

Latvian CDI: methodology, developmental trends and cross-linguistic comparison

Olga Urek, Anna Vulāne, Roberts Dargis, Agrita Tauriņa, Tija Zīriņa & Hanne Gram Simonsen

Abstract

In this article we report the results of a large-scale population study based on the Latvian adaptation of Communicative Development Inventories (CDI, Fenson et al. 2007) – a parental report tool aimed at mapping the lexical and grammatical development of children under the age of three. Two CDI forms are discussed: CDI I: “Words and Gestures” (8–16 months), and CDI II: “Words and Sentences” (17–36 months). This article discusses the internet-based methodology used for the data collection, reports the main developmental trends of the lexical development of Latvian children, and compares these trends to analogous data from American English, Norwegian, and Russian.

Keywords: CDI; Latvian; lexical development; receptive and productive vocabulary; communicative gestures; cross-linguistic comparison

1. Introduction

Communicative Development Inventories (CDI) is a parental report tool for assessing language and communicative development of infants and young children (Fenson et al. 2007). The present article is concerned with two CDI forms. The CDI “Words and

Gestures” (Infant form) is designed for use with children between 8 and 16 months of age, and measures productive and receptive vocabulary and communicative gestures and routines. The CDI “Words and Sentences” (toddler form) targets 16- to 30-month (36 for some languages) old children, and measures productive vocabulary and certain aspects of grammar. Each CDI form contains extensive and language-specific checklists of lexical items, communicative gestures, and grammatical constructions. Originally developed for American English, the CDI has since been adapted to over 60 languages and dialects. The Latvian CDI is unique among the language assessment tools currently used in Latvia in two crucial respects. First, it is the first instrument of its kind to reflect grammatical, lexical, and cultural-linguistic properties specific to the Latvian language. Second, it is the first language assessment instrument that was adjusted for the Latvian child population. It has to be mentioned that unlike most other available CDI adaptations, Latvian CDI has been normed with a child population where a non-negligible exposure to a second language (referred to as L2 in linguistics, as compared to mother tongue, L1) was presumed. Although the state language of Latvia is Latvian, it is *de facto* a bilingual nation, with ethnic Russians constituting approximately 26% of the country’s population and 34% of inhabitants mainly using the Russian language at home (results of the 2011 Population Census). It might be expected that non-negligible early exposure to Russian as L2 might affect lexical development of Latvian-speaking children. Previous studies show that bilingual toddlers attain lower CDI scores than their monolingual peers in both expressive (Poulin-Dubois et al. 2013) and receptive vocabulary (Bialystok et al. 2010) in their L1. At the same time, studies show no difference between monolingual and bilingual toddlers when both L1 and L2 vocabularies are taken into account (e.g. Junker and Stockman 2002; Pearson et al. 1993). While it remains a possibility that

the linguistic situation in Latvia might impact the acquisition progression of those Latvian-speaking children with minimal exposure to Russian, we found no evidence that children in our sample show lower vocabulary scores than their peers acquiring American English, Norwegian, and Russian.

Over the years, the CDI has proven itself as a reliable instrument across languages and target populations of children (for an overview see Fenson et al. 2007; Law and Roy 2008). It has been successfully used with typically developing children, both as an assessment method and as a screening tool to identify late talkers, i.e. children whose language development is significantly delayed relative to the norm (see Heilmann et al. 2005). It has also been employed as an evaluation, screening, and monitoring tool with different atypical populations (see Thal et al. 1999 for specific language impairment; Berglund, Eriksson, and Johansson 2001 for Down syndrome; Charman et al. 2003 for autism; Thal et al. 2007 for cochlear implants), and populations considered at risk of developing language impairment (see Feldman et al. 2003 for otitis media; Magill-Evans and Harrison 1999 for healthy preterm children; Koster et al. 2005 for familial risk of dyslexia). In addition, the availability of CDI adaptations to different languages makes it an excellent tool for the study of bilingual language acquisition (Pearson et al. 1997, De Houwer et al. 2006, Conboy and Thal 2006, Gatt, O'Toole, and Haman 2015, O'Toole et al. 2017) and cross-linguistic trends in language development. With respect to the latter, previous studies have revealed both similarities and differences in the lexical development of monolingual children across languages. An extensive survey of CDI studies has been conducted by Bleses et al. (2008), where they compared the results of the Danish CDI with 17 CDI-based studies in other languages. It has been found that a number of properties characterizing early vocabulary development are cross-linguistically constant. Such

as, for instance, a large degree of individual variation among children, a vocabulary spur in the second year of life, and asymmetry between receptive and productive vocabulary scores. At the same time, considerable differences across languages have been observed in productive and receptive vocabulary sizes at different ages.

The aim of this article is to report the results of a large-scale population study based on the Latvian adaptation of CDI I and II, and compare the observed developmental trends to the analogous data from the Russian, Norwegian, and American English CDI. The selection of CDI adaptations to compare with Latvian CDI data was mostly dictated by the typological proximity and similarities in terms of the morphological complexity of Latvian and Russian on the one hand, and Norwegian and English on the other. An additional rationale for selecting the language was the size of the norming populations, with both Latvian and Russian CDI being based on relatively restricted samples, while Norwegian and American were both normed with a large and well-balanced sample of respondents.

In Section 2, we describe the methodology used to collect the data and the demographic characteristics of the final sample. In the second part of Section 2, we discuss the Latvian adaptation of the CDI I and CDI II forms. Section 3 presents the main results of the normative study and compares them to the results obtained for American English, Norwegian, and Russian. Section 4 is the conclusion and summary of the discussion.

2. Methodology

2.1. Recruitment

In this study, we used the internet-based data collection procedure first implemented for the Norwegian CDI and described in Kristoffersen et al. (2013) and Kristoffersen

and Simonsen (2012). Participants were recruited with the help of the Office of Citizenship and Migration Affairs of Latvia (OCMA). On 16 April 2016, 6,612 families corresponding to the pre-defined criteria described below were randomly selected from the Population Registry of Latvia and sent a registered letter prepared by the research team. The letter contained a brief description of the CDI and the normative study, information on voluntary participation and withdrawal, contact details of the research team, an invitation to participate in the study, the web address of the CDI forms and an automatically-generated access code. To protect the anonymity of the respondents, all letters were sent by OCMA. Prior to data collection, an information campaign was launched in Latvian mass media (in printed media, on the radio, and online) to raise awareness of the study.

2.2. Pre-selection criteria

In order to be included in the respondent pool, participants had to meet a number of pre-defined criteria pertaining to the age of their children, geographical location and language. Since the norming study was concerned with CDI I and II, we only targeted families with children aged between 7.5 and 35.5 months on the date when data collection started. A total of 228 families of the targeted age per month were included in the respondent pool, and children of each gender were represented in equal proportions. The selected families were evenly distributed across four geographical areas: Riga, the Latgale region, the Vidzeme region and the Kurzeme region (1,653 families in each). To maximize the likelihood of internet access being readily available to respondents, we targeted primarily large cities and towns in each region.

Special care was taken to maximize the proportion of monolingual Latvian-speaking children in the respondent pool. According to the results of the 2011

Population Census, the proportion of respondents who use Latvian as the main language in their household was 56%, while the proportion of those who mainly use Russian was 34%. The proportions also vary considerably by geographical area: for instance, only 38% of the respondents in Riga indicated that they mainly use Latvian at home. Since the current study is concerned with mapping typical monolingual communicative development, the data obtained from bilingual children could not be included in the analysis. While the background questionnaire included as a part of the CDI contained a question about exposure to languages other than Latvian, our concern was that we would have to exclude too many respondents if participants are not pre-selected based on their first language. Since the Latvian Population Registry does not register information about an individual's native language(s), we used self-declared ethnicity information as a proxy when selecting a sample of families to target. Only families where both parents declared their ethnicity as "Latvian" were included in the respondent pool.

2.3. Electronic forms

After CDI I and II were adapted for the Latvian language (Vulāne, Urek, and Markus 2016), electronic versions of both forms were created and tested by the research team. After they logged in, participants were prompted to enter their child's gender, date of birth, and their relation to the child (mother, father, or "other"), and they were then automatically directed to either CDI I (for children aged between 8 and 16 months) or CDI II (for children aged between 17 and 36 months). Upon completion of the CDI form, participants were asked to fill out a background questionnaire containing demographic related questions (area of residence, birth order and sibling status of the child, parental education) and a number of questions to determine if the child met the

exclusion criteria (exposure to languages other than Latvian, health issues that might affect language development, parental concerns). In addition, respondents could fill out their contact information if they wanted to receive the norms upon completion of the study.

2.4. Final sample

Within four months of dispatching invitation letters, a total of 538 people (approximately 8.1%) filled out the forms. Since the response rate was much lower than we anticipated based on the results of the Norwegian and Danish normative studies (Kristoffersen et al. 2013), we decided to make CDI forms available to the general public, and added the functionality allowing participants not originally included into the study to request an access code. In an attempt to recruit more respondents, a second information campaign was launched in printed media in Latvia and on social media. This resulted in an additional 340 responses, bringing the overall response rate to 13.3% (cf. 37% in the Norwegian norming CDI study and 34% in the Danish norming CDI study; Kristoffersen et al. 2013; Bleses et al. 2008).

Of the 878 completed forms, 131 had to be excluded from the final sample for various reasons. The majority of those were excluded due to a considerable exposure to languages other than Latvian (N=124). Given the fact that the vast majority of Latvian children grow up in a bilingual environment, we set the threshold for “considerable exposure” at 24 hours per week (cf. the American CDI norming study, where the threshold was set at 12 hours per week, Fenson et al. 2007, 55). Additionally, seven children were excluded based on the information provided by the parents in the “parental concerns” section: innate intellectual disability (N=3), deafness (N=2), and lack of exposure to Latvian (N=2). Other parental concerns (e.g.

articulation accuracy, limited expressive vocabulary, non-language-related concerns) did not constitute a basis for exclusion. Thus, our final sample consisted of 747 respondents, of which 208 completed CDI I: “Words and Gestures” and 539 completed CDI II: “Words and Sentences.” The distribution of children by age and gender is indicated in (1):

(1)

a. [Table 1 CDI I: “Words and Gestures”]

b. [Table 2 CDI II: “Words and Sentences”]

The analysis of the educational background information revealed that our respondent sample is considerably skewed in the direction of higher parental education, both for fathers and mothers (see Figure 1 and 2). Approximately 74% of mothers and 46% of fathers in our sample have attained at least an undergraduate degree. In comparison, only 36% of women and 20% of men aged between 20 and 54 have higher education (including those holding master’s and doctoral degrees) in the overall population.

[Figure 1: Distribution of women by education level]

[Figure 2: Distribution of men by education level]

Skewedness of normative CDI samples in the direction of higher parental education has been previously reported for other CDI studies as well (see Kristoffersen et al. 2013 for Norwegian; see Bleses et al. 2008 for Danish; see Fenson et al. 2007 for American English). In the Danish normative sample, for instance, the proportion of respondents with university education (including those holding advanced degrees)

was 46%, compared to 24% in the child family population of Denmark. Similarly, in the Norwegian CDI normative sample, the parents holding at least an undergraduate degree were also overrepresented (66% vs. 53% in the child family population of Norway). As will be discussed in Section 4 below, however, the statistical analysis revealed no correlation between maternal education and child performance on vocabulary measures in either CDI I or CDI II.¹ The only measure analyzed here that positively correlated with maternal education was the overall gestural score in CDI I, and even there the effect was only marginally significant ($p=0.0859$).

2.5. Latvian CDI vocabulary checklists

2.5.1. Adaptation

The adaptation of CDI I: “Words and Gestures” and CDI II: “Words and Sentences” for the Latvian language was completed in 2016 as a part of the LAMBA project (see Vulāne, Urek, and Markus 2016). When the adaptation process was taking place, the Latvian Child Speech Corpus (LCSC) (Auzina et al. 2016; see below) was still under construction, so the available longitudinal diary studies (Rūķe-Draviņa 1992, 1993; Markus 2003, 2007) and other observational studies (Dzintere, Stangaine, and Augstkalne 2014; Tauriņa 2015; Vulāne 2008, 2015; Freidenfelds, Lapāne, and Markus 2009) were used as the main resource.

On the one hand, the goal of the adaptation was to keep the vocabulary checklists as close as possible to the American CDI in order to facilitate cross-linguistic comparison. At the same time, however, care was taken to include the items that reflect the daily reality of Latvian children. For this reason, Latvian CDI I and CDI II differ from the American versions both in the composition of checklists and, in some sections, also in the number of items included under different categories.

Vocabulary sections like “Games and Routines,” “Food and drink,” and “Outside things” were the ones containing the largest number of culture-specific items. To give just a few examples, under “Food and drink” Latvian CDI includes nouns like *biezpiens* “quark,” *krējums* “sour cream,” *putra* “porridge,” *sieriņš* “milk curd snack,” and does not have items like *cheerios*, *applesauce*, *jello*, and *peanut butter*. Similarly, “outdoor things” in the Latvian version of the CDI includes *čiekurs* “conifer cone,” *dobe* “planting bed,” *kastanis* “chestnut,” *ozolzīle* “acorn,” but not *lawn mower*, *sprinkler*, etc. Adjustments also had to be made in cases where distinct lexical categories and/or constructions are used to express a given meaning in the two languages. The most prototypical example of such lack of correspondence is “hungry” and “thirsty,” frequent English adjectives included in the American CDI. While Latvian has semantically equivalent forms – the past participles *izsalcis* and *izslāpis* – these are not typically used in children’s directed speech. Instead, the most natural way of formulating a question directed at a young child is *Gribi ēst/dzert?* “Do you want to eat/drink?” For this reason, the decision was made not to include the equivalent items in the Latvian CDI. Just like in other CDI adaptations, in the case of synonyms of comparable frequency both items were included into the checklist, e.g. *rotaļlieta/mantiņa* “toy,” *bučot/skūpstīt* “to kiss.” In some synonym pairs, one item belonged to the literary norm, while the other one was colloquial, e.g. *šorti/īsbikses* “shorts.” In some cases, the same approach was also used in the “sound effects and animal sounds” section, e.g. *ai/au/vai/auč* “ouch,” *br/brm/brum* “sound that a car makes.”

(2) [Table 3: Vocabulary section in Latvian, American, Norwegian and Russian CDI]

2.5.2. Validity

In order to see to what extent the lexical items included in the Latvian CDI forms are representative of Latvian children’s vocabulary, we compared them with nouns, verbs, and adjectives appearing in the Latvian Child Speech Corpus – a longitudinal orthographically and morphologically annotated corpus collected as part of the LAMBA (“Latvian language in monolingual and bilingual acquisition”) project (Auzina et al. 2016). LCSC contains approximately 190 hours of spoken parent-child interaction and includes data collected over the course of 16 months from three monolingual Latvian-speaking children and one bilingual Latvian-Russian speaking child. All monolingual children were recorded weekly for 30 minutes, while the bilingual child had two half-hour weekly sessions, one in each of his languages. At the time when the corpus collection started, the youngest participant was aged 17 months, and at the end of the data collection the oldest participant was 48 months old. Therefore, the children in LCSC are representative of the target population of CDI II: “Words and Sentences.”

For the comparison, we extracted all noun, verb, and adjective lemmas produced by LCSC children in the period between 17 and 36 months. Due to the rich derivational morphology of Latvian, certain adjustments were made to the resulting list in order to facilitate the comparison. For nouns and adjectives, diminutives and citation forms were counted as one lemma, and their frequencies were combined. For example, *draugs* “friend” and *draudziņš* “friend” (diminutive), or *mazs* “small” and *maziņš* “small” (diminutive) were counted as one item. For verbs, only stems, but not derivational prefixes, were considered where the meaning of the prefixed form was compositional. For instance, *ie-lik* “put in,” *at-lik* “put aside,” and *lik* “put” were counted as the same item, while *at-teikt* “refuse” and *teikt* “say” were not.

The final list consisted of 597 items, of which approximately 58% were nouns, 31% were verbs, and the remaining were adjectives (see Table 4). These items were then compared to the content words included in the vocabulary section of CDI II. Of all the content words produced by LCSC children, approximately 47% were also included in CDI II, while of the words produced more than once, 59.6% were included. Out of the lemmas that appeared at least five times in the child corpus, 77.7% were also included in the Latvian CDI II. In other words, only about one-fifth of the frequently used content words were not represented in the CDI vocabulary section. This result is comparable to the Norwegian and Danish CDI validity reports, where the correspondence between CDI items and frequent items in child corpus data was estimated at 78.6% and 91.2% respectively (Kristoffersen et al. 2013, Bleses et al. 2008). Note, however, that both of these studies also included function words in the validity analysis.

(3) [Table 4: Content words in Latvian CDI II and LCSC]

2.5.3. *Analysis*

The distribution of different CDI scores across age groups is presented for selected quantiles (0.1, 0.25, 0.5, 0.75, 0.9), fitted using quantile regression with monotonicity restrictions (gcrq function in `quantregGrowth` package in R, Muggeo 2013, Muggeo et al. 2013). Only the forms where at least one question has been answered in each subsection were included in the analysis of the respective sections. The quantile scores reported in the text are rounded to whole numbers. Raw data for languages other than Latvian was downloaded from CDI Wordbank (Frank et al. 2017; see Simonsen et al. 2014 for Norwegian; Vershinina et al. 2011, and Eliseeva and

Vershinina 2009 for Russian; Fenson et al. 2007 for American English). In order to examine the effects of age and other demographic variables on the scores attained in different sections, two-way ANOVAs were run on the individuals' raw scores in "Words and Gestures" and "Words and Sentences" samples in R. To examine the effect of language (Latvian, Russian, Norwegian, English) on different scores, multiple linear regressions were performed. Results for boys and girls are combined in all samples.

3. Results

In this section, we present the results of Latvian CDI I: "Words and Gestures" and CDI II: "Words and Sentences," and compare them to the results of Norwegian, Russian and American English CDI.

3.1. CDI I: "Words and Gestures"

In this section, we examine the results of vocabulary and gesture measures of the Latvian CDI I: "Words and Gestures" and compare some selected scores with the results of CDI I obtained in Norwegian, Russian, and American English norming studies.

3.1.1. CDI I: Latvian results

Figure 3 in (4) illustrates the distribution of receptive vocabulary scores by age for the selected quantiles in the Latvian CDI I. Here, receptive vocabulary scores are the combined total number of responses "understands" and "understands and uses" in the vocabulary section of the Latvian CDI I. As evident from Figure 3, the distribution of receptive vocabulary scores is characterized by a steady growth over age, with the

median score increasing from 39 words at the age of eight months to 209 words at the age of 16 months. Quantile curves above 0.25 indicate a rapid growth in receptive vocabulary size between 9 and 11 months. For the median score, the increase amounts to 65 words: from 58 at the age of nine months to 123 at the age of 11 months. Another period of rapid growth occurs for all quantiles below 0.9 between 12 and 14 months, when the median score increases from 124 to 207 words understood. Although the increase with age in the total number of words understood can be seen for all quantiles, the total number of words added to the passive vocabulary in the span from 8 to 16 months varies from 108 in the 0.1 quantile to 211 in 0.9 quantile. The ANOVA analysis with age, maternal education, and weekly exposure to other languages as independent variables revealed a highly significant main effect of age ($F(8, 128)=8.46, p=3.45e-09$), but no significant effect of either language exposure ($p=0.586$) or maternal education ($p=0.856$).

(4) [Figure 3: WG: receptive vocabulary size by age]

Figure 4 in (5) illustrates the distribution of Latvian productive vocabulary scores by age for children aged between 8 and 16 months. Productive vocabulary scores are calculated as the total number of “understands and uses” responses in the Latvian CDI I. As in the case of the receptive vocabulary scores reported in Figure 3, the trend of increase with age is also evident in (8). As expected, however, productive vocabulary scores are considerably lower than receptive vocabulary scores for all quantiles. Thus, the median productive vocabulary score grows from 0 words produced at eight months to 23 words at 16 months, while the score in the 0.9 quantile reaches only four words at eight months, and 123 words in the oldest group tested. The ANOVA analysis with age, maternal education, and weekly exposure to other languages as

independent variables revealed a highly significant main effect of age ($F(8,128)=2.696$, $p=0.00896$) and a marginally significant effect of language exposure ($F(1,128)=3.562$, $p=0.0614$), but no effect of maternal education ($p=0.14$).

(5) [Figure 4: WG: productive vocabulary size by age]

Finally, let us consider the distribution of gestural scores for children aged 8 to 16 months. Here, the “gestural score” is the sum of all positive responses (i.e. “sometimes” and “often” responses taken together) across all gestures and actions measures of CDI I: “Words and Gestures.” In Latvian CDI I: “Words and Gestures” the maximal achievable gestural score is 68 gestures/actions (Figure 5). We can observe a steady increase with age in the number of used gestures in all quantiles, with the period of stability starting around 13-14 months. Interestingly, even in the oldest group tested children are not yet at the ceiling (with 0.9 quantile score being 57). ANOVA analysis revealed a highly significant effect of age ($F(8,159)=32.06$, $p<2e-16$) and a marginally significant effect of maternal education ($F(7,159)=1.825$, $p=0.0859$), but no effect of exposure to languages other than Latvian. Unfortunately, no data for a cross-linguistic comparison is available to us at this point.

(6) [Figure 5: WG: gesture score by age]

3.1.2. CDI I: Cross-linguistic comparison

Let us now compare the results of the Latvian CDI I with the results obtained by American, Norwegian, and Russian questionnaires. The figure in (7) illustrates the distribution of median scores over age for Latvian, Russian, American, and

Norwegian CDI: “Words and Gestures.” What is immediately apparent on Figure 6 is that Russian and Latvian on the one hand and Norwegian and English on the other hand appear to be very close together. While it is tempting to attribute this pattern to the typological similarity between the languages in the respective pairs, it has to be noted that the tendency does not hold for all measures analyzed here. In order to compare the receptive vocabulary scores of children aged between 8 and 16 months across languages, we ran a multiple linear regression with “language” and “age group” as independent categorical variables. As expected, the model revealed that, controlling for age, Latvian participants score significantly higher than both American participants ($\beta=-43.74$ (SE=6.19), $t=-7.066$, $p=1.94e-12$) and Norwegian participants ($\beta=-63.15$ (SE=6.01), $t=-10.5$, $p<2e-16$), while the difference in performance between Latvian and Russian children does not reach statistical significance ($p=0.332$).

(7) [Figure 6: WG: median receptive vocabulary score by age and language]

Relatively high median scores in Latvian and Russian, especially in the youngest age group, call for some further examination. At least in the case of Russian, high receptive vocabulary scores might be attributable to a higher number of sound imitations and animal sounds included in the CDI I vocabulary section – 37 items in Russian, versus 14 in Latvian, 12 in American, and 11 in Norwegian. Indeed, the examination of per item scores in the 8-months-olds group (data downloaded from CDI WordBank, Frank et al. 2017; see Simonsen et al. 2014 for Norwegian, Vershinina et al. 2011, and Eliseeva et al. 2009 for Russian, Fenson et al. 2007 for American English) revealed a slight over-representation of sound imitations and animal sounds among the 50 items most frequently marked as “understands” or

“understands and uses” in Russian compared to other languages (see (8)). However, no such pattern was found for Latvian. After excluding sound imitations and animal sounds from the vocabulary score calculation, median receptive vocabulary scores at 8 months dropped for all languages (from 41 to 33 for Russian, from 39 to 37 for Latvian, from 10 to 9 for Norwegian, and from 17 to 16 for American English); however, Latvian and Russian participants still scored considerably higher than Norwegian and American ones.

(8) [Table 5: Sound imitations in CDI I and in 50 most frequently used vocabulary items]

Figure 7 plots median productive vocabulary scores in the four languages against age. As expected, median scores in all four languages start at 0 at the age of 8 months, and gradually grow as age increases. Interestingly, median scores in the oldest age group tested are very close for Latvian, Russian, and Norwegian (23, 24, and 23 words respectively), while American participants have a median productive vocabulary size of 40 words at the same age. In order to investigate whether productive vocabulary scores attained by children in the “Words and Gestures” form differ across languages, a multiple linear regression with “language” and “age group” as categorical predictors was performed. The model revealed that, when age is controlled for, Latvian participants score significantly higher than both Norwegian ($\beta=-7.4981$ (SE=2.67), $t=-2.803$, $p=0.0051$) and Russian participants ($\beta=-5.9$ (SE=2.84), $t=-2.07$, $p=0.0379$). No difference in performance, however, was found between Latvian and American children ($p=0.4814$).

(9) [Figure 7: WG: median productive vocabulary score by age and language]

3.2. *CDI II: “Words and Sentences”*

In this section, we will consider the results of the vocabulary sub-section of CDI II: “Words and Sentences.”

3.2.1. *CDI II: Latvian results*

Let us first examine the results obtained with the Latvian version of CDI I. Figure 8 illustrates the distribution of the productive vocabulary scores in Latvian children aged between 17 and 36 months for the selected quantiles. Productive vocabulary scores are calculated as the total number of words that parents marked as “produces” in the vocabulary section of CDI II. The growth over age trend is evident for all quantiles. The median score increases from 32 words at the age of 17 months to 633 words at 36 months. The differences among the quantiles are large both in the youngest and in the oldest group tested. Thus, at the age of 17 months, the scores vary from 7 words for children in the 0.1 quantile to 160 words for the children in the 0.9 quantile, while at the age of 36 months their productive vocabularies include 212 and 712 words respectively. The ANOVA with age, maternal education, and weekly exposure to other languages as independent variables revealed a highly significant main effect of age ($F(19,382)=17.794, p<2e-16$), while the effect of other predictors did not reach statistical significance.

(10) [Figure 8: WS: active vocabulary size by age]

3.2.2. *CDI II: Cross-linguistic comparison*

Finally, let us turn to the cross-linguistic comparison. The figure in (11) shows the distribution of median productive vocabulary scores in Latvian, Russian, American, and Norwegian CDI: “Words and Sentences” for children aged between 17 and 36 months (17 to 30 months for American English). Although the median scores for the four languages appear to be very close together, a multiple linear regression revealed that, when age is controlled for, Latvian children score overall lower than American ($\beta=32.32$ (SE=8.558), $t=3.776$, $p=0.00016$) and Norwegian ($\beta=41.702$ (SE=7.8), $t=5.346$, $p=9.26e-08$) children on this measure, while no difference in performance between Latvian and Russian children was found ($p=0.29$). Just as in the case of receptive vocabulary in CDI: “Words and Gestures,” Latvian and Russian participants pattern together. While receptive vocabulary scores in CDI: “Words and Gestures” were higher for Latvian and Russian participants than for American and Norwegian ones, however, for the productive vocabulary scores in CDI: “Words and Sentences” the pattern is reversed. Interestingly, while Latvian children got overall higher productive vocabulary scores than Norwegian children in CDI: “Words and Gestures,” in CDI: “Words and Sentences” the Norwegian participants significantly outperform the Latvian participants. It remains to be seen whether this apparent lack of continuity reflects an actual characteristic of the population or is an artifact of the sampling process. With respect to the former, it might be the case that Latvian children indeed start out by acquiring vocabulary more rapidly than Norwegian children, and then slow down at some older age. The relatively small sample that Latvian CDI scores are based on in combination with a high degree of individual variation characteristic of young children mean, however, that sampling error cannot be dismissed as a possible explanation for the observed cross-linguistic differences across measures.

(11) [Figure 9: WS: median vocabulary score by age and language]

To sum up, the analysis of the results of the Latvian CDI I and II revealed that performance on all measures analyzed improves steadily with age in all quantiles. For CDI I, it was shown that productive vocabulary scores are significantly lower than receptive vocabulary scores overall. A marginally significant effect of L2 exposure on productive vocabulary has been found for CDI I, while maternal education was not found to affect the performance on any of the measures analyzed. As for the cross-linguistic comparison, it was found that Latvian children score significantly higher than American and Norwegian peers on the receptive vocabulary, even after sound imitations and animal sounds were excluded. The tendency was not preserved in the productive vocabulary measure, where there was no difference between Latvian and American CDI I scores, while Latvian children scored higher than both Norwegian and Russian peers. In CDI II, Latvian children scored lower than both American and Norwegian peers on the productive vocabulary measure.

4. Conclusion and Discussion

In this article, we have presented the adaptation of CDI I: “Words and Gestures” and CDI II: “Words and Sentences” into Latvian, reported the results of the norming study of the Latvian CDI and compared those to the results obtained for American English, Norwegian, and Russian CDI.

The main challenges of the Latvian CDI study were encountered during the norming stage. While the Norwegian CDI normative study using the analogous internet-based procedure enjoyed the response rate of approximately 37%, in our

study only about 8% of the targeted families completed the questionnaires, which necessitated a second round of data collection. It has to be noted that the difference in response rates is not commensurate with the difference in internet accessibility in the two countries. In 2008, i.e. the year when data collection for the Norwegian CDI norming was conducted, the overall proportion of households with internet access in Norway was 84% (Eurostat; see also Kristoffersen et al. 2013). In 2016, when the data collection for Latvian CDI took place, the same indicator for Latvia was at 77% (Eurostat). Besides, the proportion of active internet users is higher in our target age group. Thus, according to the data of the Central Statistical Bureau of Latvia, 96.7% of respondents aged between 25-34 years and 87.8% of respondents aged between 35-44 years reported using internet at least once a week. In the hopes of collecting a larger sample over time, we decided to keep the Latvian CDI forms available online at www.lamba.lv, where all those interested can request an access code and fill out the forms. Low response rates contributed to the skewedness of the final sample in the direction of higher parental education, which was somewhat larger in the Latvian sample than what was reported for Norwegian, Danish, and American English. A marginally significant effect of parental education, however, has only been found on gestural scores in CDI II in the present sample. Nevertheless, it is possible that the effect of parental education would be more evident in a larger sample where parents with lower education level are proportionally represented.

Another challenge at the norming stage was a high proportion of bilinguals in the population, which led to high exclusion rates. In the population census of 2011, for example, 34% of all respondents reported that Russian is the main language they use at home (vs. 56% who named Latvian). In Riga, the proportion is even larger: 49% of respondents named Russian as their primary language, while only 38% said

they mainly spoke Latvian at home (data provided by the Central Statistical Bureau). While the normative sample only included the children with weekly exposure to other languages not exceeding 24 hours a week, a marginally significant effect of language exposure ($F(1, 128)=3.562, p=0.0614$) on the receptive vocabulary scores has been found in CDI I: “Words and Gestures.”

As expected, the analysis of the CDI results showed considerable variation among children, which, however, was comparable to that observed for other languages. Interestingly, Latvian CDI scores very closely corresponded to those found in the Russian normative sample when it comes to the receptive vocabulary scores in the span between 8 and 16 months, and productive vocabulary scores in the span between 17 and 36 months. While it is tempting to attribute this to the typological similarity between the two languages, the tendency does not hold for the productive vocabulary scores in CDI I: “Words and Gestures” population, where Latvian and American children pattern together.

Being the first instrument of its kind normed with the Latvian child population, Latvian CDI represents a useful resource for both researchers and practitioners, and provides a good starting point for further inquiry into the communicative development of Latvian-speaking children. The study has however, important limitations that need to be kept in mind. First, the study is based on a relatively limited sample of children. Second, most children in our sample come from urban areas. Third, the sample is skewed in the direction of higher parental education. This means that the scores should be interpreted with caution when used for clinical purposes, as it still remains to be seen how Latvian CDI fares with respect to child populations with different socioeconomic status, and the effect of societal

bilingualism on communicative development of Latvian-speaking children needs to be examined more closely.

Acknowledgements

The research leading to these results has received funding from the Norwegian Financial Mechanism 2009-2014 under Project Contract No NFI/R/2014/053. We also gratefully acknowledge the funding from the Research Council of Norway (Centres of Excellence funding scheme, project no.223265).

References

- Auzina, I., K. Levane-Petrova, G. Rabante-Busa, R. Dargis, and A. Fabregas. 2016. "Designing an Annotated Longitudinal Latvian Children's Speech Corpus." In *Human language technologies—the Baltic perspective: Proceedings of the seventh international conference Baltic HLT 2016*. Amsterdam: IOS Press.
- Berglund, E., Eriksson, M. & Johansson, I. 2001. "Parental reports of spoken language skills in children with Down syndrome". *Journal of Speech, Language, and Hearing Research* 44: 179–191.
- Bialystok E, Luk G, Peets KF, Yang S. 2010. "Receptive vocabulary differences in monolingual and bilingual children". *Bilingualism: Language and Cognition* 13: 525–531.
- Bleses, D., W. Vach, M. Slott, S. Wehberg, P. Thomsen, T. Madsen, and H. Basbøll. 2008. "Early vocabulary development in Danish and other languages: A CDI-based comparison." *Journal of Child Language* 35: 619–50.

- Central Statistical Bureau of Latvia. 2011. Population Census 2011. <<https://www.csb.gov.lv/lv/statistika/statistikas-temas/iedzivotaji/tautas-skaitisana>>
- Charman, T., Drew, A., Baird, C., & Baird, G. 2003. "Measuring early language development in pre-school children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form)". *Journal of Child Language* 30: 213–236.
- Conboy, B. T, Thal D. J. 2006. "Ties between the lexicon and grammar: Cross sectional and longitudinal studies of bilingual toddlers". *Child Development* 77: 712–735.
- De Houwer, A., Bornstein, M. H. & De Coster, S. 2006. Early understanding of two words for the same thing: A CDI study of lexical comprehension in infant bilinguals. *International Journal of Bilingualism* 10: 331–347.
- Dzintere, D., I. Stangaine, and D. Augstkalne. 2014. *Bērna komunikatīvās kompetences attīstība*. Riga: RaKa.
- Eliseeva, M. B. and E. A. Vershinina. 2009. *Nekotorye normativy rechevogo razvitiya detei ot 8 do 18 mesyatsev (po materialam MakArturovskogo oprosnika)*. *Materialy mezhdunarodnoi konferentsii 17-19 iyunya 2009 g.* St. Petersburg: "Yazyki slavyanskoi kul'tury."
- Eurostat. 2007-2016. Households with internet access (as % of all households). <http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_ci_in_h&lang=en>
- Feldman HM, Dollaghan CA, Campbell TF, Colborn DK, Janosky J, Kurs-Lasky M, et al. 2003. Parent-reported language skills in relation to otitis media during

- the first 3 years of life. *Journal of Speech Language, & Hearing Research* 46: 273–287.
- Fenson, L., V. A., Marchman, D. J. Thal, P. S. Dale, J. S., Reznick, and E. Bates. 2007. *MacArthur-Bates communicative development inventories: User's guide and technical manual*. 2nd ed. Baltimore: Brookes Publishing.
- Frank, M. C., M. Braginsky, D. Yurovsky, and V. A. Marchman. 2017. “Wordbank: An open repository for developmental vocabulary data.” *Journal of Child Language* 44 (3): 677–94. doi: 10.1017/S0305000916000209
- Freidenfelds, I. J., D. Lapāne, and D. Markus. 2009. *Bērnu valodas vārdnīca*. Riga: Izglītības soli.
- Gatt, D., C. O’Toole, and E. Haman. 2015. “Using parental report to assess early lexical production in children exposed to more than one.” In *Assessing multilingual children: Disentangling bilingualism from language impairment*, edited by S. Armon-Lotem, J. de Jong, and N. Meir. 149–93. Bristol: Multilingual Matters.
- Heilmann, J., Weismer, S. E., Evans, J., Hollar, C. 2005. “Utility of the MacArthur Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers.” *American Journal of Speech-Language Pathology* 14: 40–51.
- Junker, D., & Stockman, I. 2002. “Expressive vocabulary of German-English bilingual toddlers”. *American Journal of Speech-Language Pathology* 11: 381 – 394.
- Koster, C., Been, P., Krikhaar, E., Zwarts, F., Diepstra, H., & van Leeuwen, T. 2005. “Differences at 17 months: productive language patterns in infants at familial

- risk for dyslexia and typically developing infants”. *Journal of Speech, Language, and Hearing research* 48: 426–438.
- Kristoffersen, K. E., and H. G. Simonsen. 2012. *Tidlig språkutvikling hos norske barn: MacArthur-Bates foreldrerapport for kommunikatív utvikling*. Oslo: Novus.
- Kristoffersen, K. E., H. G. Simonsen, D. Bleses, S. Wehberg, R. N. Jørgensen, E. A. Eiesland, and L. Y. Henriksen. 2013. “The use of the Internet in collecting CDI data—an example from Norway.” *Journal of Child Language* 40 (3): 567–85.
- Law, R. & Roy, P. 2008. “Parental report of infant language skills: A review of the development and application of the Communicative Developmental Inventories”. *Child and Adolescent Mental Health* 13: 198–206.
- Magill-Evans, J. & Harrison, M. J. 1999. “Parent-child interactions and development of toddlers born preterm”. *Western Journal of Nursing Research* 21: 292–307.
- Markus, D. 2003. *Bērna valoda: No pirmā kļiedziena līdz pasakai*. Riga: Rasa ABC.
- Markus, D. 2007. *Bērnus runā kultūras pasaulē*. Riga: Rasa ABC.
- Muggeo, V. 2013. quantregGrowth: Growth charts via regression quantiles. R package version 0.1-1.
<<http://cran.rproject.org/web/packages/quantregGrowth/>>
- Muggeo, V., M. Sciandra, A. Tomasello, S. Calvo. 2013. “Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology”. *Environmental and Ecological Statistics*: 20: 519–531
- O’Toole, C., D. Gatt, T. M. Hickey, A. Miękisz, E. Haman, S. Armon-Lotem, and S. Kern. 2017. “Parent report of early lexical production in bilingual children: A

- cross-linguistic CDI comparison.” *International Journal of Bilingual Education and Bilingualism* 20 (2): 124–45.
- Pearson, B., Fernandez, S., Lewedeg, V., Oller, D. K. 1997. “The relation of input factors to lexical learning by bilingual infants”. *Applied Psycholinguistics* 18: 41–58.
- Pearson, B. Z., Fernandez, S. C , & Oiler, D. K. 1993. “Lexical development in bilingual infants and toddlers: Comparison to monolingual norms”. *Language Learning* 43: 93- 120.
- Poulin-Dubois D., Bialystok E., Blaye A., Polonia A., & Yott J. 2013. “Lexical access and vocabulary development in very young bilinguals”. *International Journal of Bilingualism* 17: 57–70.
- Rūķe-Draviņa, V. 1992. *No 5 mēnešiem līdz 5 gadiem*. 2nd ed. Riga: Zvaigzne.
- Rūķe-Draviņa, V. 1993. *Latviešu meitene apgūst savu pirmo valodu*. Riga: Dio Noordik.
- Simonsen, H. G., K. E. Kristoffersen, D. Bleses, S. Wehberg, and R. N. Jørgensen. 2014. “The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences.” *First Language* 34 (1): 3–23. doi: 10.1177/0142723713510997
- Tauriņa, A. 2015. “Latviešu bērnu izmantotās uzrunas formas multikulturālā vidē.” In A. Vulāne and D. Markus, eds. *Bērnu valoda Latvijā 21. gadsimtā*, 223–46. Riga: Zinātne.
- Thal, D. J., DesJardin, J., Eisenberg, L. 2007. “Validity of the MacArthur-Bates Communicative Development Inventories for measuring language abilities in children with cochlear implants”. *American Journal of Speech-Language Pathology* 16: 54–64.

- Thal, D. J., O'Hanlon, L., Clemmons, M., & Fralin, L. 1999. "Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment". *Journal of Speech, Language, and Hearing Research* 42: 482–496.
- Vershinina, E. A., M. B. Eliseeva, T. S. Lavrova, V. L. Ryskina, S. N. Tseitlin. 2011. *Nekotorye normativy rechevogo razvitiya detei ot 8 do 18 mesyatsev. Spetsial'noe obrazovanie: traditsii I innovatsii. Sbornik nauchno-metodicheskikh trudov s mezhdunarodnym uchastiem*. St. Petersburg: Rossiiskii gosudarstvennyi pedagogicheskii universitet.
- Vulāne, A. 2008. "Lietvārdu skaitļa kategorijas semantika mūsdienu latviešu valodā: problēmas pieteikums." In *Letonikas otrais kongress*, edited by B. Bušmane, 165–75. Riga: LU Latviešu valodas institūts.
- Vulāne, A. 2015. *Vecvecāku nosaukumi latviešu bērnu spontānajā runā*. In *Bērnu valoda 21. gadsimtā*, edited by A. Vulāne and D. Markus, 297–319. Riga: Zinātne.
- Vulāne, A., O. Urek, and D. Markus. 2016. "Makartura-Beitas komunikatīvās attīstības testa (KAT) piemērošana Latviešu valodai." In *Proceedings of the International Scientific Conference Society. Integration. Education* (2), 487–502. Rēzekne: Rēzeknes Tehnoloģiju akadēmija.

Notes

¹ As noted by one reviewer, the lack of correlation between maternal education and child performance on Latvian CDI is likely due to the fact that the normative sample is highly skewed in the direction of higher parental education.