Kazakovskaya (2017) and Kazakovskaya and Voeikova (2021). An attempt to compare both affixation and compounding with the data of the compound-rich Estonian language was made in Argus and Kazakovskaya (2013, 2018) and Kazakovskaya and Argus (2021). Systematic study of derivational morphology in the comparative–Russian–German–standpoint is being undertaken for the first time.

For the analysis in each CS a) the proportion of nominal derivatives was determined, b) the degree of their diversity (in lemmas) and frequency (in tokens) was established, c) the number of new (i.e. first-appearing) nominal derivatives and old (i.e. repeated) ones⁴, as well as the ways of word-formation (affixation vs compounding), “working” morphemes and their semantics within the new derivatives were taken into account, d) occasionalistic derivatives and other specific phenomena accompanying the acquisition of word-formation such as *simplex–derivative* pairs, chains and families were revealed.

A chi-square test (statistical significance threshold of $p < 0.05$) was used for the statistical analysis.

3. Results and their discussion

3.1. SIMILARITIES IN THE ACQUISITION OF NOMINAL DERIVATION

The results indicating similarities in the development of Russian nominal word-formation in early mono- and bilingualism are presented first.

⁴ This analysis was carried out by eliminating repetitions of derivative lemmas documented in previous recordings.

3.1.1. THE PROPORTION OF DERIVATIVES IN NOUNS

The analysis shows that the proportion of nominal derivative tokens in the speech of both children is comparable since the differences between them are not statistically significant ($p>0.05$). As Table 2 demonstrates, more than a third of nouns are such derivatives, namely 30.7% in Filipp's speech and 32.6% in Anna's. The percentage of derivative lemmas also exceeds this value in both CS, but to varying degrees (see Section 3.2).

Table 2. *Nominal derivatives (lemma/token)*

	Nouns	Derivatives	Derivatives among nouns (%)
Filipp	874/3,803	575/1,168	65.8/30.7
Anna	982/2,771	388/902	39.5/32.6

Another indicator of the development of derivation mechanisms can be the size of new lemmas in relation to those already documented in CS, that is, to the old ones. Table 3 shows that in the speech of both subjects, the percentage of new derivatives is high for lemmas and for their tokens.

Table 3. *New nominal derivatives (lemma/token)*

	All derivatives	New derivatives	New derivatives among all derivatives (%)
Filipp	575/1,168	280/456	48.7/39.0
Anna	388/902	210/355	55.4/39.4

Nevertheless, when making a comparison with the data presented in Table 2, it becomes obvious that Anna, with an equal share of new tokens ($p>0.05$), has a higher percentage of new lemmas than Filipp ($p<0.05$). Therefore, the proportion of her repetitions is lower. We believe that the high repeatability of nominal derivatives by Filipp can be explained by his general strategy of language acquisition which is more repetitive than creative (e.g. Voeikova 2015).

3.1.2. NOMINAL PATTERNS: AFFIXATION VS COMPOUNDING

In both situations—monolingual and bilingual—the vast majority of nominal derivatives are formed by affixes, while compounds are equally rare ($p > 0.05$). Table 4 shows different patterns of affixal derivatives and their number, along with compounds, the majority of which have at least one noun member, and further shows the proportion of each group among all derivatives in the CS under observation.

Table 4. *Affixal derivatives and compounds (based on new lemmas)*

	Noun + affix	Verb + affix	Adjective + affix	Compounding
Filipp	244 (84.7%)	29 (10.1%)	7 (2.4%)	8 (2.8%)
Anna	168 (78.1%)	38 (17.7%)	4 (1.9%)	5 (2.3%)

In both CS, the proportion of denominal derivatives represented by the “noun+affix” pattern is especially high and comparable ($p > 0.05$): *ruč-k(a)* ‘hand-DIM’ (Filipp 1;8) ← *ruk(a)* ‘hand’, *korabl-ik* ‘ship-DIM’ (Anna 3;0) ← *korabl* ‘ship’. Conversely, deadjective derivatives (based on the “adjective+affix” pattern) are equally rare ($p > 0.05$): *čern-ik(a)* ‘blueberry’ ← *čern(yj)* ‘black’ (Filipp 2;3), *slad-ost(i)* ‘sweets’ ← *sladk(ij)* ‘sweet’ (Anna 3;4). The middle position of frequency within affixal derivatives is occupied by deverbal nouns (the “verb+affix” pattern): *peč-en’j(e)* ‘biscuit’ ← *peč* ‘to bake’ (Filipp 2;0), *lej-k(a)* ‘watering can’ ← *lit* ‘to pour’ (Anna 3;0). However, in the speech of bilingual Anna, their share is almost twice as high, which is significant ($p < 0.02$) (see Section 3.2). In general, the distribution obtained reflects the system-linguistic property of nominal derivation in Russian (Švedova 2005). That is, the formation of nominal derivatives occurs mainly from nouns despite the fact that almost any class of word to the point of prepositions can serve as the stem for word-formation.

The dominant affix in nominal word-formation is the suffix (see all examples mentioned above). As for the order of emergence in the

speech of both subjects, suffixation⁵ precedes prefixation (derivatives like *po-drug(a)* ‘girlfriend’ ← *drug* ‘friend’ (Anna 3;5) documented very rare), as well as precedes their simultaneous implementation—*pod-osin/ov-ik* ‘boletus’ ← *osin(a)* ‘aspen’ (Filipp 2;1)—and compounding—*tr+e+ugol’+nik* ‘triangle’ ← *tr(i)* ‘three’ + INTRF + *ugol* ‘angle’ (Anna 4;2)—including the synthetic types of the former, e.g. *mux+o+mor-Ø*⁶ ‘fly agaric’ ← *mux(a)* ‘fly’+INTERF+ *mori(t’)* ‘to starve’ (Filipp 2;1), see more in Dressler *et al.* (2019).

The proportion of compound lemmas was small. In the speech of either child, it does not exceed 3% (see Table 4). Thus, the assumption concerning the influence of the acquired compound-rich German, which would be expressed by the appearance of more compounds in the Russian speech of bilingual Anna, has not been confirmed. At the same time, in each corpora, compounds serve as the stems for derivatives. This is most often observed within diminutivisation, e.g. *samoljot-ik* ‘aeroplane-DIM’ ← *sam+o+ljot-Ø* ‘aeroplane’ (Filipp 1;9, Anna 3;4), *os’minož-ek* ‘octopus-DIM’ ← *os’m/i+nog* (lit. *eight legs*) ‘octopus’ (Anna 4;1).

3.1.3. SUFFIX INVENTORY

In both CS, the inventory of morphemes with which derivatives are formed is quite wide. More than 40 suffixes (excluding their allomorphs) were recorded in Filipp’s speech and about 30 in Anna’s; 21 of the suffixes were used by both children. The vast majority of these suffixes are productive in the modern Russian language.

Their distribution in patterns is as follows.

1. Suffixes *-k*, *-ik*, *-ok/ek*, *-yšk/ušk*, *-onok*, *-nik*, *-ess* (see examples presented within the text), *-ušek*: *vorob-ušek* ‘sparrow-DIM’ ← *vorobej* ‘sparrow’ (Filipp 2;3), *-c/ic*: *zermal’-c(e)* ‘mirror-DIM’ ← *zermal(o)* ‘mirror’ (Anna 3;10), *-očk/ečk*: *mam-očk(a)*

⁵ See “early positional salience” in Slobin (1973).

⁶ Hereafter the sign “Ø” will be used for a zero suffix.

‘mom-DIM’ ← *mama* ‘mom’ (Anna 3;3), -š: *xrju-š(a)* ‘piggy’ ← *xrjukat* ‘to oink’ (Filipp 2;1), -en’k/in’k: *za-in’k(a)* ‘hare-DIM’ ← *zajac* ‘hare’ (Filipp 2;6), -nic: *bol’-nic(a)* ‘hospital’ ← *bol’* ‘pain’ (Anna 3;7), -čik: *čemodan-čik* ‘suitcase-DIM’ ← *čemodan* ‘suitcase’ (Anna 3;10) form derivatives mostly within the “noun+affix” pattern.

2. Suffixes -ux, -enij/anij (see examples presented within the text), -k: *zakol-k(a)* ‘hairpin’ ← *zakolot* ‘to pin up (hair)’ (Anna 3;10), Ø: *pricep-Ø* ‘trailer’ ← *pricepit* ‘to attach’ (Filipp 2;0) are used within the “verb+affix” pattern.
3. The suffix -ost’: *slad-ost(i)* ‘sweets’ (Anna 3;4, see above) is used within the “adjective+affix” pattern.

The most frequent morphemes are suffixes with diminutive semantics serving diminutivisation (e.g. Savickienė, Dressler 2007): *pal’čik* ‘finger-DIM’ ← *palec* ‘finger’ (Filipp 1;8), *jabloč-k(o)* ‘apple-DIM’ ← *jablok(o)* ‘apple’ (Anna 3;0). The number of diminutives exceeds the number of those which can be covered by the notion of non-diminutives. In the former, following *Academic Russian Grammar* (Švedova 2005), we include the nominations of

- a) animal babies: *utj-onok* ‘duscling’ ← *utk(a)* ‘duck’ (Filipp 2;1), *l’vj-onok* ‘lion cub’ ← *lev* ‘lion’ (Anna 3;3),
- b) females: *zajč-ix(a)* ‘hare-FEM’ ← *zajac* ‘hare’ (Filipp 2;2), *princ-ess(a)* ‘princess’ ← *princ* ‘prince’ (Anna 4;0),
- c) singulatives: *goroš-in(a)* ‘(one) pea’ ← *gorox* ‘pea’ (Filipp 2;8), *snež-ink(a)* ‘snowflake’ ← *sneg* ‘snow’ (Anna 3;4), as well as words that are
- d) stylistic (colloquial) modifications (SM): *okošk(o)* ‘window-SM’ ← *okn(o)* ‘window’ (Filipp 2;8), *kolen-k(a)* ‘knee-SM’ ← *kolen(o)* ‘knee’ (Anna 3;11), interpreted in some papers as diminutives.

The proportion of diminutives in relation to all nominal derivatives documented in both CS (see Table 5) is comparable ($p > 0.05$). Nearly two-thirds of the early derivative lemmas and their tokens consist of diminutives. Their share among all nouns is expectedly

lower. At the same time, it is exactly the same for both subjects in terms of the number of tokens ($p>0.05$).

Table 5. *Diminutives (lemma/token)*

	Deriva- tives	Diminu- tives	Diminutives among derivatives (%)	Diminutives among nouns (%)
Filipp	575/1,168	413/840	71.8/71.9	47.2/22.1
Anna	388/902	262/619	67.5/68.6	26.7/22.3

A high proportion of diminutives is an important characteristic of a particular corpus. It determines the degree of intensity of their usage by the child and, thus, whether the corpus belongs to the diminutive-rich or diminutive-poor ones. Recent studies have shown that despite the acquisition of a diminutive-rich language (which Russian is considered to be, opposite to, for example, German or Estonian), the speech of Russian children and their caregivers may not reflect this property. Specifically, the corpus of monolingual Filipp, being the basis for the current comparison, is diminutive-rich (Kazakovskaya, Argus 2021; Kazakovskaya, Voeikova 2021). Consequently, the corpus of bilingual Anna having a similar number of diminutives can also be characterised as diminutive-rich. Thereby the high proportion of diminutives is an essential feature of both CS under observation.

It is also important to note that the proportion of diminutives may not always depend on child gender. In particular, in the earlier studies conducted on different monolingual data it was claimed that there were more diminutives in girls' speech than in boys' (Gleason *et al.* 1990; Protassova, Voeikova 2007; Kazakovskaya, Argus 2021). However, in the speech of bilingual Anna, the percentage of diminutive lemmas in relation to all nouns is lower than in Fillip's ($p<0.001$).

3.1.4. DEVELOPMENT OF THE DERIVATION SYSTEM: WORD-FORMATION PAIRS, CHAINS AND FAMILIES OF NOMINAL DERIVATIVES

The analysis of the development of connections within nominal word-formation in both CS revealed the following common trends. As Table 6 shows, based on the most frequent denominal pattern, the proportion of derivatives having an appropriate simplex pair (i.e. paired derivatives) in CS like *ded* ‘grandfather’ (1;5) → *ded-ušk(a)* ‘grandfather-DIM’ (Filipp 2;1) or *pčel(a)* ‘bee’ (3;0) → *pčel-k(a)* ‘bee-DIM’ (Anna 3;1) is quite high. The paired derivatives approach the half in Anna’s speech and exceed this figure in Filipp’s ($p < 0.05$).

Table 6. *Derivatives and their simplexes*
(based on the “noun+affix” pattern)

	All derivatives	Derivatives with simplexes (% among all derivatives)	Simplexes precede derivatives (% among paired derivatives)	Simplexes and derivatives appear simultaneously (% among paired derivatives)	Derivatives precede simplexes (% among paired derivatives)
Filipp	244	141 (57.8)	72 (51.1)	29 (20.6)	40 (28.4)
Anna	168	80 (47.6)	36 (45)	19 (23.75)	25 (31.25)

In addition, the pairs where a simplex precedes a derivative (as in both examples above) or appears simultaneously with it—*dyr(a)* ‘hole’ → *dyr-k(a)* ‘hole-SM’ (Filipp 1;5), *život* ‘belly’ → *život-ik* ‘belly-DIM’ (Anna 3;4)—also make up the majority of paired derivatives in both CS ($p > 0.05$). The consistent appearance of the derivative and, in general, the presence of a ‘simplex – derivative’ pair can be interpreted as an indicator showing productive rather than lexicalized use of the derivative by a child. The former also accompanies the acquisition of derivational morphology at the early stages of language acquisition.

Another important indicator of productivity is the presence of word-formation chains and families. These chains consist of more than two single-root words (i.e. more than one pair), e.g. *dyr(a)* ‘hole’

koz(a) ‘goat’ → *kozlj-onok* ‘goatling, goat baby’ → *kozlj/onoč-ek*
‘goatling-DIM’
→ *koz-očk(a)* ‘goat-DIM’
→ *kozj-ol* ‘goat-MALE’ → *kozl-ik* ‘goat.MALE-DIM’
(Filipp),
kupat’sja ‘to bathe’ → *kup-anij(e)* ‘bathing’
→ *kupa-l’nik* ‘swimsuit’ (Anna).

jož ‘hedgehog’ → *jož-ik* ‘hedgehog-DIM’
 → *jež-onok* ‘hedgehog baby’
 → *jež-ix(a)* ‘hedgehog-FEM’ (Filipp),
nos ‘nose’ → *nos-ik* ‘nose-DIM’
 → *nos+o+rog* ‘rhinoceros’ (Anna).

Occasionalistic nominal derivatives are equally infrequent in both CS. Their percentage proportion of the total number of nouns is very low. At the same time, all occasionalisms documented are built according to productive models. These are denominal diminutives in Filipp's speech, e.g. *det-ik**⁷ 'child-DIM' ← *det(i)* 'children' (1;10), *garmoš-išk(a)** 'accordion-DIM' ← *garmon* 'accordion' (2;1) as well as deverbatives in Anna's speech, e.g. *lep-enij(e)** instead of *lep-k(a)* 'modelling' ← *lepit* 'to mould from plasticine' (3;9), *sčita-nij(e)** ← *sčitat* 'to calculate' (3;8).

⁷ Hereafter the asterisk “*” will be used for marking children’s occasionalisms.

A wide look at Anna's speech production shows that her inflection innovations are more frequent than word-formation ones. Erroneous grammar forms most often occur during the formation of number and case forms, especially in non-declinable (e.g. *sal'to* 'somersault'), material (e.g. *xleb* 'bread') or pluralia tantum nouns (e.g. *nožnicy* 'scissors'), see more in Gagarina and Reichel (2013).

3.1.6. SEMANTIC DOMAINS OF DERIVATIVES: FREQUENCY AND EMERGENCE

A comparative analysis of the semantics of nominal derivatives in CS has revealed, first, the frequency of different semantic categories and, second, the order of their emergence. In addition to high-frequency diminutives (see Table 5), the next position in frequency is occupied by stylistic (colloquial) modifications of nouns (8–6%), as well as designations of various types of activities and/or their results (6–8%).

The names of instruments, animal babies and different objects like *otkryt-k(a)* 'greeting card' ← *otkryt* 'to open' (Filipp 2;2) or *sneg/ov-ik* 'snowman' ← *sneg* 'snow' (Anna 3;4) also have some frequency (importantly, comparable in both corpora) (6.5–3.5%).

Derivatives denoting females and males, e.g. *pet-ux* 'rooster' ← *pet* 'to sing' (Anna 4;2), singulatives (see examples above), locations, e.g. *skvoreč-nik* 'birdhouse, lit. house for starlings' ← *skvorec* 'starling' (Anna 3;0), agents, e.g. *pomošč-nik* 'helper' ← *pomošč* 'help' (Anna 4;0) and abstract notices like *nastroj-enij(e)* 'mood' ← *nas-troit* 'to tune' (Anna 3;6) are equally infrequent (3–0.5%).

The inventory of semantic groups of derivatives also matches so each child has 12 of them. Almost half of the categories (namely five, three of which are frequently used by both children) appear in the same order (see Table 7). In general, the emergence of subjects and objects precedes that of activities or their results along with the abstract names. The appearance of less frequent derivatives in CS and the intensity of enrichment of the derivative repertoire is individual (see Section 3.2.4).

Table 7. *Semantic domains in order of emergence (similarities)*

	Filipp (age)	Anna (age)
Diminutives, stylistic modifications, objects	1;5–1;8	3;0–3;1
Activities/results	2;2–2;5	3;4–3;5
Abstract notions	2;6–2;8	3;6–3;7

The same order in the emergence of the semantic domains represented by derivatives, as well as their frequency in both CS, may indicate that semantics is associated more with the cognitive development and its mechanisms rather than with purely linguistic ones.

3.2. DIFFERENCES IN THE ACQUISITION OF NOMINAL DERIVATION

Together with the presence of prominent similarities in the acquisition of nominal derivation in mono- and bilingual situations, some differences were noted.

3.2.1. NOUN TOKENS

The most obvious differences include the proportion of nouns in CS (see Table 8). The percentage of noun tokens in the monolingual data is almost twice as high as in the bilingual data ($p < 0,001$).

Table 8. *Noun tokens*

	All words	Nouns	Nouns among all words (%)
Filipp	16,486	3,803	23.1
Anna	21,455	2,771	12.9

However, this finding requires some clarification. As this table shows, though the number of nouns in Anna's speech is indeed lower than in Philipp's (2,771 vs 3,801), the total number of words in the bilingual corpus exceeds the corresponding value in the monolingual one.

Meanwhile, a careful analysis of the dialogue recordings from each CS shows that this excess is largely due to the peculiarities of bilingual communication in the early stages. So, Russian and German are switched or mixed in Anna's speech, and there occur hesitation pauses supplemented by various fillers. Also, functional words along with discourse markers are often used. Such words and markers are much less evident in Filipp's speech. Their scarcity, together with the lack of code switching, reduces the overall length of the dialogue and thereby increases the proportion of nouns. Thus, the quantitative discrepancy mentioned above can be associated not only with the smaller Russian vocabulary of Anna (as expected), but also with the specifics of adult-child communication in a bilingual situation.

3.2.2. DERIVATIVE LEMMAS

The next dissimilarity exists in a smaller number of derivative lemmas in relation to nouns in Anna's speech (see Table 2 above). There are almost 50% fewer of them than in Filipp's speech ($p < 0.001$). This result may indicate a smaller size of the bilingual child's active vocabulary, on the one hand, and a different speed in the development of derivational processes on the other. We connect its slowdown in the bilingual situation with the simultaneous development of the derivational relations in German, where a different way of word-formation (*viz.* compounding) prevails.

In particular, a higher proportion of verbal derivatives in Anna's speech reflects the results obtained in the study of word-formation based on the data of spontaneous speech of German-speaking children aged 1;9 to 3;0 (Schipke, Kauschke 2011). It was found that they produce more verbal than nominal derivatives and their compounds are based more on verbs than on nouns. Moreover, the results showed simultaneous development of compounding and derivation.

3.2.3. DEVELOPMENT OF NOMINAL DERIVATION

Anna's speech contains a smaller number of paired derivatives ($p < 0.05$), that is, those derivatives that have the corresponding simplex in the data (see Table 6). The result may indicate a higher degree of lexicalization of the derivatives used by her. This is especially noticeable in the first recordings of her dialogue with her mother. After 3;6, this trend clearly changes. Almost all derivatives in the girl's speech appear after the simplex, e.g. *čaj* 'tea' (3;0) → *čaj-nik* 'teapot' (3;7), *karandaš* 'pencil' (3;1) → *karandaš-ik* 'pencil-DIM' (3;11) or, at least, simultaneously with it, e.g. *mašin(a)* 'car' (3;7) → *mašin-k(a)* 'car-DIM' (3;7), *počt(a)* 'post office' (3;7) → *počt-aljon* 'postman' (3;7).

A smaller number of word-formation chains and families were also documented in Anna's speech. She has a little more than ten of them, while Filipp has twice as many. We tend to explain this result with a lower intensity of the derivational development in a predominantly affixal Russian language, which is carried out against the background of the development of a compound-rich German language. At the same time, it may seem curious that, on one hand, there is the small number of Russian compounds mentioned above and, on the other hand, a high proportion of diminutive lemmas and their tokens, which are not typical for German. Thus, the typological features of Russian in terms of nominal word-formation do not seem to undergo noticeable changes in contact interaction with German.

Perhaps the influence of the German word-formation system should be seen in Anna's production of so-called childish compounds such as *mama-kurica* 'mom-hen' (3;9), *kaljaka-maljaka* '~scribble' (3;11), as well as in a certain number of reduplications not documented in Filipp's speech. We are talking not only about the conventional doubling of adjectives like *bol'šoj-bol'šoj* 'big-big' (3;0)⁸,

⁸ This adjective was documented later with the usual prefix *pre-*: *bol'šoj-pre/bol'šoj* (3;7).

verbs like *šla-šla-šla* 'went-went-went' (3;7), *lezu-lezu* 'climb-climb' (3;11), adverbs like *daleko-pre/daleko* 'far far away' (3;9), *bystro-pre/bystro* 'quickly-very quickly' (3;10), *bol'no-bol'no-bol'no* 'hurt-hurt-hurt' (4;0) or even onomatopoeias from baby talk like *bum-bum* 'boom-boom' (3;3) or *njam-njam* 'yum-yum' (3;8), which are used to indicate the intensification and/or duration of some attribute or action, but about the repetition of nouns. And if the context of the earliest reduplication *ryby-ryby* 'fishes-fishes' (3;3) does not have an unambiguous interpretation, the later ones indicate that Anna attempted to strengthen the corresponding semantics with the help of repetition. To illustrate, the girl said *mjač-mjač* 'ball-ball' (3;7) at the moment when she rolled out plasticine and wanted to make a very big ball out of it. Or she said *mizinec-mizinec* 'pinky-pinky' (3;10) in the conversation with her mother, remembering what the smallest finger on her hand is called in Russian. In the first instance, this reduplication can be related with the development of augmentative semantics, and in the second one, with a diminutive one.

3.2.4. SEMANTIC DIFFERENCES

Finally, characterising the features of the semantic development of derivatives, the intensity of the process with which this occurs in a bilingual child should be noted. So, in Filipp's speech all semantic domains appeared sequentially during 16 months, whereas in Anna's speech this process was carried out twice as fast. And whereas Filipp's earlier derivatives turned out to be denominal ones (viz. quite simple diminutives and animal babies), Anna's speech consisted of more complex domains, such as locatives and instruments. The former are built on a less frequent and more complex deverbal pattern. Moreover, as mentioned, according to this pattern all of Anna's occasionalisms were created. This finding can be associated with her higher level of cognitive development as a child acquiring two language systems, and to some extent with her age.

4. Concluding remarks

Despite the fact that these results have the status of a case study to date, there undoubtedly are more similarities than differences in the acquisition of nominal derivation in mono- and bilingualism. Similar conclusions were made when describing the acquisition of spatial prepositions by Anna compared to the monolingual Russian-speaking girl Toma (Jakovleva 2016).

Meanwhile, the quantitative analysis of different facets of nominal word-formation—from the number of lemmas to the inventory of suffixes—indicates that bilingual Anna is somewhat lagging behind monolingual Filipp in this component of system-language competence. In general, this confirms the well-known trend that, on the one hand, there is the superiority of bilinguals in “cumulative” language development (and, according to the recent evidence, cognitive), but on the other hand, they lag behind monolingual peers in each of their languages (Białystok 2009; Miller *et al.* 2018 among others). Our results are consistent with those studies that point to the cognitive advantages of bilinguals. However, there is also contrary evidence (Nicoladis 2018). The question of why the results in the cognitive domain are different remains unanswered.

The specific research question of the present study was to examine how a building block model of complexity (Dziubalska-Kończak 2014; Zurek 1990) can be applied to the emergence of nominal derivatives in the course of development of a bilingual child. According to this model, a child should start with simple stems and the derivational complexity should increase during development. That is, children should start to use derived nouns only after they have already acquired the corresponding simplexes. Our study showed that despite the different proportion of nominal derivatives in relation to all nouns (Table 2), in terms of new derivative lemmas, the proportion of paired derivatives in the speech of both subjects was equally high ($p < 0.05$) and derivatives with the preceding simplex were quite frequent (Table 6). The analysis of word-formation chains

and the general sequence in the appearance of derivatives indicated that in most instances suffixes were first used with non-derived stems (or roots), after which they were attached to different derived stems, derivatives and compounds, including synthetic ones. Particularly, both subjects under observation, first, use suffixes (mainly diminutive and stylistic ones) with simplexes like *dyr-k(a)* ‘hole-SM’ (Filipp 1;5), *det-k(i)* ‘children-DIM’ (Anna 3;9) and then with derivatives *grib/oč-ek* ‘mushroom-DIM-DIM’ (Filipp 1;8), *det/išk(i)* ‘children-DIM’ (Anna 4;2). After that both children began to produce compounds: *magnit+o+fon* ‘tape recorder’ (Filipp 1;11), *nose+o+horn* ‘rhinoceros’ (Anna 3;8). However, in Filipp’s speech, compounding was accompanied by suffixation, including the zero one like *sam+o+ljet-Ø* ‘aeroplane’ slightly later (at 2;0), whereas in Anna’s speech, different types of compounds appeared during one recording session and much later (at 3;8). This circumstance, as the fact that some nouns appear in CS as derivatives (and without simplexes documented to the end of observations), does not allow us to confirm completely a building block model of complexity in early bilingualism. This hypothesis was also confirmed only partially on the monolingual data of Russian and Estonian (Argus, Kazakovskaya 2018: 34–35).

The following briefly outlined main prospects conclude the present Russian-German investigation. One of the objectives, in addition to increasing the data analysed, was the study of adjective and verb derivatives in a bilingual situation, which has been conducted so far only for Russian L1 (Kazakovskaya, Voeikova 2021). The next important aim is to analyse the child-directed speech, that is, the linguistic input children receive. We plan to study the following problems: firstly, the influence of input on the development of a child’s language system; secondly, the mechanisms involved in this process, specifically fine-tuning (Snow 1995); and thirdly, the peculiarities of caregiver communicative strategies with mono-, bi- and plurilingual children, see the initial experience of analysing Russian-Italian-Norwegian trilingualism in Kazakovskaya and Khačaturjan (2015).

Of particular interest is the reactive tactics of caregivers towards children's code mixing or switching, as well as children's errors, the study of which has been actively conducted in recent decades (e.g. Lanza 2001; Kilani-Schoch *et al.* 2009; Kazakovskaya 2021).

REFERENCES

- Argus, Reili; Kazakovskaya, Victoria V. 2013. Acquisition of compounds in Estonian and Russian: frequency, productivity, transparency and simplicity effect. – *Eesti Rakenduslingvistika Ühingu aastaraamat 9* [Estonian Papers in Applied Linguistics]. Tallinn: Eesti Rakenduslingvistika Ühing, 23–42.
- Argus, Reili; Kazakovskaya, Victoria V. 2018. Acquisition of noun derivations in Estonian and Russian. – *Eesti Rakenduslingvistika Ühingu aastaraamat 14*. [Estonian Papers in Applied Linguistics]. Tallinn: Eesti Rakenduslingvistika Ühing, 23–39.
- Bayram, Fatih; Miller, David; Rothman, Jason; Serratrice, Ludovica 2018. Studies in bilingualism: 25 years in the making. – *Bilingual cognition and language. The state of the science across its subfields. Studies in Bilingualism 54*. Eds. David Miller, Fatih Bayram, Jason Rothman, Ludovica Serratrice. Amsterdam: John Benjamins, 1–12.
- Białystok, Ellen 2009. Bilingualism: The good, the bad, and the indifferent. *Bilingualism*. – *Language and Cognition 12* (1), 3–11.
- Ceitlin, Stella N. 2009. *Očerki po slovoobrazovaniju i formoobrazovaniju v detskoj reči* [Essays on derivation and inflection in child speech]. Moscow: Znak.
- Clark, Eve; Berman, Ruth 1984. Structure and use in the acquisition of word-formation. – *Language 60* (3), 542–590.
- De Houwer, Annick 1990. *The acquisition of two languages from birth: A case study*. Cambridge: Cambridge University Press.
- De Houwer, Annick; Ortega, Lourdes (eds.) 2019. *Cambridge handbook of bilingualism*. Cambridge: Cambridge University Press.
- Döpke, Susanne 1992. *One Parent – One Language. An interactional approach*. *Studies in Bilingualism 3*. Amsterdam: John Benjamins.
- Döpke, Susanne (ed.) 2000. *Cross-linguistic structures in simultaneous bilingualism*. *Studies in Bilingualism 21*. Amsterdam: John Benjamins.

- Dressler, Wolfgang U. 1999. Ricchezza e complessità morfologica. – Fonetologia e Morfologia dell'Italiano e dei Dialetti d'Italia. Atti del 21. Congresso SLI. Eds. Paola Benincà, Alberto Mioni, Laura Vanelli. Roma: Bulzoni, 587–597.
- Dressler, Wolfgang U. 2007. Introduction. – Typology Perspectives on the Acquisition of Noun and Verb Morphology. Eds. Sabine Laaha, Steven Gillis. Antwerpen, 3–9.
- Dressler, Wolfgang U.; Kilani-Schoch, Marianne; Ketrez, F. Nihan (eds.) 2017. Nominal Compound Acquisition. Language Acquisition and Language Disorders Series 61. Amsterdam: John Benjamins.
- Dressler, Wolfgang U.; Sommer-Lolei, Sabine; Korecky-Kröll, Katharina; Argus, Reili; Dabašinskienė, Ineta; Kamandulitė-Merfeldienė, Laura; Johansen, Ijäs J.; Kazakovskaya, Victoria V.; Laalo, Klaus; Thomadaki, Evangelia 2019. First-Language Acquisition of Synthetic Compounds in Estonian, Finnish, German, Greek, Lithuanian, Russian and Saami. – Morphology 29, 409–429.
- Dressler, Wolfgang U.; Mattiello, Elisa; Korecky-Kröll, Katharina; Noccetti, Sabrina; Dabašinskienė, Ineta; Kamandulytė-Merfeldienė, Laura; Kazakovskaya, Victoria V. 2022. Communication with diminutives to young children vs. pets in German, Italian, Lithuanian, Russian, and English. Stem-, Spraak- en Taalpathologie 27, 53–68.
- Dziubalska-Kończak, Katarzyna (ed.) 2014. Special volume on linguistic complexity. Papers and Studies in Contemporary Linguistics 50 (2).
- Gagarina, Natalia V. 2008. Stanovlenie grammatičeskix kategorij russkogo glagola v detskoj reči [Development of grammar categories of the Russian verb in children's speech]. Saint Petersburg: Nauka.
- Gagarina, Natalia V.; Reichel, Aleksandra 2013. Zajcy rasplavjat kogti i žiš ty xir? Innovacii rusko-nemeckojazyčnyx detej doškol'nogo vozrasta [Hares will melt their claws and *žish ty xir*? Innovations of Russian-German preschool children]. – Saint Petersburg School of ontolinguistics. Saint Petersburg: Zlatoust, 195–212.
- Gagarina, Natalia; Posse Dorothea; Düsterhöft, Stefanie; Golcher, Felix; Topaj, Nathalie 2017. Bilingual lexicon development in German in preschool children with the home languages Russian and Turkish. – Dynamics of Linguistic Diversity. Eds. Hagen Peukert, Ingrid Gogolin. Amsterdam: John Benjamins, 125–142.

- Gleason, Berko J.; Perlmann, Rivka; Ely, Richard; Evans, David W. 1990. The baby talk register: Parent's use of diminutives. – *Handbook of Research in Language Development using CHILDES*. Eds. Jeffrey L. Sokolov, Catherine E. Snow. Hillsdale, NJ: Lawrence Erlbaum Associates, 50–70.
- Gvozdev, Aleksandr N. 1961. *Voprosy izučenija detskoj reči* [Issues of the investigation of children's speech]. Moscow: Izdatel'stvo Akademii pedagogičeskix nauk RSFSR.
- Jakovleva, Irina V. 2016. *Usvoenie sistemy predlogov russkogo jazyka v uslovijax monojazyčija i dvujazyčija* [Acquisition of the system of Russian prepositions in monolingualism and bilingualism]. – *Faces of bilingualism*. Ed. Stella N. Ceitlin. Saint Petersburg: Zlatoust, 46–61.
- Jurjeva, Nadežda M. 2006. *Problemy rečevogo ontogeneza: Proizvodnoe slovo. Dialog. Eksperimental'nye issledovanija* [Problems of speech ontogenesis: A derivative. A dialogue. Experimental studies]. Moscow: Institut jazykoznanija RAN.
- Kazakovskaya, Victoria V. 2019. *Vopros i otvet v dialoge 'vzroslyj – rebjonok'*. *Psicholingvističeskij aspekt* [Questions and answers in adult-child dialogue. Psycholinguistic aspect]. Moscow: LIBROCOM.
- Kazakovskaya, Victoria V. 2020. Epistemic modality in the speech of Russian young children and adolescents. – *Philologia Estonica 5. Languages, orderings and successions*. Eds. Reili Argus, Annika Hussar. Tallinn: Tallinn University Print, 134–166.
- Kazakovskaya, Victoria V. 2017. Acquisition of nominal compounds in Russian. – *Nominal Compound Acquisition. Language Acquisition and Language Disorders Series 61*. Eds. Wolfgang U. Dressler, Marianne Kilani-Schoch, F. Nihan Ketrez. Amsterdam: John Benjamins, 63–90.
- Kazakovskaya, Victoria V. 2021. Feedback in verb acquisition: Evidence from Russian 'adult – child' interaction. – *Russian Grammar: System – Language Usage – Language Variation. Potsdam Linguistic Investigations*. Ed. Vladislava Warditz. Berlin, Bern, Bruxelles, New York, Oxford, Warszawa, Wien: Peter Lang, 245–270.
- Kazakovskaya, Victoria V.; Argus, Reili 2021. Acquisition of diminutives in typologically different languages: Evidence from Russian and Estonian. – *Eesti Rakenduslingvistika Ühingu aastaraamat 17* [Estonian Papers in Applied Linguistics]. Tallinn: Eesti Rakenduslingvistika Ühing, 81–97.

- Kazakovskaya, Victoria V.; Voeikova, Maria D. 2021. Acquisition of derivational morphology in Russian. – *The Acquisition of Derivational Morphology. A cross-linguistic perspective*. Language Acquisition and Language Disorders 66. Eds. Veronika Mattes, Sabine Sommer-Lolei, Katharina Korecky-Kröll, Wolfgang U Dressler. Amsterdam: John Benjamins, 170–195.
- Kazakovskaya, Victoria V.; Khačaturjan, Elizaveta V. 2015. Kommunikativnye strategii vzroslyx v dialoge s mono- i pljurilingval'nymi det'mi [Caregiver communicative strategies in dialogues with monolingual and multilingual children]. – *Problems of ontolinguistics – 2015: Mechanisms of language acquisition and establishment of communicative competence*. Saint Petersburg: Institute for Linguistic Studies of RAS, 98–102.
- Kharčenko, Vera K.; Ozerova, Elena G. 1999. Složnye slova v detskoj reči [Compounds in children's speech]. Belgorod: Belgorod University Press.
- Kilani-Schoch, Marianne; Balčiūnienė, Ingrida; Korecky-Kröll, Katharina; Laaha, Sabine; Dressler, Wolfgang U. 2009. On the role of pragmatics in child-directed speech for the acquisition of verb morphology. – *Journal of Pragmatics* 41 (2), 219–239.
- Krasnoščekova, Sofia V. 2016. Mestoimennyy dejksis v russkoj detskoj reči [Pronoun deixis in Russian children's speech]. PhD dissertation abstract. Saint Petersburg: Institute for Linguistic Studies of RAS.
- Lanza, Elizabeth 2001. Bilingual first language acquisition: a discourse perspective on language contact in parent-child interaction. – *Trends in Bilingual Acquisition*. Eds. Jacone Cenoz, Fred Genesee. Amsterdam: John Benjamins, 201–229.
- MacWhinney, Brian 2000. *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mak, Pim; Lomako, Julia; Gagarina, Natalia; Abrosova, Ekaterina; Tribushinina, Elena 2019. Keeping two languages apart: Connective processing in both languages of Russian-German bilinguals. – *Bilingualism: Language and Cognition* 23 (3), 532–541.
- Mattiello, Elisa 2017. *Analogy in Word-Formation: A Study of English Neologisms and Occasionalisms*. Berlin; Boston, MA: De Gruyter Mouton.
- Mattes, Veronika 2018. *Derivationsmorphologie und Wortarten im Erstspracherwerb des Deutschen. Mit einem Fokus auf dem Schuleintrittsalter: Habilitationsschrift*. Graz. University of Graz.

- Mattes, Veronika 2022. Methods for evaluating the productivity of derivational morphology. – Papers, presented at Pre- and Protomorphology in Language Acquisition seminar at 18.2.2022. Austrian Academy of Sciences, Vienna.
- Mattes, Veronika; Sommer-Lolei, Sabine; Korecky-Kröll, Katharina; Dressler, Wolfgang U. (eds.) 2021. *The Acquisition of Derivational Morphology. A cross-linguistic perspective. Language Acquisition and Language Disorders* 66. Amsterdam: John Benjamins.
- Miller, David; Bayram, Fatih; Rothman, Jason; Serratrice, Ludovica (eds.) 2018. *Bilingual cognition and language. The state of the science across its subfields. Studies in Bilingualism* 54. Amsterdam: John Benjamins.
- Nicoladis, Elena 2018. Simultaneous child bilingualism. – *Bilingual cognition and language. The state of the science across its subfields. Studies in Bilingualism* 54. Eds. David Miller, Fatih Bayram, Jason Rothman, Ludovica Serratrice. Amsterdam: John Benjamins, 81–102.
- Protassova, Ekaterina 2007. Sprachkorrosion: Veränderungen des Russischen bei russischsprachigen Erwachsenen und Kindern in Deutschland. – *Kindliche Kommunikation. Einsprachig und mehrsprachig.* Hrsg. Katharina Meng, Jochen Rehbein. New York; München; Berlin: Waxmann, 299–334.
- Protassova, Ekaterina Ju.; Voeikova, Maria D. 2007. Diminutives in Russian at the early stages of acquisition. – *The Acquisition of Diminutives. A Cross-Linguistic Perspective. Language Acquisition and Language Disorders* 43. Eds. Ineta Savickienė, Wolfgang U. Dressler. Amsterdam; Philadelphia: John Benjamins, 43–72.
- Savickienė, Ineta; Dressler, Wolfgang U. (eds.). *The Acquisition of Diminutives: A Cross-Linguistic Perspective. Language Acquisition and Language Disorders* 43. Amsterdam; Philadelphia: John Benjamins, 2007.
- Schipke, Christine; Kauschke, Christina 2011. Early word formation in German language acquisition: A study on word formation growth during the second and third years. – *First Language* 31 (1), 67–82.
- Slobin, Dan 1973. Cognitive Prerequisites for the Development of Grammar. – *Studies of child language development.* Eds. Charles Ferguson, Dan Slobin. New York: Holt, Rinehart & Watson, 175–208.
- Snow, Catherine E. 1995. Issue in the study of input: fine-tuning, universality, individual and developmental differences, and necessary causes. –

- Handbook of child language. Eds. Paul Fletcher, Brian MacWhinney. Oxford: Oxford University Press, 180–193.
- Swan, Deanne 2004. How to Build a Lexicon: A Case Study of Lexical Errors and Innovations. – *The Child Language Reader*. Eds. Kate Trott, Sushie Dobbinson, Patrick Griffiths. London: Routledge, 166–178.
- Švedova, Natalia Ju (ed.) 2005 [1980]. *Russkaja grammatika T. 1* [Russian grammar V. 1]. Moscow: Nauka.
- Stadtmiller, Elizabeth; Lindner, Katrin; Süss, Assunta; Gagarina, Natalia 2022. Russian-German five-year-olds: What omissions in sentence repetition tell us about linguistic knowledge, memory skills and their interrelation. – *Journal of Child Language* 49 (5), 869 – 896.
- Tribushinina, Elena; Valcheva, Eva; Gagarina, Natalia 2017. Acquisition of additive connectives by Russian-German bilinguals: A usage-based approach. – *Usage-based Approaches to Language Acquisition and Language Teaching. Studies on Language Acquisition [SOLA] 55*. Eds. Jacqueline Evers-Vermeul, Raiser Laurent, Elena Tribushinina. Berlin: De Gruyter, 207–232.
- Voeikova, Maria D. 2015. Stanovlenie imeni: Rannie etapy usvoenija det'mi imennoj morfologii russkogo jazyka [Establishment of Name: Early phases of the acquisition of nominal morphology by children]. Moscow: Jazyki slavjanskix kul'tur.
- Xanthos, Aris; Gillis, Steven 2010. Quantifying the Development of Inflectional Diversity. *First Language* 30, 175–198.
- Xanthos, Aris; Laaha, Sabine; Gillis, Steven; Stephany, Ursula; Aksu-Koç, Ayhan; Christofidou, Anastasia; Gagarina, Natalia; Hrzica, Gordana; Ketrez, F. Nihan; Kilani-Schoch, Marianne; Korecky-Kröll, Katharina; Kovačević, Melita; Laalo, Klaus; Palmović, Marijan; Pfeiler, Barbara; Voeikova Maria D.; Dressler, Wolfgang U. 2011. On the role of morphological richness in the early development of noun and verb inflection. – *First Language* 31, 461–479.
- Zurek, Wojciech H. (ed.) 1990. *Complexity, Entropy and the Physics of Information*. Redwood City: Addison Wesley.

ABBREVIATIONS

CS – child speech

CDS – child-directed speech (input)

DIM – diminutive

INTERF – interfix

MLU – mean length of utterance

L1 – first language acquisition

SM – stylistic modification

RESÜMEE

VEENE NIMISÕNATULETUSE OMANDAMINE ÜKS- JA KAKSKEELSE SITUATSIOONIS

Victoria V. Kazakovskaya

Institute for Linguistic Studies, Russian Academy of Sciences

Uurimus keskendub ükskeelsete (vene) ja simultaansete kakskeelsete (vene-saksa) laste vene keele nimisõnade tuletusmorfoloogia omandamisele. Analüüsi ja tulemuste aluseks on pikiuuringuga kogutud lindistused, mis on transkribeeritud CHAT-vormis, kasutades CHILDES-i süsteemi. Esitatud on esimesed mallid ja nimisõnatuletuse mehhanismid koos morfeemidega, mida lapsed kasutavad. Nimisõnatuletiste produktiivsele kasutusele osutavateks teguriteks on lapse kõnes leiduvate tuletamata tüvede ja tuletise ehk sõnapere ja juhutuletiste kasutamine. Arutletakse nimisõnatuletuse omandamise, sh tuletiste semantika omandamise sarnasuste ja erinevuste üle ükskeelsetes ja kakskeelsetes situatsioonides.

Võttesõnad: nimisõnad, sõnamoodustus, tuletised, tuletamata tüvi, liitmine, vene, saksa

Victoria V. Kazakovskaya's main research interests are connected with functional grammar, Russian syntax, language acquisition, communicative competence, epistemic modality, word-formation, child-directed speech, pet-directed speech.
victory805@mail.ru

THE INSERTION OF PERSON REFERENCES MOTIVATED BY PRAGMATIC DIFFERENCES IN ESTONIAN-ENGLISH-JAPANESE FACEBOOK COMMUNICATION

Geidi Kilp

Tallinn University

Abstract.¹ Sociopragmatic differences have been examined between many languages and cultures, including English and Japanese. However, Estonian and Japanese have yet to be compared, and thus this data of Estonian-English-Japanese communication on Facebook offers a look at a type of code-switching that is caused by the sociopragmatic differences between Estonian and Japanese – i.e. the insertion of person references from Japanese to Estonian and English utterances by native Estonians.

I am using the Estonian-English-Japanese Facebook communication dataset from Kilp (2021) with new added conversations. The data consist of synchronous private Facebook messages between 2015 and 2021: a total of 7 informants, 50 conversations and 14,681 tokens. A usage-based approach and a qualitative analysis are applied to the data from individual informants and particular cases.

These data show that a perception of pragmatic differences causes the insertions of the Japanese person references, *senpai* ‘senior’ and *sensei* ‘teacher’, in various forms (affixed to the name, replacing the name, elongated, capitalised, in the Latin alphabet, in Japanese script) in Estonian and English utterances, while factors such as vertical hierarchy, horizontal solidarity and (situational) salience play an important role in facilitating insertion.

¹ **The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:** This research was funded by the Estonian Ministry of Education and Research grant “Data and corpora of Estonian children and youth multilingual communication” [grant number EKKD33].

Keywords: deixis, code-switching, vocative, usage-based approach, contact linguistics, computer-mediated communication, salience, Estonian, English, Japanese

1. Introduction

There have been no previous contact linguistic studies (to the best of my knowledge) that have compared the sociopragmatic aspects of Estonian and Japanese languages and/or cultures. Nor has an analysis of code-switching between Estonian and Japanese been done regarding person references or deixis, or from the perspective of pragmatics (except briefly in Kilp 2021). This set of data of Estonian-English-Japanese Facebook communication includes several cases of person references that have been inserted from Japanese into Estonian or English utterances due to sociopragmatic differences between the languages and cultures. Therefore, along with input from the informants through a semi-structured interview, this study offers an in-depth view into this phenomenon.

There are different terms used to describe person references. Irgens (2017) uses the term *person deixis*, which can also be called *personal deixis* (e.g. Marchello-Nizia 2006), and which is often used to refer to personal pronouns, specifically. Following Irgens's definition (while using the broader term *person references*), person references are considered here as any linguistic references to discourse participant roles, including "expressions referring to the speaker, listener and to other persons, who may or may not be present in the discourse situation" (2017: v). A distinction will be made between the vocative second person (honorific or descriptive) and third person (descriptive) usage of person references (see Section 4).

Person references may manifest in various ways across languages (Irgens 2017: v). Japanese language and culture have been compared to English, for example, from a sociopragmatic perspective (Irgens 2017), Estonian has been compared to, for example, Swedish

(Keevallik 2012), and the three languages, Estonian, English and Japanese, have been examined separately or in other combinations as well (e.g. Stivers *et al.* 2007, Schegloff 2008, Takahara 1992). This study will use information from prior research about the sociopragmatic aspects of person references in Estonian, English and Japanese for the analysis of the instances present in this set of data of Estonian-English-Japanese communication (Sections 2.2 and 4).

The approach in this study is usage-based (see Backus 2015; Verschik 2019; Zenner *et al.* 2019), applies a cognitive point of view to contact-induced change and is holistic in nature. I use a bottom-up approach with the focus that language change starts in the mind of the speaker (Weinreich 1953: 71), while a qualitative analysis is needed to examine the reasons behind individual cases, as they are highly variable and not evenly distributed among all informants (see Section 5.1).

This paper begins with an overview of the theoretical background and methodology (2), including the background of the study of person references and pragmatic differences (2.1), the sociopragmatic aspects of person references in the Estonian, English and Japanese languages and cultures (2.2), the usage-based approach and a description of the semi-structured interview (2.3). Then follows an explanation of the data and informants (3), examining the nature of the data of Estonian-English-Japanese Facebook communication that has been used for the qualitative analysis (3.1), aspects of computer-mediated communication (3.2), along with characteristics of its informants (3.3) and their input (3.4). The analysis (4) is sectioned according to the types of insertions and divided into second person (4.2) and third person (4.3). Individual cases are analysed along with the conversational backgrounds. The perceptions of the informants are analysed separately (4.1). The discussion section (5) covers the limitations of this study and its data (5.1) and the notion of grammatical correctness (5.2), and outlines some possible future research directions (5.3). Finally, conclusions are drawn from the findings.

2. Theoretical background and methodology

2.1. THE THEORETICAL APPROACH

Person references or deictics have been studied from varying points of view, including in the cases of Estonian, English, and Japanese. For example, Irgens (2017) uses a contrastive approach regarding person deixis (between Japanese and English); Pajusalu (2009) offers a typological overview of pronouns and reference in Estonian (giving choice-influencing categories); Stivers *et al.* (2007) offer a comprehensive cross-cultural overview of person reference in natural conversation (looking at different languages and cultures); Howell (2007) focuses on the use of sociolinguistic and pragmatic resources in the English subtitling of character voice (in Japanese animations); Keevallik (2012) looks at the pragmatics of Estonian heritage speakers in Sweden (specifically pragmatic interference and politeness). The cases of *senpai* 'senior' and *sensei* 'teacher', influenced by pragmatic differences, have also been briefly analysed in Kilp (2021, therein *pragmatic gaps* according to Verschik 2010) along with various other factors that contribute to insertion (therein *code-copying* according to Johanson 2002).

As the perception of sociopragmatic differences is subjective, rather than focusing on structural or 'objective' differences between Estonian, English and Japanese, this work focuses on the cognitive reasons behind the insertion of Japanese person references and the perceptions and aims of the informants, applying a cognitive approach to contact-induced change and utilising semi-structured interviews (see Sections 2.3 and 4.1). Even if there is a difference (e.g. between referencing in Estonian and Japanese, see Section 2.2.), if the user does not (consciously or subconsciously) notice it, it does not affect their usage directly. Vice versa, even if there is no 'objective' difference between the languages regarding a certain aspect, but the user perceives there to be a difference, it may affect their usage and perhaps cause them to compensate for the difference in

some manner (e.g. cause insertions, such as in this data). The specific cases will be analysed in depth in Section 4.

2.2. THE SOCIOPRAGMATICS OF PERSON REFERENCES IN ESTONIAN, ENGLISH AND JAPANESE

Various studies have noted sociopragmatic differences between Japanese and English in general (e.g. Howell 2007: 292), and also in regard to deixis (Irgens 2017: v-vi). Irgens argues that person deixis is less grammaticalized in Japanese than in English based on the fact that nominal ellipsis is widespread in Japanese (2017: v). English has verbal agreement inflection (Irgens 2017: v), as does Estonian (Pajusalu 2009). In Japanese, however, person deixis is “primarily lexically manifested in the form of “person nouns”, whose meanings vary according to different social variables” (Irgens 2017: v), and there is no verbal agreement. This means that situations where the person being referred to can only be inferred from the context and/or previous utterances are relatively common.

Regarding personal pronouns in second person, specifically, there are unmarked second person singular and/or plural pronouns in Estonian and English, while no generic second person pronoun has developed in Japanese (Takahara 1992: 119). The Estonian *sina* ‘you’ (informal, singular), *Sina* ‘you’ (polite, singular), *teie* ‘you’ (informal, plural), and *Teie* ‘you’ (polite, singular or plural) differentiate between number (singular and plural) and register (informal or polite) (see e.g. Pajusalu 2009, Pajusalu *et al.* 2010). In standard English, *you* is used for both singular and plural, and there is no distinction of politeness. However, “no generic second person pronoun has developed in Japanese”, and all of them “are marked for social status, gender, age differences as well as relative intimacy to the speaker” (Takahara 1992: 119). For example, *anata* ‘you’, commonly used as the second person equivalent in teaching Japanese as a foreign language (e.g. in the series of textbooks entitled *Minna no Nihongo*, which the informants in this data were also taught), is rarely used by

native Japanese speakers nowadays; it is used mostly between strangers, where the hierarchical differentiation is not known, in advertisements that are directed towards a wider audience, or by wives to address their husbands (Kaiser *et al.* 2013: 140–141).

In Japanese it is more common to use the person's (sur)name vocatively and most often with an honorific suffix or title (Howell 2007: 294), such as *Tanaka-sensei* 'teacher Tanaka', which is "equivalent to *you* in English" (Yui 2012: 62), although grammatically it is in third person. In this data, similarly, *senpai* 'senior' and *sensei* 'teacher' are used as direct addresses, although the suffix or title is mainly used to replace the name and sometimes together with the given name (in Japanese, honorific usage would require the surname).

There are also differences between first person pronouns (or in the case of Japanese, person nouns), but as none of those were inserted in this data, they will not be focused on.

2.3. THE METHODOLOGICAL APPROACH

The focus of this study is on individuals and their language use as "language change ultimately goes back to individual instances of language use" (see Zenner *et al.* 2019: 9). Furthermore, a qualitative approach is needed to examine the implications in individual cases as the elements that are presented in this study are not used by all informants and are used to varying degrees depending on the co-speaker (solidarity) and language ability, among other factors. It is not possible to apply statistical or diachronic methods as the amount of data is not extensive enough and this type of communication is a relatively new phenomenon largely brought about by the globalisation of Japanese media, as well as the growth of Facebook as a conversational medium.

As these insertions of person references from Japanese are, except for one humorous usage by Informant E, present in the usage of only two informants out of 7 (i.e. informants A and B), in addition to a conversational (and situational) analysis of the cases, two

semi-structured interviews were also conducted. It should be noted that my own conversations are part of the data (I am Informant A), and I interviewed Informant B, who showed consistent usage of *senpai* 'senior' towards myself (Section 4.2). Although there was another conversation pair with the senior-junior distinction (A and G), there was no usage of *senpai*, and thus the interviews can show why it is present in the repertoire of Informant B. The interviews are analysed in Section 4.1.

3. Data and informants

3.1. DATA OF ESTONIAN-ENGLISH-JAPANESE COMMUNICATION

This data is based on the data used in Kilp (2021), with new added conversations. The data consists of synchronous private messages on Facebook between the years of 2015 and 2021, with a total of 14,681 tokens. There are seven informants, seven conversation pairs between them, and a total of 50 conversations. In some cases it may be difficult to establish a base language (which can be Estonian but also English or Japanese), but this was not the goal of my research. In trilingual communication, certain social, cultural and psychological factors may assume high significance (Hoffmann 2001: 2), and thus cultural and personal backgrounds are also analysed. The data is not strictly *trilingual*, however, as some cases also include other languages, namely German, Russian, Spanish and French.

3.2. COMPUTER-MEDIATED COMMUNICATION

As the data is computer-mediated and involves languages that use different writing systems, we see some cases involving digraphia, i.e. the coexistence of the Latin alphabet and the Japanese writing system (see Kilp 2021: 184). Some sentences in Facebook communication are also very short and may be syntactically incomplete or ungrammatical; therefore, it is sometimes difficult to determine the borders of individual sentences (similar to spoken language). Participants in

computer-mediated communication also inhibit higher awareness than during speech on the level that they can correct their errors (such as spelling) (Dorleijn 2016: 11), although errors (and misunderstandings) do still occur.

Cases using the Japanese writing system have been transliterated in [square brackets] for clarity as square brackets are very rarely used in Facebook conversations otherwise. Line changes have been marked with two forward slashes //, as line changes can occur within an utterance and often replace both inter-sentential and intra-sentential punctuation. Some names and emoticons have also been marked with square brackets.

3.3. INFORMANTS

The informants in this data are coded as letters A to G (referred to here as Informant A, for example). All of them are native Estonian speakers while one of them grew up as an Estonian-Russian bilingual (but has attended Estonian-speaking schools). They were between 18 and 29 years of age at the times of the conversations. Personal characteristics have been coded in the data, such as names or nicknames (marked as [name]).

Most of the informants had studied or were studying Japanese language and culture at Tallinn University at the times of the conversations. There is one informant who has never studied Japanese but has had extensive enough contact with both Japanese language and culture to understand many everyday words and phrases and uses Japanese in communication daily (Informant E). It seems that the more extensive the contact with a language (such as Japanese), the higher the chance is for the user to start to feel pragmatic differences when using other languages, although usage often cannot be explained by the extent of contact and language ability alone. For example, Informant B uses honorifics very often when addressing Informant A in Estonian or English, while other informants rarely use any honorifics (despite having similar levels of language ability and similar input).

3.4. INPUT FROM JAPANESE

As almost all of the informants in this data have studied the Japanese language and culture at Tallinn University (except Informant E, who has never studied Japanese), a certain amount of their input is from language courses at the university. Tallinn University offers Japanese language courses from A1 level to C1 level (according to CEFR levels), and in addition also classical Japanese and practical translation courses (from Japanese to Estonian, e.g. classical literature). At university, students are generally taught standard Japanese in polite style (*teinei-tai*, also polite expressions and honorific suffixes) and at higher levels also polite speech (*keigo*, divided into humble and deferential speech), while colloquial usage, slang and dialects are generally not taught (with the exception of certain verb forms).

All of the informants also have contact with Japanese media, such as animations, movies, TV drama series, manga (Japanese comics) and music, to varying degrees. Media enables them to encounter colloquial language and dialects, which are generally not taught in schools. There may also be slang and jargon, and also ungrammatical and unconventional usage, depending on the particular series. Standard Japanese (which is based on Kantō dialect) is used most commonly, but Kansai and Kyūshū dialects are also popular in media representing characters that are from those regions. Some informants also have Japanese friends (e.g. exchange students), although not too many native Japanese are permanently living in Estonia. Some of the informants have also visited Japan.

4. Insertion of person references in Estonian-English-Japanese Facebook conversations

As the intentions behind the cases of insertion of person references are important, this section will start with the semi-structured interviews with Informant B, wherein their reasons for using *senpai* and

sensei are discussed, along with their perceptions of the underlying meanings and attitudes. Then follow analyses of particular cases, divided between usage in second person (honorific or descriptive), and usage in third person (descriptive). The numbers of occurrences are provided in Section 5.1.

4.1. PERCEPTION OF THE INFORMANTS

In the two semi-structured interviews (conducted in January and March, 2022, in Facebook Messenger), Informant A (who is also the author and interviewer) and Informant B discussed the usage of *senpai* and *sensei* in their conversations. Informant B was asked specific questions, and the viewpoint of Informant A (the author/interviewer) will be provided alongside the questions.

The languages in the interview were mainly Estonian and English, and thus the Estonian text has been translated to English, indicated with [square brackets]. The only Japanese elements used in the interviews were the related terms of person references. No comments were made on the languages used in the interviews, and no restrictions were given, either. Only the relevant questions and answers are shown here, while some irrelevant parts (e.g greetings, agreements, other topics) have been omitted.

Question 1: “[Where did the idea come from to call me *senpai*? If you know.]”

Informant B (towards Informant A/author/interviewer):

“[I don’t remember for certain anymore, but it felt natural], cuz you were a year above me in uni so you were a literal *senpai* 😊
[I don’t remember if it was before going to Japan or after] 🤔 but it felt like you know so much more about Japanese but also kind of about life? And I guess having seen the whole *senpai* concept in Japanese culture/media it felt natural but also cool to bring it into our communication. It was definitely a in the moment idea/ thought flash that just stuck with usage not really something I consciously planned/though of. Does that make sense? ☺”

From this answer, it can be seen that Informant B did not consciously decide to use *senpai*, but it came naturally as they felt a connection to native Japanese usage (situational salience) and the (hierarchical) similarity in their relationship with Informant A (*a year above me* and *literal senpai*). The discussion of whether they should use *senpai* or *sensei* is seen in Examples 2a and 2b (March 2016, see Section 4.2), while they had already used *senpai* in several cases before that (throughout 2015), without any specific discussions about the usage. As there were many different instances of *senpai* in Informant B's usage, in various forms, I asked a follow-up question. As the question was longer and with clarifications, here is a condensed version:

Question 2: I asked Informant B about the different forms that *senpai* can be used in (towards myself), naming *senpai*, *Geidi-senpai*, *senpai Geidi*, and also elongation and capitalization, and asked if they sense any difference, or whether they intended the meaning or attitude to be different (e.g. in politeness) depending on the choice of the particular form.

Informant B:

"[I wouldn't say there is a difference in the thought/meaning.] I naturally try to bend the language and change it up, so it would stay fun and interesting, at least I think so (haven't given it a thought before), although the intention sometimes is to sound more formal indeed, to show appreciation maybe too. [But I don't remember specifically anymore]" [...]

As a comment, at the start of these conversations, I (Informant A) was not aware of the form distinctions, either, and I had no concept for honorific and descriptive referencing. I knew what honorifics (titles and suffixes) were in general terms, but I certainly did not discern any differences in politeness depending on whether *senpai* was used alone, or before or after my (sur)name. I did, of course, sense a difference in attitude or meaning from other factors, such as that elongation and/or all capital letters may infer a higher degree of emotion, such as excitement, whining or exasperation.

As can be seen from the answers, the forms are not something that Informant B has considered. Nor was Informant A (who *senpai* was used towards) aware of them at the time of the related conversations (starting from 2015). Thus, there is certainly no conscious separation between descriptive and honorific forms within these conversations, nor the knowledge that the grammatical honorific form in the case of *senpai* is surname+suffix (not the given name, and not as a prefix), and thus grammatical correctness will not be the main focus in the analysis (see Section 5.2). It is also evident that Informant B does use *senpai* to signal *appreciation* and *to sound formal*, but this is not the case in every instance of its usage.

4.2. USAGE IN SECOND PERSON

As mentioned, the person references used in this data are *senpai* ‘senior’ and *sensei* ‘teacher’. *Senpai* and *sensei* are conventionally used as honorific suffixes (following the surname), while they may also be used instead of the name (both in descriptive and honorific usage). *Senpai* is very common in this data in the use of Informant B in reference to Informant A, although it is not seen in other conversation pairs. This may be because in most of the other conversation pairs the informants are of relatively equal hierarchical status, e.g. the same year of studies in the case of informants A, C and F, and informants B and D, respectively (while informant E had no formal education in Japanese, and thus no such context). *Sensei* is used in writing only by three informants (see Section 5.1), mostly in meta-linguistic contexts, and only in two cases in reference to a fellow student (Informant B towards Informant A, Example 2a). *Sensei* is more commonly used in speech than in writing, most often in classroom environments. No other types of Japanese honorifics or titles are seen in this particular set of data (including cases other than insertion) although many more exist.

Example (1) shows different instances of the usage of *senpai* ‘senior’ (also ‘upperclassman’, a person ahead of you, e.g. in the

same career, sport, hobby or field of studies, as opposed to *kouhai* ‘junior’), inserted from Japanese. This person reference occurs in both Estonian and English utterances (used on 24 occasions). It can be seen as a replacement for the other person’s (given) name (see Examples 1a, 1c and 1d), or also before it (see Example 1b), sometimes in capitalization (1b), in some cases with elongations (1c), and in some cases also in Japanese script (1d). *Senpai* is not seen suffixed to the name in this data.

- (1a) B: *Sai väga palju abi ;__;* *thank you **senpai***
‘I got a lot of help’ [teary eyes smiley] ‘thank you **senior**’ [heart emoji]
- (1b) B: *PALJU ÕNNE SÜNNIPÄEVAKS KALLIS **SENP**AI [GIVEN NAME]!!!!* *Soovin sulle ainult parimat*
‘happy birthday dear **senpai** [given name]’ ‘I wish you all the best’ [heart emoji]
- (1c) B: *Aitäh kallis **senpai**!!! // **Senpaaaaaaaaaaaaai**, how are you?*
‘thank you dear **senpai**’ [heart eyes emojis] // ‘**senpai**, how are you?’
- (1d) *Aitäh 先輩!!!! Year 22 here I come!*
‘thank you [**senpai**]’ ‘year 22 here I come’

It can be inferred that Informant B senses a pragmatic difference, understanding that the honorific *senpai* is prevalent in usage in Japan(ese) when referring to someone who has more experience and/or knowledge in a field the speaker is also involved in, as they have also stated in the interview (Question 1, 4.1), while no equivalent distinction exists in Estonia(n). There do exist words for *senior* and *upperclassman* in both English and Estonian (*seeni*or ‘senior’, *vanem kolleeg* ‘older colleague’); however, they are not used in this data, perhaps, because they are not used in similar contexts in Estonian or English and do not denote this type of hierarchical (and generally) respectful attitude.

Generally the *senpai* would also be older, although in the context of hobbies, such as sports, it may not always be the case. In

this case, Informant A is two years older than informant B and two years ahead in studies (they mistakenly stated *a year above* in the interview), which matches the Japanese distinction of *senpai-kouhai* (seniors vs juniors) as Informant B themselves has noted (*you were a literal senpai*).

The conversation can be seen here in consecutive Examples 2a and 2b (although 2b is third person) from February 2016, where Informant B first asked Informant A about the appropriate way they should address them. This came up in a conversation where Informant B had referred to Informant A as *sensei* ‘teacher’ while speaking in Japanese, and Informant A had been surprised by that usage (seen in Example 2a). Informant B then proceeded to explain that they do not like choosing and asked Informant A to decide between *sensei* ‘teacher’ and *senpai* ‘senior’ (seen in example 2b; although using an imperative request *~te kudasai*, which should technically be used by someone hierarchically higher, not lower). In this case, the one who is referred to as *sensei* is not an actual teacher but a former student of Asian Studies by that time. Following a metalinguistic discussion on how to say ‘what do you think?’ in Japanese, Informant B says [...] *arigatou sensei* ‘thank you **teacher**’ (see Example 2a) as it was common for Informant B to ask for advice from Informant A regarding grammar, etc. They had been referring to Informant A as *senpai* before this incident, as well.

- (2a) B: はい, 分かります, ありがとうございます先生 [*hai, wakarimasu, arigatou sensei*]

‘Yes, I understand, thank you **teacher**’

A: 先生ってね xD [*sensei tte ne*]

‘**teacher**, huh?’ [smiley]

- (2b) B: 私は決めることが好きじゃありませんから... 先生/先輩は決めてください。

[*watashi wa kimeru koto ga suki ja arimasen kara... sensei/senpai wa kimete kudasai*]

‘I don’t like deciding, so... please choose **sensei/senpai**’

A: >-< // んじゃあ、先輩でいいかな [njaa, *senpai* de ii kana] // 正式に先生じゃないから [seishiki ni *sensei* ja nai kara] [...] [smiley] // ‘well then, I guess I’ll go with *senpai*’ // ‘as I’m not officially a *teacher*’

There was one other instance (Example 3) where Informant B addressed Informant A as *sensei* ‘teacher’, in which case there were no comments about this usage. It can be seen that this is an instance of asking for advice regarding grammar, which may have prompted the usage of *sensei*.

- (3) B: *sensei* // how do I say // I like Japan more than before” in japanese?
 ‘*teacher*’ // how do I say // I like Japan more than before” in japanese?
 A: 前より日本が好きになった [mae yori nihon ga suki ni natta] //
 ? // i think
 ‘I’ve come to like Japan more than before // ? // i think’

While in Estonia, for example, it is relatively common to agree upon whether to use polite *you* (*sina* ‘you’, second person singular, or *Teie*, second person plural, polite form) or not, depending on familiarity and preference, and also institutional culture, in Japan it is not common to ask as the usage of honorifics (and registers) is generally inferred based on objective hierarchies and their adequate assessment. It also depends on factors such as the situation (formal or informal) and emotional distance, which may change, and thus the usage of honorifics, and also polite grammatical style, can vary and change over time. Variation can also be seen in the case of informants A and B, although *senpai* is much more prevalent than *sensei* (see Table 1 in Section 5.1).

4.3. USAGE IN THIRD PERSON

The next case, Example (4) of *sensei* ‘teacher’ used by Informant A, is an instance of descriptive usage. It is used in third person to describe a language teacher at the university (of Japanese nationality), likely following the Japanese tradition of referring to teachers using this honorific (which they are also instructed to do in class), although it is used towards Japanese language teachers of Estonian nationality, as well. In speech it is seen both in usage directly with teachers (as honorific usage) and when talking about teachers without them being present (descriptive usage). As mentioned in 4.1, Informant A was not aware of the difference between honorific and descriptive usage at the time.

- (4) A: *sensei* küsis kumb on suurem, kas 巨大な [kyodai na] või mingine teine mida ma ei mäleta
 ‘teacher asked which was bigger, gigantic or some other one that I don’t remember’

In addition to referencing people known to the informants, there are also cases of metalinguistic discussions, such as in Example (5). This example is in reference to a character in the *Naruto* series, specifically talking about how in the animated version his name is spelled in two ways, depending on the translator. In Japanese script, the spelling follows the pronunciation (Maito Gai in *katakana*; マイト・ガイ), while there are two Latinized versions (Might Guy or Maito Gai).

- (5) A: *seda on tegelt suht palju // et nimesid kuuldakse valesti // ka see vahe nt kas on gai-sensei või guy-sensei [...]*
 ‘there’s quite a lot of this, actually // that names are heard wrong // also if it’s **gai-sensei** or **guy-sensei** etc’ [...]
 [...]

 B: [...] Ma olen seda **Guy sensei** asja näinud, aga vist kuidagi ei registreerinud. [...]

 ‘I have seen this **Guy sensei** thing, but I guess I haven’t registered it.’ [...]

There is also one anecdotal case (Example 6), where Informant A had been memorising politics-related vocabulary for work purposes. In this case, Informant E, who has never studied Japanese and has no knowledge of this type of vocabulary, suggests using *prime-sensei*, literally 'prime teacher', instead of the Japanese word *souridaijin* 'prime minister'. As such a word is not something that could be used with native Japanese, it can be inferred to be humorous.

- (6) A: *souridaijin* // *souridaijin souridaijin*
 'prime minister' // 'prime minister prime minister'
 [...]
 A: *prime minister xD*
 E: ***prime-sensei***
 'prime-teacher'
 A: *lol*
 'laughing out loud'

While *sensei* is more often used by students in spoken language, especially vocatively, and there are not many cases of it in this data, it is still used in different forms (both as a replacement for the name and as a suffix). The opposite order (*sensei* + name; prefixed), like in the case of *senpai*, is not seen in this data (which would also be ungrammatical).

5. Discussion

5.1. THE LIMITATIONS OF THE DATA

As this data included only seven informants and a specific situation (friendly, informal), one of the terms (*senpai* 'upperclassman') was only used by one informant (Informant B) and by one other informant in a metalinguistic context (Informant A). If more data were available, it might be possible to see *senpai* in usage by other people and in other settings. However, Informant B used the term *senpai* very consistently (in 24 occasions within 14 conversations, see Table 1 below). A metalinguistic conversation about the usage of

the term is also available in this data. They have also provided their reasons for wanting to use *senpai* towards Informant A during the semi-structured interviews (4.1), and thus we are able to look at their personal motivation and their own perception.

Table 1. *Second (2P) and third person (3P) usages of sensei and senpai*

	A	B	C	D	E	F	G
Sensei 2P	–	2	–	–	–	–	–
Sensei 3P	1 reference, 2 meta-ling.	2 meta-ling.	–	–	1 (humour)	–	–
Senpai 2P	–	8 (Jp script) 15 (Latin)	–	–	–	–	–
Senpai 3P	3 meta-ling.	1 meta-ling.	–	–	–	–	–

Additionally, the term *sensei* ‘teacher’ was mostly used descriptively (3P) as these conversations are between friends, and not between ‘teachers and students’. If, for example, classroom (spoken) usage or conversations with natives were observed, honorific usage could be seen, as well. However, within lessons it is generally encouraged to speak in monolingual Japanese, not mixing it with Estonian, for example, and thus insertion may not be very common, depending on the particular students (and the amount of moderation by their teacher).

On another note, usage of *senpai* deviates grammatically and does not comply with traditional Japanese register norms (and is technically not classifiable as honorific usage even though it is intended to be *appreciative* and *formal* by Informant B, among other factors). As Informant B, who used *senpai* frequently, does not feel a difference between the different forms, they themselves did not

differentiate between descriptive and honorific usage (see Section 4.1), and thus a binary distinction between the forms may not be applicable in this case (at least regarding the intention), and thus distinctions were made mainly between usage in second person and third person (vocative or not, respectively).

5.2. GRAMMATICAL CORRECTNESS

The instances in this data are not always used conventionally, at least in an honorific sense. Honorific usage requires the use of the surname, not the given name, or just the honorific alone. It is uncommon to use the person's surname vocatively in Estonian, and the use of the given name is preferred, which may have influenced the name choice. The honorific should also follow the name and cannot be used as a prefix. In honorific usage elongation is not socioculturally appropriate as honorifics are generally supposed to denote respect. The usages of emojis, exclamation marks and capitalizations also indicate that the usage is not conventionally respectful but rather follows the style of playful usage often seen in media. As seen from the interviews (see Section 4.1), Informant B is not aware of these factors, and according to their own perception, there was no difference in intention or meaning depending on the form that *senpai* was used in. Although their perception is different from native Japanese, this does not inhibit insertion as this is a safe environment, where linguistic play is common and where grammatical and cultural correctness is not inherently important.

5.3. POSSIBLE FURTHER RESEARCH

There are many possible future research directions, including a more general analysis that includes person deixis in this type of multilingual communication in general, not just the cases where they occur as insertions from Japanese (focusing, for example, on how Estonians use the terms when speaking in Japanese in comparison

with native speakers and whether they use terms that are sociopragmatically appropriate in Estonian but not in Japanese, such as *anata* ‘you’). Another option would be to look at the pragmatic differences from other angles, like has been done in the case of Estonian and Swedish (Keevallik 2006, 2012), for example, regarding greeting and farewell sequences, where there are certainly differences between Estonian and Japanese, as well. There are also further differences in expressing attitudes and emotion more generally, including phenomena such as language play and humour (also noted by Dezi 2022, about Russian and Italian), as well as (sentence-final) particles (see e.g. Estigarribia 2021, about Paraguayan Spanish) and others.

Conclusions

The aim of this work was to examine which person references are inserted from Japanese to Estonian and/or English utterances, in which forms, and what the reasons are behind the cases. These data show that a perception of pragmatic differences between the Japanese and Estonian (or English) languages and cultures causes the insertion of the Japanese person references *senpai* ‘senior’ and *sensei* ‘teacher’ in second and third person. The person reference *senpai* is used vocatively (in second person) in various forms (prefixed to the name, replacing the name, elongated, capitalised, in the Latin alphabet, in Japanese script) in Estonian and English utterances, while *sensei* is used mainly descriptively (in third person) replacing the person’s name or as a suffix (except for two occasions where it was used in the vocative case). Both are also seen in metalinguistic contexts.

The reasons behind the instances of *senpai*, according to the interviews, are *to show appreciation* and *to sound more formal*, although these factors were not consciously thought of or planned beforehand. The aim was also to imitate native usage, with an influence of the Japanese media, and to show how the conversation partner is a more knowledgeable *literal senpai* (vertical hierarchy). The

addition felt *natural* and *cool*, according to the interviews, and also *fun and interesting*, which indicates an element of language play. There is also an awareness of a similar experience and kinship (solidarity), while accompanied by sociocultural awareness of what is (perceived to be) appropriate in this type of relationship (situational salience). The usage of *sensei* was rare since there were no literal student-teacher relationships between any of the informants, and *sensei* is most often seen in spoken language.

While this work covers a rare, understudied phenomenon, more qualitative, usage-based research is needed to get a broader understanding of the cognitive reasons and intentions behind code-switching (in broader terms) that is motivated by sociopragmatic differences.

REFERENCES

- Backus, Ad 2015. A usage-based approach to code-switching: The need for reconciling structure and function. – Code-switching between structural and sociolinguistic perspectives. Eds. Gerald Stell, Kofi Yakpo. Berlin, München, Boston: Mouton de Gruyter, 19–37.
- Dezi, Alessandra 2022. “Рефлексивная” языковая игра в интернет-дискурсе русскоязычных жителей Италии. [Reflexive language play in the Internet discourse of Russian speakers living in Italy]. – *Studia Slavica XIX*. Eds. Jekaterina Jabloкова, Inna Adamson. Tallinn: Tallinn University Press, 235–248.
- Dorleijn, Margreet 2016. Introduction: Using Multilingual Written Internet Data in Code-Switching and Language Contact Research. – *Journal of Language Contact* 9, 5–22.
- Estigarribia, Bruno 2021. A Speech Planning Account of Guarani Grammatical Borrowings in Paraguayan Spanish. – *Journal of Language Contact* 13 (3), 663–696.
- Hoffmann, Charlotte 2001. Towards a Description of Trilingual Competence. – *International Journal of Bilingualism* 5, 1, 2001, 1–17.
- Howell, Peter 2007. Character voice in anime subtitles. – *Perspectives* 14 (4), 292–305.

- Irgens, Benedicte Mosby 2017. Person Deixis in Japanese and English – A Contrastive Functional Analysis. PhD Dissertation. University of Bergen.
- Johanson, Lars 2002. Contact-induced change in a code-copying framework. – *Contributions to the Sociology of Language* 86, 285–314.
- Kaiser, Stefan; Ichikawa, Yasuko; Kobayashi, Noriko; Yamamoto, Hilofumi 2013. Japanese: a comprehensive grammar. Routledge comprehensive grammars. 2nd edition. Routledge.
- Keevallik, Leelo 2006. Keelekontakt ja pragmaatika [Language contact and pragmatics] [conference paper]. – *Teoreetiline keeleteadus Eestis II* [Theoretical Linguistics in Estonia II]. Toim. Ilona Tragel, Haldur Õim. Tartu: Tartu Ülikooli Kirjastus, 85–96. <https://liu.se/ikk/medarbetare/leelo-keevallik/dokument/1.413469/Kontakt.pdf> (25.10.2022).
- Keevallik, Leelo 2012. Pragmatics of the Estonian heritage speakers in Sweden. – *Finnisch-ugrische Mitteilungen*, 1–22.
- Kilp, Geidi 2021. Code-copying in Estonian-English-Japanese Facebook Communication: a Usage-Based Approach. – *Multilingual Practices in the Baltic Countries. Acta Universitatis Tallinnensis*. Ed. Anna Verschik. Tallinn: Tallinn University Press, 176–211.
- Marchello-Nizia, Christiane 2006. From personal deixis to spatial deixis. – *Space in languages: linguistic systems and cognitive categories*. Amsterdam, Philadelphia: John Benjamins, 103–120.
- Pajusalu, Renate 2009. Pronouns and reference in Estonian. – *Language Typology and Universals* 62 (1–2), 122–139.
- Pajusalu, Renate; Vihman, Virve; Klaas, Birute; Pajusalu, Karl 2010. Eestlaste ja venelaste suhtluskäitumine: *sina*, *teie* ja keegi veel. – *Eesti Rakenduslingvistika Ühingu aastaraamat 6* [Estonian papers in applied linguistics]. Tallinn: Eesti Rakenduslingvistika Ühing, 207–224.
- Schegloff, Emanuel Abraham 2007. Categories in action: person-reference and membership categorization. – *Discourse Studies* 9 (4), 433–461. <https://doi.org/10.1177/1461445607079162>
- Stivers, Tanya; Enfield, Nick J.; Levinson, Stephen C. 2007. Person reference in interaction. – *Person reference in interaction: Linguistic, cultural, and social perspectives*. Cambridge: Cambridge University Press, 1–20.

- Takahara, Kumiko 1992. Second person deixis in Japanese and power semantics. – *Intercultural Communication Studies* 11 (1), 117–128.
- Verschik, Anna 2010. Ethnolect debate: Evidence from Jewish Lithuanian. – *International Journal of Multilingualism* 7 (4), 285–305.
- Verschik, Anna 2019. English-Estonian code-copying in blogs: combining contact linguistic and cognitive approach. – *Cognitive Contact Linguistics*. Eds. Eline Zenner, Ad Backus, Esme Winter-Froemel. Berlin, Boston: Mouton de Gruyter, 51–80.
- Weinreich, Uriel 1953. *Languages in Contact*. The Hague: Mouton.
- Yui, Megumi 2012. The function of address terms in English and Japanese: Analysis using scenarios. – *IAPL Journal* 1, 59–73.
- Zenner, Eline; Backus, Ad; Winter-Froemel, Esme 2019. Placing usage, meaning and mind at the core of contact-induced variation and change. – *Cognitive Contact Linguistics*. Eds. Eline Zenner, Ad Backus, Esme Winter-Froemel. Berlin, Boston: Mouton de Gruyter, 1–20.

RESÜMEE

PRAGMAATILISTEST ERINEVUSTEST AJENDATUD ISIKUVIIDETE SISESTAMINE EESTI-INGLISE-JAAPANI FACEBOOKI VESTLUSTES

Geidi Kilp

Tallinna Ülikool

Sotsiopragmaatilisi erinevuseid on uuritud erinevate keelte ja kultuuride vahel, kaasa arvatud inglise ja jaapani. Küll aga ei ole varem võrreldud eesti ja jaapani keelt. Antud uurimus eesti-inglise-jaapani Facebooki suhtluse kohta täidab seda lünka, analüüsides jaapanikeelsete isikuviidete sisestamist eesti- ja ingliskeelsetesse lausungitesse eestlaste poolt, mis tuleneb sotsiopragmaatilistest erinevustest nende keelte ja kultuuride vahel.

Kasutatud materjalis on Facebooki privaatvestlused vahemikus 2015–2021, kokku 50 vestlust ja 14,681 sõnet. Materjalis on 7 keelejuhti, ja igas vestluses on kaks osalejat. Lähenemine on kasutuspõhine ja analüüs on kvalitatiivne, keskendudes üksikisikutele ning nende tajule, arvestades nii nende endi kui ka vestluste tausta. Kahe osalejaga on ka läbi viidud poolstruktureeritud intervjuud.

Materjalist ning intervjuudest võib järeldada, et pragmaatilised erinevused põhjustavad jaapanikeelsete isikuviidete *senpai* 'seenior' ja *sensei* 'õpetaja' sisestamist erineval kujul (eel- ja järelliitena, nime asendusena, pikendatult, läbiva suurtähega, ladina tähestikus, jaapani kirjasüsteemis). Olulised kognitiivsed mõjufaktorid on vertikaalne hierarhia, horisontaalne solidaarsus ning (situatsioonipõhine) esilduvus.

Võtmesõnad: deiksis, koodivahetus, vokatiiv, kasutuspõhine lähenemine, kontaktlingvistika, netipõhine suhtlus, esilduvus, eesti keel, inglise keel, jaapani keel

Geidi Kilp is a doctoral student at Tallinn University whose research interests include Estonian-Japanese language contacts, computer-mediated communication and code-switching.
geidi.kilp@tlu.ee

THE BILINGUAL ADVANTAGE: PERFORMING THE NON-WORD REPETITION TEST

Eglė Krivickaitė-Leišienė, Ineta Dabašinskiė

Vytautas Magnus University

Abstract. The study investigated the accuracy of non-word production by bilingual and monolingual children. The participants (125 children in total) belonged to two groups of bilingual children with different language repertoires and one group of monolingual Lithuanians. The analysis revealed that the overall performance of both bilingual groups was better than in the monolingual group. The bilingual children demonstrated more accurate and statistically significant results in repeating longer and structurally more complex non-words. The findings of this study suggest that the bilinguals being acquainted with two phonological systems had a greater experience with diverse phonology, which ensured a more precise performance of the task.

Keywords: bilingual and monolingual children, language acquisition, Lithuanian, English, Russian

Introduction

Research in the field often reports that bilingual children with migration experiences have some disadvantages. For example, they often do not reach the developmental milestones in their linguistic competence of L1 at the same pace as monolingual children or have difficulties acquiring L2, the dominant language of the society (Paradis 2010). It is also observed that, compared to monolingual children, bilinguals often perform linguistic tasks more poorly (Gibson, Jarmulowicz, Oller 2019).

Recently bi- or multilingual literacy acquisition at the primary school has become a focus of extensive research, and oral language

proficiency at school entry has been reported to be a crucial indicator of literacy development in bilinguals well before they start to read (Bialystok 2002; Silven, Lunden 2011). Although research on bilingual language acquisition has raised many questions, the primary aim of this study is to contribute to the controversial debate on bilingual advancement. For that purpose, we will present the results of a specific linguistic task (non-word repetition test, NWR) performed by two groups of sequential bilingual and one group of monolingual children.

The relevant studies emphasise differences between bilingual children with regard to their profile. Montrul (2013) distinguishes three profiles: (1) simultaneous bilinguals (i.e. those exposed to the heritage and the majority language before the age of 5); (2) sequential bilinguals or child L2 learners (i.e. the ones exposed to the heritage language at home until the age of 4–5 and to the majority language once they start preschool); and (3) late child L2 learners (i.e. children monolingual in the heritage language who received some elementary schooling in their home country and immigrated around 7–11 years of age) (Montrul 2013: 284). However, the classification is not always clear-cut. The age of acquisition and the type and amount of L1 at home and L2 outside are relevant variables for understanding the linguistic abilities of bilinguals. However, it is not always obvious how to measure the degree of languages a child is exposed to. The languages used at home and outside influence the type of bilingualism and literacy skills and, ultimately, academic achievements. The data obtained from different studies (Pearson 2007; Leseman, van Tuijl 2006) propose that the balanced use of languages and a child's regular involvement in joint reading or other interactional activities may have a long-term impact on academic attainments and personal satisfaction.

In contrast, due to restricted home settings in early childhood, L1 speakers may have reduced access to L2 and experience difficulties in linguistic abilities compared to titular language speakers (Kondo-Brown 2004; Montrul 2011; O'Grady 2011). Moreover, in

language development paths, diverse and deviant or nonstandard performance can be observed not only in L2 but also in L1; however, this result should not be considered as a deficit in language acquisition but as a unique stage in language development (cf. Gathercole 2013; Paradis *et al.* 2011). Thus, it is essential to compare bilinguals and monolinguals not only to confirm deviations from the monolingual “norm” repeatedly but also to determine if these groups demonstrate unique or specific characteristics when performing certain tasks.

The non-word repetition test is considered an important measure in monitoring the child’s language development. At a young age, the child’s ability to repeat a new polysyllabic word that she/he hears for the first time shows her/his ability to learn new words later and extend the lexicon (Gathercole 2006). Our study attempts to investigate linguistic performance in two groups of sequential preschool bilinguals (i.e. Russian-Lithuanian and Lithuanian-English) and one monolingual Lithuanian group to identify specific patterns characteristic of these groups in the completion of pronunciation tasks.

1. Non-word repetition test: theoretical assumptions

The non-word repetition test is an experimental method when the respondent is asked to repeat non-words¹. In order to be able to repeat the word which is heard for the first time and does not have any meaning, linguistic-cognitive abilities (phonological processing, short-term memory, articulation abilities, etc.) are necessary (Rispen, Parigger 2010). Each word that the child has heard for the first time some time ago sounded unusual and strange, similar to the words in this test (Chiat, Roy 2007). The results of longitudinal

¹ A non-word is a phonological sequence of sounds which corresponds to phonotactic rules of a specific language and do not have any meaning and function in a sentence.

research have demonstrated that the children who performed better in the test after a year had a broader lexicon than those who performed worse (Gathercole 1995; Baddeley *et al.* 1998).

The research conducted in different languages and on a different population (monolingual vs bilinguals) does not demonstrate a straightforward result. Some studies show rather similar bilinguals' and monolinguals' non-word repetition performance. For instance, one study compared English non-word repetition accuracy of 7-year-old monolingual English and bilingual Korean–English, Chinese–English, and Spanish–English children. The results demonstrated similarity between the monolinguals and bilinguals – there was no statistically significant difference in performance (Lee, Gorman 2012). Another study comparing children aged from 3 to 5 years old (30 Korean–English sequential bilinguals and 30 Korean monolinguals) also did not find any statistical difference between the two groups (Lee, Kim, Yim 2013). Russian–Hebrew bilingual children (4;5–6;6), Hebrew monolingual children (4;6–6;6), and Russian monolingual children (4;0–6;0) were tested with the same task but did not show any differences either (Armon-Lotem, Chiat 2012). A study of migrant children in preschool- and school-age (mean age 9;4) with L2 German (different L1 languages: Russian, Turkish, and Urdu) confirmed the same performance of monolingual and bilingual children (Grimm, Hübner 2016). French-speaking monolingual children in grades 3 and 6 and bilingual children have exhibited the same tendency in accuracy results in the non-word repetition test (Thordardottir, Reid 2022). A study on school-age (around 11 years old) bilinguals who live in Iceland and attend Icelandic schools but who speak a language other than Icelandic at home (different L1 languages: Polish, Lithuanian, Ukrainian, etc.) demonstrated very high scores on an Icelandic non-word repetition test (Thordardottir, Juliusdottir 2012).

Other studies report bilingual advantage in repeating non-words in first language (L1). Greek children learning English as a second language were more accurate repeating non-words in their

native language (Greek) than in the second one (English) (Masoura, Gathercole 1999). Additionally, Summers *et al.* (2010) found that Spanish-English bilingual children aged 4;6 and 6;5 produced the Spanish-like non-words more accurately than the English-like non-words. The study (Gibson *et al.* 2014) of 52 English-Spanish five-year-old children (26 Spanish-dominant and 26 English-dominant) conducted for English and Spanish showed that the Spanish-dominant group performed better than the English-dominant group for both Spanish and English non-words. The authors claim that not only language experience but also phonological structure has effects.

There are also studies showing worse performance on non-word repetition tests by bilinguals than monolinguals. Researchers have found that bilingual Spanish-English children aged 7;10–13;11 (Kohnert *et al.* 2006) and 6;0–11;6 (Windsor *et al.* 2010) performed significantly below monolinguals. Messer with colleagues (2010) found that Turkish-Dutch 4-year-olds had lower scores than their Dutch monolingual peers in a Dutch non-word repetition test, but higher scores on a Turkish test, reflecting differences in language experience within the two groups. Another study compared the results of 44 bilingual children with various European languages as their L1 and the performance of the Luxembourgish non-word repetition test by monolingual children. The data showed that the monolinguals performed significantly better than the bilingual group (Pascale 2011). As we observe, the differences in performance of diverse populations are mainly related to the language experience and phonological sensitivity (familiarity); however, we believe, there are many more factors influencing the performance of non-word repetition.

2. The structure of non-word repetition test

Comparative research using non-word repetition tests designed for different languages allows distinguishing three main factors which influence the accuracy of non-word repetition:

1. The complexity of non-words (CV vs CCV). Non-words with consonant clusters are repeated less accurately than non-words without consonant clusters (Kavitskaya *et al.* 2011). Word-medial and word-final clusters are repeated less accurately than word-initial consonant clusters (Marshall, van der Lely 2009; Krivickaitė 2014; 2017).
2. The non-word length (the number of syllables in the word). Non-word repetition accuracy declines with the increasing number of syllables; one- or two-syllable words are uttered more accurately than three- or four-syllable non-words (Chiat, Roy 2007). It is related to the ability to keep phonological information in short-term memory. The length effect has been identified in a variety of languages, such as English (Dollaghan, Campbell 1998; McDonald, Oetting 2019), Italian (D'Odorico *et al.* 2007; Piazzalunga *et al.* 2019; Farabolini *et al.* 2021), Spanish (Girbau, Schwartz 2007; Windsor *et al.* 2010), Swedish (Radeborg *et al.* 2006), Dutch (Messer *et al.* 2015), Cantonese (Stokes *et al.* 2006), Gulf Arabic (Shaan 2020), Czech (Sileo, Tyčová 2019), Lithuanian (Krivickaitė 2014, 2017).
3. The age of participants: the older the group, the more accurately both shorter (one–two syllable) and longer (three–four syllable) non-words are repeated (Santos *et al.* 2006; Park, Scarz 2012). Older children have a larger and more developed lexicon; they are also more exposed to and experienced with different sound clusters and thus can produce various consonant clusters more accurately (Munson *et al.* 2005).

2.1. THE STRUCTURE OF THE LITHUANIAN NON-WORD REPETITION TEST²

The Lithuanian non-word repetition test was designed following the structural characteristics of Lithuanian words (word length and syllable structure) (see Kazlauskienė 2007; Kazlauskienė, Raškinis 2008a, 2008b; Kazlauskienė 2010; Girdenis, Karosienė 2010). The test consists of 24 non-words with a different structure: eight non-words have two syllables (4–6 phonemes), eight non-words have three syllables (6–7 phonemes), and eight non-words have four syllables (7–8 phonemes). There are two non-words without consonant clusters and six non-words with consonant clusters in each group (see Table 1). In terms of word length and syllable structure, each non-word is associated with a Lithuanian true word equivalent.

Table 1. *Non-word items and their syllable structure*³

2-syllable non-words	3-syllable non-words	4-syllable non-words
k e m u CV.CV	g e l o f a CV.CV.CV	s u l e r i t e: CV.CV.CV.CV
d o j æ CV.CV	ʃ i r u t a CV.CV.CV	ž a d e v i n a CV.CV.CV.C
s k i m o CCVCV	ʃ k u l i n e: CCV.CV.CV	s n a l i d i n a CC.CV.CV.CV
ʃ v e l a CCV.CV	p l e m u t a CCV.CV.CV	s p i r a t u f a CCV.CV.CV.CV
g a : p r e: CV.CCV	m a : s p u l e: CV.CCV.CV	n i s p a r i m a CV.CCV.CV.CV
g i t v a CV.CCV	l a s m u v i CV.CCV.CV	m a g v u n o l e: CV.CCV.CV.CV
s m i n t o CCVC.CV	s p a : d æ k i CCV.CV.CV	s t a l i g o s a CCV.CV.CV.CV
k l e s t a CCV.CCV	p a : s v a p i CV.CCV.CV	g o s a k l u : n i CV.CV.CCV.CV

² The Lithuanian non-word repetition test (Dabašinskienė, Krivickaitė 2013) was developed while participating in the COST project IS0804 Language Impairment in a Multilingual Society: Linguistics Patterns and the Road to Assessment (2009–2013).

³ Syllables are separated by dots; C – consonant, V – vowel.

The Lithuanian non-word repetition test was presented in a short (lasting only three–four minutes), easy, friendly and attractive game format using MS PowerPoint⁴. The child was introduced to the main character of the game, a monkey, who wants to get some bananas and has to complete the tasks. The child was asked to help the monkey and do the tasks. In each step, the child heard a recorded non-word, which she/he had to repeat. The performances were recorded. Additionally, the protocol was used to mark and comment on the child's utterances.

2.2. THE SCORING METHODOLOGY

Considering debates and arguments on the diversity of scoring methodologies of the non-word repetition tests, we decided to focus on three main elements to register inaccuracies in a child's productions.

1. *The whole item or general accuracy.* Each item was scored as either correct or incorrect. Any child's production deviant from the original (in regard to length and structure), such as the omission, addition or replacement of a sound or a syllable, was scored as incorrect. The answers were considered to be wrong if an additional sound was added, for example, *g r a: p r e:* (instead of *g a: p r e:*), or if a sound was substituted, for example, *g e g o f a* (instead of *g e l o f a*), etc. The answer was regarded to be correct only if the word was repeated absolutely precisely.

2. *Word length.* Each item of two-, three- and four-syllable stimuli was scored as correct if a child produced the same number of syllables as in the target word. The answers were treated as wrong if (1) the word became one syllable shorter because of an omitted sound, for example, *f k u l n e:* (instead of *f k u l i n e:*); (2) if the whole syllable was omitted, e.g. *s p a r i ma* (instead of *n i s p a r i ma*);

⁴ The visual design of non-word repetition test produced by Kunnari, Tolonen, and Chiat (2011).

or (3) an additional syllable was added, as in *dɔ l u j æ* (instead of *dɔ: j æ*), etc.

3. *Syllable structure (consonant clusters)*. Each item was scored correct if the cluster was repeated as in the original word. For example, *t α: p r e:* (instead of *g a: p r e:*) was counted as a correct answer because of the production of the consonant cluster; however, *k ɪ m o* (instead of *s k ɪ m o*) was counted as an incorrect because one element of the cluster was omitted.

To carry out the quantitative analysis, the data were coded manually and analysed using the SPSS 22 (*Statistical Package for the Social Sciences*) programme. In order to check statistically significant differences, an analysis of variance and the post hoc criterion were applied. The level of statistical significance was set at 5% (0.05).

2.3. PARTICIPANTS

Three groups of children (mean age 6 year) participated in the current study. Group 1 ($n = 50$) was comprised of monolingual Lithuanian children from the city of Kaunas. The members of this group attended a state kindergarten daily. The children from Group 2 ($n = 50$) were sequential bilinguals (L1 – Russian, L2 – Lithuanian) living in Kaunas and Vilnius who attended a state kindergarten for minority children with Russian as the main language of instruction and had 3–4 hours of weekly Lithuanian classes. As reported by the parents, all bilingual children used Russian as their first language, and this language was dominant at home. Group 3 ($n = 25$) were sequential bilinguals (L1 – Lithuanian, L2 – English) born in the UK or taken to London at around one year of age. They went to a state kindergarten with English as the main language and attended a Lithuanian school on Saturdays (3–6 hours per week).

All the children were typically developing (TD) and were selected for the study with their teachers' help; none of the children had records of language delay or impairment.

Table 2. *Participants*

Group	Number	Gender		Mean age
		Male	Female	
MO (LT)	50	26	24	6;1
BI (RU-LT)	50	27	23	6;3
BI (LT-EN)	25	9	16	6;0

3. Results

The test results demonstrated that both bilingual groups repeated non-words better than the monolingual participants: the bilinguals' accuracy of repeating non-words was 75%–76%. In comparison, the accuracy of the monolinguals was 69% (see Figure 1). The statistical analysis shows that the bilingual children repeated non-words significantly better than the monolinguals ($p=0.004$).

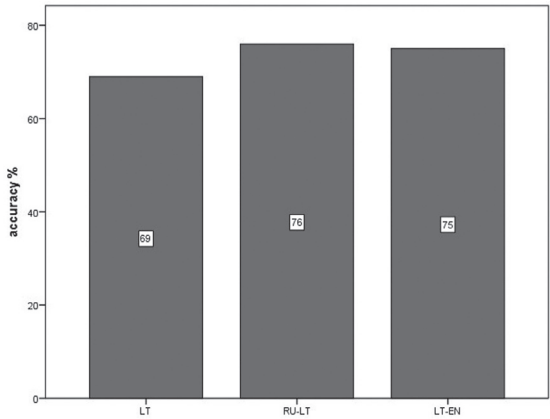


Figure 1. *The general results of the non-word repetition test: MO (LT) vs BI (RU-LT) vs BI (LT-EN)*

3.1. THE WORD STRUCTURE: GENERAL ACCURACY, WORD LENGTH, AND SYLLABLE STRUCTURE (CLUSTERS)

3.1.1. GENERAL ACCURACY

The general analysis of accuracy looks at a child's performance from two angles: first, it investigates how precisely a child can retain the number of syllables in a word and second, how she/he manages to produce more complicated structures, such as consonant clusters. Thus, we expected a child to be most precise when uttering the original word.

As regards the word length, in general, the bilingual children repeated two-, three, and four-syllable non-words better than the monolingual (see Figure 2). The data analysis indicates that all the groups repeated two-syllable non-words with 89%–93% accuracy. Longer non-words were more difficult to repeat correctly than the shorter ones for all the groups. Statistically, three- and four-syllable non-words were repeated notably worse than two-syllable non-words ($p=0.000$). The most significant difference in the results is seen in the production of two-syllable and three-/four-syllable non-words in all participant groups. The RU-LT group repeated three- and four-syllable non-words with similar accuracy (70%–73%), while the monolingual and LT-EN groups repeated four-syllable non-words much worse than three-syllable non-words. The monolinguals produced them with 56% and 75% accuracy and LT-EN bilinguals with an accuracy of 65% and 72%, respectively.

The repetition of three-syllable non-words displays similar results between the monolingual and bilingual groups: the monolinguals repeated non-words with 75% accuracy, while the bilinguals repeated them with 73%–72% accuracy (see Figure 2). So, it can be noted that there is no statistically significant difference between the groups.

The analysis demonstrates that the bilinguals repeated four-syllable non-words more precisely than the monolinguals: the RU-LT group's accuracy was 70%, the LT-EN group's score was 65%, and

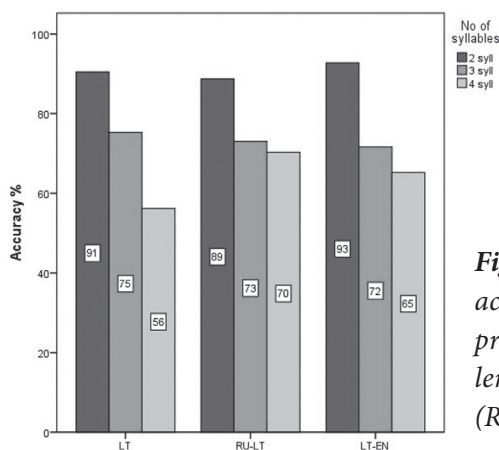


Figure 2. The general accuracy of the non-word production (structure and length): MO (LT) vs BI (RU-LT) vs BI (LT-EN)

the monolingual group repeated four-syllable non-words with 56% accuracy (see Figure 2). A statistically significant difference between the monolinguals and the bilingual RU-LT group ($p=0.000$) was found, while the difference in the results between the monolinguals and the LT-EN bilingual group was not statistically significant.

In order to pronounce words with a more complicated structure, as a rule, children look for ways of facilitating their pronunciation; for instance, they omit consonants with more complicated pronunciation patterns or substitute them with other consonants that are easier to pronounce. Inaccurate pronunciation of sounds or their substitution by other sounds is a natural development of children's language, demonstrating the cognitive processes when learning new words (Dodd *et al.* 2003: 623; Santos *et al.* 2006: 372). However, we registered such changes in pronunciation as incorrect.

3.1.2. THE LENGTH OF NON-WORDS

This section looks only at one parameter – the retainment of the word structure, namely, the child's ability to produce a word in all its length, with all the required syllables. The accuracy in pronunciation of consonant clusters was not considered and measured here

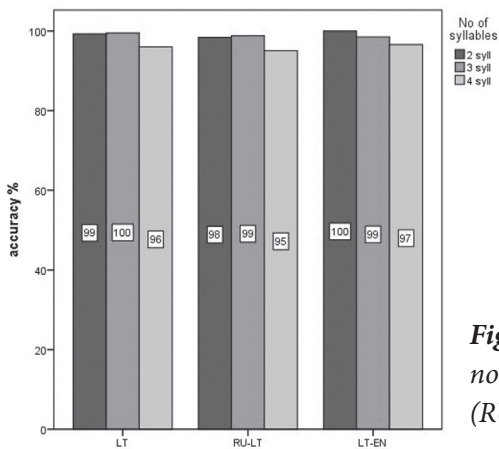


Figure 3. *The length of non-words: MO (LT) vs BI (RU-LT) vs BI (LT-EN)*

(omission, changes, additions). None of the participant groups found it difficult to retain the number of syllables in a word: the length of two- and three-syllable words was retained with the accuracy of 99%–100%, and the length of four-syllable words was retained with the accuracy of 95%–97% (see Figure 3).

3.1.3. COMPLEXITY (CONSONANT CLUSTERS):

INITIAL VS MEDIAL POSITION

The sample included very few words without a consonant cluster: 2 two-syllable words and 2 three- and four-syllable words. Thus, we will only analyse words with a consonant cluster to observe the children's ability to articulate more complex structures.

Clusters usually appear in a word in diverse positions, and our analysis focuses on initial and medial positions. It was observed that clusters in the initial position were repeated significantly more accurately than those in the medial position ($p=0.000$). The results of consonant clusters in the initial position ranged in the interval of 85%–91%, and consonant clusters in the medial position were produced with an accuracy of 72%–85% (see Figure 4).

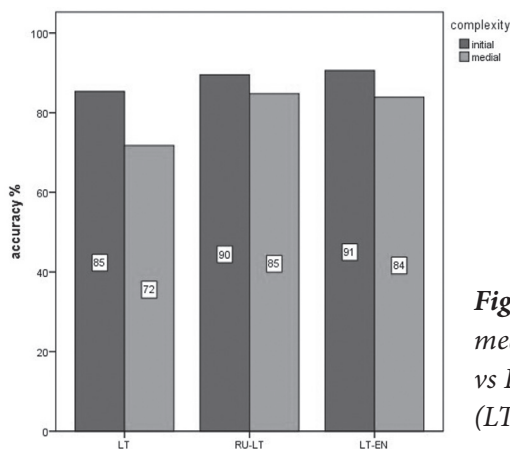


Figure 4. *Initial and medial clusters: MO (LT) vs BI (RU-LT) vs BI (LT-EN)*

Both bilingual groups repeated *initial* and *medial* consonant clusters better than the monolingual group. Thus *initial* clusters were repeated with the accuracy of 85% by the monolinguals and with the accuracy of 90%–91% by the bilingual groups. As for the *medial* clusters, they were articulated with the accuracy of 72% by the monolinguals and with the accuracy of 84%–85% by the bilinguals (see Figure 4). The statistical analysis revealed that clusters in the *medial* position were repeated significantly better by both bilingual groups ($p=0.000$) compared to the monolingual group. However, a statistically significant difference was not registered in the production of initial clusters by all groups.

3.1.4. CLUSTER POSITION AND WORD LENGTH

The general tendency identified in the analysis is that the longer the word, the more difficult it was for all the participants to repeat consonant clusters in both initial and medial positions accurately.

The monolinguals repeated initial clusters in two- and three-syllable non-words similarly (97% accuracy). As regards the bilinguals, their accuracy while uttering the *initial* cluster in two-syllable non-words was 97%–98%, and the respective percentage for three-syllable

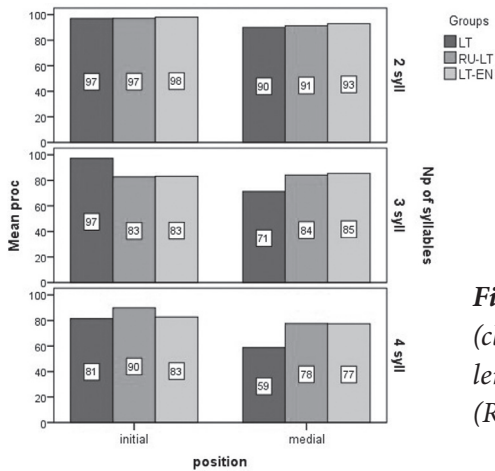


Figure 5. Complexity (cluster position) and word length: MO (LT) vs BI (RU-LT) vs BI (LT-EN)

non-words was 83%. The statistical analysis confirmed that the monolingual group repeated *initial* clusters in three-syllable non-words significantly better ($p=0.019$) than the bilinguals. The RU-LT group repeated *initial* clusters in four-syllable non-words more accurately than in three-syllable non-words (90% and 83%, respectively). The LT-EN group's performance of *initial* clusters in three- and four-syllable non-words was similar (the accuracy of 83%).

Medial clusters in two-syllable non-words were repeated with a similar accuracy by the monolinguals and bilinguals, showing 90%–93% accuracy. The monolingual group repeated *medial* clusters worse than the bilinguals in three- and four-syllable non-words. Thus clusters in three-syllable non-words were produced with the accuracy of 71% by the monolinguals and with the accuracy of 84%–85% by the bilingual participants. As for clusters in four-syllable non-words, their production accuracy was only 59% for the monolinguals and 77%–78% for the bilinguals. The statistical analysis revealed that the monolinguals repeated *medial* clusters significantly worse than the bilinguals in three-syllable ($p=0.002$) and four-syllable non-words ($p=0.000$).

4. Conclusion

To summarise, the results of the Lithuanian non-word repetition test revealed the most complicated aspects of the acquisition of phonotactics by monolingual and bilingual children:

- The performance of all groups was affected by item length and structural complexity.
- It was difficult to accurately repeat non-words longer than two syllables (i.e. three- or four-syllable non-words). In order to perceive and repeat more complex words, more linguistic efforts and additional memory were necessary.
- The cluster position in a word is an important parameter. Clusters in the medial position were repeated worse than those in the initial position, especially in longer non-words (mostly consisting of four syllables).
- The monolingual children scored below the bilingual children in most of the tasks.
- The bilinguals demonstrated better performance of longer and structurally more complex non-words than monolinguals.

As our results provide a bilingual advantage tendency, we would like to briefly discuss some of the particular findings.

The results in this study are mainly interpreted by statistical analysis; however, a closer qualitative analysis is necessary in the future to explain particular cases. As it was presented, the findings revealed that both groups of bilingual children performed better than monolingual children in many parameters, especially the longer words (except for the case of better performance of the MO group in three-syllable words with the initial cluster position). The parameters that were identified as important were the length of the word and consonant clusters. Both bilingual groups repeated four-syllable words with clusters better than the monolingual group, indicating that the bilinguals processed the most difficult structures easier. The results of the word complexity have demonstrated that both bilingual groups repeated consonant clusters more accurately than the

monolinguals, and the results for both parameters were statistically significant.

There were a few cases, mainly in producing two-syllable words, where the results of the monolingual group were similar to the bilingual ones. We assumed that bilingual Lithuanian-English children with Lithuanian L1 would show similar results to monolingual Lithuanians but not higher. It appears that Lithuanian-English bilinguals were exposed to more or less balanced use of both languages (Dabašinskienė *et al.* 2014) as Lithuanian families have demonstrated rather positive attitudes to linguistic integration and heritage language maintenance. The children used Lithuanian at home and attended Lithuanian language classes organised by the Lithuanian community several times per week. Moreover, they have a good knowledge of English as they attend English schools. The Russian-Lithuanian group used Russian at home and attended Russian schools (see Dabašinskienė, Krivickaitė-Leišienė 2019) and had, therefore, little exposure to Lithuanian but performed equally or even better than Lithuanians. As it was already reported by many studies, performance on non-word repetition test is associated with both the structure of a language and a speaker's experience with that language and predicts a child's performance on non-word repetition test (Thordardottir, Juliusdottir 2012; Armon-Lotem, Chiat 2012; Thordardottir, Reid 2022 *etc.*). Despite the fact that Russian was strongly dominating in our RU-LT group due to much less experience in Lithuanian, the results support a bilingual advantage in terms of the children's experiences with both languages. Thus, we assume that bilinguals possess two phonological systems (despite the level of a language competence) and have more diverse experience with phonology, which ensures better performance of the task. This study demonstrates the bilingual advantage only in the very particular non-word repetition test. However, grammar, which is very language-specific, is more difficult to acquire; thus, more erroneous productions are registered in bilinguals' performance (see Dabašinskienė, Krivickaitė-Leišienė 2019).

Language acquisition is affected by differences in the socio-economic, cultural characteristics, the language attitudes of bilingual communities, and the language status of children's L1 and L2. Moreover, children's age, the length and intensity of exposure to their L2 play an important role (Chiat 2015). The study has some limitations, especially, the size of the sample, but also other sociolinguistic and linguistic parameters have to be taken into consideration when conducting future research.

REFERENCES

- Armon-Lotem, Sharon; Chiat, Shula 2012. How do Sequential Bilingual Children Perform on Nonword Repetition Tasks? – *Bilingualism: Language and Cognition* 18 (3), 551–560.
- Chiat, Shula; Roy, Penny 2007. The Preschool Repetition Test: An Evaluation of Performance in Typically Developing and Clinically Referred Children. – *Journal of Speech, Language, and Hearing Research* 50, 429–443. Bristol: Multilingual Matters.
- Chiat, Shula 2015. Non-Word Repetition. – *Assessing Multilingual Children. Disentangling Bilingualism from Language Impairment*, 125–150.
- Baddeley, Alan; Gathercole, Susan; Papagno, Costanza 1998. The Phonological Loop as a Language Learning Device. – *Psychological Review* 1, 158–173.
- Bialystok, Ellen 2002. Acquisition of literacy in bilingual children: A framework for research. – *Language Learning* 52, 159–199.
- Dabašinskienė, Ineta; Krivickaitė, Eglė 2013. Language delay, incomplete learning or SLI: a non-word repetition test for Lithuanian. – Poster in IX International Bilingualism Symposium.
- Dabašinskienė, Ineta; Balčiūnienė, Ingrida; Kalninytė, Agnė 2014. Links between lexicon and grammar in narrative production: Lithuanian as L1 and L2 in bilingual children. – Paper given at 13th International Congress for the Study of Child Language, Amsterdam.
- Dabašinskienė, Ineta; Krivickaitė-Leišienė, Eglė 2019. Lithuanian as L2: a case study of Russian minority children. – *Multilingualism in the Baltic States: societal discourses and contact phenomena*, 205–237.

- D'Odorico, Laura; Fasolo, Micro 2006. The prosody of early multi-word speech: Word order and its intonational realization in the speech of Italian children. – *Enfance* 3, 317–327.
- Dodd, Barbara; Holm, Alison; Hua, Zhu; Crosbie Sharon 2003. Phonological development: a normative study of British English-speaking children. – *Clinical Linguistics & Phonetics* 17 (8), 617–643.
- Dollaghan, Chris; Campbell, Thomas F. 1998. Nonword Repetition and Child Language Impairment. – *Journal of Speech, Language, and Hearing* 41, 1136–1146.
- Donegan, Patricia; Stampe, David 2009. Hypothesis of Natural Phonology. – *Poznan Studies in Contemporary Linguistics* 45 (1), 1–31.
- Farabolini, Gianmatteo; Rinaldi, Pasquale; Cristina, Caselli Maria; Cristia, Alejandrina 2021. Non-word repetition in bilingual children: the role of language exposure, vocabulary scores and environmental factors. – *Speech, Language and Hearing*. <https://doi.org/10.1080/2050571X.2021.1879609>.
- Gathercole, Susan 2006. Nonword Repetition and Word Learning: The Nature of the Relationship. – *Applied Psycholinguistics* 27, 513–543.
- Gathercole, Susan 1995. Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. – *Memory & Cognition* 23 (1), 83–94.
- Gathercole, Virginia C. Mueller 2013. Solutions for the Assessment of Bilinguals. Bristol: Multilingual Matters.
- Gibson, Todd, A; Jarmulowicz, Linda; Oller, Kimbrough, D 2019. Bilingual and monolingual children's articulation rates during nonword repetition tasks. – *International Journal of Bilingualism*, 23 (5), 1159–1179.
- Gibson, Todd A; Summers, Connie; Pena, Elizabeth; Bohman, Thomas M 2014. The role of phonological structure and experience in bilingual children's nonword repetition performance. – *Bilingualism: Language and Cognition* 18 (3), 551–560.
- Girbau, Dolors; Schwartz, Richard 2007. Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). – *International Journal of Language & Communication Disorders* 42, 59–75.

- Grimm, Angela; Hübner, Julia 2016. Nonword repetition by bilingual learners of German: The role of language-specific complexity. – *Bilingualism and Specific Language Impairment*. Amsterdam: Benjamins.
- Girdenis, Aleksas; Karosienė, Vida 2010. Bendrinės lietuvių kalbos statistinė struktūra: fonologijos dalykai. Vilnius: Vilniaus universiteto leidykla.
- Kavitskaya, Darya; Babyonushchev, Maria; Walls, Theodore; Grigorenko, Elena 2011. Investigating the effects of syllable complexity in Russian-speaking children with SLI. – *Journal of Child Language* 38, 979–998.
- Kazlauskienė, Asta 2007. Intervokaliniai priebalsiai: vienanarės ir dvinarės grupės. – *Kalbų studijos* 11, 36–42.
- Kazlauskienė, Asta; Raškinis, Gailius 2008a. Lietuvių kalbos fonologinio skiemens struktūrinių modelių dažnumas. – *Žmogus ir Žodis* 1, 24–31.
- Kazlauskienė, Asta; Raškinis, Gailius 2008b. Intervokaliniai priebalsiai: trinarės, keturnarės ir penkianarės grupės. – *Kalbų studijos* 13, 51–59.
- Kazlauskienė, Asta 2010. Lietuvių kalbos žodžių foneminės struktūros dėsniumai. – *Žmogus ir žodis* 1, 35–41.
- Kondo-Brown, Kimi 2004. Investigating interviewer-candidate interactions during oral interviews for child L2 learners. – *Foreign Language Annals* 37, 602–615.
- Kohnert, Kathryn; Windsor Jennifer; Yim, Dongsun 2006. Do Language-Based Processing Tasks Separate Children with Language Impairment from Typical Bilinguals? – *Learning Disabilities Research & Practice* 21 (1), 19–29.
- Krivickaitė, Eglė 2014. The Research of Phonotactics Acquisition of Typically Developing Children and Children with Specific Language Impairment by Non-Word Repetition Test. – *Special Education* 2 (31), 105–125.
- Krivickaitė, Eglė 2017. Vaikų taikomos žodžių tarimo paprastinimo strategijos. Eksperimentinis tyrimas. – *Taikomoji kalbotyra* 9, 46–72.
- Kunnari, Sari; Tolonen, Anna-Kaisa; Chiat, Shula 2011. Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment. <http://www.bi-sli.org> (01.11.2022).
- Lee, Susan Ann; Gorman, Brenda K. 2012. Nonword repetition performance and related factors in children representing four linguistic groups. – *International Journal of Bilingualism* 17 (4), 479–495.

- Leseman, P. M. Paul; Van Tuijl, Cathy 2006. Cultural diversity in early literacy. – *Handbook of Early Literacy Research* 2, New York: The Guilford Press, 211–228.
- Masoura, Elvira V; Gathercole, Susan, E 1999. Phonological Short-term Memory and Foreign Language Learning. – *International journal of psychology* 34 (5/6), 383–388.
- Marshall, Chloe; Van der Lely, Heather K. J 2009. Effects of word position and stress on onset cluster production: evidence from typical development, specific language impairment, and dyslexia. – *Language* 85 (1), 39–57.
- McDonald, Janet L; Oetting, Janna B 2019. Nonword repetition across two dialects of English: Effects of specific language impairment and non-mainstream form density. – *Journal of Speech, Language, and Hearing Research*, 62 (5), 1381–1391.
- Messer, Marielle; Leseman, Paul P M; Boom, Jan; Mayo, Aziza Y 2010. Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. – *Journal of Experimental Child Psychology* 105 (4), 306–323.
- Messer, Marielle H; Verhagen, Josje; Boom, Jan; Mayo, Aziza Y; Leseman, Paul P. M 2015. Growth of Verbal Short-Term Memory of Nonwords Varying in Phonotactic Probability: A Longitudinal Study with Monolingual and Bilingual Children. – *Journal of Memory and Language* 85, 24–36.
- Montrul, Silvina 2011. Morphological errors in Spanish second language learners and heritage speakers. – *Studies in Second Language Acquisition* 33, 155–161.
- Montrul, Silvina 2013. Bilingualism and the Heritage Language Speaker. – *The Handbook of Bilingualism and Multilingualism*. Blackwell Publishing.
- Munson, Benjamin; Kurtz, Berth A; Windsor Jennifer 2005. The Influence of Vocabulary Size, Phonotactic Probability, and Wordlikeness on Nonword Repetitions of Children with and Without Specific Language Impairment. – *Journal of Speech, Language, and Hearing Research* 48, 1033–1047.

- O'Grady, William 2011. Interfaces and Processing. – *Linguistic Approaches to Bilingualism* 1 (1), 63–66.
- Paradis, Johanne 2010. Bilingual Children's Acquisition of English Verb Morphology: Effects of Language Exposure, Structure Complexity, and Task Type. – *A Journal of Research in Language Studies* 60 (3), 651–680.
- Paradis, Johanne; Genesee, Fred; Crago, Martha B 2011. Dual language development and disorders: A handbook on bilingualism and second language learning (2nd Edition). Baltimore, MD: Brookes.
- Park, Hyejin; Schwarz, Isa 2012. Nonword Repetition Performance of Early School-Age Children in Korea: A Preliminary Study. – *Asia Pacific Journal of Speech, Language, and Hearing* 15 (3), 178–196.
- Pascale, M J Engel de Abreu 2011. Working memory in multilingual children: is there a bilingual effect? – *Memory* 5, 529–537
- Pearson, Barbara Zurer 2007. Social factors in childhood bilingualism in the United States. – *Applied Psycholinguistics* 28, 399–410.
- Piazzalunga, Silvia; Previtali, Lisa; Pozzoli, Raffaella; Scarponi, Letizia; Schindler, Antonio 2019. An articulatory-based disyllabic and trisyllabic Non-Word Repetition test: reliability and validity in Italian 3- to 7-year-old children. – *Clinical Linguistics & Phonetics* 33 (5), 1–20.
- Radeborg, Karl; Barthelom, Eleonora; Sjorberg, Monika; Sahlen, Birgitta 2006. A Swedish non-word repetition test for preschool children. – *Scandinavian Journal of Psychology* 47, 187–192.
- Rispens, Judith; Parigger, Esther 2010. Non-word repetition in Dutch-speaking children with specific language impairment with and without reading problems. – *British Journal of Developmental Psychology* 28, 177–188.
- Santos F. H, Bueno, Orlando, F A; Gathercole, Susan E 2006. Errors in non-word repetition: bridging short- and long-term memory. – *Brazilian Journal of Medical and Biological Research* 39, 371–385.
- Sileo, Roberto; Tyčová, Dana 2019. Czeck-Rep: Introducing A Nonword Repetition Task for Czech. – *Studie z Aplikované Lingvistiky* 9, 101–107.
- Silven, Maarit; Lunden, Maija 2011. Balanced communication in mid-infancy promotes early vocabulary development: effects of play with mother and father in mono- and bilingual families. – *International Journal of Bilingualism* 15 (4), 535–559.

- Shaalán, Saleh 2020. Non-word Repetition Skills in Gulf Arabic-Speaking Children With Developmental Language Disorder. – *Journal of Speech, Language, and Hearing Research* 63 (11), 3700–3713.
- Stokes, Stephania F; Wong, Anita M-Y; Fletcher, Paul; Leonard, Laurence 2006. Non-word Repetition and Sentence Repetition as Clinical Markers of Specific Language Impairment: The Case of Cantonese. – *Journal of Speech, Language, and Hearing Research* 49, 219–236.
- Summers, Connie; Bohman, Thomas, M; Gilliam, Ronald, B; Oena, Elizabeth, D; Bedore, Lisa, M 2010. Bilingual performance on non-word repetition in Spanish and English. – *International Journal of Language & Communication Disorders* 45(4), 480–493.
- Thordardottir, Ellin; Juliusdottir, Anna Gudrun 2012. Icelandic as a second language: a longitudinal study of language knowledge and processing by school-age children. – *International Journal of Bilingual Education and Bilingualism* 64 (4), 411–435.
- Thordardottir, Elin; Reid, Heather 2022. Nonword repetition to identify DLD in older school-age children. A French nonword repetition test with word-like and nonword-like items from 2 to 8 syllables. – *Enfance* 1 (1), 59–79.
- Windsor, Jennifer; Kohnert, Kathryn; Lobitz, Kelann F; Pham, Giang T 2010. Cross-Language Nonword Repetition by Bilingual and Monolingual Children. – *American Journal of Speech-Language Pathology* 19, 298–310.

RESÜMEE

KAKSKEELSUSE EELIS: VÄLJAMÕELDUD SÕNADE KORDAMISE TEST

Eglė Krivickaitė-Leišienė, Ineta Dabašinskiene

Vytautas Magnus University

Uurimus käsitles kaks- ja mitmekeelsete laste väljamõeldud sõnade produktiooni. Osalejad (kokku 125 last) kuulusid kahte (erinevate keelerepertuaaridega) kakskeelsete laste gruppi ja ühte ükskeelsete leedulaste gruppi. Analüüsist tuli välja, et kakskeelsete gruppide sooritusel olid ükskeelsest grupist paremad. Kakskeelsed lapsed näitasid täpsemaid ja statistiliselt olulisi tulemusi pikemate ja struktuurilt keerukamate väljamõeldud sõnade kordamisel. Uurimuse tulemused näitavad, et kakskeelsetel lastel on laialdasem kogemus erineva fonoloogiaga, kuna neil on kokkupuude kahe fonoloogilise süsteemiga, mis aitas neil ülesannet täpsemalt sooritada.

Võttesõnad: kaks- ja mitmekeelsed lapsed, keeleomandamine, leedu, inglise, vene

Ineta Dabašinskiene's research interests include, but are not limited to language policy and multilingualism, monolingual and bilingual second language acquisition, and grammar of spoken language.
ineta.dabasinskiene@vdu.lt

Eglė Krivickaitė-Leišienė's research interests focus on child language acquisition, early bilingualism/ multilingualism, and child language disorders.
egle.krivickaite-leisiene@vdu.lt

„MIS KEELES MA RÄÄGIN, *I DON'T KNOW*“. EESTI SISULOOGJATE INGLISE KEELE KASUTUSEST YOUTUBE'IS

Kristiina Praakli, Mari-Liis Korkus, Aive Mandel,
Elisabeth Kaukonen, Annika Kängsepp, Triin Aasa,
Kristel Algvere, Helen Eriksoo, Marion Mägi,
Getri Tomson, Liina Lindström
Tartu Ülikool

Annotatsioon.¹ Artiklis käsitleme YouTube'i platvormil videoid loovate juutuuberite keelepruuki fookusega inglise keele kasutusel. Analüüsisime kaheksa eestikeelse sisulooja inglise keele kasutuse määra ja põhjusi. Valisime küsitluse põhjal välja juutuuberid, kelle videoid gümnaasiumiealised noored enim jälgivad ning kes avaldavad neid YouTube'is regulaarselt. Keeleainestikuks on litereeritud video(lõigu)d, mille põhjal analüüsisime ingliskeelsete sõnede hulka ning koodivahetuse põhjusi. Tulemustest selgus, et kõik sisuloojad kasutasid oma videotes inglise keelt, ent selle esinemismäär oli erinev. Inglise keelt kasutati enim n-ö erialaterminoloogiana, mis kajastab globaalses sisuloomes kasutatavat sõnavara (26% inglise keele esinemise juhtudest). Lisaks kasutati inglise keelt ingliskeelse popkultuuriga seotud nähtustele viitamisel (16%) ning emotsioonide väljendamisel (12%), sage oli ka kinnistunud väljendite kasutamine (6%). Üpris suure osa (23%) moodustasid juhud, mille puhul oli raske näha inglise keele eelistamise põhjusi.

Võtmesõnad: YouTube, osalusmeedia, sisuloojad, noorte keel, koodivahetus, inglise keel

¹ Artikkel on valminud riikliku programmi „Eesti keel ja kultuur digiajastul“ (EKKD) rahastatava projekti „Teismeliste keel Eestis“ (EKKD3) raames, selle valmimist on toetanud ka Euroopa Liidu Regionaalarengu Fond (Eesti-uuringute Tippkeskus).

I. Sissejuhatus

Üleilmastumine ning uued osalusmeediaplatvormid kujundavad otseselt noorte keelekasutust, mistõttu on ka mitmekeelsed suhtluspraktikad üha levinumad. Inglise keelel on selles kanda tähenduslik roll: globaalsed trendid ja ingliskeelne meediasisu avaldavad üha enam mõju eesti noortele, kes on aktiivsed meediakasutajad ning oskavad inglise keelt enamasti hästi. Püsijuurdepäas ingliskeelsele meediasisule algab eesti noortel varakult (vt Kalmus *et al.* 2009; Siibak 2020). Kaasaja keelekontaktid leiavad aset peaaesjalikult meedia kaudu, tehniliselt vahendatult ning ruumist sõltumata. Et eesti noortel on inglise keelega üha varasemad ning aktiivsemad kontaktid, on nende Est/i/o/nglishiks (nt Oder 2011; Niineste 2017²) nimetatud keelekasutus pälvinud avalikkuses laiemat tähelepanu, enamasti küll negatiivses võtmes. Sisuloojad ja nende jälgijaskond moodustavad praktikakogukonna (ingl *community of practice*), kes tegutseb YouTube'is kui ühises sotsiaalses ruumis. Ühise osaluse kaudu luuakse ühine sisuloojate keskkond. Selle jagatud normid ja praktikad jällegi kujundavad keeleliste ressursside valikut ning suhtlus- ja eneseesitlusviise. Teisalt võib sisuloojate suuresti performatiivne esituslaad ollagi platvormipõhine (vt Siibak *et al.* 2012) ega pruugi tingimata väljendada nende keelekasutust laiemalt (nt väljaspool osalusmeedia platvormi). Seega võiks sisuloojate suulist eneseväljendust pidada selles kontekstis kasutuskeskseks keelevariandiks.

Teaduslik huvi eesti sisuloojate keelelise käitumise vastu on alles algamas. Senistes uuringutes on vaatluse alla võetud eneseesitlusviisid (Muuli 2017), meediakuvand (Liiver 2019), persoonibränding (Auser 2018), samuti teismeliste tarbimisharjumused YouTube'is (Perm 2017; Orgmets 2018; Miil 2019; Kask 2020). Sisuloojate

² Niineste (2017) kirjutab: „Õuervervelming on kuulata, kuidas noored räägivad eesti keelt inglise keele toortõlgete ja käibefraasidega pooleks [...]“

keelelist käitumist aitavad mõista Silja Ratti (2017) ning Helin Kase (2021) uurimused.

Siinses artiklis keskendume sisuloojate inglise keele kasutusele vestlustes, mille põhikeel on eesti keel. Selleks võtame vaatluse alla YouTube'i keskkonnas tegutsevad eesti sisuloojad ning nende loodud videod. Arvestades YouTube'i kui keskkonna ning seal tegutsevate sisuloojate populaarsust ja kõrget jälgitavust noorte hulgas, võime oletada, et juutuuberite keelevalikud mõjutavad olulisel määral ka nende jälgijate, enamasti noorte keelekasutust. Meid huvitab, 1) mil määral ja mis ülesannetes kasutatakse sisuloomes inglise keelt; 2) kuivõrd on inglise keele kasutamine tingitud osalusmeedia žanrispetsiifilistest asjaoludest; 3) milline on nende keeleline reper tuaar üldisemalt. Kuna juutuuberitel on jälgijaskonnale suur mõju (Murumaa-Mengel, Siibak 2020) ja nad on sellest teadlikud (Muuli 2017), võib oletada järgmist:

1. Populaarsete juutuuberite keelevalikud mõjutavad noore jälgijaskonna keelekasutust, aga ka vastupidi.
2. Juutuuberid arvestavad oma sihtrühma keelekasutuse ja keeleliste ootustega.
3. Juutuuberite inglise keele kasutus on seotud üleilmse noortekultuuriga ja on žanrispetsiifiline.

Artikli põhifookus lasub inglise keele kasutuse ulatusel ja põhjustel. Teises peatükis võtame vaatluse alla YouTube'i sisuloomes ja sisuloojate tegevuse lähemalt, peatume eesti ja inglise keele kontaktituatsioonil ning vaatleme koodivahetust kui üht mitmekeelset suhtlusnähtust. Kolmandas tutvustame ainekust ja meetodit, neljandas esitleme analüüsi põhitulemusi nii kvantitatiivselt kui kvalitatiivselt.

2. Taust

2.1. YOUTUBE JA SELLE SISULOOLAJAD

YouTube'i videojagamisplatvormist (loodud 2005, omanik Google) on praeguseks kujunenud populaarseim osalusmeedia keskkond ning multimeediumiteenuste pakkuja (Anderson, Jiang 2018). YouTube'i külastab iga päev ligi 122 miljonit kasutajat, kes tarbivad kokku üle miljardi video (Omnicores Agency 2022). Platvormil tegutsejad võib laias laastus jagada kaheks: ühed, kes loovad sisu (ingl *user-generated content*, *UGC*) ja teised, kes seda tarbivad (vt Torjesen 2021: 169).

YouTube on sisuloojate seas populaarne. Kuigi videote avaldamist võimaldavad ka teised sotsiaalmeediakanalid (nt Instagrami Video, Facebook Watch, Twitch, Vimeo, TikTok), on YouTube'i eelis selle tehnilised võimalused (nt videote pikkusel puudub ajaline piirang) ning funktsionaalsus. Kuigi viimastel aastatel on järsult kasvanud TikToki populaarsus, on Eesti noorte seas selle platvormi tarbimine alles hoogustumas (Diktor 2022). Ka mitmed Eesti sisuloojad on laiendanud oma tegevust TikTokki, kuid kasutavad põhi-kanalina endiselt YouTube'i.

YouTube'i sisuloojaks (juutuuberiks; ingl *YouTuber*, *content creator*) nimetatakse selles keskkonnas regulaarselt videote postitajat, kes on saavutanud tänu platvormile tuntuse ning kel on stabiilne jälgijaskond (Himma-Kadakas *et al.* 2018). Artiklis kasutame mõisteid *sisulooja* ja *juutuuber* sünonüümselt.³ Igal juutuuberil on oma jälgijaskond, kellega ta pidevalt suhtleb: olgu need siis tavajälgijad, fännid või antifännid (Murumaa-Mengel, Siibak 2020). Jälgijaskonna hoidmiseks kasutatakse erinevaid võtteid: avaldatakse regulaarselt uusi videoid, kutsutakse üles oma jälgijaid kanali vaatajaskonnaga liituma (ingl *subscribe*), uue video ilmumisel sellest

³ Internetiuuringuis on kasutatud teisigi mõisteid, nt mikrokuulsused (ingl *micro celebrities*), sotsiaalmeedia mõjutajad/suunamudijad (ingl *social media influencers*) jne (vt lähemalt Abidin 2018; Murumaa-Mengel, Siibak 2020).

esimesena teavitust saama (ingl *bell notification*), videotele meeldivust väljendama ehk *laikima* (ingl *like*) ja kommenteerima (ingl *comment*), esitama küsimusi, väljendama arvamust, andma tagasisidet.

Juutuuberite videod on žanriliselt mitmekesised. Sotsiaalmee-
dia kiire kasvuga on tekkinud uued sisulomežanrid (vlogid kui
videopäevikud, reaktioonivideod, *challenge*'id ehk väljakutsed,
humoorikad sketšid, mänguvideod, *how-to*-õpetused, *haul*'id ehk
ostututvustused jne), mida on juba omajagu uuritud (nt Muuli,
Perm 2017; Himma-Kadakas *et al.* 2018). Eesti juutuuberite popu-
laarseimad žanrid on mistahes (aga eriti koostööpartnerite) too-
dete/teenuste reklaam (ingl *advertisement*) ning üleilmseid trende
järgivad väljakutsed (ingl *challenge*); seevastu Euroopa ja üleilmsete
sisuloojate seas mänguvideod (ingl *gaming*) ja omalooming (ingl
original performance) (Himma-Kadakas *et al.* 2018). Kuigi iga juu-
tuuberi sisu võib olla temaatiliselt erinev (nt elustiil, videomängud,
muusika), läbib igaühe videolomeprotsess enne vaatajaskonna ette
jõudmist samad etapid (teemavalik; käsitus- ja edastusviis, sh kee-
leline; turundamine). Žanrianalüüsi vaatepunktist on analüüsitud
keelekasutust, nt iluvlogijate videotest (Riboni 2017).

2.2. EESTI JA INGLISE KEELE KONTAKTID MEEDIASTUNUD MAAILMAS

Viimaste sajanditega globaalseks maailmakeeleks kujunenud ing-
lise keel on oluliselt muutnud keelekeskkondi üle ilma. Nii meedias-
tumine (vt ka Hepp 2018) kui inglise keele kasutusala laienemine on
mõjutanud sellega aktiivses kontaktis olevate keelekasutajate hoia-
kuid, keeleoskust ning suhtluspraktikaid.

Eesti keele kõnelejate praeguseks rohkem kui paarsada aastat
kestnud kokkupuuted inglise keelega on olnud erineva intensiivsuse
ja kaaluga (vt Jõgi 2014; Zabrodskaja, Kask 2017). Inglise keele rolli
muutumist viimase sadakonna aasta jooksul peegeldavad muuhul-
gas noorte keelekasutuse (põhiliselt sõnavara hõlmavad) andmed.
Näiteks Vaigla (1928: 97; vt ka Palmeos 1935: 513) andmeil ei leidu

20. sajandi esikümnendite üliõpilassõnavarast „erilisi jälgi ei inglise ega soome keelest, peamised laenuandjakeeled on saksa, ladina ja prantsuse“⁴. Pool sajandit hiljem on noorte keelelistes valikutes olulisimad vahendajakeeled inglise, Tallinnas ka soome keel (vt Loog 1988, 1990), eriti ilmseks saab inglise keele osa noortekeeles 2000ndatel (vt Vainola, Kaplinski 2003). Nagu ka praegu, hõlmab laensõnavara noortele tähendusrikkaid vestlusteemasid (nt omavahelised suhted, meelelahutus, vaba aeg).

Praeguse positsiooni dominantseima vahendaja- ning võõrkeelena kinnistas inglise keel pärast taasiseseisvumist. Seda on kujundanud Eesti ja eestlaste avatus maailmale, majanduslik ja poliitiline lõimumine (sh sellega kaasnenud keele- ja hariduspoliitilised otsused) ning info- ja meediaajastu. Inglise keel on koolinoorte seas üha atraktiivsem, mis ilmneb mh nende positiivsetes hoiakutes inglise keele suhtes (Tammemägi, Ehala 2012: 257–259).

Eesti-inglise keelekontaktidest johtuvate muutuste ning mitmekeelsete suhtluspraktikate analüüsimine kõnetab ka uurijaid üha rohkem. Üksikasjaliku analüüsi vanematest inglise päritolu keelenditest (kirjalike allikate näitel kuni 1970ndateni) ning nende eesti keelde ülekande- ja kohanemismallidest esitab Jõgi (2014, varasemaid uurimusi *ibid.* 47–50). Uuema laensõnamaterjali põhjal on inglise laensõnu põhjalikumalt uurinud Leemets (2003) ja Pedaja (2006). Mõlemad käsitlevad keelendite kohandamist, kirjeldavad teemavaldkondi ja peatuvad kasutusel. Üha rohkem tähelepanu on pööratud ka inglise keele kasutusele uutes meediaplatvormides ja kiirsuhtlusrakendustes (nt Vaba 2010; Igav 2013; Ratt 2017; Kask 2021).

⁴ Vaigla (1928: 97) selgitab seda järgmiselt: „Kultuuriline kontakt inglise ja soome kultuuriga on meil ainult esimesi vagusid ajav; meie kokkupuuted nimetatud maade üliõpilastega on vaid külastamiskokkupuuted, – seepärast pole erilisi jälgi ei inglise ega soome keelest me üliõpilaskeeles.“

2.3. MÕNDA INGLISE KEELE FUNKTSIOONIDEST SISULOOJATE SUULISES KEELES

Siinses artiklis on mitmekeelse suhtluse eri nähtustest (tõlkelaenud, koodivahetus, koodivaheldumine, mitmekeelse vestluse mallid) vaatluse all koodivahetus. Mõistame selle all kahe või enama keele või keelevariandi kasutamist suhtluses (lähemalt Gumperz 1982; Verschik 2004). Mitmekeelse suhtluse uuringuis on koodivahetusele lähenetud mitmeti, nt grammatilisest ja pragmaatilisest vaatenurgast (vt Gumperz 1982; Myers-Scotton 1993, 2002; Auer 1995). Viimase lähenemise puhul on fookuses koodivahetuse funktsioonid vestluses, mis on vaatluse all ka siinses artiklis.

Koodivahetusuuringutes on funktsioonide määratlemisele lähenetud mõnevõrra erinevalt (vt nt Zabrodskaja 2006: 233). Üldiselt eristuvad järgmised funktsioonid, mis on mõneti kattuvad ja/või üksteisega seotud (lähemalt Grosjean 1982; Auer 1995; Appel, Muysken 2005):

1. Vestluses osalejatega seotud muutused (sh kaasamine, välis-tamine, rollivahetus).
2. Temaatilised ja tegevuslikud viited.
3. Ekspressiivsus (sh emotsioonide, hinnangute väljendamine).
4. Tsiteerimine ja matkimine.
5. Leksikaalse lünga täitmine (sh puudulik sõnavara).
6. Öeldu kordamine eri keeltes.
7. Täpsustamine, rõhutamine ja kommenteerimine (sh meta-kommentaarid).
8. Keelemängud ja humoorikad märkused.
9. Gruppide eristamine („meie“- ja „nende“-koodi vastandamine, vt ka Gumperz 1982).

Artiklis keskendume inglise keele kasutuse funktsioonidele muidu eestikeelses suhtluses. Funktsioonide määratlemisel läh-tusime andmetest joonistuvatest mustritest. Lähemalt kirjeldame koodivahetusjuhtude märgendamise ja kodeerimise põhimõtteid alapeatükis 4.3.

3. Ainestik ja meetod

Artikli aluseks on 2022. a alguses tehtud pilootuurimus, mille käigus kogusime keeleandmeid inglise keele kasutuse kohta YouTube’is tegutsevate sisuloojate näitel. Selleks valisime välja kaheksa juutuuberit, kes on tuntud meediapersoonid ning kelle videod on avalikusele kättesaadavad.

Juutuuberite valikul lähtusime sellest, et nad postitaksid regulaarset sisu (vähemalt 10 videot viimasest aastast) ning oleksid noorte hulgas piisavalt tuntud. Selleks tegime viies üldhariduskoolis (Tallinnas, Tartus, Antslas, Kuressaares ja Jõhvis) veebiküsitluse. Küsitlusele vastas 232 noort vanuses 15–20 eluaastat: neist 146 (63%) olid tüdrukud, 79 (34%) poisid, ülejäänud (3%) määratlesid end teisiti või ei soovinud oma sugu avaldada. Küsitlusega soovisime saada teada, kui palju, milliseid ja mis keeles kõnelevaid juutuubereid noored jälgivad.

Rohkem kui pooled vastanuist (55%) jälgivad eesti juutuubereid iga päev ja/või mitu korda päevas; veerand (26%) vähemasti kord nädalas. Juutuubereid ei jälgi või jälgib harva u 10% vastanuist (vt tabel 1).

Tabel 1. Vastused küsimusele „Kui sageli sa jälgid juutuubereid ja muid mikrokuulsusi?“

Sugu	Pidevalt / peaaegu iga päev	Vähe- malt kord nädalas	Vähe- malt kord kuus	Peaaegu mitte kunagi	Mitte kunagi	Ei oska öelda	Kokku
M	55	14	2	6	1	1	79
N	68	45	15	12	4	2	146
Muu	5	1	0	0	0	1	7
Kokku	128	60	17	18	5	4	232
Kokku %	55	26	7	8	2	2	100

Ligi pooled vastanuist (48%) jälgivad nii eesti- kui ingliskeelseid (või ka mingit kolmandat keelt – enamasti vene keelt kasutavaid) juutuubereid (vt tabel 2). Seejuures on palju rohkem neid, kes jälgivad ainult ingliskeelseid (35%), kui neid, kes jälgivad ainult eesti keelt kasutavaid sisuloojaid (6%). Seega võib öelda, et noorte tarbimiseelised juutuuberite jälgimise näitel kalduvad tugevasti mitmekeelse (eriti ingliskeelse) osalusmeediaruumi suunas. Järelikult on küsitluses osalenud noored ka ise mitmekeelsed või vähemalt saavad inglise keelest piisavalt hästi aru, et tarbida ingliskeelset sisu.

Tabel 2. Vastused küsimusele „Mis keeles Sinu jälgitavad juutuuberid kõnelevad?“

Keel	Vastajaid	Vastajate %
eesti	14	6
inglise	80	35
eesti, inglise (ja muu)	111	48
eesti, vene	2	1
muud kombinatsioonid (ilma eesti keelela)	22	10
Kokku	229	100

Ühtekokku mainiti 20 erinevat juutuuberit; nimetatuid valisime populaarsuse ja muude kriteeriumite põhjal edasiseks analüüsiks välja kaheksa inimest. Üks oluline kriteerium oli eestikeelsete videote postitamine 2021. aastal, st analüüsist jäid kõrvale need, kelle varasemad videod olid küll eestikeelsed, kuid kes olid nimetatud perioodil üle läinud ingliskeelse sisu tootmisele. Keeleandmete kogumise konteksti mõttes oli ka tähtis, et eestikeelsed videod oleksid avaldatud piisava järjepidevusega (vähemalt kord kuus), et nende loojate keelekasutuse kohta oleks võimalik teha laiemaid järeldusi.⁵

Žanriliselt leidis vaadeldud materjali hulgas enim vlogisid, väljakutseid (ingl *challenge*), Q&A'sid (küsimused-vastused)

⁵ Analüüsitava keeleainestiku moodustavad kaheksalt eesti sisuloojalt valitud videod (AZ, RV, MR, HK, MH, ST, LP, KL).

ja reaktsioonivideoid; otseselt tooteturundusega seotud videoid (ingl *advertisement*) oli vähe. Igaühelt valisime välja kaks umbes 10-minutilist videot või videolõiku (kokku 160 minutit). Videote valikul lähtusime sellest, et analüüsitava lõigus kõneleks peaaesjalikult sisulooja ise (ja mitte niivõrd külalised), lõik moodustaks suhtluse ülesehituse ning teemakäsitluse mõttes tervikliku üksuse (hõlmates vaatajaskonna tervitamist, teemakäsitlust, kokkuvõtet, hüvastijättu), peegeldaks suhtlusolukorda ja konteksti, kõneleja isikukeelt ning keelelisi vahendeid võimalikult mitmekülgsest.

Kõik videod litereerisime käsitsi; litereeringud on siinses artiklis peamine uurimismaterjal⁶. Ainestikus märkisime ära kõik kõneleja öeldud ingliskeelse päritoluga ühikud (sõnad, väljendid, lausungid) ning töötasime välja märgendussüsteemi nende funktsioonide analüüsiks. Igal tekstil oli kaks märgendajat, lisaks kontrollis ja ühtlustas kõik tekstid seda ülesannet täitnud projekti liige. Kodeerimissüsteemi väljatöötamisel keskendusime vaid koodivahetusnähtustele. Selleks määratlesime ja märgendasime:

- 1) koodivahetuse üksused ning koodivahetuse piirid;
- 2) ühesõnaliste koodivahetuste puhul sõnaliigid;
- 3) kõikide koodivahetusjuhtude puhul vestluse teemad ning koodivahetuse funktsioonid (ülesanded selles vestlusolukorras).

Lisaks tegime juutuuberitega poolstruktureeritud intervjuud, milleks nad meelsasti nõusoleku andsid. Videote pikkus varieerus 30–90 minuti vahel. Vestlesime kolmel põhiteemal: sisulooje olemus ja tähendus, videolooje liigid ja tehniline teostus, arvamused ja hinnangud inglise keele kasutamise kohta videotes. Koroonajaastule omaselt toimusid intervjuud videokõnerakenduste vahendusel (Zoom, Teams). Artiklis kasutame intervjuuainestikku vajaliku taustainfona, et mõista paremini sisulooje telgitaguseid ning tehnilis-kunstilisi nüansse.

⁶ Kogutud ainestiku põhjal valmis ka juutuuberite sõnastik.

Kogu ingliskeelne aines on litereeritud häälduspäraselt, kasutades eesti keele ortograafiat. Arusaadavuse huvides on näidete juurde lisatud ingliskeelne algvorm (NomSg). Keelenäited on varustatud kõneleja initsiaalidega, intervjuu tsitaatidest on isikut tuvastav info eemaldatud.

4. Tulemused

4.1. SISULOOGJATE KEELEKASUTUSE ÜLDISELOOMUSTUS

Vaadeldud keeleandmed peegeldavad nii eesti suulise kõne vane- maid kihistusi kui ka uusi ressursse. Juutuuberite suhtluses kasutatavad vahendid sõltuvad suuresti sellest, millistele keeltele, mil määral ja millistes kontekstides on kõnelejal olnud juurdepääs (nt formaalne keeleõpe, rändekogemus, kuulumine praktikakogukonnadesse). Sestap peegeldab juutuuberite keel isikupäraseid keelelisi erisusi, praktikakogukonnana joonistub aga selgesti välja keeleline ühisosa, selle normid, vormid ning nende aktsepteerimine suhtluses.

Sisuloojate suuline keel on keeleliste vahendite poolest mitmekesine. Nende keelelisse pagasisse kuulub mitmeid keeli – keskeim positsioon on inglise keelel, ent vähesel määral kasutatakse ka soome (*jumalauta, kaikki*), hispaania (*bueno, adios*), saksa (*Auf Wiedersehen, Tschüss*) ja vene keelt (*vot, davai*). Osa neist keelenäidetest on üldises kõnekeeles juurdunud ning kasutusel põlvkondadeüleselt ega pole tingimata seotud noorte keelekasutusega (nt *davai, okei, tšau, sorri* ja nende variandid). Kõnelejade mitmekeelsus väljendub leksikaalsetes valikutes, tervitus- ja hüvastijätuvormelites, kultuurispetsiifiliste tekstide folkloorse ainese ja muude kinnisväljendite kasutamises. Paljuhäälsust toetavad audio- ja multimodaalsed vahendid (nt ekraanitekst, muusika, žestid), ainest saadakse üleilmsest ja kohalikust popkultuurist. Üldistest kõnekeelsustest on aktiivses kasutuses mitmesugused partiklid (*nagu, vä*) ja partiklitaolised sõnad/määratlejad (*mingi, si(h)uke, nisuke*), adverbid (*suvakalt, aint*), intensiivistajad (*mega, ilge, täiega, räigelt, haige, nii*), *mida*-algulised emotsionaalsed väljendid (*mida vittu, mida*

hullu) ning lühendamised (*lihts*, *suht*). Esitatud näidetest võib leida ühisosa teistes uurimustes sedastatuga – nt on leitud, et publiku kaasahaaramise eesmärgil kasutatakse erinevaid keelelisi võtteid (tervitused, hüvastijätud, pöördumised, kindel sõnavara) (vt Fägers-ten 2017; Riboni 2017).

Juutuuberite mitmekeelsus saab põhiosas ainekse inglise keelest. Inglise keel avaldub keelekontaktide mitmesuguste tulemitena, nt koodivahetusena. Koodivahetused on suures osas ühesõnalised või fraasilised üleminekud inglise keelele; pikemad üleminekud inglise keelele ehk nn koodivaheldumised on harvad. Inglisekeelseid elemente integreeritakse eesti keele morfosüntaksis mitmel viisil. Ainekseks esineb terviklikke ülekandeid, nt *meie hambl* (*humble* 'tagasihoidlik') *Pärnu peatuspaik*. Sageli kohtab ka mitmesuguseid häälikulisi ja morfofonoloogilisi kohandamisi, nt *see on kõige eksai-tingum* (*exciting* 'põnev') *osa minu päevast*. Samuti leidub (küll harvemal määral) kahe keele baasil loodud kompromissvorme, nt *ma kõlan nagu tõeline raamatunõõrd* (*nerd* 'nohik') *reaalselt*.

Eraldi nähtusena tõuseb esile inglise keele kõnelemine teise keelele ning inglise keele häälduslikud viisid. On selge, et sisuloojate keeles kannavad eri hääldusviisid eri tähendust. Mainida võib kahte eripära: rahvusvaheliste mikrokuulsuste keelepruukide võimalikult täpset jälgendamist ning inglise keele kõnelemist teadlikult eestipärast hääldust jälgendades (avaldub kõne aeglustumises, sõnade ükshaaval hääldamises, alveolaarse *r*-i kasutamises).

Eestlasliku häälduse rõhutamist võib vaadata kui omaette registrit, milles teatud keelejooni kasutades seostatakse neid (siin eesti keele) kõnelejate või mingite nende omadustega. Ka selle võtte kaudu tuuakse vestlusesse mitmeid hääli, kus kõneleja seob algupärase hääle kokku lokaalse kontekstiga, nt *kõige selle hasseli*⁷ (*hassle* 'jama, tüli') *peale on nii palju aega läinud*. Stereotüübitamine nähtub ka olukordades, kus kõnelejad jälgendavad neile kättesaadavate

⁷ Kõneleja hääldab sõna ootuspärase *ä*-hääliku asemel eestikeelse kirja *p*ildi järgi *a*-häälikuga.

vahenditega teiste keelte kõnelejaid. Järgnev näide markeerib kõneleja ettekujutust prantsuse keele hääldusest, ise seda keelt oskamata: *ei ma ei mõista seda eesti keelt üldse ma tulin siia from Fräänts Parii* (from France Paris ‘Prantsusmaalt Pariisist’). Juutuuberite (mitmekeelne) keeleline loovus ja innovatiivsus avalduvad nende julges „keeletükikeste“ (ingl *bits of language*; Blommaert 2010) kasutamises võõrkeeltest, milles neil puudub sügavam keelepädevus. Näiteks hüvastijätuvormel *Auf Wiedersehen, Tschüss* ‘hüvasti, head aega’ koosneb kõneleja sõnul neist vähestest saksakeelsetest sõnadest, mis talle kooliajast meelde on jäänud.

Juutuuberite verbaalsest võimekusest ja loovusest annavad tunnistust erinevad keelemängud ja rohke huumori kasutamine. Erinevate leksikaalsete vahenditega (nt metafooridega) luuakse uusi tähendusi, mida kasutatakse humoristlikel eesmärkidel. Rohkesti esineb iroonilist eneseesitlust, visatakse nalja nii enda, oma lähedaste kui kaassisuloojate üle. Kõnelejati tõusevad esile mitmed isikupärased väljendid, mille leksikaalne ja sotsiaalne tähendus ei pruugi väljaspool siseringi avaldudagi või mida võib pidada sisuloojate nn omasõnaks, näiteks verbi *huugama* eri tähendused AZ ja RV sisuloomes (nt *papagoi huugas õlal, Tänak huugas rallil*).

Tuginedes juutuuberitega tehtud intervjuudele, võib öelda, et nad on keeleliselt teadlikud (vt ka 4.2). Eri suhtlusviiside rakendamise kõrval avaldub see nende tähelepanekutes ja kommentaarides enda keelekasutuse kohta: kommenteeritakse keelepruuki, viidatakse inglise keele kasutusele, arutletakse eesti keele küsimuste üle (nt kas sõna on *järgepanu* või *järjepanu*). Samuti arvestavad nad oma auditooriumiga – nagu mitmeid neist intervjuudes sedastasid – ning monteerivad vajadusel välja videolõigud, kus keelepruuk pole vaatajaskonda arvestades tingimata eakohane või sobilik (nt roppused). Seega tegelevad sisuloojad aktiivselt videote järeltöötusega enne auditooriumi ette astumist.

4.2. SISULOOGJATE INGLISE KEELE KASUTUS: KVANTITATIIVNE VÕRDLUS

Esmalt vaatlesime sisuloojate ingliskeelsete sõnade ja väljendite kasutust kvantitatiivselt: kui palju on videotes inglise keelt kasutatud ning kui võrd sarnased või erinevad nad selle poolest on. Kvantitatiivne võrdlus on tehtud suhteliselt väikese hulga litereeritud tekstide põhjal: igalt juutuuberite litereerisime umbes kaks 10minutilist videot või videolõiku. Videoid ühiselt loovate juutuuberite puhul vaatlesime kolme umbes 10minutilist videolõiku ning eristasime mõlema kõneleja teksti. Siiski on tekstide maht üsna erinev, mistõttu on tabelis 3 ning joonisel 1 esitatud andmed normaliseeritud sageduste kohta (baas 1000 sõnet). Kogu materjal sisaldas 26 333 tekstisõna, neist 1047 olid ingliskeelsed. Ainestikus esinenud üldlevinud platvormide nimetusi pole koodivahetusena arvestatud (nt Instagram, Reddit jms); samuti on analüüsist välja jäetud eesti keeles põlistunud laensõnad (peamiselt *okei*). Seda laadi aines eemaldati põhianalüüsist kui eesti keeles juurdunud elemendid, mille puhul ei taju kasutaja ega vastuvõtja tõenäoliselt nende võõrapärasust. Küll aga vääriks *okei* ja selle variandid, kasutus ja funktsioonid laiemat analüüsi. Pikemate kui ühesõnaliste väljendite puhul arvestati iga ingliskeelset sõna eraldi.

Näeme, et juutuuberite ingliskeelsete sõnade-väljendite kasutamise määr on üpris erinev, ulatudes 1,5%-st kuni 9,5%-ni. Inglise keele osakaal sõltub paljuski juutuuberi vanusest ja taustast, ent ka videos käsitletavast temast ning sealsetest tegevustest. Nii võib sageduse määra kasvatada pikemate ingliskeelsete tekstide ettelugemine või varasemalt inglise keeles toimunud vestluse tsiteerimine samas keeles. Mõnes teises videos võib sama sisulooja inglise keele maht olla jällegi väiksem. Seega võib inglise keele kasutamise määr olla sama kõneleja puhul ka videoti erinev. Siiski vastab saadud tulemus muljele, mis saadi suurema hulga videote läbivaatamisel: kõik nad kasutavad ingliskeelseid sõnu ja väljendeid, ent osa neist regulaarsemalt ja sagedamini kui teised.

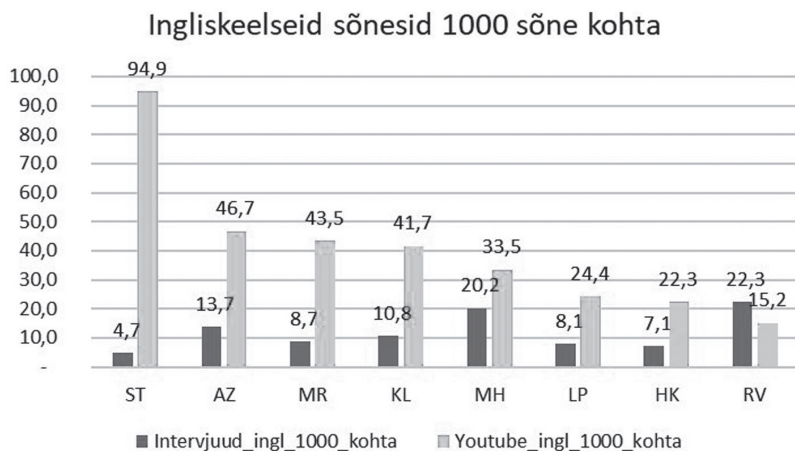
Tabel 3. *Ingliskeelsete sõnade kasutussagedus eri sisuloojatel*

Nimi	Sõnesid tekstis kokku	Ingliskeelseid sõnesid tekstis	Ingliskeelsete sõnade normaliseeritud sagedused (1000 sõne kohta)	Ingliskeelsete sõnade osakaal tekstis
ST	2003	190	94,9	9,5
AZ	3899	182	46,7	4,7
MR	4647	202	43,5	4,3
KL	4097	171	41,7	4,2
MH	4413	148	33,5	3,4
LP	2948	72	24,4	2,4
HK	2284	51	22,3	2,2
RV	2042	31	15,2	1,5

Kõrvutades sisuloojate inglise keele kasutust intervjuuainestikuga, näeme, et intervjuudes kasutati inglise keelt palju vähem (vt joonis 1). Üldistavalt öeldes püüti intervjuusituatsioonis inglise keelt vältida ning intervjuueerijatega ingliskeelseid sõnu-väljendeid üldjuhul ei kasutatud. Vaid üks neist kasutas meiega kõneledes enam ingliskeelseid sõnu kui vaadeldud videotes, ent see tulenes peamiselt arutlusest ingliskeelsete sõnade üle.

Juutuuberitega inglise keele kasutamise üle arutledes mõnsid mitmeid neist kasutuse sagedust, põhjendades seda nt alateadliku valiku ning automaatse kasutusega, nagu selgitab üks autoreist: „Paratamatult ma arvan alateadvuses mul on mingid terminid, et kui ma tahan midagi kirjeldada juba, automaatselt mulle tuleb ainult see üks ingliskeelne termin meelde.“ Inglise keele aktiivset kasutust tingib ka keele „sissepõimitus“, nagu kirjeldab teine intervjuueritu: „Ma kasutangi neid ingliskeelseid väljendeid, sest et inglise keel on nii ... noh, tänapäeval sisse põimunud.“ Samuti nenditakse auditooriumi ja selle ootustega arvestamist.

Kokkuvõttes võib öelda, et kõik vaadeldud juutuuberid kasutasid inglise keelt eri määral ja mahus. Intervjuude põhjal võime öelda,



Joonis 1. *Ingliskeelsete sõnade kasutussagedus eri sisuloojatel vaadeldud videotel ning meiega peetud intervjuudes*

et sisuloojad on oma keelekasutusest teadlikud ning vältisid intervjuusituatsioonis reeglina inglise keele kasutamist (v.a üksikud). See tähendab, et nad arvestavad auditooriumi ja selle ootustega. Kuna nende jälgijad on vanuseliselt üldiselt noored inimesed, võib öelda, et inglise keele kasutamine on üks noore publiku kõnetamise vahendeid. Kuigi keeleandmete piiratud maht ei luba teha kaugeleulatuvaid järeldusi, on need oluliseks sisendiks edasisteks uuringuteks.

4.3. JUUTUUBERITE EESTI-INGLISE KOODIVAHETUSE FUNKTSIOONID

Koodivahetuse vestluspragmaatiliste funktsioonide tuvastamiseks analüüsisime 633 videoainestikus fikseeritud koodivahetusjuhtu. Selleks tuvastasime inglise keelele üleminekud ning märgendasime iga koodivahetuse puhul inglise keelele ülemineku võimaliku põhjuse konkreetses kõnevoorus. Peale selle vaatlesime inglise elementide morfofonoloogilist vormi ehk seda, kas ingliskeelset sõna või väljendit kasutati vestluse põhikeele (eesti keele) morfosüntaktilise struktuuriga mugandatud või mugandamata kujul.

Analüüsist nähtub, et videoainestiku koodivahetusjuhtudest u pool (50,1%) esines inglise morfofonoloogilist struktuuri järgival kujul, peaaegu samas mahu (49,9%) esines eesti keele morfofonoloogilise struktuuriga mugandatud inglise keelendeid, tüüpiliselt kohandatud eesti käände- (*vlogimine*, *biuutiplenderiga*, *poodkästis*, *klämmim*) ja pöördesüsteemiga (*romantisaisida*, *käntseldan*, *kreivivad*), samuti ühinesid inglise ja eesti keel liitsõnamoodustuses (*fitnessipõhine*, *vaheapdeit*, *raamatunöörd*).

Koodivahetuse põhjused ei ole alati selged, samuti võib neid tõlgendada mitmeti. Määrasime igale koodivahetusjuhule (kui see oli võimalik) ühe funktsioonimärgendi, ligi kolmandikule ainekust (199 koodivahetusjuhtu) kaks funktsioonimärgendit. Seega on märgendatud koodivahetusjuhtude koguarv eri funktsioonide lõikes kokku 832⁸. Materjalis tõusis esile 11 funktsiooni, vt tabel 4.

Tabelist on näha, et kõige selgemalt tõusevad inglise keelele ülemineku põhjustena esile üleilmse sisuloo sõnavara tarvitamine, mida siinses analüüsis nimetame erialaterminoloogiaks (26%) ja inglise keelendi eelistamine (23%). Kokku on selliselt märgendatud ligi pooled (49%) koodivahetustest. Suhteliselt sagedasteks (vahemikus 10–16%) funktsioonideks osutusid viitamine, emotsioon/hinnang ja kinnisväljend, mõnevõrra tõusis esile ka lünga täitmine (6%). Tunduvalt vähem (1–2%) esines näiteid refereerimise/tsiteerimise, metakommentaari, kordamiste, diskursusmarkerite ja pöördumiste kohta.

Märgendi **erialaterminoloogia** alla koondasime eriala- või oskuskeelena määratlemist võimaldavad sõnad ja väljendid. Juutuuberite tegevuse kontekstis on selleks esmajoones sotsiaalmeedia-platvormide, nt videoloomega seotud mõistestik (näide 1). Eriaterminoloogiaks võib lugeda ka mis tahes muu valdkonna (IT, fotograafia, sport vm) sõnavara, siinses materjalis paistis eeskätt silma ingliskeelne kosmeetika- ja moemõistestik (näide 2).

⁸ Arv sisaldab neid koodivahetusjuhte, millel oli märgendatud kaks funktsiooni.

- (1) *kui sa oled näiteks nagu **vlooger** (vlogger ‘videoblogija’) onju* (LP)
- (2) *järgmiseks sammuks on mul alati **pronser** (bronzer ‘päikesepuuder’)* (KL)

Eelistamisena määratlesime juhud, kus ingliskeelse sõna või väljendi kohta on olemas eestikeelne vaste, mis on eeldatavasti teada ka sisuloojale (nagu näite 3 puhul *lõbus* või *tore*, näites 4 *tervishoid*, *tervisekindlustus*, näites 5 *seisukoht*, *arvamus asjast*). Põhjused, miks kõneleja eelistab (kas teadlikult või teadvustamatult) ingliskeelset vastet, lasuvad muus kui leksikaalse tühimiku täitmises (nagu see on lünga täitmise funktsiooni puhul, vt allpool) ning nende eristamine jääb siinses uurimuses oletuste tasandile. Ühel või teisel põhjusel võib keelend olla esimesena kättesaadav just inglise keeles (näites 4 räägib KL elust USAs, mis võib tingida ingliskeelse termini kasutuse), tunduda ekspressiivsem, kanda kõneleja jaoks teist/täpsemat tähendusnüanssi, olla kinnistunud kõneleja idiolektis või suhtlusvõrgustikus. Inglise keele kasutamine võib olla otsene viide praktikakogukondadesse ning sisevõrgustikesse kuulumisele – sel viisil luuakse inglise keele kultuuriline ja sotsiaalne ühisosa.

- (3) *juunis tellite saate terve suve sõita on **fann** (fun ‘lõbus’)* (MH)
- (4) *seal ei ole väga head mingid **helthkeerid** (healthcare ‘tervishoid’)* (KL)
- (5) *nüüd mis on minu **stānds on the māter** (stands on the matter ‘hoiakud, vaated asjaolu suhtes’)* (MR)

Viitamiseks lugesime sellised näited, kus ingliskeelne aines pärineb globaalsest ingliskeelsest popkultuurist. Selles kategoorias on filmidest, kirjandusest või mujalt pärit pealkirjad, lend- ja reklaamlaused (näide 6). Peale selle määratlesime viitamistena mängude või toodete nimed (näide 7), aga ka videos ette loetavad ingliskeelsed tekstid, nupud, teated jms (näide 8):

- (6) *ma ei suudaks enam seda nagu läbi teha ma arvan **never sei never** (never say never ‘ära iial ütle iial’) **ohvkoors** (of course ‘muidugi’) aga kindlasti ma ei võta uut koera* (KL)

Tabel 4. Funktsioonimärgendite liigid, osakaal ja näited

Märgend	KV arv	KV %	Näide
Erialatermino- loogia	215	26%	siis päevaga on seal mingi kümme tuhat subskraiberit (subscriber 'jälgija') (AZ)
Eelistamine	194	23%	meie esimene step (step 'samm') on tehtud (MR)
Viitamine	135	16%	ärge unustage lugeda atoomik häbitsit (Atomic Habits 'Aatomharjumused') (MR)
Emotsioon/ hinnang	102	12%	vat te hell (what the hell 'mida kuradit') reaalselt (MR)
Kinnisväljend	52	10%	ma olen hull autofriik (freak 'hull') nagu nagu näha (KL)
Lünga täitmine	49	6%	olin lihtsalt nii kuidas see sõna on ouvervelmd (overwhelmed 'ülekoormatud') eelmise nädala lõpuks (LP)
Refereerimine/ tsiteerimine	23	2%	ise mingi Räpinast pärit ma mingi brou vat de fakk (bro what the fuck 'vend mida kuradit') (AZ)
Metakommentaar	19	2%	miks see on nii ai du nat nou (I do not know 'ma ei tea') (KL)
Kordamine	15	1%	need inimesed, kes nagu halvustavad tšaadživad (judge 'hukka mõistma') mind sellepärast (MR)
Diskursuse marker	15	1%	eniueis (anyways 'igatahes') mul on siukene pisikene mingi Oriflame'i pintsil (KL)
Pöördumine	13	1%	me oleme lausa kuus korda stuudiosse jõudnud uau brou (bro 'vend') (AZ)
Kokku	832	100%	

- (7) *kui mingi trend tuleb välja nagu näiteks see **among as** (Among Us) ee trend mis mulle väga meeldis (MH)*
 (8) *mis see **khaar** (car 'auto') nupp teeb siis (HK)*

Selgelt eristuv rühm näiteid seostub **emotsioonide** (viha, rõõm, kurbus jne) või **hinnangute** (nagu heakskiit või hukkamõist) väljendamisega. Selles moodustavad suure osa hüüatused (*oumaigaad, vau, kamoona*), aga ka emotsiooni/arvamust väljendavad omadus- ja määrsõnad (9), intensiivistajad ning ropendamised (10) jm lausungid (11).

- (9) *ja ma isegi ei loe neid asju tagasi tihti sest see on lihtsalt **krindž** (cringe 'ebamugavustunnet tekitav') vahel (MR)*
 (10) *sest sa ise tantsid nagu nii **faking** (fucking 'kuradi') hästi (RV)*
 (11) *kõik okei tsau **ai laav juu** (I love you 'ma armastan sind') (MH)*

Suhteliselt sageli esinevate **kinnisväljendite** all peame silmas selliseid laene, mis on kas eesti keeles mугanenud või laialdaselt kasutuses (näited 12–13). Siia alla grupeeruvad ka igasugused globaalse levikuga trendi- ja meemiväljendid.

- (12) *oleme siin Stockholmis Nilsiga alati käinud muuseumites või **šoppamas** (shop 'poodlema') (LP)*
 (13) *ma olen räme jõulu **fänn** (fan 'austaja') onju (KL)*

Lünga täitmiseks (näited 14–15) lugesime juhud, kus (erinevalt üleval kirjeldatud eelistamise funktsioonist) ingliskeelset sõna või väljendi asendamine eestikeelsega ongi ühel või teisel põhjusel keeruline – eestikeelne vaste kas puudub või on tähendusnüansid liiga erinevad või pole vaste üldkasutatav (nt kui on tegu on kitsama erialatermini või uudissõnaga, mida keelekogukond pole täielikult omaks võtnud). Sellisel juhul võib inglise keele kasutust tõlgendada kui leksikaalse tühimiku täitmist teise keele materjaliga. Tabelis 4 toodud näites markerib leksikaalset lünka sõnaotsing (nt formuleeringus *kuidas see sõna on ...*).

- (14) *ma üritan nagu anda mingit **haipi** (hype ‘elevus’) anna andeks* (MH)
- (15) *miks sul Kalev Cramo särk seljas on võibolla selline väike **tiiser** (teaser ‘vihje’) äkki* (AZ)

Väiksema hulga näidete põhjal eristub veel mitmeid funktsioone. Inglise keelt kasutati paarikümnel juhul varem toimunud vestluse, tsitaadi, lausungi, arutelu vms sisu **refereerimiseks** või varem kõneldu **tsiteerimiseks** (näide 16). Koodivahetuses avaldub ka **metakommentaari**de ehk mitmesuguste inglise keeles tehtud kõrvalmärkuste (näide 17) ning **diskursusemarkerite** (näide 18) kasutuses. Vähem sagedateks funktsioonideks on **kordamine**, mil kõneleja kordab öeldut teises keeles (näide 19), ja **pöördumine** adressaadi (auditooriumi, vestluspartneri) poole (näited 16, 20).

- (16) *ise mingi Räpinast pärit ma mingi **brõu vat te fakk** (bro what the fuck ‘vend mida kuradit’) sa oled mingi Räpinast pärit* (AZ)
- (17) *vend tõmbas ühe sõõriku sisse ja nüüd paneb viisteist tundi niimoodi **nõu problem** (no problem ‘pole probleemi’)* (LP)
- (18) ***oolsou** (also ‘samuti’) ma olen nüüd jõudnud selle kassi kõrvade joonistamise juurde* (MR)
- (19) *nagu lihtsalt katsetage lihtsalt andke sellele proov **giv it a trai** (give it a try ‘proovi järgi’)* (MR)
- (20) ***suiidi** (sweetie ‘kallike’) minu jaoks on see et mida rohkem katvust seda parem* (KL)

Eristasime koodivahetusjuhtude puhul vaid üks-kaks funktsiooni, ent mitmete näidete puhul ei tähendanud see sugugi valikuvõimaluste ammendumist. Ilmselt võiks kõigi või vähemasti mitme viimase näiteploki koodivahetusjuhtumi puhul argumenteerida, et need on eelistamise funktsiooni juhud, kuivõrd kõneleja on mingil põhjusel eelistanud alternatiivsele eestikeelsele vastele (nt näites 17 *pole probleemi*) ingliskeelset. Eriti avar ongi eelistamise ja lünga täitmise funktsiooni ühine haare, kuivõrd emma-kumma neist võiks lisada igale näitele (konkureeriv eestikeelne vaste kas on või

see puudub), sh paljudele erialaterminoloogiaks liigitatud näidetele. Eelistamine ise on siinses käsitluses jäänud sisemiselt heterogeenseks funktsiooniks, mida saaks teistsuguse materjalikäsitluse põhjal vaadelda mitme erineva funktsioonina.

Ka ei ole siinne funktsioonide loetelu mõistagi lõplik. Sisuloojate seisukohast üks selgelt eristuv, kuid siinsete funktsioonide lõikes eri märgendite alla jagunev keelendite rühm on globaalset noortekultuuri väljendav sõnavara. Paljud ülaltoodud näidetest on mõisteta-
vad igale inglise keele oskajale, ent kitsamalt noortekultuuri peegeldav sõnavara võib olla sihtrühmast väljapoole jäävale kuulajale ähmane (näited 21–23).

- (21) *me lihtsalt tõime laua siia ja **oudžiid** (OG; original gangster 'kaua-aegne jälgija, tegija') teavad mis lauaga tegu on ma arvan (AZ)*
- (22) *see on täpselt see nagu see eesti parima filmi **taip biit** (type beat 'tüüpi, moodi, hõnguga') noh (AZ)*
- (23) *see on reaalselt juba nii palju **buužim** (boujee 'peen, uhke') (ST)*

5. Arutelu ja kokkuvõte

Vaatlesime artiklis kaheksa YouTube'is tegutseva eesti sisulooja inglise keele kasutust kvantitatiivsest (inglisekeelsete sõnade hulk tekstis, koodivahetuste arv) ja kvalitatiivsest (inglise keelele lülitumise põhjused) perspektiivist.

Materjali põhjal võib öelda, et kõik vaadeldud sisuloojad kasutasid oma videotest mingil määral inglise keelt, ent nende vahelised keelelised erinevused olid üsna suured. Intervjuude põhjal võib öelda, et nad olid üldjuhul oma inglise keele kasutusest teadlikud ja kasutasid seda intervjuusituatsioonis palju vähem või üldse mitte (ühe erandiga). See tähendab, et nad arvestavad videoid luues oma auditooriumiga (kelleks on reeglina noored) ning näevad inglise keele kasutust YouTube'i sisulooja loomuliku osana, ent ei kanna seda üle teistesse suhtlusolukordadesse (nt intervjuusituatsiooni).

Selles artiklis keskendusime ennekõike inglise keele funktsioonidele. Ainestikust joonistus välja 11 erinevat koodivahetuse

funktsiooni: ingliskeelse erialaterminoloogia kasutamine (26% kasutusjuhtudest), ingliskeelse sõna/väljendi eelistamine eestikeelsele (23%), viitamine ingliskeelsele popkultuurile (16%), emotsiooni või hinnangu väljendamine (12%), kinnisväljendid (10%), lünga täitmine eestikeelse sõna puudumisel (6%), refereerimine ja tsiteerimine (2%), metakommentaariid (2%), diskursuspartiklid (1%), sõna kordamine teises keeles (1%) ja pöördumised (1%). Ehkki need funktsioonid on hinnangulised ning osa neist kahetiselt tõlgendatavad, annab see jao- tus siiski aimu, miks ja kuidas sisuloojad inglise keelt kasutavad.

Siinse pilootuuringu tulemused on kõrvutatavad Kase ja Ratti uurimusega. Kase (2021) ainekust noorte ja noorte täiskasvanute inglise keele kasutuse kohta ilu-, moe- ja elustiili teemalistes blogides ja vlogides nähtus, et enamiku ingliskeelsustest moodustab valdkon- naspetsiifiline (nt ilutoodete jms nimetused), aga ka osalusmeediaga seotud sõnavara. Inglise keele kasutamise peamiste põhjustena nime- tab Kask lisanüansi olemasolu kõneleja jaoks või eesti keeles sobiva vaste puudumist. Ratti (2017) analüüsist näeme, et enamasti tingib inglise keelele ülemineku keelendi semantiline spetsiifilisus või sti- listilised erinevused. Need tendentsid tulid välja ka meie materjalist.

Lisaks esildusid vaadeldud materjalist suunatus auditooriumile ning performatiivsus. Inglise keelt eelistatakse suhtluskeskonnas (üleilme platvorm, lai auditoorium), kus seda ka soodsalt vastu võe- takse. Oluline on meeldida noorele publikule, kes on sisuloojate olu- lisim jälgijaskond. Sisuloojad on pidevas interaktsioonis oma audi- tooriumiga ning toimub vastastikune mõjutamine, sh keeleliselt, sest ka jälgijaskond on inglise keele suhtes positiivselt meelestatud. Keelelised valikud tehakse lähtuvalt sellest, mida auditoorium akt- septeerib. Meie materjalist kajastus see kõige otsesemalt funktsioo- nis, mida nimetasime *eelistamiseks*. Juutuuberite keelelised valikud on ühtlasi osa eneseesitlemise ja isikubrändi loomise vahendeist.

Ehkki koodivahetuse põhjusi on sageli keerukas määratleda, on üldisemad jooned hästi nähtavad. Juutuuberid kasutavad ing- lise keelt enim lähtuvalt enda professionaalsetest vajadustest, viita- maks sotsiaalmeedias ja sisuloomes laiemalt käibivale ingliskeelsele

terminoloogiale. Suur osa YouTube'is tegutsevatest Eesti sisuloojatest jälgib eelkõige ingliskeelseid kanaleid, nagu nende auditooriumgi. Seega moodustab ingliskeelne sõnavara sisuloojate ja nende jälgi-jate, laiemalt aga kogu globaalse sotsiaalmeedia ja üksikisiku vahe-lise ühisosa. Kõige silmapaistvama osa koodivahetusest tingib juu-tuuberite töökeskkond – rahvusvaheliste platvormide, sh YouTube'i sisuloomeformaadid.

KIRJANDUS

- Anderson, Monica; Jiang, Jingjing 2018. Teens, Social Media & Technology 2018. YouTube, Instagram and Snapchat are the most popular platforms among teens. Fully 95% of teens have access to a smartphone, and 45% say they are online 'almost constanly'. https://www.pewinternet.org/wp-content/uploads/sites/9/2018/05/PI_2018.05.31_TeensTech_FINAL.pdf (06.10.2022).
- Abidin, Crystal 2018. Internet Celebrity: Understanding Fame Online. (= SocietyNow.) Bingley: Emerald Publishing Limited.
- Appel, René; Muysken, Pieter 2005. Language Contact and Bilingualism. Amsterdam: Amsterdam University Press.
- Auer, Peter 1995. The pragmatics of code-switching: A sequential approach. – One speaker, two languages: Cross-disciplinary perspectives on code-switching. Ed. Lesley Milroy, Pieter Muysken. Cambridge: Cambridge University Press, 115–135. <https://doi.org/10.1017/CBO9780511620867.006>.
- Auser, Merili 2018. Persoonibränding sotsiaalmeedias Eestis tegutsevate *youtuber*'ite, sportlaste ja idufirmade juhtide näitel. Magistritöö. Tartu Ülikool, majandusteaduskond. Tartu.
- Blommaert, Jan 2010. The Sociolinguistics of Globalization. Cambridge: CambridgeUniversityPress. <https://doi.org/10.1017/CBO9780511845307>.
- Diktor 2022 = Elisa blogi. <https://diktor.geenius.ee/blogi/elisa-blogi/tiktok-naitab-eestis-meeletut-populaarsuse-kasvu-ja-soob-youtubei-turuosa/> (06.10.2022).
- Fägersten, Kristy Beers 2017. The role of swearing in creating an online persona: The case of YouTuber PewDiePie. – Discourse, Context & Media 18, 1–10. <https://doi.org/10.1016/j.dcm.2017.04.002>.

- Grosjean, François 1982. *Life with Two Languages: An Introduction to Bilingualism*. Cambridge, Mass: Harvard University Press. <https://doi.org/10.1017/CBO9780511609787>.
- Gumperz, John Joseph 1982. *Discourse Strategies*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511611834>.
- Hepp, Andreas 2018. Meediakultuur. Meediastunud maailmade kultuur. Tlk Katrin Kaugver. (= Bibliotheca mediorum et communicationis.) Tallinn: Tallinna Ülikooli Kirjastus.
- Himma-Kadakas, Marju; Rajavee, Allan; Orgmets, Maarja-Liis; Eensaar, Linda; Kõuts-Klemm, Ragne 2018. The food chain of YouTubers: engaging audiences with formats and genres. – *Observatorio (OBS*)*, 54–75.
- Igav, Reet 2013. Inglise-eesti koodikopeerimine Facebooki vestlustes. Magistritöö. Tallinna Ülikool, eesti keele ja kultuuri instituut. Tallinn.
- Jõgi, Aino 2014. Inglise päritolu sõnad eesti keeles. – *Töid antropoloogilise ja etnolingvistika vallast* 9. Toim. ja koost. Urmas Sutrop. Tallinn: Eesti Keele Sihtasutus.
- Kalmus, Veronika; Pruulmann-Vengerfeldt, Pille; Runnel, Pille; Siibak, Andra 2009. Mapping the terrain of „Generation C“: places and practices of online content creation among Estonian teenagers. – *Journal of Computer-Mediated Communication* 14 (4), 1257–1282. <https://doi.org/10.1111/j.1083-6101.2009.01489.x>.
- Kask, Helin 2021. English-Estonian code-copying in Estonian blogs and vlogs. Väitekirj. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Kask, Kert 2020. Uudissisu tarbimise võimalused Youtube’is ja Eesti noorte sellekohased ootused. Bakalaureusetöö. Tartu Ülikool, ühiskonnateaduste instituut. Tartu.
- Leemets, Tiina 2003. Inglise laenud sajandivahetuse eesti keeles. – *Keel ja Kirjandus* 8, 571–584.
- Liiver, Mari 2019. Mikrokuulsuste meediarepresentatsioon traditsioonilise meedia veebiväljaannetes Eesti juutuuberite näitel. Bakalaureusetöö. Tartu Ülikool, ühiskonnateaduste instituut. Tartu.
- Loog, Mai 1988. Soome mõjusid Tallinna noorte keelekasutuses. – *Keel ja Kirjandus* 2, 83–89; 3, 144–148.
- Loog, Mai 1990. Laensõnad Tallinna koolinoorte slängis ehk ida pole ilma-kaar. – *Keel ja Kirjandus* 3, 167–169.

- Miil, Marget 2019. 9–13-aastaste Eesti noorte vaatamiseelistused Youtube’is ja eetilisel probleemse sisu tõlgendused. Bakalaureusetöö. Tartu Ülikool, ühiskonnateaduste instituut. Tartu.
- Murumaa-Mengel, Maria; Siibak, Andra 2020. From Fans to Followers to Anti-Fans. Young Online Audiences of Microcelebrities. – Reimagining Communication: Meaning. Ed. Michael Filimowicz, Veronika Tzankova. London: Routledge, 228–245.
- Myers-Scotton, Carol 1993. Duelling Languages: Grammatical Structure in Codeswitching. Oxford: Oxford University Press.
- Myers-Scotton, Carol 2002. Contact Linguistics: Bilingual Encounters and Grammatical Outcomes. Oxford: Oxford University Press.
- Niineste, Mart 2017. Estonglish on eestlaste destini osa. – Eesti Päevaleht, 15. juuli. <https://epl.delfi.ee/artikkel/78869794/estonglish-on-eestlaste-destini-osa> (06.10.2022).
- Muuli, Liisi Maria 2017. Eesti *youtuber*’ite eneseesitlus, auditooriumi ja privaatsuse taju. Bakalaureusetöö. Tartu Ülikool, ühiskonnateaduste instituut. Tartu.
- Muuli, Liisi Maria; Perm, Kaari 2017. YouTube’i keskkonnas kasutatavate mõistete ja žanrite definitsioonid. Kodutöö kursusel „Contemporary Issues of Information Age“. Tartu: Tartu Ülikool.
- Oder, Tuuli 2011. Estinglish – eesti keele võit või kaotus? – Postimees, 23. aprill. <https://arvamus.postimees.ee/423067/tuuli-oder-estinglish-eesti-keele-voit-voi-kaotus> (06.10.2022).
- Omnicores Agency 2022 = YouTube by the Numbers: Stats, Demographics & Fun Facts. <https://www.omnicoreagency.com/youtube-statistics/> (06.10.2022).
- Orgmets, Maarja-Liis 2018. Köitev sisu Youtube’i videotes 15–18aastaste õpilaste näitel. Bakalaureusetöö. Tartu Ülikool, ühiskonnateaduste instituut. Tartu.
- Palmeos, Paula 1935. Tartu üliõpilaste erikeelest. – Üliõpilasleht 12, 506–514.
- Pedaja, Kati 2006. Uute laensõnade morfoloogiline kohanemine. – Keel ja Kirjandus 10, 784–796.
- Perm, Kaari 2017. *Youtuber*’ite jälgimise põhjused ja harjumused 15–17 aasta vanuste tüdrukute hulgas. Bakalaureusetöö. Tartu Ülikool, ühiskonnateaduste instituut. Tartu.

- Ratt, Silja 2017. Inglise-eesti koodikopeerimine Maria Rannavälja vlogides. Bakalaureusetöö. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Riboni, Giorgia 2017. The YouTube makeup tutorial Video. A preliminary linguistic analysis of the language of „makeup gurus“. – *Lingue e Linguaggi* 21, 189–205.
- Siibak, Andra; Forsman, Michael; Hernwall, Patrik 2012. Employing Creative Research Methods with Tweens in Estonia and Sweden: Reflections on a Case Study of Identity Construction on Social Networking Sites. – *Journal of Technology in Human Services* 30 (3/4), 250–261.
- Siibak, Andra 2020. Digipõlvkonnast sotsiaalmeedia põlvkonnaks: põlvkondlikku enesemääratlust kujundavad trendid Eesti noorte internetikasutuses. – *Methis. Studia humaniora Estonica* 21 (26), 17–34. <https://doi.org/10.7592/methis.v21i26.16908>.
- Vaigla, A. 1928. Mõningaid jooni üliõpilaskeekest. – *Üliõpilasleht* 7, 94–97.
- Tammemägi, Anni; Ehala, Martin 2012. Eesti õpilaste keelehoiakud 2011. aastal. – *Keel ja Kirjandus* 4, 241–260.
- Torjesen, Aleksander 2021. The genre repertoires of Norwegian beauty and lifestyle influencers on YouTube. – *Nordicom Review* 42 (2), 168–184.
- Vaba, Maarja 2010. Eesti-inglise koodikopeerimine Skype'i Tallinna kontori näitel. Magistritöö. Tallinna Ülikool, eesti keele ja kultuuri instituut. Tallinn.
- Vainola, Kätlin; Kaplinski, Lemmit 2003. Eesti slängi sõnaraamat. Tallinn: Aule Kirjastus.
- Verschik, Anna 2004. Koodivahetus meil ja mujal. – *Keel ja Kirjandus* 1, 25–45.
- Zabrodsckaja, Anastassia 2006. Vene-eesti koodivahetuse funktsioonid Kohtla-Järve venekeelsete laste vestluses. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 2, 231–251. <https://doi.org/10.5128/ERYa2.16>.
- Zabrodsckaja, Anastassia; Kask, Helin 2017. Eesti keele kontaktid. – Eesti rändeajastul. Eesti inimarengu aruanne 2016/2017. Toim. Tiit Tammaru. Eesti Koostöö Kogu, 177–185. <https://www.2017.inimareng.ee/ranne-identiteedi-ja-keelemuutused/eesti-keele-kontaktid/> (06.10.2022).

RESÜMEE

„MIS KEELES MA RÄÄGIN, *I DON'T KNOW*“. ON THE USE OF ENGLISH BY ESTONIAN YOUTUBERS

Kristiina Praakli, Mari-Liis Korkus, Aive Mandel,
Elisabeth Kaukonen, Annika Kängsepp, Triin Aasa,
Kristel Algvere, Helen Eriksoo, Marion Mägi,
Getri Tomson, Liina Lindström
University of Tartu

This article focuses on English use in the example of Estonian-speaking YouTubers. Altogether, we analysed videos from eight content creators, each well-known among high-school-aged viewers who post regular videos in Estonian. The dataset consists of videos (or video excerpts) in which we look into the proportional share of English words or phrases and explore potential functions of code-switching. The results show that while all eight YouTubers use English in multiple videos, the usage frequencies differ significantly and reflect individual differences. English emerged in platform-specific contexts where the words were directly related to content creation (26% of all code-switching cases). Occasionally, the speakers referred to English pop culture phenomena (16%), expressed emotions (12%) and used loanwords or other (embedded) elements (6%). For numerous cases (23%), it was hard to determine why they preferred using an English word or phrase instead of its Estonian equivalent.

Keywords: YouTube, participatory media, content creators, youth language, code-switching, English

Kristiina Praakli on Tartu ülikooli rakenduslingvistika kaasprofessor, kelle uurimisteemad seonduvad teismeliste keele ja mitmekeelse suhtlusega.
kristiina.praakli@ut.ee

Mari-Liis Korkus on Tartu ülikooli doktorant, kelle uurimisteemadeks on teismeliste mitmekeelne suhtlus ja eesti-rootsi koodivahetus.
mari-liis.korkus@ut.ee

Aive Mandel on projekti „Teismeliste keel Eestis“ koordinaator. Ta uurib teismeliste keelt.
aive.mandel@ut.ee

Elisabeth Kaukonen on Tartu ülikooli doktorant, kes uurib sooliselt markeeritud sõnavara eesti keeles.
elisabeth.kaukonen@ut.ee

Annika Kängsepp on Tartu ülikooli doktorant, kes uurib indefiniitpronoomenite käändevormide varieerumist.
annika.kangsepp@ut.ee

Triin Aasa on Tartu ülikooli magistrant, kes uurib intensiivistajaid.
triin.aasa@ut.ee

Kristel Algre on Tartu ülikooli doktorant, kes uurib keelelist ebakindlust.
kristel.algre@ut.ee

Helen Eriksoo, Marion Mägi ja **Getri Tomson** on Tartu ülikooli eesti ja üldkeeleteaduse instituudi üliõpilased.

Liina Lindström on Tartu ülikooli tänapäeva eesti keele professor, kelle uurimisteemad seonduvad keele varieerumisega.
liina.lindstrom@ut.ee

KAS NOORED ON INGLISE KEELEGA OBSESSED? MILLEST RÄÄGIVAD KORPUSANDMED?

Virve-Anneli Vihman, Kristiina Praakli,
Maarja-Liisa Pilvik, Mari-Liis Korkus
Tartu Ülikool

Annotatsioon. Artiklis analüüsime inglise keele kasutust projekti „Teismeliste keel Eestis“ (TeKE)¹ raames 9–18-aastaste noorte suulisest ja tähtsusest kogutud keeleandmetes. Andmeid koguti rahvateaduse meetodil. Nagu TeKE korpuse sagedusandmed ja neist kooruvad mustrid viitavad, tasub teismeliste keelekasutuses pöörata tähelepanu inglise keelele. Artiklis esitame esmase analüüsi inglise keele kasutusest ja osakaalu projekti kaasatud noorte suulise ja tähtsusest andmetes. Tulemused osutavad, et inglise keele osakaal varieerub kõnelejati, ei ole seostatav sooga vanusega, vaid sõltub pigem muudest vestlusega seotud asjaoludest, sh vestlusteemast ja registrist. Küll aga kooruvad soolised ja vanuselised erinevused välja noorte ingliskeelse sõnavara analüüsist.

Võttesõnad: noortekeel, suuline keel, tähtsuse, korpusanalüüs, eesti keel, inglise keel

1. Sissejuhatus

Dominantse maailmakeelena on inglise keel võtnud viimastel aastakümnetel ka Eestis varasemast teistsuguse rolli. Suhtluskeskkondi, kus inglise keel on nähtav ja kuuldav, on oluliselt juurde tulnud veebis, meedias ja avalikus ruumis. Inglise keel on enim õpitud võõrkeel üldhariduskoolides; kõrgharidustasemel peetakse

¹ Projekti „Teismeliste keel Eestis“ (EKKD3) rahastab riiklik programm „Eesti keel ja kultuur digiajastul“ (EKKD).

inglise keelt pigem baasoskuseks kui võõrkeeleks (vt nt Tartu ülikooli keelepõhimõtted 2009²). Nii veebis kui väljaspool on inglise keel atraktiivne ning produktiivne suhtlust vahendav keel (vt lähemalt Zabrodsckaja, Kask 2017; Praakli, Koreinik 2020). Lisaks inglise keele ainetundidele üldhariduskoolis saavad noored mitmekeelsist keelelist sisendit neile tähenduslikest suhtlusvõrgustikest, sh sotsiaalmeediast. Nende lähedane suhe inglise keelega annab neile juurdepääsu üleilmale.

Eestlaste ja eriti noorte väga head inglise keele oskust tõstetakse esile eri raportites (nt EF English Proficiency Index³). Teisalt kostub Eesti avalikkuses arvamusi inglise keele negatiivsest mõjust emakeele oskusele ning keelelisele lohakusele. Nii väitis kultuuriajakirja Akadeemia peatoimetaja 2019. aasta arvamusfestivalil, et inglise keele õpetamine koolides tuleks üleüldse lõpetada, sest „see on kui umbrohi, see tuleb inimestele ise külge“ (Voltri 2019). Muu hulgas on meedias hämmeldust väljendatud noorte inglise keelest mõjutatud keelekasutuse üle (nimetatud ka *Estonglish*’iks), nt: „*Estonglish* jäi igal juhul kõrva kriipima ja ilusast emakeele kasutamisest on selline kõnepruuk kaugel“ (Kull 2019). Eesti keele seisundi uuring (Lukk *et al.* 2017) jällegi vahendas emakeeleõpetajate muret noorte eestikeelse sõnavara muutumise ja kahanemise ning täiskasvanute keelepruugist eemaldumise üle, teisalt tõdedes, et põhjalikud uurinud selle kohta siiski puuduvad.

Noorte spontaanset suhtlust käsitlevaid uuringuid Eesti keeleteaduses napib. Noorte keelepruugile pole teadlastel vanuselistel ja eetilistel asjaoludel ka sugugi lihtne ligi pääseda. Sestap on oluline ja teretulnud noorte kui enda keeles vaieldamatult asjatundjate aktiivne kaasalöömine noortekeele uurimisel, sh nii andmekorjes kui analüüsimisel. Eesti keeleteaduses ongi suurem osa noortekeele uuringuist kirjutatud eeskätt (üli)õpilaste teadustööde vormis

² Inglise keele oskus on eristatud teiste võõrkeelte oskusest: „[...] akadeemilises maailmas on inglise keel kujunenud *lingua franca*’ks.“ Vt <https://www.ajakiri.ut.ee/artikkel/2923> (04.05.2022).

³ EF Education First. <https://www.ef.com/wwen/epi/> (04.05.2022).

(nt Ruven 1999; Rannamäe 2006; Happonen 2010; Lehtpuu 2016). Ka õpilaste teadustööde riiklikule konkursile⁴ (2002–2022) esitatud keealastest uurimistöödest nähtub, et üks noori kõnetavaid teemasid on eakaaslaste sõnavara, mida on vaadeldud nii suulise suhtluse, veebikeskkondade kui kiirsuhtlusäppide keeleandmete põhjal.

„Teismeliste keel Eestis“ (edaspidi TeKE; 2019–2022)⁵ projekt on eesti noorteele uurimisloos esimene, mis analüüsib laiapõhjaliselt korpusandmete toel inglise keele kasutust noortevahelises suhtluses. Projekti raames kogusime keeleandmeid 9–18-aastaste noorte omavahelisest suulisest ja tätsisuhtlusest⁶. Siinses artiklis otsime vastuseid järgmistele küsimustele:

1. Kui palju kasutavad eesti emakeelega teismelised ingliskeelseid sõnu omavahelises eestikeelses vestluses?
2. Milliseid sõnu kasutatakse enim?
3. Kas ja kuidas erineb inglise keele kasutus suulises ja tätsisuhtluses?
4. Kas ja kuidas erineb teismeliste inglise keele kasutus vanuse ja soo lõikes?

Järgmises peatükis võtame vaatluse alla noorte keele uurimise ning koodivahetuse kui keskseima mitmekeelse suhtluse nähtuse noorteelees, eriti praeguses meediastunud maailmas, kus just tehnilistel platvormidel on noorte suhtluses üha suurem roll. Kolmandas peatükis kirjeldame ainekikku ja meetodit ning loodavat korpus. Neljas peatükk tutvustab uuringu tulemusi: anname ülevaate

⁴ Eesti Teadusagentuur. <https://etag.ee/tegevused/konkursid/opilaste-teadustoodet-konkurss/varasemad-konkursid/> (08.08.2022).

⁵ Vt <https://sisu.ut.ee/teke/>.

⁶ Siin ja edaspidi räägime kitsamalt *tätsikeelest* ja *-suhtlusest* – see on üks veebikeele variante, mille aines pärineb kiirsuhtlusäppide vahendusel peetud vestlustest (vrd netikeele mõistega, mis pakuti välja ajal, kui tehniliselt vahendatud suhtlus toimus jututubades, e-kirjades ja/või kommentaarides; vt Hennoste 2000). Veebikeel on heterogeenne keelevariant, mis osaliselt sõltub rakenduste tehnoloogilistest funktsioonidest ning teksti- ja vooruvahetuse kiirusest (vt McCulloch 2019; ka Androutso-poulos 2015).

inglise keele osakaalust suulistes ja tähtsustlustes, koodivahetuse avaldumisest ning sagedasematest ingliskeelsetest sõnadest. Järgnevad arutelu ja kokkuvõte.

2. Millest kõneleb noorte keelekasutus?

Teismelistelt kogutud keeleandmed kui väljavõtted noorema vanuserühma keelelisest käitumisest peegeldavad keele varieerumist, varasemate uurimisandmetega kõrvutavalt ka võimalikke keele muutusi, nt sõnavaras; teisalt ka ühiskonnas toimuvaid protsesse. Noorte keelekasutust kirjeldatakse märksõnadega nagu innovaatiivne, loov, trenditeadlik (vt Eckert 1997, 2000; Jonsson *et al.* 2019). Noorte keeleline minapilt saab olulisi täiendusi ja mõjutusi murdekeeleks, mil muutub olulisemaks just eakaaslastelt saadav sisend. Nende keelekasutuse põhjal saab aimu neid kõnetavate prestiižsete keelte, suhtluses aktsepteeritud mitmekeelsete suhtlusnähtuste ja noori kõnetavate vestlusteemade kohta.

Ühiskondlikus kontekstis peegeldavad noorte keelelised repertuaarid nii kohalikke kui üleilmselt haaratud muutusi ja protsesse. Praegusaja noorte puhul hõlmab meedia ja meediastumine (Hepp 2018) nende kõiki igapäevapraktikaid: kooliskäimist, vaba aja veetmist ning omavahelist läbikäimist. Kooliealistel on COVID-19 pandeemia ja sellega kaasnenud e-õppe rohkus seda tendentsi veelgi süvendanud. Eeskätt meediastumisest ning aktiivsest sotsiaalmeedias toimetamisest tulenevalt saavad noored keelelist sisendit ka üle ajaliste ja ruumiliste vahemaade neile tähenduslikest võrgustikest ning vahendavad seda enda võrgustikesse. Noorte hea inglise keele oskus toetab võrgustikele juurdepääsu, ühtlasi saades neilt toetust ja sisendit. Meediastumine ja platvormide kättesaadavus toob noorte keelepruugi nähtavamale ja kuuldavamale kui eales varem. Siiski pole noortekeele puhul põhjust kõnelda diskreetsest keelekujust, kuna see on varieeruv ja kiiresti muutuv ning sisaldab erinevaid kontaktmõjusid.

Teismeliste keelepruugis avalduv suhtlusvahendite mitmekesisus nähtub noortekeele uurimustes (vt nt Rampton 1995; Kern,

Selting 2011; Nortier, Svendsen 2015; Tagliamonte 2016). Suhtluse sotsiolingvistika raamistikus on eeskätt Euroopa suurlinnade näitel uuritud noorte mitmekeelseid keelekujusid ning suhtlusnähtusi (nt *code-switching*, *crossing*), sageli ühendatult teise keele õppimise ja omandamise küsimustega (vt nt Jørgensen 2002, 2008; Källström, Lindberg 2011; Lehtonen 2015; Jonsson *et al.* 2019). Omaette valdkonna moodustavad nn slängisõnavara uuringud, sh ka tabu- ja vandesõnade uuringud (vt Stenström 2020). Digiajastu on suulise keele kõrval toonud fookusesse suhtluspraktikad kiirsuhtlus- ja mängurakendustes (vt Durkin *et al.* 2010).

Eesti noortekeele kogumine ja teaduslik uurimine sai alguse eelmise sajandi esikümnenditel, kuid järjepidev uurimistöö selles vallas on siiski lünklik, põhiosa moodustavad gümnasistide ja üliõpilaste uurimused (vt ülevaated Tender 1994; Hennoste 2000; Praakli 2022). Uurijate rõhuasetus on põhiliselt lasunud noorte sõnavaral (lähemalt Praakli, ilmumas), kus keelekontaktidest johtuvaid keelemuutusi on näha laensõnades ja mitmekeelsetes suhtluspraktikates (nt Loog 1988, 1990). Alates nullindatest on üha rohkem tähelepanu pälvinud inglise keele kasutus, seda nii suuliste (intervjuud, argivestlused, vlogid; nt Lehtpuu 2016; Ratt 2017; Korkus 2021) kui digisuhtluse andmete (tšätid, blogid; nt Kask 2021) valguses.

2.1. KOODIVAHETUSE UURIMISEST

Mitmekeelsusuuringute fookus lasub kahe või enama keele(variandi) kooskasutamise ja suhtluskontaktidest johtuvate muutuste kirjeldamisel. Üks mitmetest kontaktsituatsioonidest ilmnevatest ja noortekeelele iseloomulikest nähtustest, mis ka siinse artikli keskmes, on koodivahetus. Põhitähenduses tähistab koodivahetus mitmekeelset suhtlusviisi, kus lausungis või vestluses kasutatakse vähemalt kahte keelt või keelevarianti (Verschik 2004; Myers-Scotton 2017).

Koodivahetusuuringud on üldiselt keskendunud kas selle sotsio-pragmatilistele funktsioonidele (nt Auer 1998) või lausungisisestele eri keele elementide morfosüntaktilise integreerimise küsimustele

(nt Myers-Scotton 1993; Deuchar 2020). Artiklis käsitleme koodivahetusilminguid eeskätt leksikaalsest vaatest, keskendudes sisestatud ingliskeelsetele elementidele eestikeelses diskursuses.

Koodivahetuse kirjeldamisel on oluline eristada, milline on vestluskontekstis põhi-, milline sisestuskeel (vt nt Myers-Scottoni 1993 maatrikskeele mudel). Siinses artiklis loeme keeleandmetest lähtuvalt vestluste põhikeeleks vaikimisi eesti keele. Poplack (1980) eristab koodivahetusilmingute puhul kolme põhitüüpi. Esimeseks on lausesisene koodivahetus (ingl *intrasentential* CS), kus keeltevaheline üleminek toimub ühe lausungi või lause sees (näide 1⁷). Teise tüübina eristab Poplack lauseülest koodivahetust (ingl *inter-sentential* CS, ka *code alternation*), kus keeltevaheline üleminek toimub lausungite või lausete vahel (näide 2⁸). Viimasena eristab Poplack süntaksiväliste diskursuspragmaatiliste üksuste (nt partiklite, hüüdsõnade) ülekannet (ingl *tag-switching*). Need paiknevad tavaliselt lausungi või lause alguses (näide 3) või lõpus (näide 2, *right* 'õigus, eks') mistahes positsioonis.

1. **CHA10M:** *right clicki* mu nime peale
2. **CHA14M_1:** su emakeel on eesti keel *right*?
CHA14M_2: *nah* hispaania keel
CHA14M_1: *sirius*
CHA14M_2: *ofc* its eesti keel
CHA14M_1: *what* *languses* *u* *speak* and *how* *bad* or *good*
CHA14M_2: inglise keel suht hea
CHA14M_2: vene keel suht sitt
3. **SUU14N_1:** ja siis kõik arvasid et me k- ee nagu meeldime
 üksteisele ja käime
SUU14N_1: aga me (.) me ei käinud

⁷ Igas näites sisaldab lausungile eelnev kood korpus (SUU/CHA), kõneleja vanust, sugu ja (vajadusel) järjenumbrit. Kalkkiri tähistab sisestuskeele elemente.

⁸ Siinne näide ilmestab muuhulgas vajadust noorte keelehoiakute väljendamise lähemaks analüüsiks.

SUU14N_2: *vau*[_wow_] tõesti vä[või]

SUU14N_1: jaa (...) jah

SUU14N_2: *oo_mai_kaad*[_oh_my_god_] ma arvasin et sa nagu täiega käisid temaga

Kvantitatiivne lähenemine võimaldab eristada suure hulga inglise keele näidete põhjal ka kasutatavate keelte määra. Kuna lühikeste lausungite puhul on väiksem võimalus kasutada kahte keelt koos, siis eristame andmeanalüüsis kuni kahekeelset (ühikeelset) lausungit pikematest.

Morfofonoloogilise tasandi kirjeldamisel on mitmekeelsusuuringuis rakendatud ka Lars Johanson koodikopeerimise mudelit (Johanson 2002; Verschik 2008; Kask 2021). Mudel eristab kolme kopeerimistüüpi: täielikud (ingl *global copy*; näide 4, milles ingliskeelsed sõnad on üle kantud terviklikul kujul, häälduse ja tähendusega), valikulised (ingl *selected copy*; näide 5a–b, milles kõnelejad kasutavad ingliskeelsete sõnade omadusi valikuliselt ning ortograafia viitab eestipärasele hääldusele) ja segakopiiad (ingl *mixed copy*; näide 6, milles liitsõna sisaldab mõlema keele elemente):

4. SUU16N: see oli päris *lol*[_LOL_] ikka
SUU16N: aga no ma olin et *jolo*[_YOLO_]
5. a. CHA17M: valgus *is veri* good
b. CHA17N: *Yesyes iz uki dunt vörri* [it's ok don't worry]
6. CHA16N: nagu ma realselt tegelen õppimisega ja siis unustan täiesti ära, et *livetunnid* on kaa

Koodivahetuses esinevate keelte eristamine toetub traditsioonilisele vaatele, mille kohaselt keeled on eraldiseisvad süsteemid. Uuema kontseptsioone (nt *metrolingualism*, *translanguaging*, *superdiversity*, *polylingualism*; vt tähenduslikest erinevustest Pennycook 2016) ilmestab arusaam, et mitmekeelse kõneleja keeled moodustavad ühe süsteemi, samuti ei sõltu koodivahetus kõneleja ühe või teise keele oskusest, vaid tema suhtluseesmärkidest (Androutsopoulos 2015). Seda väidet toetab nähtus, kus noored kasutavad eri keelte sõnu või

väljendeid ilma, et nad keelt valdaksid, nt inglise keelt kõnelevad ameerika noored hispaaniakeelset sõna *mamacita* või eesti noorte-keeles vene keelest mugandatud *taipohh*.

Artiklis loome ülevaatliku kvantitatiivse aluse, millest hiljem liikuda lähemale kvalitatiivsele analüüsile. Kõik ülal nimetatud koodivahetuse avaldumisvormid on kaasatud ka siinse artikli kvantitatiivsesse analüüsi, koodivahetuse struktuuralse integreerimise ja eri tüüpide analüüs järgneb edasistes uurimustes.

2.2. (SOTSIAAL)MEEDIASTUNUD KEELE REGISTRID⁹

Kaasaja noortekultuuris mängib olulist rolli sotsiaalmeedia.¹⁰ Noored on selle regulaarsed tarbijad eelkõige nutiseadmete kaudu (Lauricella *et al.* 2014). Selle valguses nimetab Siibak tänapäeva noori mitte digi- vaid sotsiaalmeediapõlvkonnaks (2020: 30), kuna nooremates vanuserühmades on aktiivne sotsiaalmeedia kasutus kujunenud sotsiaalseks normiks ning põlvkondliku identiteedi vahendiks.

Just üleilmastumisel (Bucholtz, Skapoulli 2009; Otsuji, Pennycook 2010) ja meediastumisel (Leppänen 2007) on noorte keelekasutuses keskne roll. Nende mõju nähtub mitmekihiliselt:

1. Mitmeregistrilisus: kasutatakse akronüüme jm veebipõhiseid keelendeid (nt *OMG* ehk 'oh my god', *XD* 'iksdee' ehk visuaalselt tekkinud naerunägu, *DM* 'dii emm' ehk 'direct message'; vt ka Crystal 2008);
2. Mitmekeelsus: kasutatakse noortekeeles üleilmselt tuntud sõnu või väljendeid (nt *hot girl summer* 'kuuma tüdruku suvi', *on fleek* 'tasemel, ideaalne', *vibe* 'hõng, meeleolu') ning jälgendatakse nende hääldust (vrd nt sõna *meme* erinevaid

⁹ Mõistame *registri* all kasutuskeskseid varieeruvaid keelekujusid, keelevariante; siinse artikli raamides suulist argikõnet ning tähtkeelt.

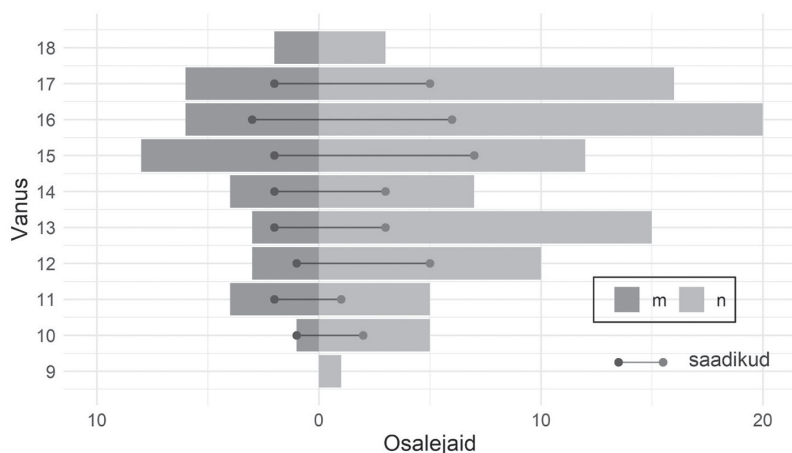
¹⁰ Võib ka väita, et sotsiaalmeedia mõju on laiem ning läbinähtav eri ühiskonna kihtides, seetõttu räägitakse ka tänapäeval platvormiühiskonnast (ingl *Platform Society*; vt lähemalt van Dijck *et al.* 2018).

hääldusvariante: eestipärane 'meem', ortograafiapõhine 'meme', ingliskeelne 'miim');

3. Mitmeallikalisus: viidatakse erinevatele allikatele ja tsiteeritakse filmides/raamatutes/sarjades/veebikeskkondades kuulnud-nähtut (vt Sierra 2021).

Veebivõrgu laiemasse kasutusse jõudmine ning infoajastu tehnilised võimalused on laiendanud suhtluskeskkondi, neis rakendatavaid vahendeid (nt e-kirjad, foorumipostitused, kommentaarid, *online*-mängude sõnumivahetused jms) ning mitmekesistanud suhtlus- ja tekstiloomi viise. Ühtlasi on tehniliselt vahendatud suhtlus loonud keelekontaktidele uue pinnase ning muutnud seeläbi kontaktis olemise kättesaadavust, viise ja võimalusi. Veebikontaktid haaravad laiemat kõnelejaskonda, kontaktid on intensiivsemad, suhtlusesse kaasatakse erinevaid keeli ning muid elemente (emotikonid, videolingid jm). Teaduslik huvi veebisuhtluse ning selle keelevariantide vastu sai alguse 1990ndatel (nt Ferrara *et al.* 1991; Donath *et al.* 1999; Crystal 2001). Ka eesti veebisuhtluse kohta kirjutati esimesed uuringud samal ajal (nt Salla 1999), hilisemalt on – peamiselt üliõpilastööde vormis – vaadeldud nii eri keskkondade suhtlusvahendeid kui ka mitmekeelset suhtlust nt blogides-vlogides, tsättides (nt Igav 2013; Kilp 2017; Ratt 2017; Kask 2021).

Nii nagu suulises keeles, esineb ka tsätikeeles üsna palju varieeruvust (vt lähemalt McCulloch 2019). Tagliamonte ja Denis (2008) toovad välja, et kiirsuhtlusvestlustes kombineerivad noored formaalseid (nt kirjakeelseid vorme *shall* ja *must*) ja mitteformaalseid (nt emotikonid, lühendid) keeleelemente. Platvormide mitmekülgne funktsionaalsus võimaldab kasutajatel korraka tegutseda mitmes rakenduses ning sellest tulenevalt on nende eneseesitlus ja tähendusloome platvormi erinev (Deakin, Wakefield 2014; Boczkowski *et al.* 2018; Tagg, Lyons 2021). Uurimishuvi veebikeele (ja kitsamalt ka tsätikeele) vastu on jõudsalt kasvamas, eriti võib seda märgata (üli)õpilasuuringutes (ehk noored uurivad noori; vt ka 2).



Joonis 1. Kõigi projektis osalenud noorte vanus ja sugu (keeleaadikud ja kaasatud sõbrad)

3. Ainestik ja meetod

Analüüsitavad keeleandmed kogusime projekti „Teismeliste keel Eestis“ raames. Selle põhieesmärk on luua noortekeele korpus, mis võimaldab uurida noorte spontaanset keelekasutust suulistes ja tähtvestlustes. Selleks kutsusime eri piirkondade üldhariduskoolide 4., 6., 8. ja 10. klassides õppivaid õpilasi projekti nn keeleaadiku-teks.¹¹ Vastava koolituse läbinuna kaasasid keeleaadikud projekti enda sõpru, salvestasid nendega suulisi vestlusi (keskmiselt kaks tundi aadiku kohta), lisaks kogusid ja vahendasid uurimisrühmale tähtides peetud jutuaajamisi. Kõik artiklis kasutatavad keeleandmed ja -näited pärinevad TeKE korpusest.

Kokku osales andmekorjes 131 noort vanuses 9–18 eluaastat. Joonis 1 annab ülevaate korpuse vanuselisest ja soolisest koosseisust. Kuigi sihtisime vanuselist, soolist ja piirkondlikku tasakaalu, on valimis ülekaalukalt naissoost osalejaid (94 tüdrukut, 37 poissi).

¹¹ Projekti rahvateaduse meetodil põhinevast andmekorjest lähemalt vt Mandel *et al.* (2022).

Keelesaadikud pärinesid Tallinna, Tartu, Saaremaa ja Võrumaa koolidest, kuid nende kaasatud sõprade kaudu on projektis esindatud 11 maakonda (lisaks üks välisriigis elav õpilane).

3.1. KORPUSTE KIRJELDUS

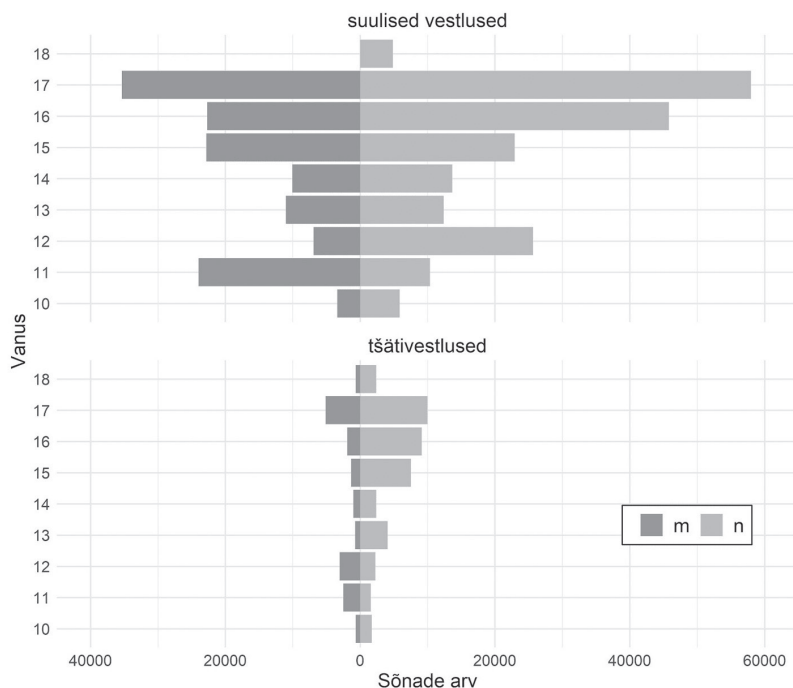
Andmekorjega kogusime 97 tundi helisalvestusi ning 111 521 sõna mahus tšätte (valdavalt Messengerist, mõnevõrra Discordist). Lisaks küsisime noortelt veebiküsitluse kaudu nende (võõr)keeleoskuse ja -kasutuse ning internetitarbimise kohta. Kogutud materjalist on artikli kirjutamise hetkel (mai 2022) suulistest andmetest transkribeeritud kujul analüüsivalmis 64 kõneleja suulised vestlused (kokku 39 tundi, 335 822 sõna¹²) ning tšätivestlused 97 osalejalt (mahus 59 688 sõna¹³). Keskmine sõnade arv kõneleja kohta suulistes vestlustes on 5247 ($sd = 3058$), tšätivestlustes 585 sõna ($sd = 531$), ent kõnelejatevaheline varieeruvus on väga suur. Valimi sõnade arvu vastavalt kõneleja soole ja vanusele kuvab joonis 2.

Suuliste vestluste helifailide transkribeerimiseks kasutasime annotatsioonitarkvara ELAN¹⁴. Iga lausungi puhul märgistasime selle piirid, transkribeerisime iga kõneleja kõnet eraldi transkripsiooniridadel ning märgendasime kõnekeelsusi ja kasutatud keele. Kõik isikuandmed ja muu isiku tuvastamist võimaldav materjal on muudetud. Keelejuhtide kõne transkribeerimisel lähtusime kohendatud kirjakeele ortograafia põhimõtetest: kirjutasime üles kõik keeleüksused, sh üneemid, häämitsused, poolikud sõnad, oletatavad sõnad (st sõnad, mille täpne kuju pole eristatav, kuid on kontekstist võimalik oletada). Kirjakeele ortograafiat ei kasutatud juhtudel, kui kõneleja keelepruuk sellest süsteemselt erines: näiteks üldises kõnekeeles laialt levinud (nt *keka*, *õps*, *põhimõtst*, *kakskend*) või kirja-keelest lahknevad (nt *videosid*, *praegult*) vormid. Sellistel puhkudel

¹² Sõnadeks on arvatud ka (nt eneseparanduste tõttu) pooleli jäänud sõnad, üneemid ja partiklid.

¹³ Sõnadeks on arvatud ka lingid, pildid, emotikonid ja (sõnumi)reaktsioonid.

¹⁴ ELAN (versioon 6.2). <https://archive.mpi.nl/tla/elan> (03.05.2022).



Joonis 2. Valimi sõnade arv vastavalt kõneleja soole ja vanusele

lisasime sõna järele nurksulgudesse sõna kirjakeelse vormi (nt *kakskend*[*kakskümmend*], *praegult*[*praegu*]), seda eeskätt järgmises etapis rakendatava automaatse morfoloogilise analüüsi tarvis. Nurksulud ja alakriipsud eristavad võõrkeelseid sõnu ja pikemaid võõrkeelseid fraase (nt *dõup*[*_dope_*], *õu_mai_gaad*[*_oh_my_god_*]). Eraldi keelereal on märgendatud teise keele kasutust (nt inglise, vene, hispaania; ka eesti murded). Kõik transkribeeritud failid vaatas üle, parandas ja ühtlustas veel vähemalt üks transkribeerija.

Tšätiaainestiku edastasid keelesaadikud Wordi-failidena. Nendest eemaldasime sõnumisaatja nime sisaldavad read (nt *Maarja sent*), samuti muutsime ära tekstides esinevad isikunimed ning isiku tuvastamist võimaldava sisu. Kõnelejate voorud eristasime failis eri taustavärvidega. Wordi-failid konverteerisime xml-failideks, mis

võimaldas html-värvikoodide abil eristada üksteisest eri kõnelejate voorud ning sisuteksti (sh pildid, lingid, emotikonid, (sõnumi) reaktsioonid) igasugusest metainfost (kuupäevad, profiilipildid, teavitustekstid; nt *You received a call from*). Eraldatud info koondasime tabelkujul tsv-failidesse, mille ridadel oli iga tähti tekstirida ning tulpades seda iseloomustav metainfo (sisutekst/metatekst, kuupäev ja kellaaeg, osaleja kood). Sisutekstile lisasime märgendustööriista INCEption¹⁵ (Klie *et al.* 2018) abil info võõrkeelte, lühendite (nt *mdea* 'ma ei tea', *idk* 'I don't know') kasutuse ning konfidentsiaalse teabe kohta.

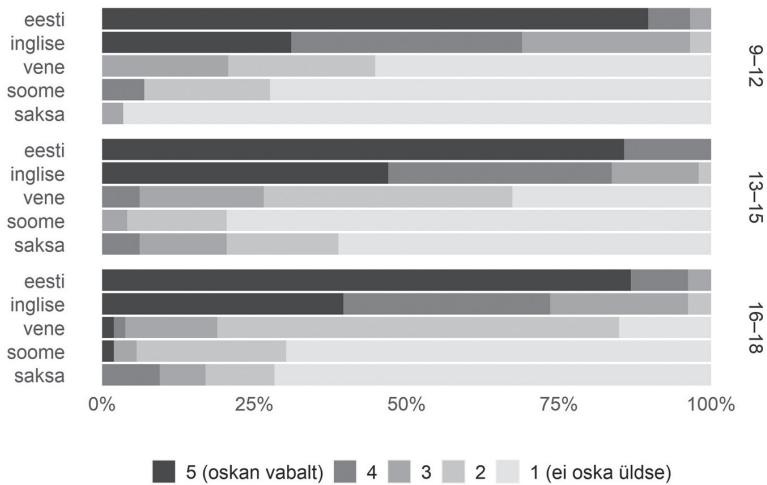
3.2. ÜLEVAADE OSALEJATEST

Projektis osalenud noorte taustainfo kogusime veebiküsimustiku kaudu. Siinses artiklis on relevantssed noorte antud hinnangud oma keeleoskuse kohta (vt joonis 3) ning kiirsuhtlusäppide kasutusharjumused (vt joonis 4).

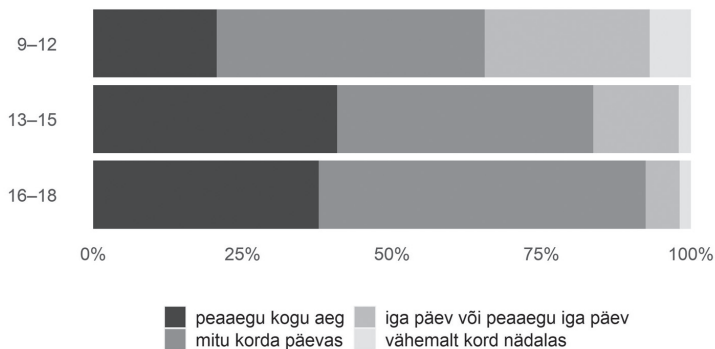
Keeleoskuse andmed põhinevad vastajate hinnangutel. Selleks paluti märkida viiepalliskaalal, kuidas nad iga keele puhul oma keeleoskust hindaksid. Ootuspäraselt hindavad kõik vanuserühmad kõige kõrgemaks eesti keele oskust. Ka inglise keele oskusele antud hinnang on väga kõrge (viiepalliskaalal hinded 4 või 5), st noored on arvamusel, et nad oskavad inglise keelt „vabalt“ või „väga hästi“. Ühtlasi näeme, et inglise keele oskuse hinnangud sarnanevad pigem eesti keele kui teiste küsitud võõrkeelte (vene, soome, saksa) oskuste hinnangule. Inglise keele väga hea oskus võib olla seotud mitme teguriga, ent põhilisena nähtub tugev seos osalusmeediaga (vt Siibak 2020), kus märkimisväärne osa argisuhtlusest ning rahvusvahelisest noortekultuurist on ingliskeelne.

TeKE projektis osalenud teismeliste vastustest nähtus, et enamik suhtleb kiirsuhtlusäppides kas peaaegu kogu aeg või mitu korda päevas, seejuures on kasutus tihedam vanemate osalejate (13–15- ja

¹⁵ INCEption (versioon 22.4). <https://inception-project.github.io> (03.05.2022).



Joonis 3. Osalejate hinnangud enda keeleoskusele



Joonis 4. Osalejate vastused küsimusele „Kui sageli suhtled kiirsuhtlusäppides?“

16–18-aastaste) seas. Vaid marginaalne osa (alla 5%) vastanuist kasutab suhtlusäppe harvem kui kord päevas. Populaarseimad kiirsuhtlusäpid on Messenger (54%), Snapchat (27%) ja Discord (10%), viimane eelkõige poiste hulgas. TeKE projektile edastatud sõnumi-vestlused peeti valdavalt Messengeris, väiksem osa Discordis.

4. Tulemused

4.1. INGLISE KEELE OSAKAAL SUULISTES JA TŠÄTIVESTLUSTES

Kvantifitseerisime koodivahetuse intensiivsust noortekeeles võrdlemisi lihtsal viisil ning hindasime seda kui ingliskeelsete sõnade osakaalu¹⁶ kõigist kõneleja sõnadest. Olgu öeldud, et see, mis on korpuses märgitud inglise keeleks, on kohati vaieldav, kuna hulk ingliskeelseks loetud sõnu on eesti keeles võrdlemisi hästi kohane-nud (nt *sorri*, *tšätt*, *vaib*). Korpuses on märgendatud võõrkeeli pigem inklusiivselt, et koodivahetuse uurijal oleks võimalik luua andmete alamhulki enda vajadustest lähtuvalt.

Ingliskeelsete sõnade protsenti võib siinjuures pidada võrdlemisi konservatiivseks hinnanguks koodivahetuse intensiivsusele, kuna suulistes vestlustes on sõnadeks loetud kõik semantilised ja grammatilised sõnad, lisaks üneemid, poolikud sõnad jm suulise kõne komponendid, millest osadel keelemärgendit ei ole. Tšättides omakorda lähevad sõnadena arvesse ka muud semantilist sisu kandvad elemendid, nagu pildid, emotikonid ja lingid, millel samuti keelemärgendid puuduvad.

Leidsime, et suulistes vestlustes on ingliskeelsete sõnade protsent keskmiselt 3,3, tšätivestlustes aga märgatavalt kõrgem – 7,3. Eesti- ja ingliskeelsete sõnade osakaalu soo ja vanuserühmade lõikes illustreerib tabel 1.

Ingliskeelsete sõnade esinemise kohta selgeid mustreid ei soo ega vanuserühmade lõikes välja ei joonistu. Võib eeldada, et noored omandavad inglise keelt juba varakult (nagu viitab joonis 3) ning näeme, et ingliskeelsete sõnade osakaal vanusega ei kasva. Vanusest olulisemat rolli näivad mängivat kõneleja individuaalsed valikud

¹⁶ Edasised analüüsid põhinevad suuliste vestluste transkriptsioonide lausungi- ja sõnakihil, kus võõrkeelsed sõnad on eristatud eestikeelsetest, ent milles konkreetse keele kohta infot pole. Transkriptsioonide keelekihi põhjal oleme aga teinud kindlaks, et võõrkeele kasutusjuhtudest üle 99% moodustab just inglise keel, seega räägime siin ja edaspidi vaid inglise keelest. Sama kehtib tšätivestluste kohta.

Tabel 1. Eesti- ja ingliskeelsete sõnade osakaal soo ja vanuserühmade võrdluses

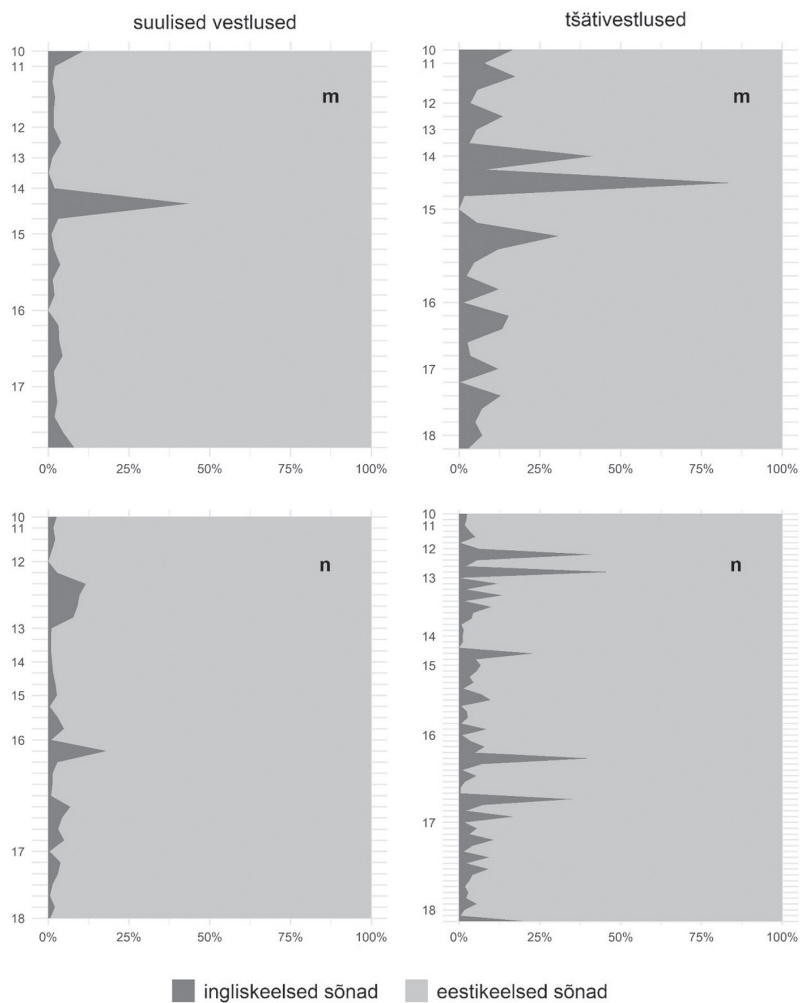
Vanus aastates ja sugu		Suulised vestlused		Tšätivestlused	
		ee % (N)	en % (N)	ee % (N)	en % (N)
9–12	m	97% (33 317)	3% (941)	90% (5586)	10% (587)
	n	95% (39 699)	5% (2198)	91% (5103)	9% (480)
	Kokku	96% (73 016)	4% (3139)	91% (10 689)	9% (1067)
13–15	m	95% (41 653)	5% (2307)	82% (2877)	18% (649)
	n	98% (48 037)	2% (991)	95% (13 359)	5% (656)
	Kokku	96% (89 690)	4% (3298)	93% (16 236)	7% (1305)
16–18	m	97% (56 081)	3% (1934)	93% (7144)	7% (518)
	n	97% (105 847)	3% (2817)	93% (21 232)	7% (1497)
	Kokku	97% (161 928)	3% (4751)	93% (28 376)	7% (2015)
Kokku		96,7% (324 634)	3,3% (11 188)	92,7% (55 301)	7,3% (4387)

(nt veebikeskkondades veedetud aeg, suhtlusvõrgustik jne), vestlusteema ja -olukord. Joonis 5, kus on esitatud eesti- ja ingliskeelsete sõnade suhtelised sagedused, illustreerib võõrkeelse keeleainese osakaalu ulatuslikku varieerumist üksikkõnelejate lõikes, aga ka registriti.

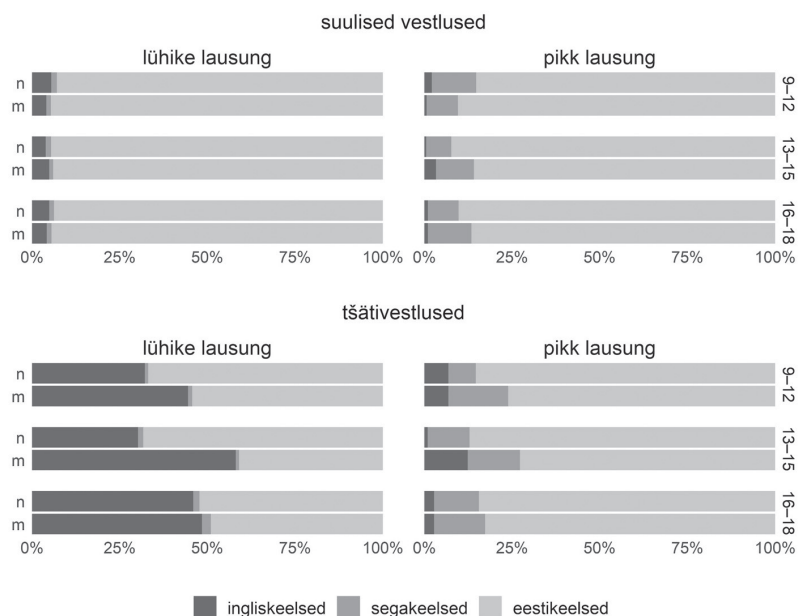
4.2. LAUSESISENE JA LAUSEÜLENE KOODIVAHETUS

Lisaks ingliskeelsete sõnade osakaalule huvitusime ka sellest, kui-võrd ulatuslikult esineb teismeliste suhtluses lausesisest¹⁷ ja lause-ülest koodivahetust. Selleks eristasime täielikult eestikeelsed lausungid, täielikult ingliskeelsed lausungid ning lausungid, milles esineb mõlema keele elemente. Suulistes vestlustes on eestikeelseid lausungeid 90,3%, ingliskeelseid 2,3% ja segakeelseid 7,4%;

¹⁷ Siia alla liigitub ka diskursuspragmatiliste üksuste ülekanne, nn sildivahetus (ingl *tag-switching*), mida meie valitud meetod ei võimaldanud eristada.



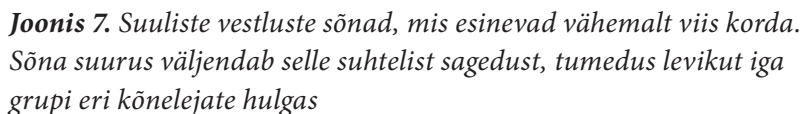
Joonis 5. Inglis- ja eestikeelsete sõnade suhtelised sagedused individuaalsete kõnelejate keelekasutuses



Joonis 6. *Inglisekeelsete, eestikeelsete ja segakeelsete lausungite suhteline sagedus suulistes ja tšätivestlustes. Kuni kahesõnalised (lühikesed) lausungid on eristatud pikematest*

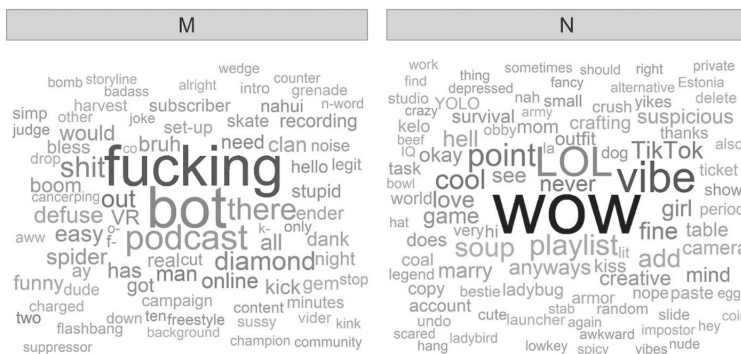
tšätivestlustes on vastavad protsendid 85,1%, 7% ja 7,9%¹⁸. Seega on täiesti võõrkeelsete lausungite osakaal tšätikorpuses märgatavalt suurem ja koodivahetuse intensiivsus üldisemalt kõrgem, olgugi et suuliste ja tšätivestluste lausungid ei ole omavahel üks ühele võrreldavad. Tšätivestlustes tulevad selgemini esile ka sugudevahelised erinevused: nooremates vanuserühmades on poisid märksa aktiivsemad koodivahetajad, mis peegeldab paljuski nende vestlusteemasid (ainestiku näitel on selleks sport ja arvutimängud).

¹⁸ Keskmine lausungi pikkus on suulistes vestlustes 6,26 sõna ($sd = 5,3$) ja tšätivestlustes 4,13 sõna ($sd = 8,26$). Viimaste puhul on kahe korpuse võrreldavuse huvides lausung võrdsustatud ühe kõneleja ühe saadetud sõnumiga.



(sageli omadussõnad, nt *cool, nice, creepy, basic*) või emotsioone väljendavad sõnad (mh intensiivistajad ja tugevamaks peetud ropp-sõnad, nt *fucking, shit, bitch, damn*), viimased eriti poiste kõnes. Suur osa nooremate kõnelejate ingliskeelsest sõnavarast seondub mängukeskkondadega (nt *block, skin, armor, bot, defuse*), kusjuures kõige nooremas vanuserühmas on seda tüüpi sõnavara levinud nii poiste kui tüdrukute keeles. Vanemates vanuserühmades annab eriti tüdrukute sõnavaras tooni erinevate sotsiaalmeedia platvormidega seostatav sõnavara (nt *selfie, playlist, chat, story*). Sotsiaalmeedia ja mängudega seotud ingliskeelne sõnavara näibki moodustavat suure osa noorte nn erialaterminoloogiast, milles peegelduvad nende igapäevapraktikad ja eelistatud ajaveetmisviisid (vt ka Praakli *et al.* samas kogumikus). Huvitava tähelepanekuna nähtub asjaolu, et kui korpuse nooremates vanuserühmades vestlevad enamasti poisid poistega ja tüdrukud tüdrukutega, siis vanemas vanuserühmas vestlevad poisid rohkem ka tüdrukutega, seetõttu on ka vanema vanuserühma ingliskeelne sõnavara sugude lõikes sarnasem.

Selleks, et võrrelda ingliskeelse sõnavara kasutust soopõhiselt veidi lähemalt, eristasime omakorda sõnad, mille suhteline sagedus ühes rühmas oli vähemalt kolm korda suurem kui teises rühmas. Joonis 8 kuvab vasakul ingliskeelsed sõnad, mis on vanuserühmade



Joonis 8. Ingelskeelsete sõnade kasutus suulises korpuses, eri soost kõnelejate hulgas

Nn mängusõnavara ning negatiivseid emotsioone väljendavad sõnad (nt *shit, fucking, dank*) on omased pigem poistele, positiivsed emotsioonid ja hinnangud (nt *cool, fine, love*) ning sotsiaalmeedia kasutamise seonduv sõnavara (nt *playlist, survival, crafting*) pigem tüdrukutele. Poistele ainuomane on ka erinevate n-ö maskuliinsete pöördumismarkerite (nt *dude, man, bruh*) kasutamine.

Sagedased ingliskeelsed sõnad tähtsates (joonis 9) jaotuvad oluliselt ühtlasemalt. Lisaks funktsioonisõnadele (nt *to, the, a, than*) annab tooni tähtsusele kui registrile iseloomulik lühendite kasutus (nt *lol* 'laughing out loud', *plz* 'please', *gn* 'good night', *idk* 'I don't know', *wym* 'what [do] you mean'). Tuleb ka arvestada, et sõnade



Joonis 9. Tšätivestluste sõnad, mis esinevad vähemalt kolm korda. Sõnaloendites on ühtlustatud suur- ja väiketähe kasutust, et hajuvas andmestikus sagedusi koondada

kirjapilt on tähtsates oluliselt varieeruvam kui transkriptsioonides, kus ei tähistatud kogu häälduslikku varieerumist, mistõttu sagedaste tähtsõnade loendisse jõuab võrdlemisi juhuslik hulk sama keelelist funktsiooni täitvatest sõnadest (nt *yay*, *yayyy*, *jei*, *jeei*, *yei*).

5. Arutelu ja kokkuvõte

Siinne uuring annab esmase ülevaate eesti teismeliste keelekorpuse andmetest. Artiklis tutvustame kvantitatiivseid andmeid teismeliste kakskeelsetest suhtluspraktikatest nendevaheliste suuliste ja tähtsõnadest näitel. Inglise keel on küll laialt levinud (inglisekeelseid sõnu leidub pea kõigilt kõnelejatelt), kuid selle osakaal ei ole ülemäära suur. Suulises suhtluses moodustab inglise keel kõikidest sõnadest ca 3%, sh ei sõltu inglise keele osakaal vanusest ega soost. Erinevused ilmnevad aga registritevahelises võrdluses: tähtsõnades moodustab ingliskeelsete sõnade maht ainekust 7,3%; ühes vanuserühmas (poisid vanuses 13–15 a) koguni ligi viiendiku (18%) kõigist sõnadest.

Sõnade sagedusloend näitab inglise keele esinemise temaatilisi ja funktsionaalseid erinevusi eri vanuses ja soost kasutajatel. Inglise keele atraktiivsus seisneb prestiižis ja valdkondade üleses ning eri suhtluseesmärkidega kasutusulatuses. Ühtlasi on inglise keel kaasaja noortekultuuri üks tunnuslikke jooni; sestap võib ka öelda, et selle justkui märkamatu kasutus eestikeelses suhtluses on osa grupikuuluvust tähistavast koodist. Teisest küljest viitavad andmed sellele, et inglise keele esinemine pole sõnamahult sedavõrd dominantne, nagu võib jääda mulje meedias avaldatud arvamustest.

Noorte sõnavaras ilmnevad selged seosed tänapäeva internetikultuuriga, mille kaudu nende keelerepertuaari jõuab sageli mitmesuguseid ingliskeelseid sõnu ja väljendeid. Võib eeldada, et sotsiaalmeedia on vahendaja, mille kaudu omandatakse varakult inglise keelt (vrd võõrkeelsete telekanalite roll möödunud aastakümnetel). Sõnavara näitab, kas ja kui palju tarbitav sisu eri vanuserühmade lõikes sarnaneb või erineb. Kuigi sagedusloenditest nähtub, et

mitmesuguseid meemiväljendid ja -sõnad on üsna levinud (nt *dank, vibe, lol, basic*), ei tähenda see, et noorte inglise keele kasutust ise-loomustavad üksnes trendikad keelendid. Kõige sagedasemad üksused (nt *oh my god, wow, fucking*) on kasutusel ka väljaspool sotsiaalmeediat. Noorte keelekasutuses esineb rohket individuaalset varieerumist, mistõttu tulevikus on oluline lisada sõnapilvedele detailsemat analüüsi, et kirjeldada ingliskeelse sõnavara kasutust ning keelelist ja sotsiaalset konteksti.

Nagu taustaandmetest nähtub, peavad noored enda inglise keele oskust väga heaks. See on ühiskondlikus mõttes väärt teadmine, sõltumata sellest, kas nende tegelik keeleoskus vastab keeletestide standarditele või mitte. Noorte kõrge enesehinnang võõrkeelte oskusele, keeleteadlikkus ning keelelised kogemused on oluline ressurss, millega võiks haridussüsteemis eri pädevuste (nt kultuuri- ja väärtus-, enesemääratlus-, suhtluspädevus) kujundamisel ka laiemalt arvestada. Ehk tuleks laiemat tähelepanu pöörata noorte mitmekeelsusele, nende keelelisse repertuaari kuuluvate suhtlusvahendite ja viiside paljukeelsele rakendamisele laiemalt. Inglise keeles ei peaks nägema ohtu eesti keelele, vaid vahendit ja võimalust noorte mitmekeelse teadlikkuse kujundamiseks.

Ühtlasi vihjavad sõnavara erinevused eri registrites ning poiste ja tüdrukute vahel sellele, et noorte sõnavara ja suhtluspraktikad tuleks vaatluse alla võtta laiemalt. TeKE korpus võimaldab lisaks mitmekeelsetele suhtluspraktikatele lähemalt uurida nii sõnade koosesinemise sagedusi, morfosüntaktilist integreerimist kui ka muid nähtusi. Plaanis on uurida ka kõneleja tasemel registri mõju keelekasutusele ja idiolektile. Tulevikus lubab korpus uurida keele varieerumise ja muutumise protsesse laiemalt.

KIRJANDUS

- Androutsopoulos, Jannis 2015. Networked multilingualism: Some language practices on Facebook and their implications. – *International Journal of Bilingualism* 19 (2), 185–205.
- Auer, Peter 1998. *Code-Switching in Conversation. Language, Interaction and Identity*. London: Routledge.
- Boczkowski, Pablo. J.; Matassi, Mora; Mitchelstein, Eugenia 2018. How young users deal with multiple platforms: The role of meaning-making in social media repertoires. – *Journal of Computer-Mediated Communication* 23 (5), 245–259. <https://doi.org/10.1093/jcmc/zmy012>.
- Bucholtz, Mary; Skapoulli, Elena 2009. Youth language at the intersection: From migration to globalization. – *Pragmatics* 19 (1), 1–16. <https://doi.org/10.1075/prag.19.1.01buc>.
- Crystal, David 2001. *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, David 2008. *Txtng: the Gr8 Db8*. Oxford: Oxford University Press.
- Deakin, Hannah; Wakefield, Kelly 2014. Skype interviewing: reflections of two PhD researchers. – *Qualitative Research* 14 (5), 603–616. <https://doi.org/10.1177/1468794113488126>.
- Deuchar, Margaret 2020. Code-Switching in Linguistics: A Position Paper. – *Languages* 5 (2), 22. <https://doi.org/10.3390/languages5020022>.
- Dijk, José van; Poell, Thomas; Waal, Martijn de 2018. *The Platform Society: Public Values in a Connective World*. Oxford: Oxford University Press.
- Donath, Judith; Karahalios, Karrie; Viégas, Fernanda 1999. Visualizing conversation. – *Journal of Computer-Mediated Communication* 4 (4), JCMC442. <https://doi.org/10.1111/j.1083-6101.1999.tb00107.x>.
- Durkin, Kevin; Conti-Ramsden, Gina; Walker, Allan J. 2010. Computer-mediated communication in adolescents with and without a history of specific language impairment (SLI). – *Computers in Human Behavior* 26, 176–185. <https://doi.org/10.1016/j.chb.2009.10.007>.
- Eckert, Penelope 1997. Why ethnography? – *Ungdomsspråk i Norden*. Eds. Ulla-Britt Kotsinas, Anna-Brita Stenström, Anna-Malin Karlsson. Stockholm: Stockholm University, 52–62. <https://web.stanford.edu/~eckert/PDF/whyethnography.pdf> (06.10.2022).
- Eckert, Penelope 2000. *Language Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Hoboken: Wiley-Blackwell.

- Ferrara, Kathleen; Brunner, Hans; Whittemore, Greg 1991. Interactive written discourse as an emergent register. – *Written Communication* 8 (1), 8–34. <https://doi.org/10.1177%2F0741088391008001002>.
- Happonen, Natali 2010. Tähendusülekanded kui sõnavara täiendamise viis eesti, soome ja vene koolinoorte kõnekeeles ja slängis – inimkesksed metafoorid. Magistritöö. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Tartu.
- Hennoste, Tiit (toim.) 2000. Eesti keele allkeeled. Tartu Ülikooli eesti keele õppetooli toimetised 16. Tartu: Tartu Ülikool.
- Hepp, Andreas 2018. Meediakultuur. Meediastunud maailmade kultuur. Tlk Katrin Kaugver. (= Bibliotheca mediorum et communicationis.) Tallinn: Tallinna Ülikooli Kirjastus.
- Igav, Reet 2013. Inglise-eesti koodikopeerimine Facebooki vestlustes. Magistritöö. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Johanson, Lars 2002. Do languages die of ‘structuritis’? On the role of code-copying in language endangerment. – *Italian Journal of Linguistics* 14, 249–270.
- Jonsson, Rickard; Årman, Henning; Milani, Tommaso M. 2019. Youth language. – *The Routledge Handbook of Linguistic Ethnography*. Ed. Karin Tusting. London: Routledge, 259–272.
- Jørgensen, Jens Normann (toim.) 2002. De unges sprog: Artikler om sproglig adfærd, sproglige holdninger og flersprogethed hos unge i Danmark. København: Akademisk.
- Jørgensen, Jens Normann 2008. Polylingual languaging around and among children and adolescents. – *International Journal of Multilingualism* 5 (3), 161–176. <https://doi.org/10.1080/14790710802387562>.
- Kask, Helin 2021. English-Estonian code-copying in Estonian blogs and vlogs. Väitekiri. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Kern, Friederike; Selting, Margret (toim.) 2011. *Ethnic Styles of Speaking in European Metropolitan Areas*. Amsterdam: John Benjamins Publishing. <https://doi.org/10.1075/silv.8>.
- Kilp, Geidi 2017. Eesti-inglise-jaapani mitmekeelne suhtlus Facebooki keskkonnas. Magistritöö. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.

- Klie, Jan-Christoph; Bugert, Michael; Boullosa, Beto; Eckart de Castilho, Richard; Gurevych, Iryna 2018. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. – Proceedings of System Demonstrations of the 27th International Conference on Computational Linguistics (COLING 2018). Santa Fe, New Mexico, USA, 5–9. <https://inception-project.github.io/publications/INCEpTION-COLING2018-Demo.pdf> (06.10.2022).
- Korkus, Mari-Liis 2021. Sotsiopragmaatiline sissevaade koodivahetusele: näiteid rootsieesti teismeliste keelest. – Eesti Rakenduslingvistika Ühingu aastaraamat 17, 141–157. <http://dx.doi.org/10.5128/ERYa17.08>.
- Kull, Kristiina 2019. Repliik. Räägid estonglish'it? – Sakala, 9. august. <https://sakala.postimees.ee/6748860/repliik-raagid-estonglish-it> (06.10.2022).
- Källström, Roger; Lindberg, Inger (toim.) 2011. Young Urban Swedish: Variation and Change in Multilingual Settings. Gothenburg: University of Gothenburg, Department of Swedish Language.
- Lauricella, Alexis R.; Cingel, Drew P.; Blackwell, Courtney; Wartella, Ellen; Conway, Annie 2014. The mobile generation: Youth and adolescent ownership and use of new media. – Communication Research Reports 31 (4), 357–364. <https://doi.org/10.1080/08824096.2014.963221>.
- Lehtonen, Heini 2015. Tyylitellen: Nuorten kielelliset resurssit ja kielen sosiaalinen indeksisyys monietnisessä Helsingissä. Väitöskirja. Helsinki: Helsingin yliopisto.
- Lehtpuu, Kairit 2016. Teismeliste suulise keelekasutuse leksikaalsed jooned. Magistritöö. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Tartu.
- Leppänen, Sirpa 2007. Youth language in media contexts: Insights into the functions of English in Finland. – World Englishes 26 (2), 149–169. <https://doi.org/10.1111/j.1467-971X.2007.00499.x>.
- Loog, Mai 1988. Soome mõjusid Tallinna noorte keelekasutuses. – Keel ja Kirjandus 2, 83–89; 3, 144–148.
- Loog, Mai 1990. Laensõnad Tallinna koolinoorte slängis ehk ida pole ilma-kaar. – Keel ja Kirjandus 3, 167–169.
- Lukk, Maarika; Koreinik, Kadri; Kaldur, Kristjan; Vihman, Virve-Anneli; Villenthal, Anneli; Kivistik, Kats; Jaigma, Martin; Pertšjonok, Anastasia 2017. Eesti keeleseisund. Tartu: Tartu Ülikool ja Balti Uuringute Instituut (IBS). https://www.hm.ee/sites/default/files/aruanne_2017.pdf (06.10.2022).

- Mandel, Aive; Praakli, Kristiina; Vihman, Virve-Anneli; Koreinik, Kadri 2022. Rahvateadusega noortekeelt kogumas. – Oma Keel 44 (1), 16–21. <https://www.emakeeleselts.ee/wp-content/uploads/2022/06/Mandel-jt.pdf> (06.10.2022).
- McCulloch, Gretchen 2019. *Because Internet: Understanding the New Rules of Language*. New York: Riverhead Books.
- Myers-Scotton, Carol 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press.
- Myers-Scotton, Carol 2017. Code-Switching. – *The Handbook of Sociolinguistics*. Ed. Florian Coulmas. Hoboken: Wiley-Blackwell, 217–237.
- Nortier, Jacomine; Svendsen, Bente Ailin (toim.) 2015. *Language, Youth and Identity in the 21st Century: Linguistic Practices across Urban Spaces*. Cambridge: Cambridge University Press.
- Otsuji, Emi; Pennycook, Alastair 2010. Metrolingualism: Fixity, fluidity and language in flux. – *International Journal of Multilingualism* 7 (3), 240–254. <https://doi.org/10.1080/14790710903414331>.
- Pennycook, Alastair 2016. Mobile times, mobile terms: The trans-superpoly-metro movement. – *Sociolinguistics: Theoretical Debates*. Ed. Nikolas Coupland. Cambridge: Cambridge University Press, 201–216. <https://doi.org/10.1017/CBO9781107449787.010>.
- Poplack, Shana 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. – *Linguistics* 18, 581–618. <https://doi.org/10.1515/ling.1980.18.7-8.581>.
- Praakli, Kristiina; Koreinik, Kadri 2020. Keelemuutus vs. sotsiolingvistiline muutus: eesti keele sotsioperioodid *re-revisited*. – *Keel ja Kirjandus* 11, 915–934.
- Praakli, Kristiina 2022. Sissevaade noortekeele uurimisse Eestis: üldist ja üksikasjalikku. – *Keel ja Kirjandus* 11, 1014–1030. <https://doi.org/10.54013/kk779a3>
- Praakli, Kristiina; Korkus, Mari-Liis; Mandel, Aive; Kaukonen, Elisabeth; Kängsepp, Annika; Aasa, Triin; Algvere, Kristel; Eriksoo, Helen; Mägi, Marion; Tomson, Getri; Lindström, Liina 2022. „Mis keeles ma räägin, *I don't know*“. Eesti sisuloojate inglise keele kasutusest YouTube'is. – *Philologia Estonica Tallinnensis* 7, 263–291. DOI: <https://doi.org/10.22601/PET.2022.07.10>.

- Rampton, Ben 1995. *Crossing: Language and Ethnicity among Adolescents*. London, New York: Longman.
- Rannamäe, Kerttu 2006. Koolinoorte släng. Bakalaureusetöö. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Ratt, Silja 2017. Inglise-eesti koodikopeerimine Maria Rannavälja vlogides. Bakalaureusetöö. Tallinna Ülikool, humanitaarteaduste instituut. Tallinn.
- Ruven, Annika 1999. Antsla, Põlva, Värska õpilaste slängist. Bakalaureusetöö. Tartu Ülikool, eesti ja üldkeeleteaduse instituut. Tartu.
- Salla, Sigrid 1999. Jututubade arvutisläng. – Artikleid ja arhivaale II. Keeleuuenduse kirjastik B3. Toim Helgi Vihma. Tallinn: Johannes Aaviku Selts, 60–70.
- Sierra, Sylvia 2021. *Millennials Talking Media: Creating Intertextual Identities in Everyday Conversation*. Oxford, New York: Oxford University Press.
- Siibak, Andra 2020. Digipõlvkonnast sotsiaalmeedia põlvkonnaks. Põlvkondlikku enesemääratlust kujundavad trendid Eesti noorte internetikasutuses. – Methis. *Studia Humaniora Estonica* 21 (26), 17–34. <https://doi.org/10.7592/methis.v21i26.16908>.
- Stenström, Anna-Brita 2020. English-and Spanish-speaking teenagers' use of rude vocatives. – *Swearing and Cursing: Contexts and Practices in a Critical Linguistic Perspective*. Eds. Nico Nassenstein, Anne Storch. Berlin, Boston: De Gruyter Mouton, 281–302.
- Tagg, Caroline; Lyons, Agnieszka 2021. Polymedia repertoires of networked individuals. A day-in-the-life approach. – *Pragmatics and Society* 12 (5), 725–755. <https://doi.org/10.1075/ps.20051.tag>.
- Tagliamonte, Sali A.; Denis, Derek 2008. Linguistic ruin? LOL! Instant messaging and teen language. – *American Speech* 83 (1), 3–34. <https://doi.org/10.1215/00031283-2008-001>.
- Tagliamonte, Sali A. 2016. *Teen Talk: The Language of Adolescents*. Cambridge: Cambridge University Press.
- Tender, Tõnu 1994. Eesti släng: olemus ja uurimislugu. – *Keel ja Kirjandus* 5–6, 293–297, 346–355.
- Verschik, Anna 2004. Koodivahetus meil ja mujal. – *Keel ja Kirjandus* 1, 25–45.

- Verschik, Anna 2008. *Emerging Bilingual Speech: From Monolingualism to Code-Copying*. London: Bloomsbury Publishing.
- Voltri, Johannes 2019. Üleskutse Arvamusfestivalilt: inglise keel on kui umbrohi, mida koolis ei peaks õpetama. – Postimees, 10. august. <https://www.postimees.ee/6750278/uleskutse-arvamusfestivalilt-inglise-keel-on-kui-umbrohi-mida-koolis-ei-peak-opetama> (06.10.2022).
- Zabrodskaia, Anastassia; Kask, Helin 2017. Eesti keele kontaktid. – Eesti rändeajastul. Eesti inimarengu aruanne 2016/2017. Toim Tiit Tammaru. Eesti Koostöö Kogu, 177–185. <https://www.2017.inimareng.ee/ranne-identiteedi-ja-keelemuutused/eesti-keele-kontaktid/> (06.10.2022).

SUMMARY

ARE ESTONIAN TEENAGERS OBSESSED WITH ENGLISH? WHAT DO CORPUS DATA SAY?

Virve-Anneli Vihman, Kristiina Praakli,
Maarja-Liisa Pilvik, Mari-Liis Korkus
University of Tartu

Abstract. This article presents findings from the project “Teen speak in Estonia” (TeKE; 2019–2022), in which a dual corpus was compiled of spoken and (online) chat messaging language used by Estonian speakers aged 9–18. The data was collected through a citizen science approach. We analysed frequency data from the corpus to investigate patterns in the use of English words in teenagers’ language choices. The results show that the amount of English used is not predicted by the speakers’ age or gender; instead, language choices show a great deal of individual variation and reflect features of speakers’ interactional context and register (spoken or chat conversations). A snapshot of their most frequently used vocabulary, however, shows differences by age and gender.

Keywords: youth language, spoken language, online language, corpus analysis, Estonian, English

Autorid kuuluvad Tartu ülikooli “Teismeliste keel Eestis” tööühma (EKKD3), mida rahastab Haridus- ja Teadusministeeriumi teadus- ja arendustegevuse programm „Eesti keel ja kultuur digiajastul 2019–2027“.

Virve-Anneli Vihman on Tartu ülikooli psühholingvistika kaasprofessor. Ta peamised uurimissuunad on seotud keele omandamise, kakskeelsuse ja keele varieerumisega. virve.vihman@ut.ee

Kristiina Praakli on Tartu ülikooli rakenduslingvistika kaasprofessor. Ta uurib teismeliste keelt ja mitmekeelset suhtlust. kristiina.praakli@ut.ee

Maarja-Liisa Pilvik on Tartu ülikooli eesti keele teadur. Tema uurimisteemad seonduvad keele varieerumisega eesti keele murretes ja situatiivsetes allkeeltes.
maarja-liisa.pilvik@ut.ee

Mari-Liis Korkus on Tartu ülikooli doktorant, kelle uurimisteemadeks on teismeliste mitmekeelne suhtlus ja eesti-rootsi koodivahetus.
mari-liis.korkus@ut.ee