LANGUAGE EXPOSURE AND COMPETENCE IN ESTONIAN-NORWEGIAN BILINGUAL CHILDREN

Adele Vaks[a], Virve-Anneli Vihman[a], Jan de Jong[b]
[a] *University of Tartu, EE*
[b] *University of Bergen, NO*
adele.vaks@ut.ee, virve.vihman@ut.ee, jan.jong@uib.no

Abstract.This paper investigates the relationship between language exposure and Estonian competence of Estonian-Norwegian bilinguals aged 5–7. The sample includes 29 children acquiring Estonian and Norwegian (16 successive and 8 simultaneous bilinguals acquiring Estonian as a heritage language in Norway, 5 simultaneous bilinguals acquiring Norwegian as a heritage language in Estonia). Their language environments are described using data from the Q-BEx questionnaire, and scores from a Sentence Repetition Task and a Cross-Linguistic Lexical Task are used to describe Estonian competence. A correlation analysis shows significant correlations between exposure quality and amount of current exposure. Conditional inferencing trees and random forest models identify the proportion of current Estonian exposure as the most important predictor of morphosyntactic competence. For lexical competence, both proportion and diversity of Estonian exposure play an important role.

Keywords: bilingual development, heritage language, language exposure, Estonian, Norwegian

DOI: https://doi.org/10.12697/jeful.2025.16.2.06

1. Introduction

Language exposure has been researched as a crucial predictor of language acquisition in both mono- and multilingual settings. While earlier seminal papers on monolingual acquisition stressed the importance of cumulative quantity of exposure (e.g. Hart & Risley 1995), methods for quantifying and describing children's language exposure have diversified notably over the years, revealing a nuanced interplay of factors like context, diversity and quantity of exposure, age of onset of acquisition and chronological age (Ågren, Granfeldt & Thomas 2014;

Tsinivits & Unsworth 2021; Unsworth et al. 2011; Hoff et al. 2012; Rowe 2012).

Given the potential variation in both exposure and outcomes, the question of *what kind* of exposure affects *what aspect* of language acquisition remains open. This study investigates the relationship between the language exposure and language competence of Estonian-Norwegian bilingual children. Regarding exposure contexts, those children represent a growing bilingual population of heritage speakers without an extensive minority language community in their country of residence. This study is the first to explore the effect of exposure on acquiring Estonian as a heritage language.

2. Background

From studies with monolingual children, it is well established that the amount (Hart & Risley 1995) and quality (Huttenlocher et al. 2010; Rowe 2012) of exposure children receive in early childhood predict later language outcomes. In multilingual settings, the amount of exposure children receive is divided between languages. Several child-internal (e.g. gender, see Kuvač-Kraljević et al. 2021; Urm & Tulviste 2016) and child-external (e.g. socioeconomic status, see De Cat 2020) variables interact with input, modulating the rate at which children acquire language. While in the literature review we will use the different terms used in the original studies (e.g. *exposure*, *experience*, *input*, and *intake*, Carroll 2017), we will use *exposure* in the current study to refer to the language the child hears and reads. The following section gives an overview of previous findings on multilingual exposure and language acquisition, specifically in heritage language contexts.

2.1. Heritage language and bilingual acquisition

A child's heritage language (HL) is acquired as a minority language in a family setting, typically in an immigration context, while the societal language (SL) is used outside the home (Polinsky 2018), and, depending on family situation, may also be used at home. We define HL speakers through the context of their exposure to the language, not by level of proficiency or differences in language use compared to the homeland, or SL, speakers (see Rothman et al. 2023, for a longer discussion). The term covers a variety of language communities and linguistic experiences and belies a great deal of heterogeneity among HL contexts, e.g. in their input and motivation to speak the language. We now introduce some of the exposure factors that have been found to significantly affect HL development.

2.1.1. Timing of exposure

Heritage speakers can be either simultaneous (acquiring both of their languages at birth) or successive bilinguals (acquiring one language at a later point, e.g. at pre-school). The cut-off point for the onset of acquisition used for distinguishing between the two groups varies (Unsworth et al. 2011), leading to a question of whether more scalar measures are more valid than binary groups.

Instead of the binary successive/simultaneous distinction, exposure timing can be described by Age of Onset of acquisition (AoO): the age at which a child starts being exposed regularly to the language. Advantages in phonological, morphological and even lexical acquisition have been found for learners with AoO at birth or in early childhood, compared to later AoO (Johnson & Newport 1989; Bylund, Hyltenstam & Abrahamsson 2021). Some studies have found that AoO makes a significant difference for structures that are typically acquired early by monolinguals, but not necessarily for more complex constructions (Ågren, Granfeldt & Thomas 2014; Unsworth et al. 2011).

However, studies considering not only Age of Onset but also Length of Exposure (LoE) indicate that the advantage of an earlier AoO may stem from the longer exposure time (also leading to increased quantity), rather than the developmental effects of starting early. Although LoE and AoO are tightly interwoven, they can have different predictive power for language outcomes. For example, LoE predicts vocabulary growth better than AoO (Smolander et al. 2021; Thordardottir 2019). Mitrofanova et al. (2018) also found LoE to best predict grammar acquisition when adjusted for the relative amount of input received in the language.

In heritage language research, the focus is often on whether a later AoO of the SL gives children an advantage in HL development. This has been found to be the case for e.g. English HL children in Israel (Armon-Lotem, Rose & Altman 2021), Arabic HL children in the US

(Albirini 2018), Russian HL children in Israel and the Netherlands (Meir & Janssen 2021). Kim & Kim (2022) investigated Russian and Chinese children residing in Korea, and found that lexical retrieval in their respective native languages was significantly influenced by their length of residence (i.e., LoE) rather than their age at arrival in Korea (i.e., AoO). Thus, it is clear that the language outcomes of HL children are influenced by something beyond the age metrics, calling for a closer look at quantity of exposure.

2.1.2. Quantity of exposure

The amount of input children receive varies, even for children with equal LoE, depending on the language of instruction at school, number of family members speaking either language, and other factors (cf. Mitrofanova et al. 2018). This calls for more fine-grained measures of exposure quantity. Two measures commonly used are *cumulative* and *current amount of exposure*. While the latter takes into account the day-to-day language environment of the child at the point of data collection (typically over the most recent year), *cumulative amount of exposure* also considers changes in language exposure over the child's lifespan, e.g., moving countries, starting daycare and school, and other events (Unsworth 2013).

Cumulative amount of exposure has been found to significantly affect both grammatical (Thordardottir 2015; Unsworth 2013) and lexical development (Thordardottir 2011, 2019). In Montreal, where two languages are SLs, Thordardottir (2011, 2015) found that English-French bilingual children perform similarly to monolinguals in the language which accounts for over half of their cumulative exposure. De Cat (2020) arrived at a similar cutoff point (about 57% of cumulative exposure) for children acquiring English as SL. Importantly, the relationship between exposure quantity and language outcomes is not linear, with some (later-acquired and more complex) structures requiring more exposure to be mastered (see Albirini 2018; Torregrossa, Flores & Rinke 2023 on cumulative exposure; Daskalaki et al. 2020; Unsworth 2013 on current exposure). In addition, the predictive power of cumulative and current exposure can be modulated by language status: De Cat et al. (2025) found cumulative exposure to be the best predictor of SL competence, but current exposure to best predict HL competence.

While it is clear why current and cumulative exposure rates may differ, it may not always be feasible to measure both. Especially for educational or diagnostic evaluation, it is important to find the measures that best explain language outcomes and reduce testing burden on parents and specialists. De Cat (2020) compared models using cumulative and current exposure and use as the main predictor of SL skills with a model that used a *combined* measure of cumulative exposure and use (i.e. exposure adjusted for use). Both the combined measure and cumulative exposure alone predicted language outcomes far better than current exposure, current use or cumulative use. However, Unsworth (2013) found both current and cumulative exposure to be significant for acquiring Dutch gender-marking – a non-transparent and relatively late-acquired feature. From these and other studies in the last two decades, it is clear that cumulative exposure rates should be considered where possible, and that the relationship between exposure and linguistic attainment is not necessarily linear.

2.1.3. Quality of exposure

Quality of exposure is a multidimensional construct that has been measured in various ways, e.g. by linguistic diversity or the sources of input. When recordings of child-directed speech are available, measures can include number of lexical items, number of inflectional forms, mean length of utterance, type-token ratio, number of rare words or decontextualised utterances (Rowe 2012). For larger participant samples, questionnaires can be used to document the amount of time a child spends in different language environments (Place & Hoff 2016).

Especially relevant for HL children is who they hear speaking the languages. The number of speakers and the proportion of native speakers among the child's conversation partners have been found to influence language competence (Place & Hoff 2011, 2016). As HL children typically receive input from adults who are similarly removed from a majority language context, it may be relevant to include whether the parents are first- or second-generation HL speakers (see Daskalaki et al. 2020; Fishman 1972). The presence of older siblings tends to increase SL input at home (Tsinivits & Unsworth 2021).

Importantly, however, SL input at home has not been found to significantly support SL acquisition (Rodina & Westergaard 2017; Sorenson

Duncan & Paradis 2018), while HL exposure at home does significantly support HL acquisition (Biedinger, Becker & Klein 2015; Rodina et al. 2020; Tao, Cai & Gollan 2021).

The quality measure used in this study is a richness score from the Q-BEx questionnaire (De Cat et al. 2022). The score includes both qualitative and quantitative measures (see Section 3.3.2 for more details). The measure has found to be a reliable representation of language exposure quality by Unsworth et al. (2024).

Exposure quantity and quality are usually correlated, making it challenging to disentangle possibly differential effects on language acquisition (see Huttenlocher et al. 2010; De Cat 2020). However, some findings confirm that exposure quality and quantity can have differing effects on language outcomes over time. Looking at monolingual parent-child dyads, Rowe (2012) found that when quantity was controlled for, input quality in parental speech to 18-, 30- and 42-month-olds explained variation in the children's vocabularies a year later. Quantity was the most important predictor during the second year of life, while quality became the most important predictor from the third year of life.

To what extent the effects of quality and quantity can be disentangled in older children remains to be studied more thoroughly. As mentioned above, there is evidence for different aspects of grammar having different sensitivity to exposure (Tsimpli 2014; Unsworth et al. 2011; Ågren, Granfeldt & Thomas 2014). De Cat (2020: 315) found that errors in functional categories in the Sentence Repetition Task were better predicted by amount of exposure than lexical errors. The observation that the acquisition of grammar is more sensitive to amount of exposure than lexical acquisition could point to the declarative/procedural model (Ullman 2001), which claims that vocabulary learning relies more on declarative memory (related to fact and event knowledge) while grammatical acquisition relies more on procedural memory (related to skill and pattern acquisition). Procedural memory requires repeated exposure, while declarative memory, while benefitting from repetition, can also lead to rapid learning, opening this mechanism to benefitting from diversity in current exposure (Ullman 2016). This may explain findings that grammar learning is more sensitive to exposure amount than lexical learning.

2.2 Estonian-Norwegian bilingualism

2.2.1. Estonian-Norwegian community

Reports estimate between 5000–6000 first- and second-generation Estonian immigrants in Norway (Integratsiooni Sihtasutus 2022; Statistics Norway 2025). The families in this study are quite representative for the contemporary diaspora, having moved abroad after Estonia joined the EU in 2004. All Estonian-speaking parents of the children in the study are first-generation immigrants. Most of the families have relatives in Estonia, and they spend time there during school holidays. Estonian media is easily accessible online, but Estonian-medium formal education is limited to additional classes intended to support the understanding of class content learned in Norwegian (2024: secs. 3–6). Estonian language classes are organised as 'Sunday schools' by the diaspora communities in Oslo and Trondheim; a children's choir gathers in Stavanger.

The Norwegian community in Estonia is much smaller than the Estonian community in Norway, with 238 people identifying as Norwegian in the latest census (Statistics Estonia 2021). A relatively small community translates into fewer opportunities for learning Norwegian. However, Norwegian literature and other media are accessible, and the families travel to Norway and/or receive visitors regularly.

2.2.2. Estonian and Norwegian

Estonian (Finno-Ugric) is known for its rich inflectional morphology and flexible word order. Norwegian (Germanic) is more analytic, with a relatively more rigid word order. Due to shared historical influences, there is some overlap in vocabulary in the two languages (see Section 3.3.3).

Estonian-Norwegian bilingualism has been researched in student theses on conversation strategies in a bilingual family (Limmer 2008) and Estonian adults learning Norwegian as L2 (e.g. Peterson 2020; Piksarv 2013). A previous study has described the use of noun phrases by children acquiring Estonian and Norwegian (Vaks & Vihman 2022). Data from the current sample has also been used to investigate cross-linguistic influence (Vaks & Vihman 2025).

Heritage acquisition of Estonian has been researched in various diaspora contexts. After World War II, the largest communities of displaced Estonians formed in Sweden, Canada, the USA, Germany and Australia. These diaspora communities became the loci of Estonian child language research in the late 20th century, documenting Estonian acquisition in the USA (Lipp 1977; Osterreich 1977; M. M. Vihman 1982, 1985), Australia (Salasoo 1995, 2010) and Sweden (Oksaar 1971). More recent studies include explorations of Estonian-Finnish bilingualism (Hassinen 2002; Jokela & Paulsen 2010; Teiss 2006), as Finland has become the top emigration destination (Integratsiooni Sihtasutus 2022).

Most of the above research did not quantify children's language exposure. Laar (1990) used parental questionnaires to explore Estonian use among the families attending the Estonian Kindergarten in Toronto – home to one of the most concentrated Estonian communities in the diaspora. Laar found that 50% of the fathers and 60% of mothers spoke only Estonian to their children, and 20% of the children were reported to speak only Estonian to other children. Although transmission of Estonian language and ethnocultural identity was important to all parents, Laar concludes that the community was undergoing a language shift (as described by Fishman 1972), with children using less Estonian than their parents. The language environments of children in the 21st century diaspora remain to be studied.

2.3. This study

This study investigates the language exposure of Estonian-Norwegian children, and the relationship between exposure and Estonian competence. Exposure quantity is operationalised by amounts of current and cumulative exposure to each language, and exposure quality is operationalised as the richness scores derived from the Q-BEx questionnaire. Language competence is measured with a sentence repetition task (morphosyntactic competence) and a vocabulary task (lexical competence). Exposure and competence measures are introduced in detail in Section 3.3.

In the context of heritage language acquisition, quantity of exposure is limited by the everyday logistics of the time family members can spend together. Although a greater amount of exposure allows for more diversity in the input, quality of exposure can be said to be more under the parents' control. Quality-related factors like the number of interlocutors (Place & Hoff 2016) and availability of schooling in the HL (Rodina et al. 2020) can also contribute to a larger proportion of exposure in the HL, linking exposure quality and quantity. We thus examine the relationship between the two measures, as set out in RQ1.

> RQ1: What is the relationship between the quantity and the quality of the language exposure of Estonian-Norwegian children?

> H1: There is a positive correlation between the quantity (measured as proportion) and the quality of language exposure (richness score) within each of the languages.

Further, we explore the relationship between language exposure and competence, the latter represented here by morphosyntactic and lexical knowledge. Amounts of current and cumulative exposure may have independent effects on morphological development (Unsworth 2013), and quality and quantity can have different effects at different points in development (Rowe 2012). We explore how the two dimensions of exposure affect language competence in our sample, as expressed in RQ2.

> RQ2: How do quantity and quality of exposure affect morphosyntactic and lexical competence in Estonian for Estonian-Norwegian children?

> H2: Both quantity and quality of Estonian exposure affect the competence scores, but quantity will have a more significant effect than quality.

3. Method

3.1. Recruitment

The sample consists of 29 Estonian-Norwegian bilingual children. Prior to recruitment and data collection, ethics approval was obtained from the Research Ethics Committee of the University of Tartu (protocol 364-T17). The invitation to participate was shared on social media, in mailing lists, and via personal contacts. The initial recruitment resulted in a group of mostly successive and some simultaneous bilingual children acquiring Estonian as HL in Norway. An additional group of five simultaneous bilingual children residing in Estonia was recruited to allow for comparisons with the Norwegian simultaneous group. Identifying and accessing qualifying families was difficult because of the size of the population in question, and the time-intensive testing, so this variation in bilingual contexts was allowed.

Parents of all participating children signed an informed consent form. Children were familiarised with the study and asked to give verbal assent before starting the language tasks. Children over 7;0[1] also signed an age-appropriate consent form. The families received a children's book in Estonian as thanks, and children received stickers during the testing as a reward.

While efforts were made to recruit as large a sample as possible, we acknowledge that the participant pool introduces a sampling bias. The recruitment channels may have resulted in a sample skewed toward families actively maintaining ties with the Estonian community. Moreover, since the study involves language production, a prerequisite for participating children was minimal competence in spoken Estonian. Hence, our sample excludes children who have very little competence or only passive knowledge of the heritage language.

3.2. Participants

The sample in the study is made up of three groups. The children residing in Norway with Estonian-speaking parents will be referred to as successive bilinguals (N-suc). Children with one Estonian-speaking and one Norwegian-speaking parent, or simultaneous bilinguals, are referred to as N-sim (residents of Norway) and E-sim (residents of Estonia).

For logistical reasons, a full dataset was not obtained from all children. The analyses make use of different subsets of participants, based on how many children provided the necessary data points for the analysis in question. The size of each dataset is specified for each analysis.

All N-suc and N-sim children were either born in Norway or moved there before their second birthday. The Estonian parent(s) of Norway-born children were all born in Estonia and migrated in adulthood or

1 The children's ages are hereafter presented as years;months.

adolescence. All E-sim children were born in Estonia and have one Norwegian-speaking parent.

Two of the E-sim children were siblings with one L1 Estonian parent who spoke Norwegian to the children at home. The family had spent a year in Norway, where children attended Norwegian daycare. The children's performance did not differ from the rest of the group in the Norwegian competence measures, and so they were included in the sample. Any differences in the quantity or quality of their Norwegian exposure compared to other children are reflected in the exposure measures.

Two of the children in Norway were taking part in weekly, Estonian-medium additional classes at school (cf. Section 2.2). Five children partook in (bi-)monthly Sunday schools or choirs, one child took weekly online Estonian classes. No Estonian residents took part in any formal or semi-formal Norwegian education.

Table 1 provides a summary of the participants. Parental education level indicates the highest level of education obtained by either parent.

Table 1. Demographic overview of the participants.

	Successive bilinguals	**Simultaneous bilinguals**	
Country of residence	Norway	Norway	Estonia
N	16	8	5
Mean age in months (SD), range	76 (13), 57–94	75 (7), 64–82	81 (13), 66–95
Gender	7 female, 9 male	5 female, 3 male	3 female, 2 male
Parental education	1 secondary education 4 post-secondary training 11 higher education	2 post-secondary training 6 higher education	5 higher education

This study is a part of a PhD project, in which data were collected using three language tests, based on the LITMUS tasks (Armon-Lotem, de Jong & Meir 2015), in each language: the Sentence Repetition Task, the Cross-Linguistic Lexical Tasks and the Multilingual Assessment Instrument for Narratives (MAIN). The former two are used in this study and described in more detail below.

3.3. Data collection and analysis

3.3.1. Language competence measures

Language competence is operationalised through scores from two tasks: the Sentence Repetition Task and Cross-Linguistic Lexical Tasks. Both were developed as a part of the LITMUS test battery (initiated within COST Action IS0804, Armon-Lotem, de Jong & Meir 2015) and have been adapted into both Estonian and Norwegian. Thanks to the common protocols used for developing the tasks across languages, the linguistic content of the tasks is as comparable as possible regarding age of acquisition, linguistic complexity, and familiarity.

The Sentence Repetition Task (Marinis & Armon-Lotem 2015) is used to assess the acquisition of grammatical structures. Pre-recorded sentences are presented to the child auditorily, and the child is asked to repeat them as accurately as possible. The tasks include a mix of universal and language-specific structures. Language universal structures (included in all languages) include both syntactically simple structures that serve as controls and more complex structures that involve embedding and/or syntactic movement. Language-specific structures pose challenges to children acquiring the specific language and depend on its typological features (Marinis & Armon-Lotem 2015: 101). The sentences are of sufficient length to ensure that repetition cannot rely solely on short-term phonological memory, but requires comprehension of syntax (Polišenská, Chiat & Roy 2015). Thus, successful repetition indicates that the child can process the structure. The vocabulary is controlled for, using early acquired items to ensure that repetition relies as little as possible on vocabulary size (Marinis & Armon-Lotem 2015).

The Estonian Sentence Repetition Task (henceforth E-SRT), developed by Vihman, Padrik and Hallap, consists of 33 sentences (see Padrik, Vihman & Fil 2022 for list of structures). The Norwegian Sentence Repetition Task (henceforth N-SRT), developed by Bome and Vargen, currently consists of 60 sentences (see Bome & Vargen 2017 for full list of structures). To mitigate the potential effects of the difference in length between E-SRT and N-SRT, percentages of the maximum score are used here, rather than raw scores of the tests. A preliminary analysis showed no significant drop in accuracy in the N-SRT toward the end of the test.

Vocabulary knowledge is represented by scores from the Cross-Linguistic Lexical Task. The task contains four subtests measuring noun and verb comprehension and production, presented as picture recognition and naming tasks, respectively. The Estonian Cross-Linguistic Lexical Tasks (henceforth E-CLT) were developed by Vaks, Padrik & Vihman (2025). This study uses the pilot version of E-CLT. The Norwegian Cross-Linguistic Lexical Tasks (henceforth N-CLT) were developed by Gram Simonsen, Hansen and Łuniewska.[2]

3.3.2. Language exposure measures

Information on the children's language background was collected with Q-BEx (De Cat et al. 2022), an online questionnaire for parents. The questionnaire contains two obligatory modules, *Background Information* and *Risk Factors*, providing demographic data and an overview of the child's linguistic environment: who lives with the child, what languages they speak to the child, what languages the child speaks to them, as well as information on early linguistic milestones and parental concerns. Additionally, three optional modules were included. In *Language exposure and use*, the parent detailed who the child spends time with during schooldays, on weekends and holidays. In *Language Proficiency*, the parent evaluated the child's proficiency in each language, and *Richness of linguistic experience* detailed frequency of activities in the languages, number and proficiency of regular language contacts in each language, and parental education levels.

The web questionnaire was filled out by the researcher during a Zoom call[3] with one or both of the (Estonian-speaking) parents. The call typically took place 1–2 weeks before the language data collection. Three measures from the Q-BEx backend calculator were later used in analyses:

1. **Current language exposure**, measured in hours in the past year. The number of hours is estimated based on information about what languages the child hears in what contexts during a typical week.

---

2 Many thanks to the authors for permission to use the test, and to the CLT team, particularly Magdalena Krysztofiak, for generous help with the CLT application.

3 Two parents filled out the questionnaire without the researcher for logistical reasons. The researcher was available for questions via email and telephone.

2. **Cumulative language exposure**, measured in months over the child's lifetime. The number of months is estimated based on information about changes in the living situation, preschool and school start dates, and estimates of languages heard during the preverbal period.
3. **Richness score**, calculated from the frequency of activities and number and proficiency of regular contacts in each language. This is a composite score between 0 and 1, indicating the proportion of potential richness of the language experience. Originally, parental education was included in both languages' scores, but to avoid this inflating the correlation between Estonian and Norwegian richness, the scores were recalculated without parental education (see Unsworth et al., 2024). In statistical models exploring RQ2, parental education was included as a separate predictor.

For current and cumulative language exposure, percentages of relative exposure to each language were used (Norwegian exposure + Estonian exposure = 100%). All language exposure data is based on parental estimates. This is a rougher and more subjective measure than, e.g., diary notes or recordings of child-directed speech. However, Verhoeven et al. (2024) found that Q-BEx data and data from day-long recordings using LENA correlated equally strongly with children's vocabulary outcomes, suggesting that Q-BEx data is as reliable for assessing exposure quantity as observational data.

3.3.3. Procedure

Language data were collected during two meetings with the family at their home or at a public space (library or café). E-SRT and E-CLT were conducted by the first author, a native Estonian speaker. N-SRT and N-CLT were conducted by a student assistant, a native Norwegian speaker or a native Estonian speaker speaking Norwegian exclusively during the visit. For five children, tests were conducted by the first author in both languages.

The researcher and parent(s) spoke the testing language to facilitate language activation. The data collection visits were 1–30 days apart, and the order of the testing was as balanced as possible within the described logistical constraints. For children tested by the same researcher in both languages, the less dominant language was tested first, to minimise the tendency to switch to the dominant language. Language dominance was determined by percentage of current language use obtained from the Q-BEx data. For children with roughly equal use of both languages (40–60% in the more dominant language), the societal language was used. 17 children were tested in Estonian first, and 12 children were tested in Norwegian first.

E-SRT and N-SRT were presented to the child with audio stimuli, together with a motivational picture series showing a bear on an adventure. The child's responses were recorded with a voice recorder and later transcribed and scored, using the 0–3 system described by Marinis & Armon-Lotem (2015). Each sentence is scored based on the number of substitutions, omissions and additions compared to the original stimulus: 0 points for over 4 errors, 1 point for 2–3 errors, 2 points for 1 error and 3 points for identical repetition. Allowances are made for colloquial, dialectal and motherese lexical substitutions.

The vocabulary tests (E-CLT and N-CLT) were presented to children in the web interface of the CLT app developed at the University of Warsaw.[4] In the comprehension tasks, children were presented with a forced choice between four images, and a question prompting them to point at one of them (e.g. "Where is the cat?", "Who is eating?"). For production tasks, children were presented with a single image and a prompt asking them to name the referent ("What is that?") or the action ("What is she doing?"). The responses were recorded, transcribed and coded according to the guidelines by Łuniewska et al. (2021), with a few more allowances: regional and colloquial synonyms were also coded as correct. The production tasks contained some loans with common origins that bilingual children may have stored as near-homophones: *bukse* (NO)/ *püksid* (ET) 'trousers', *appelsin* (NO) / *apelsin* (ET) 'orange', *tiger* (NO) / *tiiger* (ET) 'tiger', and *gitar* (NO) / *kitarr* (ET) 'guitar'. Cases of uncertainty as to which language the child used were interpreted in the child's favour.

Two additional tasks were conducted to control for variation in children's short term and working memory and nonverbal reasoning. For short-term and working memory span, the forwards and backwards digit span test from the Norwegian WISC-II battery was used (Wechsler

4 See https://multilada.pl/en/projects/clt/

1974). For nonverbal intelligence, Raven's progressive matrices, sets A-C (Raven 1976), were used. Those tasks were carried out in the child's dominant language.

3.3.4. Analysis methods

Data was analysed and visualised in R (R Core Team 2023), using the *dplyr* (Wickham et al. 2023), *partykit* (Hothorn & Zeileis 2015), *ggplot2* (Wickham 2016) and *randomForest* (Liaw & Wiener 2002) packages. Correlation tests were run to answer RQ1. As the small sample size did not lend itself well to linear regression models, conditional inference trees (Hothorn, Hornik & Zeileis 2006) and random forests were used to explore RQ2 and RQ3. These methods are explained and justified in more detail in the relevant sections in Results.

3.3.5. Data availability

Supplementary materials are available at https://osf.io/u7b2w. This includes the full dataset and analysis script used in this article, as well as the dataset for Norwegian SRT scores, and an analysis of stimuli ordering effects (cf. Section 3.3.1).

4. Results

4.1. Describing language experience

For RQ1, we aim to describe the language experience of the children acquiring Estonian and Norwegian, investigating possible correlations between exposure quantity (percentage of current and cumulative exposure in both languages) and quality (the richness score from Q-BEx).

Table 2 provides a summary of the language exposure measures in the sample. While children were predominantly bilingual, 11 parents reported use of and exposure to English via media and/or English-speaking relatives and friends. As this did not concern all participants, and the average exposure to English did not exceed 1.5% of either current or cumulative exposure, the English exposure has been excluded from the dataset.

Table 2. Exposure measures in the sample: mean (SD), range.

		Estonian	Norwegian
Successive bilinguals (N = 16)	Cumulative exposure (%)	53.92 (9.65), 35.81–74.26	44.67 (8.33), 25.74–60.45
	Current exposure (%)	54.13 (9.34), 39.39–69.77	43.64 (7.72), 29.24–56.79
	Richness	0.34 (0.08), 0.19–0.50	0.62 (0.17), 0.22–0.92
Simultaneous bilinguals in Norway (N = 8)	Cumulative exposure (%)	42.72 (10.58), 28.54–59.16	56.28 (11.53), 36.74–71.46
	Current exposure (%)	32.61 (14.22), 18.29–55.41	67.06 (14.42), 43.74–81.71
	Richness	0.32 (0.13), 0.19–0.53	0.70 (0.15), 0.50–0.92
Simultaneous bilinguals in Estonia (N = 5)	Cumulative exposure (%)	62.49 (5.64), 54.11–67.91	37.51 (5.64), 32.09–45.89
	Current exposure (%)	75.78 (5.57), 69.82–83.34	24.22 (5.57), 16.66–30.18
	Richness	0.76 (0.14), 0.56–0.89	0.37 (0.07), 0.28–0.47

The successive bilingual group has fairly balanced rates of Estonian and Norwegian, both in cumulative and current exposure. For the simultaneous group, the amount of exposure is skewed towards the societal language, with a stronger bias among the Estonian than Norwegian residents. Richness scores for Estonian input are similar among N-suc and N-sim, while Norwegian input shows slightly higher average richness scores for the N-sim than N-suc group. Seeing that the richness score includes frequency of activities that are typically parent-led or -regulated, e.g. book-reading and media consumption, this could stem from the simultaneous families undertaking activities in both languages more equally, while successive families might prioritise Estonian activities at home. The SL richness scores are slightly higher for E-sim (Estonian as SL) than N-sim (Norwegian as SL).

Correlation analysis was run on all groups as a single sample (N = 29). Having a heterogenous sample with variation in the exposure measures allows us to explore the relations between the different exposure measures. As the sample is rather small, dividing it up further would reduce the reliability of the analysis. However, to see whether analysing a more homogenous sample would yield clearer results, the same analysis was also run on the N-suc group (N = 16), but not the simultaneous groups, which were considered too small (N < 10).

The exposure data was not normally distributed, and the different exposure measures were in a mostly monotonic, but not linear, relationship. Hence, Kendall's tau was chosen as the correlation coefficient, allowing assessment of the strength of a monotonic relationship between two numeric variables. The correlations between language exposure measures for the whole sample are reported in Figure 1, and for the N-suc group in Figure 2. Note that the current and cumulative exposure rates are proportions of the child's overall language exposure, meaning that there is, by definition, an inverse correlation between Estonian and Norwegian current and cumulative exposure. However, the correlations are not exactly –1 because some parents reported exposure to English, omitted from this analysis.

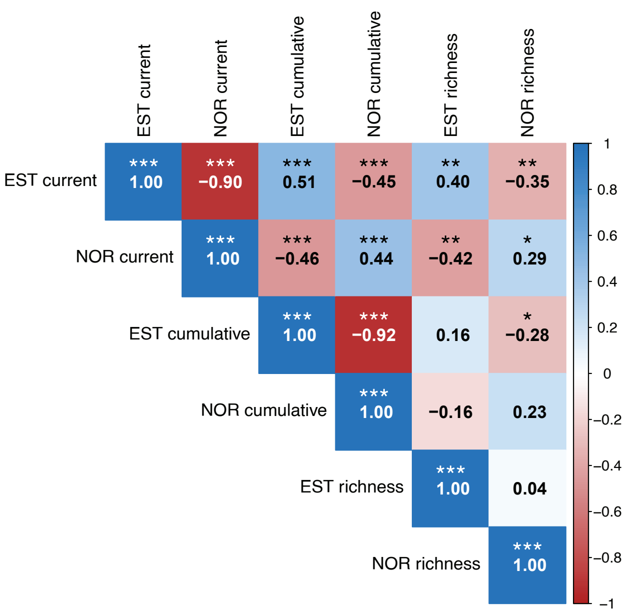

Figure 1. Correlation coefficients between exposure measures in the whole sample (N = 29). * indicates $p < .05$, ** indicates $.001 < p < .01$, *** indicates $p < .001$.

Richness scores in both languages correlate significantly with the amount of current exposure in the same language, with a moderate positive correlation for Estonian (τ = 0.40, $p < .01$) and a weak positive correlation for Norwegian (τ = 0.29, $p < .05$). These correlations are mirrored by negative correlations with the amount of current exposure in the other language (τ = –0.42, $p < .01$ for Estonian, τ = –0.35, $p < .01$ for Norwegian). Norwegian richness also has a significant negative correlation with cumulative Estonian exposure (τ = –0.28, $p < .05$), while Estonian richness does not correlate significantly with cumulative Norwegian input.

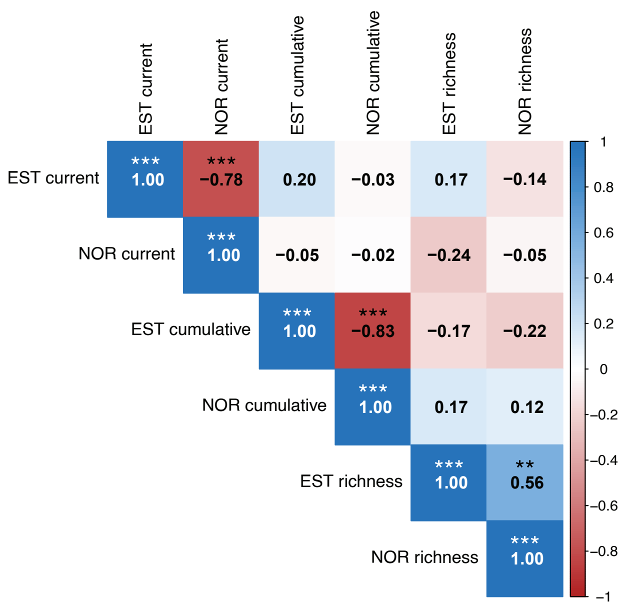

Figure 2. Correlation coefficients between exposure measures among the N-suc group (N = 16).

Analysing only the successive bilinguals, no exposure quantity measure shows any significant correlation with the quality measures, as shown in Figure 2. All correlation coefficients have low values (< .20), and positive and negative correlations alternate, suggesting little monotonicity between the quality and quantity measures. The alternation between positive and negative correlations may be an artefact of the small sample or high variability. However, this analysis shows a significant positive correlation (τ = 0.56, $p < .01$) between the richness scores in the two languages in the successive bilingual sample. This is an interesting finding, to which we return in the discussion.

4.2. Language experience and language competence

Table 3 presents the children's scores from the Sentence Repetition Tasks (E-SRT and N-SRT), and Cross-Linguistic Lexical Tasks (E-CLT and N-CLT) in both languages. All scores are percentages of the maximum score.

Table 3. Language competence scores by group: mean (SD), range.

	Estonian SRT	Estonian CLT	Norwegian SRT	Norwegian CLT
N-suc (N = 16)	73.6 (15.7), 48.5–98.0	80.7 (10.2), 57.5–95.0	54.2 (21.3), 20.0–94.4	75.2 (14.6), 31.7–92.5
N-sim (N = 8)	57.4 (24.9), 23.2–91.9	69.9 (19.1), 32.5–95.8	75.8 (11.2), 64.4–95.0	84.2 (4.8), 77.5–90.0
E-sim (N = 5)	99.6 (0.6), 98.0–100.0	98.3 (1.3), 96.7–100.0	57.1 (23.1), 21.1–80.0	73.3 (13.1), 50.8–84.2

In Norwegian, the N-suc and E-sim scored very similarly in both SRT and CLT. Although the scores are not necessarily comparable across languages, it is worth noting that the successive bilinguals score higher in the Estonian than the Norwegian tasks. The simultaneous groups score higher in their respective societal languages than in their heritage languages, with E-sim at ceiling in the Estonian tasks. Statistical tests checking for significant differences between groups were not performed due to the simultaneous groups being too small.

Scores from Raven's progressive matrices (Raven 1976) and the WISC-II digit span (Wechsler 1974) were included in the analyses as predictors. Table 4 presents these scores.

Table 4. Scores on Raven's progressive matrices and WISC-II digit span test: mean (SD), range.

	Raven's progressive matrices	WISC-II digit span
Entire sample (N = 28)	13.74 (8.28), 1–34	6.42 (2.28), 2–12
N-suc (N = 16)	12.23 (6.82), 1–26	5.92 (2.40), 2–9
N-sim (N = 8)	10.57 (3.69), 6–15	5.50 (1.05), 4–7
E-sim (N = 5)	20.0 (11.73), 6–34	8.00 (2.55), 5–12

The N-sim group is the most uniform in both tests, with the smallest standard deviations, while scores in the E-sim group vary the most. E-Sim scored noticeably higher than the Norwegian residents in both tests. This is likely due to the group's higher average age (cf. Table 1) and the presence of a few outliers with unusually high scores, skewing the mean of the small group. They were still included, in order to use data from as many children as possible.

To answer RQ2, the 'tree and forest' approach was used, combining conditional inference trees and random forest models (Klavan, Pilvik & Uiboaed 2015; Strobl, Malley & Tutz 2009). This is a non-parametric method, suitable for datasets with non-normal distributions and highly correlated variables (Klavan, Pilvik & Uiboaed 2015; Strobl et al. 2008), both characteristics of this dataset. While a conditional inferencing tree yields intuitively interpretable and easily visualised results, single tree models are unstable and sensitive to small changes in the dataset. To combat this instability, a random forest model aggregates an ensemble of inferencing trees (Klavan, Pilvik & Uiboaed 2015: 210–211; Strobl, Malley & Tutz 2009: 330–331). The training dataset is randomly sampled, and a random set of variables is chosen to select from for each split. This aggregation of tree models allows one to explore the influence of weaker predictors that would be missed by using a single decision tree (Klavan, Pilvik & Uiboaed 2015: 210–211). As the current dataset is small and contains predictor variables that are correlated, but not normally distributed and linearly related to each other, the 'tree and forest' approach offers insight into the relationships between the predictors.

Separate conditional inference (CI) tree models were built with each of the language competence variables (E-SRT, E-CLT, N-SRT and N-CLT) as the dependent variable. Models were built on the successive bilingual group separately, and the sample as a whole, in order to explore whether groups with varying levels of homogeneity yield different results. Due to small sample sizes, the simultaneous bilinguals were not modelled separately. Independent variables included current and cumulative exposure and richness scores in the relevant language, child age, gender, parental education level, scores from the digit span test and Raven's matrices. Richness score, age, digit span and Raven's matrices scores were scaled to support comparability of the numerical variables. For analyses using the whole sample, country of residence

(Estonia/Norway) and family language environment (simultaneous/ successive) were also included, to investigate whether including these binary categories alongside numerical exposure data is informative. When building the random forest models, predictors were added one by one to find the best-fitting model.

4.2.1. Exposure and SRT scores

Conditional inferencing tree models were built and visualised for both language scores (E-SRT and N-SRT), excluding children who did not complete the test in question. No predictors proved significant when looking at the N-suc sample separately from the simultaneous groups, possibly because of the sample size. When modelling scores from the N-SRT, no predictors proved significant, looking at either the N-suc sample or the whole sample. All CI tree models can be found in the supplementary materials, but further discussion in this section will focus on the model for E-SRT scores with the whole sample.

Figure 3 presents the CI tree for E-SRT scores, including the entire sample. Each numbered node represents a statistically significant split in the dataset, including the *p*-value of the difference between the resulting subsamples, as well as the value of the statistically significant variable at the split. The boxplots under each split present the scores of the resulting subsamples.

The only predictor that significantly affected E-SRT scores was current Estonian exposure ($p = .023$): children with over 49.3% of their current language exposure in Estonian scored significantly better than children with less Estonian exposure.

As the only SRT scores with any significant predictors, E-SRT scores were explored further with random forest models. Models including different sample sizes (N-suc/whole sample), predictors and ways of handling missing values (imputing/omitting) were tested. By comparing means of squared residuals and percentage of variance explained by the model, the best model (with lowest mean of squared residuals and highest percentage of variance explained) proved to be one where the entire sample was included (N = 28), and missing values were omitted. Predictors in the best model were current and cumulative Estonian exposure, Estonian richness, Raven's matrices, digit span, family environment and country.

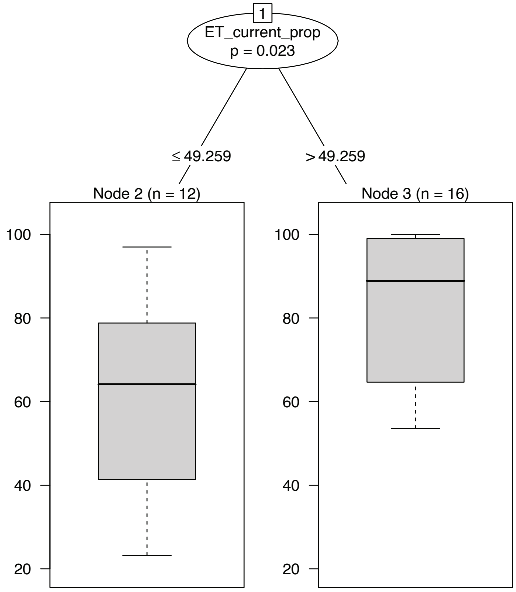

Figure 3. Significant predictor of E-SRT scores (N = 28):[5] Current proportion of Estonian exposure.

Parental education, age and gender were omitted from the model, as they contributed more to noise in the data than to explaining variance in SRT scores. The model was constructed with 500 trees and two variables at each split. The number of variables to include at each split was chosen by dividing the number of variables in the model (7) by three (Cutler et al. 2007). The default number of trees (500) was used – given the small number of observations in the dataset, increasing the number of trees did not result in significant changes in the prediction stability. The model's mean squared residual was 248.39, and it explained 46.91% of variance in E-SRT scores. These measures showed little variance when running the model on different seedings.

To compare the conditional inferencing tree and the random forest analysis, R^2 and root mean squared error (RMSE) were calculated for both models. R^2, although calculated differently than the percentage of variance explained internally by the *random.Forest* function, gives an estimate of how well the models capture variance in the data. For the

5 One child was excluded from the analysis due to not completing the language tasks.

random forest model, this was 0.55, and for the tree model, this was 0.24, indicating that the models explain 55% and 24% of the variance, respectively. To evaluate prediction accuracy, RMSE were compared as well – a lower RMSE indicates that the model makes more accurate predictions about the data. RMSE was 15.09 for the forest model, and 18.93 for the tree model, indicating better predictions in the forest model.

The variable importance plot in Figure 4 illustrates each predictor's contribution to the model, shown in %IncMSE, or increase in Mean Squared Errors (MSE), if the variable is randomly shuffled. MSE measures the average squared difference between actual and predicted values, and thus a smaller MSE indicates better model performance. The more the model's MSE increases when the variable is shuffled, the more important the variable is for making accurate predictions. Hence, a higher %IncMSE indicates greater importance of the variable.

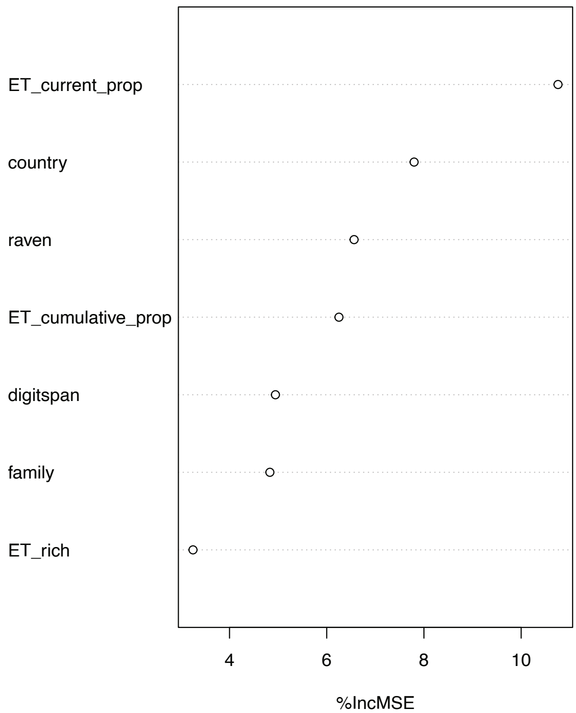

Figure 4. Importance of variables for E-SRT scores (N = 28).

Here, as in the tree model, current Estonian exposure proves to be the most important variable for explaining E-SRT scores. In the single tree model, no predictor had significant impact on the split datasets, but the forest model reveals country of residence and Raven's matrices scores as the next most important variables. Cumulative Estonian exposure is close to Raven's matrices in importance. Exposure richness is the least important among the variables included in the model.

4.2.2. Exposure and CLT scores

We also explored the relationship between CLT scores and exposure measures, to see whether the exposure factors affecting morphosyntactic knowledge and processing affect lexical knowledge in a similar manner. Although it is well documented that both exposure quality and quantity are crucial for both domains of language (Thordardottir 2011; 2015), CLT scores rely less on linguistic processing than SRT. Additionally, domain complementarity, e.g., using one language for speaking about home and family matters, and another for school, may influence the vocabulary children know in each language (Bialystok et al. 2010).

Visualising CI trees showed that the only analysis to yield any significant predictors had E-CLT as the dependent variable and included the whole sample. Figure 5 shows the resulting CI tree model.

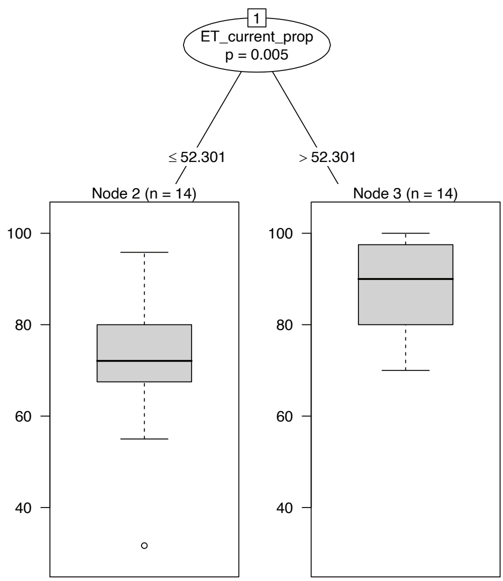

Figure 5. Significant predictor of E-CLT scores among the whole sample (N = 28): Current proportion of Estonian exposure.

Similarly to the SRT models, proportion of current Estonian exposure was the only statistically significant predictor (p = .005). Children who had over 52.3% of their current exposure in Estonian scored significantly better in E-CLT than children with less current Estonian exposure.

The CLT scores are comprised of a comprehension score and a production score, which were also modelled separately to explore whether any differences emerge. The resulting CI trees are shown on Figures 6a (comprehension) and 6b (production).

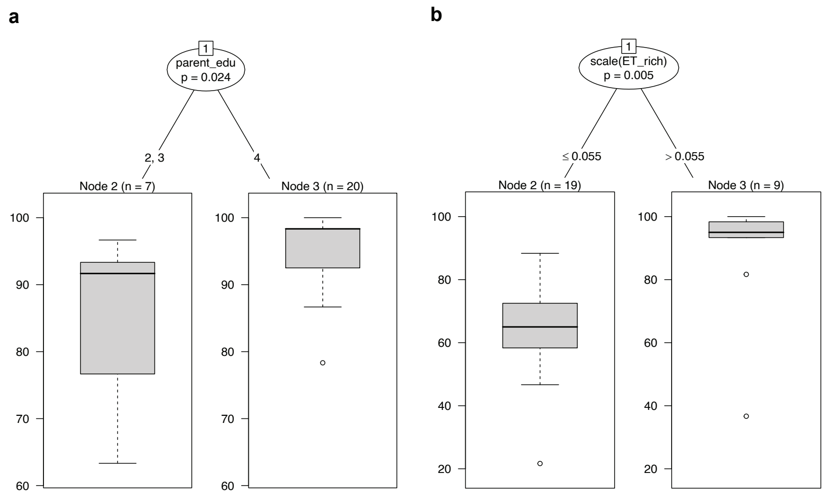

Figure 6. Significant predictors of subsections of the E-CLT: comprehension scores (panel a, N = 27) and production scores (panel b, N = 28).

Both models yield a single significant predictor, likely because the remaining split datasets have too little or too much variation internally to single out significant predictors. In word comprehension, children whose parents have higher education (coded as 4), had significantly higher scores than children whose parents had other post-secondary (coded as 3), or secondary (coded as 2), education. In word production, children with an Estonian richness score of over 0.417 (0.055 after scaling) scored significantly higher than children with a lower richness score.

A random forest model was built to further explore variables influencing E-CLT scores, analogously to the E-SRT model. The independent variables included current and cumulative Estonian exposure, Estonian richness, Raven's matrices, digit span, family type and country. The model was built with 500 trees and two variables at each split. It achieved a mean squared residual of 127.57, explaining 46.17% of variance in E-CLT scores.

R^2 for the forest model was 0.49 and 0.31 for the single tree model. RMSE was 11.29 for the forest model, and 12.92 for the tree model, again indicating the forest model to be best fitted for explaining the E-CLT scores.

Figure 7 presents the importance of variables for E-CLT scores, shown in %IncMSE.

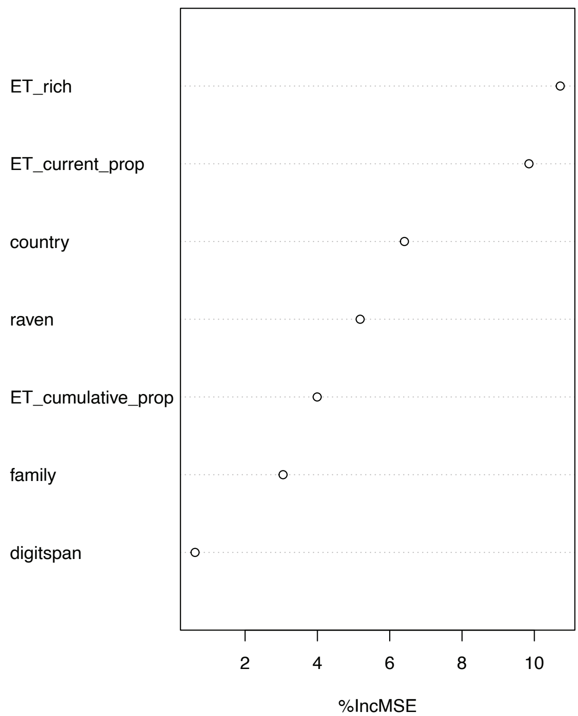

Figure 7. Importance of variables for E-CLT scores (N = 28).

As attested by the tree model, current Estonian exposure is again among the most important variables. However, the forest model shows the Estonian richness score to be slightly more important to the overall variability in CLT scores, illustrating that the first split in a tree model is *one of* the most important variables, but not necessarily the most important variable (Strobl, Malley & Tutz 2009). Aside from richness rising from the least to most important variable, other variables pattern similarly to the forest model for E-SRT scores.

5. Discussion

This study provides an overview of the language exposure and competence of a sample of Estonian-Norwegian bilingual children. In addition to being the first detailed look at this population, the study adds to research on modern Estonian diaspora communities and contributes to understanding bilingual populations with smaller, less easily accessible HL speaker communities more broadly. We now turn to the implications of the results.

First, we hypothesised a positive correlation between the quantity and quality of language exposure of Estonian-Norwegian children, language-internally. This hypothesis was confirmed in part, as significant positive correlations emerged between the amount of current exposure and richness of exposure within each language. Since the richness scores rely on estimates of current rather than cumulative exposure, this is to be expected. More exposure gives more opportunities for diverse input; likewise, the availability of diverse exposure environments is likely to mean more exposure quantitatively.

The only significant correlation between amount of cumulative exposure and richness was the weak, negative correlation between cumulative Estonian exposure and Norwegian richness. It is worth noting that although the correlations between current and cumulative exposure in each language are significant, they are moderate in effect size. These findings support the argument for collecting data on both current and cumulative exposure when possible, as one cannot necessarily act as a stand-in for the other (Unsworth 2013; De Cat 2020).

To explore potential group differences in a subsample with a more homogenous language environment, we calculated the correlation coefficients for the successive sample separately. There, correlations between exposure quality and quantity measures were not significant, alternating between positive and negative values, indicating that overall, children who get more exposure do not necessarily have higher richness scores. This analysis also revealed a moderate positive correlation between richness scores across the two languages for this subsample.

Considering that the Norwegian richness score is somewhat higher, by virtue of being the societal language, this correlation, if robust, primarily reflects the families' efforts to offer varied Estonian input. However, Norwegian richness scores vary more in the successive subsample than among the simultaneous bilinguals, so the correlation might be interpreted to suggest that families who are committed to offering language-rich activities do so in both languages. The robustness of this relationship can be explored in future research by collecting Q-BEx data from a larger sample of diaspora families in different countries.

Next, we asked how exposure quantity and quality affect language competence. For this, we analysed data from the entire sample as one group, using conditional inference trees and random forests models. No predictors proved significant for explaining Norwegian outcomes in the CI tree models. The only variable to significantly predict E-SRT scores was current Estonian exposure. Children whose current language exposure was over 49.3% Estonian performed significantly better than children with less Estonian exposure ($p = .025$). A random forests analysis also showed current exposure as the most important variable, with 10.76% IncMSE. However, richness did not seem to contribute significantly to explaining the variation in E-SRT scores, with no significant splits in the tree model, and just 3.25% IncMSE in the forest model. H2 was partially confirmed for morphosyntactic competence: current exposure quantity is the most significant predictor of SRT scores, while exposure quality did not prove to be an important predictor in our analyses.

For lexical competence, proportion of current Estonian exposure again proved to be the only statistically significant predictor in the CI tree. Children with more than 52.3% of their current language exposure in Estonian performed significantly better than children with less Estonian exposure ($p = .005$). However, when exploring the word comprehension and word production scores separately, two other, quality-related predictors proved significant: parental education for comprehension scores ($p = .024$), and Estonian richness for production scores ($p = .005$). Parental education, which can be interpreted as a proxy for socioeconomic status, has been found to be significant for vocabulary acquisition in monolingual Estonian children (Urm & Tulviste 2016). We can speculate that highly educated parents have more opportunities to offer language resources and activities in Estonian, resulting in more and greater diversity of exposure. Estonian richness emerging as significant for productive vocabulary also highlights the importance of exposure diversity for lexical competence, and particularly the greater motivation to make productive use of the language which comes with diversity of experiences in it.

These results were supported by the random forests model, identifying Estonian richness and current Estonian exposure as the most important variables for predicting E-CLT scores, followed by country of residence. Despite the seeming discrepancy between the tree model singling out current exposure, and the forest model identifying richness as the most important variable, the two models complement each other. Estimated goodness of fit identified the random forests model as more robust and reliable, and as mentioned above, the first split in a tree model is not always the most important variable. Some variability may be due to the inherent randomness in the 'tree and forest' models, although the same seeding was used for all models. H2 is partly confirmed for lexical competence: both exposure quantity and quality play an important role, but it is not straightforward to claim that quantity is more significant than quality.

It thus appears that exposure quality affects lexical more than grammatical competence. One explanation for this is the link between grammatical acquisition and procedural memory (Ullman 2001): mastering the syntactic structures necessary for repeating sentences in the SRT requires that the syntactic structures be familiar, well-rehearsed and accessible to online processing. Greater cumulative exposure is likely to have provided more repeated exposure to these structures, hence leading to better familiarisation and entrenchment in procedural memory. Greater current exposure is likely to boost levels of current activation, making the structures more easily accessible. The structures in the Estonian SRT are common enough to be accessible and repeatable for typically developing, monolingual 5-6-year-olds (Padrik, Vihman & Fil 2022), so it is likely that most of the structures are found in HL Estonian exposure at home, making variation in exposure contexts relatively less important.

Lexical acquisition, which has been linked to declarative memory in first and second language acquisition, as well as in online processing (see Ullman 2016), requires exposure to diverse lexical items. Undeniably, repetition in exposure facilitates acquiring lexical items, but more diverse exposure will give better opportunities to encounter more lexical items in the first place. In addition, richness scores were significantly correlated with the amount of current Estonian exposure, implying that the higher CLT scores are driven not only by exposure quality, but by current exposure that is both ample and diverse.

Task structure can also play a role: while the SRT contains mostly easier, early-acquired vocabulary, CLT aims to balance both earlier- and later-acquired vocabulary, giving a clearer advantage to children with more diverse exposure contexts, where less frequent vocabulary is encountered (e.g. book-reading, Sunday school language classes). Moreover, no prior activation of the language material is provided in the CLT, unlike the SRT, which necessarily involves providing the stimulus sentence to the child. Children with greater overall activation of Estonian (i.e. more current exposure) therefore had an advantage.

Cumulative Estonian exposure was not among the most important variables in either forest model. De Cat (2020) who used the same exposure measures from Q-BEx, found cumulative exposure and use to be a far better predictor of SL English skills (measured, among other methods, in LITMUS SRT scores) than current exposure or use. Thordardottir (2011, 2015) also found cumulative exposure to be the best predictor of language skills among languages with relatively equal status (English and French in Canada). A more recent Q-BEx-based study (De Cat et al. 2025) similarly found cumulative exposure to be the most informative variable for predicting SL skills. However, HL skills were best predicted by current HL exposure at home. Due to the larger size and heterogeneity in their sample, De Cat et al. (2025) used a parental assessment of HL skills as a measure of HL outcomes. Our results, using more direct measures of language competence, confirm their findings: current exposure best predicts language competence in HL. As De Cat et al. (2025) suggest, the effects of cumulative and current exposure seem to be mediated by language status: for HL children schooled in SL, the current dominance might "override" effects of cumulative exposure arising from early HL exposure, and thus current exposure becomes crucial for developing and maintaining HL skills.

This effect of language status may also partially explain the importance of country of residence in our models (7.8% IncMSE for SRT, 6.4% IncMSE for CLT), as this predictor differentiates HL Estonian children from SL Estonian children. One explanation for 'country' surpassing cumulative exposure in importance is that 'country' reflects differences in both cumulative and current exposure, which are both connected to the HL/SL status. Family type (simultaneous or successive), which could also be expected to influence the proportion of exposure to Estonian, did not stand out this way.

Instead of the binary measures, the scalar measure of current Estonian exposure was among the top predictors of language competence in all models. The tree models illustrate how the split between higher- and lower-scoring children does not go neatly between simultaneous and successive bilinguals, or between residents of Estonia and Norway. Rather, children with roughly over half of their exposure in Estonian score better. This recalls the findings of Thordardottir (2015), who found that English-French bilingual children performed similarly to monolingual peers in the language comprising over 50% of their input. Methodologically, this finding supports the growing evidence that describing language experience as a scale (e.g. of proportion of exposure in either language), rather than a binary category of successive vs. simultaneous bilinguals, or heritage vs. societal speakers of the language (Wiese et al. 2022) is more adequate for describing and analysing language exposure.

An important caveat for interpreting the random forests models is that the models explained just under half of the variance in the scores (46.91% for SRT and 46.17% for CLT). The rest of variance must be explained by individual differences not considered here, or contextual differences too subtle for our measures to capture. It is possible that a larger proportion of the variance would be explained with a larger sample, but this may also increase the variance. Insufficient in-group variance can be the reason why no significant predictors emerged when analysing the Norwegian test scores, and when looking at the N-suc subsample separately.

The most important limitation of this study is the small sample size, limited by the population of interest being relatively small and difficult to access for a researcher primarily based in Estonia. Given these limitations, we included all participants who answered our criteria, resulting in a heterogeneous group. However, heterogenous samples are typical of research on multilingual populations. By treating this heterogenous sample as a single group in quantitative analyses, including potentially meaningful predictors to control for the variability, we have arrived at conclusions which point to both the importance of exposure data and the differences between the measures. These results can be probed further with larger, though doubtless also heterogeneous, samples.

6. Conclusion

In conclusion, this first overview of the language environment of Estonian-Norwegian bilingual children confirms that ample, diverse exposure in the heritage language is both necessary for language development and possible to achieve. Both simultaneous and successive children in Norway have quite balanced exposure in the two languages, a testament to the families' efforts to support Estonian exposure. Current exposure correlated significantly with the richness of exposure in both languages. Yet, zooming in on a smaller subsample of successive bilinguals showed a positive correlation between the richness scores in the two languages, implying that exposure quality is not a zero-sum game.

Both amount and diversity of exposure proved important for predicting morphosyntactic and lexical competence. While current amount of exposure proved highly important for both subskills, diversity of exposure played a more prominent role for lexical than morphosyntactic competence. The mechanisms behind this difference remain to be studied with larger samples, more targeted methods, and both quantitative and qualitative analysis.

Acknowledgements

Data collection for this study was funded by the Kadri, Nikolai and Gerda Rõuk research fund at the Institute of Estonian and General Linguistics, University of Tartu, and a Dora Pluss mobility grant. We thank University of Bergen for hosting the first author during data collection, Arve Asbjørnsen for help with the cognitive tasks, Caroline Rowland for helpful discussion in early stages of the study, and Piia Taremaa for help with data analysis. This study would not have been possible without the help of student assistants collecting and coding Norwegian data, and naturally, without the participating families. We thank two anonymous reviewers and the editor for insightful comments on the draft of this paper.

References

Albirini, Abdulkafi. 2018. The Role of Age of Exposure to English in the Development of Arabic as a Heritage Language in the United States. *Language Acquisition* 25(2). 178–196. https://doi.org/10.1080/10489223.2016.1234620.

Armon-Lotem, Sharon, Jan de Jong & Natalia Meir (eds.). 2015. *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (Communication Disorders across Languages). Vol. 13. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137.

Armon-Lotem, Sharon, Karen Rose & Carmit Altman. 2021. The development of English as a heritage language: The role of chronological age and age of onset of bilingualism. *First Language* 41(1). 67–89. https://doi.org/10.1177/0142723720929810.

Bialystok, Ellen, Gigi Luk, Kathleen F. Peets & Sujin Yang. 2010. Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition* 13(4). 525–531. https://doi.org/10.1017/S1366728909990423.

Biedinger, Nicole, Birgit Becker & Oliver Klein. 2015. Turkish-language ability of children of immigrants in Germany: which contexts of exposure influence preschool children's acquisition of their heritage language? *Ethnic and Racial Studies* 38(9). 1520–1538. https://doi.org/10.1080/01419870.2015.1005641.

Bome, Camilla & Ingrid Kongtorp Vargen. 2017. *The development and pilot of a Norwegian adaptation of the LITMUS Sentence Repetition task*. Bergen: University of Bergen Master's thesis.

Bylund, Emanuel, Kenneth Hyltenstam & Niclas Abrahamsson. 2021. Age of acquisition – not bilingualism – is the primary determinant of less than nativelike L2 ultimate attainment. *Bilingualism: Language and Cognition* 24. 18–30. https://doi.org/10.1017/S1366728920000188.

Carroll, Susanne E. 2017. Exposure and input in bilingual development. *Bilingualism: Language and Cognition* 20(1). 3–16. https://doi.org/10.1017/S1366728915000863.

Cutler, Richard D, Thomas C. Edwards Jr, Karen H. Beard, Adele Cutler, Kyle Hess, Jacob Gibson & Joshua Lawler. 2007. Random forests for classification in ecology. *Ecology* (Appendix A. Technical Details and Additional Capabilities of Random Forests.) 88. 2783–2792. https://doi.org/10.1890/07-0539.1.

Daskalaki, Evangelia, Elma Blom, Vasiliki Chondrogianni & Johanne Paradis. 2020. Effects of parental input quality in child heritage language acquisition. *Journal of Child Language* 47(4). 709–736. https://doi.org/10.1017/S0305000919000850.

De Cat, Cécile. 2020. Predicting language proficiency in bilingual children. *Studies in Second Language Acquisition* 42(2). 279–325. https://doi.org/10.1017/S0272263119000597.

De Cat, Cecile, Arief Gusnanto, Draško Kašćelan, Philippe Prévost, Ludovica Serratrice, Laurie Tuller & Sharon Unsworth. 2025. How detailed do measures of bilingual language experience need to be? A cost–benefit analysis using the Q-BEx questionnaire. Bilingualism: Language and Cognition 1–12. https://doi.org/10.1017/S1366728925100497.

De Cat, Cécile, Draško Kašćelan, Philippe Prevost, Ludovica Serratrice, Laurie Tuller & Sharon Unsworth. 2022. Quantifying Bilingual EXperience (Q-Bex): questionnaire manual and documentation. OSF. https://doi.org/10.17605/OSF.IO/V7EC8.

Fishman, Joshua. 1972. *The Sociology of Language*. Massachusetts: Newbury: Rowley.
Hart, Betty & Todd R. Risley. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore, MD, US: Paul H Brookes Publishing.
Hassinen, Sirje. 2002. *Simultaaninen kaksikielisyys. Läheiset sukukielet viro ja suomi rinnakkain*. Oulu: University of Oulu Doctoral thesis. https://oulurepo.oulu.fi/bitstream/handle/10024/37343/isbn951-42-6648-X.pdf?sequence=1. (6 February, 2025).
Hoff, Erika, Cynthia Core, Silvia Place, Rumiche Rosario, Melissa Señor & Marisol Parra. 2012. Dual language exposure and early bilingual development. *Journal of child language* 39(1). 1–27. https://doi.org/10.1017/S0305000910000759.
Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674. https://doi.org/10.1198/106186006X133933.
Hothorn, Torsten & Achim Zeileis. 2015. partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research* 16.
Huttenlocher, Janellen, Heidi Waterfall, Marina Vasilyeva, Jack Vevea & Larry V. Hedges. 2010. Sources of variability in children's language growth. *Cognitive Psychology* 61(4). 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002.
Integratsiooni Sihtasutus. 2022. Global Estonians: stats & numbers. *Global Estonian*. https://globalestonian.com/et/global-estonians-in-numbers. (11 June, 2023).
Johnson, Jacqueline S. & Elissa L. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21(1). 60–99. https://doi.org/10.1016/0010-0285(89)90003-0.
Jokela, Hanna & Geda Paulsen. 2010. Võtke minust eesmärki! Monikielinen lapsi rakenteita yhdistämässä. *Lähivõrdlusi. Lähivertailuja* 19. 61–75. https://doi.org/10.5128/LV19.04.
Kim, Kitaek & Hyunwoo Kim. 2022. Sequential bilingual heritage children's L1 attrition in lexical retrieval: Age of acquisition versus language experience. *Bilingualism: Language and Cognition* 25(4). 537–547. https://doi.org/10.1017/S1366728921001139.
Klavan, Jane, Maarja-Liisa Pilvik & Kristel Uiboaed. 2015. The Use of Multivariate Statistical Classification Models for Predicting Constructional Choice in Spoken, Non-Standard Varieties of Estonian. *Finnish Journal of Linguistics* 28. 187–224.
Kuvač-Kraljević, Jelena, Antonija Blaži, Astra Schults, Tiia Tulviste & Suvi Stolt. 2021. Influence of internal and external factors on early language skills: A cross-linguistic study. *Infant Behavior and Development* 63. 101552. https://doi.org/10.1016/j.infbeh.2021.101552.
Laar, Kadri-Ann. 1990. Estonian Ethnic Identity: Language Maintenance and Shift. *Journal of Baltic Studies* 21(3). 239–258. https://doi.org/10.1080/01629779000000121.
Liaw, A & M Wiener. 2002. Classification and Regression by randomForest. *R News* 2(3). 18–22.
Limmer, Karoliine. 2008. *"Mis see eesti keeles on, ei tea?" ("Vet du hva det er på estisk?"). En undersøkelse av språksituasjonen og samtalestrategier hos en trespråklig familie med tospråklige estisk-norske barn*. Tartu: University of Tartu Bachelor's thesis.

Lipp, Ellen. 1977. The acquisition of Estonian inflections. *Journal of Child Language* 4(2). 313–319. https://doi.org/10.1017/S0305000900001689.

Łuniewska, Magdalena, Marta Wójcik, Joanna Kołak, Karolina Mieszkowska, Zofia Wodniecka & Ewa Haman. 2021. Word knowledge and lexical access in monolingual and bilingual migrant children: Impact of word properties. *Language Acquisition* 29(2). 1–30. https://doi.org/10.1080/10489223.2021.1973475.

Marinis, Theodoros & Sharon Armon-Lotem. 2015. Sentence repetition. In Sharon Armon-Lotem, Jan de Jong & Natalia Meir (eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (Communication Disorders across Languages 13), 95–121. Bristol: Multilingual Matters.

Meir, Natalia & Bibi Janssen. 2021. Child heritage language development: An interplay between cross-linguistic influence and language-external factors. *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.651730.

Mitrofanova, Natalia, Yulia Rodina, Olga Urek & Marit Westergaard. 2018. Bilinguals' Sensitivity to Grammatical Gender Cues in Russian: The Role of Cumulative Input, Proficiency, and Dominance. *Frontiers in Psychology* 9. https://doi.org/10.3389/fpsyg.2018.01894.

Oksaar, Els. 1971. Code Switching as an Interactional Strategy for Developing Bilingual Competence. *WORD* 27(1–3). 377–385. https://doi.org/10.1080/00437956.1971.11435633.

Osterreich, Helgi Kasak. 1977. *The Acquisition of Locatives in Estonian*. Albuquerque: University of New Mexico Unpublished doctoral thesis.

Padrik, Marika, Virve-Anneli Vihman & Olga Fil. 2022. Keelepuude märkamine kakskeelsetel lastel: eestikeelse lausete järelekordamise testi prooviuuring. *Eesti Haridusteaduste Ajakiri* 10(2). 305–333. https://doi.org/10.12697/eha.2022.10.2.11.

Peterson, Betti Marie. 2020. *Problematiske sider ved innlærig av norsk for estiske studenter*. Tartu: University of Tartu Bachelor's thesis. http://hdl.handle.net/10062/67177. (14 August, 2024).

Piksarv, Karoliine. 2013. *„Frosken, hvor er du?" – syntaktisk kompleksitet i skriftlige fortellinger av estiske innlærere av norsk som fremmedspråk*. Tartu: University of Tartu Bachelor's thesis. http://hdl.handle.net/10062/35346. (14 August, 2024).

Place, Silvia & Erika Hoff. 2011. Properties of Dual Language Exposure That Influence 2-Year-Olds' Bilingual Proficiency. *Child Development* 82(6). 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x.

Place, Silvia & Erika Hoff. 2016. Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition* 19(5). 1023–1041. https://doi.org/10.1017/S1366728915000322.

Polinsky, Maria. 2018. *Heritage Languages and their Speakers* (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107252349.

Polišenská, Kamila, Shula Chiat & Penny Roy. 2015. Sentence repetition: What does the task measure? *International Journal of Language & Communication Disorders* 50(1). 106–118. https://doi.org/10.1111/1460-6984.12126.

R Core Team. 2023. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Raven, J C. 1976. *Standard Progressive Matrices*. San Antonio: Harcourt Assessment.

Rodina, Yulia, Tanja Kupisch, Natalia Meir, Natalia Mitrofanova, Olga Urek & Marit Westergaard. 2020. Internal and external factors in heritage language acquisition: evidence from heritage Russian in Israel, Germany, Norway, Latvia and the United Kingdom. *Frontiers in Education* 5. https://doi.org/10.3389/feduc.2020.00020.

Rodina, Yulia & Marit Westergaard. 2017. Grammatical gender in bilingual Norwegian–Russian acquisition: The role of input and transparency. *Bilingualism: Language and Cognition* 20(1). 197–214. https://doi.org/10.1017/S1366728915000668.

Rothman, Jason, Fatih Bayram, Vincent DeLuca, Grazia Di Pisa, Jon Andoni Duñabeitia, Khadij Gharibi, Jiuzhou Hao, et al. 2023. Monolingual comparative normativity in bilingualism research is out of "control": Arguments and alternatives. *Applied Psycholinguistics* 44(3). 316–329. https://doi.org/10.1017/S0142716422000315.

Rowe, Meredith L. 2012. A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. *Child Development* 83(5). 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x.

Salasoo, Tiiu. 1995. Morfoloogiliste tunnuste esmakasutamine ühe lapse arenevas keeles. *Keel ja Kirjandus* 38(4). 239–252.

Salasoo, Tiiu. 2010. Laste esimesed eestikeelsed omadussõnad. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 1(1). 57–80. https://doi.org/10.12697/jeful.2010.1.1.04.

Smolander, Sini, Marja Laasonen, Eva Arkkila, Pekka Lahti-Nuuttila & Sari Kunnari. 2021. L2 vocabulary acquisition of early sequentially bilingual children with TD and DLD affected differently by exposure and age of onset. *International Journal of Language & Communication Disorders* 56(1). 72–89. https://doi.org/10.1111/1460-6984.12583.

Sorenson Duncan, Tamara & Johanne Paradis. 2018. How does maternal education influence the linguistic environment supporting bilingual language development in child second language learners of English? *International Journal of Bilingualism* 24(1). 46–61. https://doi.org/10.1177/1367006918768366.

Statistics Estonia. 2021. RL21442: Population by Ethnic Nationality, Mother Tongue, Citizenship, Sex, Age Group and Place of Residence (Settlement Region), 31 December 2021. *Statistical Database*. https://andmed.stat.ee:443/pxweb/en/stat/stat__rahvaloendus__rel2021__rahvastiku-demograafilised-ja-etno-kultuurilised-naitajad__rahvus-emakeel/RL21442.px/. (30 July, 2025).

Statistics Norway. 2025. 05183: Immigrants and Norwegian-born to immigrant parents, in total, by sex and country background 1970–2025. Statbank Norway. *Statbank Norway*. https://www.ssb.no/en/system/. (30 July, 2025).

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9(1). 307. https://doi.org/10.1186/1471-2105-9-307.

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14. 323–348. https://doi.org/10.1037/a0016973.

Tao, Lily, Qing Cai & Tamar H. Gollan. 2021. Effects of cumulative language exposure on heritage and majority language skills: Spanish and Mandarin heritage speakers in the USA. *Linguistic Approaches to Bilingualism* 11(2). 168–191. https://doi.org/10.1075/lab.18044.tao.

Teiss, Kristiina. 2006. Suhtlusstrateegiad eesti-soome kakskeelsust omandava lapse segakoodis. *Eesti Rakenduslingvistika Ühingu aastaraamat* 2. 251–269. https://doi.org/10.5128/ERYa2.17.

Thordardottir, Elin. 2011. The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism* 15(4). 426–445. https://doi.org/10.1177/1367006911403202.

Thordardottir, Elin. 2015. The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology* 17(2). 97–114. https://doi.org/10.3109/17549507.2014.923509.

Thordardottir, Elin. 2019. Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism* 23(1). 236–255. https://doi.org/10.1177/1367006917722418.

Torregrossa, Jacopo, Cristina Flores & Esther Rinke. 2023. What modulates the acquisition of difficult structures in a heritage language? A study on Portuguese in contact with French, German and Italian. *Bilingualism: Language and Cognition* 26(1). 179–192. https://doi.org/10.1017/S1366728922000438.

Tsimpli, Ianthi Maria. 2014. Early, late or very late?: Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3). 283–313. https://doi.org/10.1075/lab.4.3.01tsi.

Tsinivits, Danai & Sharon Unsworth. 2021. The impact of older siblings on the language environment and language development of bilingual toddlers. *Applied Psycholinguistics* 42(2). 325–344. https://doi.org/10.1017/S0142716420000570.

Ullman, M. T. 2001. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30(1). 37–69. https://doi.org/10.1023/a:1005204207369.

Ullman, Michael T. 2016. The Declarative/Procedural Model. In *Neurobiology of Language*, 953–968. Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00076-6.

Unsworth, Sharon. 2013. Assessing the role of current and CUMULATIVE exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition* 16(1). 86–110. https://doi.org/10.1017/S1366728912000284.

Unsworth, Sharon, Froso Argyri, Leonie Cornips, Aafke Hulk, Antonella Sorace & Ianthi Tsimpli. 2011. On the Role of Age of Onset and Input in Early Child Bilingualism in Greek and Dutch. In Mihaela Pirvulescu, Cristina Cuervo Mara, Jeffrey Steele, Nelleke Strik & Ana Teresa Pérez Leroux (eds.), *Selected proceedings of the 4th Conference on Generative Approaches to Language Acquisition North America (GALANA 2010)*, 249–265. Sommerville, MA: Cascadilla Proceedings Project. https://doi.org/10.1017/S0142716412000574.

Unsworth, Sharon, Arief Gusnanto, Draško Kašćelan, Philippe Prevost, Ludovica Serratrice, Laurie Tuller & Cecile De Cat. 2024. Unpacking language richness as a predictor of bilingual children's language proficiency. OSF. https://doi.org/10.31219/osf.io/rquvc.

Urm, Ada & Tiia Tulviste. 2016. Sources of individual variation in Estonian toddlers' expressive vocabulary. *First Language* 36(6). 580–600. https://doi.org/10.1177/0142723716673951.

Vaks, Adele, Marika Padrik & Virve Vihman. 2025. Sõnavaratesti väljatöötamine arengulise keelepuude tuvastamiseks kakskeelsetel lastel. *Eesti Rakenduslingvistika Ühingu aastaraamat* 21. 327–343. https://doi.org/10.5128/ERYa21.18.

Vaks, Adele & Virve-Anneli Vihman. 2022. Nimisõnafraas eesti-norra kakskeelsete laste keelekasutuses. *Eesti Rakenduslingvistika Ühingu aastaraamat* 18. 305–323. https://doi.org/10.5128/ERYa18.17.

Vaks, Adele & Virve-Anneli Vihman. 2025. Bilingual Acquisition of Morphology: Norwegian and Russian Influence on Children's Sentence Repetition in Estonian. *Language and Speech*. 00238309251327209. https://doi.org/10.1177/00238309251327209.

Verhoeven, Emma, Merel van Witteloostuijn, Ora Oudgenoeg-Paz & Elma Blom. 2024. Comparing Different Methods That Measure Bilingual Children's Language Environment: A Closer Look at Audio Recordings and Questionnaires. *Languages* 9(7). 231. https://doi.org/10.3390/languages9070231.

Vihman, Marilyn May. 1982. The acquisition of morphology by a bilingual child: A whole-word approach. *Applied Psycholinguistics* 3(2). 141–160. https://doi.org/10.1017/S0142716400006676.

Vihman, Marilyn May. 1985. Language differentiation by the bilingual infant. *Journal of Child Language* 12(2). 297–324. https://doi.org/10.1017/S0305000900006450.

Wechsler, David. 1974. *Wechsler Intelligence Test for Children - Revised*. (Trans.) John Olav Undheim. Norsk utgave. New York: The Psychological Corporation.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis* (Use R!). New York: Springer-Verlag. 978-3-319-24277-4.

Wickham, H, R François, L Henry, K Müller & D Vaughan. 2023. dplyr: A Grammar of Data Manipulation. https://github.com/tidyverse/dplyr, https://dplyr.tidyverse.org.

Wiese, Heike, Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Maria Martynova, et al. 2022. Heritage Speakers as Part of the Native Language Continuum. *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.717973.

Ågren, Malin, Jonas Granfeldt & Anita Thomas. 2014. Combined effects of age of onset and input on the development of different grammatical structures: A study of simultaneous and successive bilingual acquisition of French. *Linguistic Approaches to Bilingualism* 4(4). 462–493. https://doi.org/10.1075/lab.4.4.03agr.

2024. *Lov om grunnskolen og den vidaregåande opplæringa (opplæringslova) [Law of primary and secondary education (education law)]*. https://lovdata.no/dokument/NL/lov/2023-06-09-30. (14 October, 2025).

Kokkuvõte. Adele Vaks, Virve-Anneli Vihman, Jan de Jong: Keelesisend ja -oskus eesti-norra kakskeelsete laste seas. Artiklis kirjeldame 5–7-aastaste eesti-norra kakskeelsete laste keelesisendit ning keelesisendi suhet eesti keele oskusega. Valimis on 29 last: 16 suktsessiivselt kakskeelset ja kaheksa simultaanselt kakskeelset Norras, ning viis simultaanselt kakskeelset last Eestis. Keelekeskkonna kirjeldamiseks kogusime andmeid laste keelekogemuse kohta Q-BEx ankeediga. Eesti keele oskuse kirjeldamiseks kasutame lausete järelekordmistesti ning sõnavaratesti skoore. Korrelatsioonianalüüsist selgub, et keelesisendi mitmekesisus korreleerub oluliselt ja positiivselt viimase aasta eestikeelse sisendi osakaaluga. Keelesisendi ja -oskuse suhte analüüsimiseks kasutasime tingimuslikke otsustuspuid ning juhumetsa mudeleid. Mudelite põhjal mõjutab laste morfosüntaksi oskust kõige enam käesoleva aasta eestikeelse sisendi osakaal. Sõnavarateadmiste jaoks olid olulised nii eestikeelse sisendi osakaal kui ka sisendi mitmekesisus.

Märksõnad: kakskeelsus, pärandkeel, keelesisend, eesti keel, norra keel