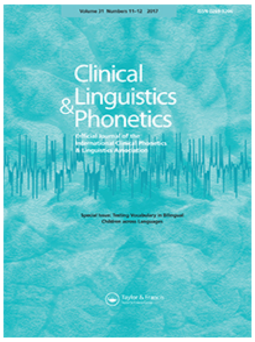


Clinical Linguistics & Phonetics

ISSN: 0269-9206 (Print) 1464-5076 (Online) Journal homepage: https://www.tandfonline.com/loi/iclp20

# Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT)

**Ewa Haman, Magdalena Łuniewska, Pernille Hansen, Hanne Gram Simonsen, Shula Chiat, Jovana Bjekić, Agnė Blažienė, Katarzyna Chyl, Ineta Dabašinskienė, Pascale Engel de Abreu, Natalia Gagarina, Anna Gavarró, Gisela Håkansson, Efrat Harel, Elisabeth Holm, Svetlana Kapalková, Sari Kunnari, Chiara Levorato, Josefin Lindgren, Karolina Mieszkowska, Laia Montes Salarich, Anneke Potgieter, Ingeborg Ribu, Natalia Ringblom, Tanja Rinker, Maja Roch, Daniela Slančová, Frenette Southwood, Roberta Tedeschi, Aylin Müge Tuncer, Özlem Ünal-Logacev, Jasmina Vuksanović & Sharon Armon-Lotem**

**To cite this article:** Ewa Haman, Magdalena Łuniewska, Pernille Hansen, Hanne Gram Simonsen, Shula Chiat, Jovana Bjekić, Agnė Blažienė, Katarzyna Chyl, Ineta Dabašinskienė, Pascale Engel de Abreu, Natalia Gagarina, Anna Gavarró, Gisela Håkansson, Efrat Harel, Elisabeth Holm, Svetlana Kapalková, Sari Kunnari, Chiara Levorato, Josefin Lindgren, Karolina Mieszkowska, Laia Montes Salarich, Anneke Potgieter, Ingeborg Ribu, Natalia Ringblom, Tanja Rinker, Maja Roch, Daniela Slančová, Frenette Southwood, Roberta Tedeschi, Aylin Müge Tuncer, Özlem Ünal-Logacev, Jasmina Vuksanović & Sharon Armon-Lotem (2017) Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT), Clinical Linguistics & Phonetics, 31:11-12, 818-843, DOI: 10.1080/02699206.2017.1308553

**To link to this article:** https://doi.org/10.1080/02699206.2017.1308553



Published online: 25 Apr 2017.

CLINICAL LINGUISTICS & PHONETICS
2017, VOL. 31, NOS. 11–12, 818–843
http://dx.doi.org/10.1080/02699206.2017.1308553



# Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT)

Ewa Haman[a], Magdalena Łuniewska[a], Pernille Hansen[b], Hanne Gram Simonsen[b], Shula Chiat[c], Jovana Bjekić[d], Agnė Blažienė[e], Katarzyna Chyl[a], Ineta Dabašinskienė[e], Pascale Engel de Abreu[f], Natalia Gagarina[g], Anna Gavarró[h], Gisela Håkansson[i], Efrat Harel[j], Elisabeth Holm[b], Svetlana Kapalková[k], Sari Kunnari[l], Chiara Levorato[m], Josefin Lindgren[n], Karolina Mieszkowska[a], Laia Montes Salarich[h], Anneke Potgieter[o], Ingeborg Ribu[b], Natalia Ringblom[p], Tanja Rinker[q], Maja Roch[m], Daniela Slančová[r], Frenette Southwood[o], Roberta Tedeschi[a], Aylin Müge Tuncer[s], Özlem Ünal-Logacev[t], Jasmina Vuksanović[d] and Sharon Armon-Lotem[u]

[a]Faculty of Psychology, University of Warsaw, Warsaw, Poland; [b]MultiLing, University of Oslo, Oslo, Norway; [c]City University London, London, UK; [d]Institute for Medical Research, University of Belgrade, Serbia; [e]Vytautas Magnus University, Kaunas, Lithuania; [f]Language and Cognitive Development Group, University of Luxembourg, Luxembourg; [g]Research Area Language Development and Multilingualism (FB II), Leibniz-ZAS Berlin, Berlin, Germany; [h]Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain; [i]Lund University, Lund, Sweden; [j]Kibbutzim College of Education, Technology and Arts, Tel-Aviv, Israel; [k]Comenius University in Bratislava, Bratislava, Slovakia; [l]Research Unit of Logopedics, University of Oulu, Oulu, Finland; [m]University of Padua, Padua, Italy; [n]Uppsala University, Uppsala, Sweden; [o]Department of General Linguistics, Stellenbosch University, Stellenbosch, South Africa; [p]Stockholm University, Stockholm, Sweden; [q]Department of Linguistics, University of Konstanz, Konstanz, Germany; [r]Faculty of Arts, Prešov University, Prešov, Slovakia; [s]Health Sciences Faculty, Health Sciences Faculty, Anadolu University, Eskişehir, Turkey; [t]School of Health Science, Istanbul Medipol University, Istanbul, Turkey; [u]Bar-Ilan University, Ramat-Gan, Israel

**ABSTRACT**

This article investigates the cross-linguistic comparability of the newly developed lexical assessment tool Cross-linguistic Lexical Tasks (LITMUS-CLT). LITMUS-CLT is a part the Language Impairment Testing in Multilingual Settings (LITMUS) battery (Armon-Lotem, de Jong & Meir, 2015). Here we analyse results on receptive and expressive word knowledge tasks for nouns and verbs across 17 languages from eight different language families: Baltic (Lithuanian), Bantu (isiXhosa), Finnic (Finnish), Germanic (Afrikaans, British English, South African English, German, Luxembourgish, Norwegian, Swedish), Romance (Catalan, Italian), Semitic (Hebrew), Slavic (Polish, Serbian, Slovak) and Turkic (Turkish). The participants were 639 monolingual children aged 3;0–6;11 living in 15 different countries. Differences in vocabulary size were small between 16 of the languages; but isiXhosa-speaking children knew significantly fewer words than speakers of the other languages. There was a robust effect of word class: accuracy was higher for nouns than verbs. Furthermore, comprehension was more advanced than production. Results are discussed in the context of cross-linguistic comparisons of lexical development in monolingual and bilingual populations.

**ARTICLE HISTORY**
Received 30 July 2015
Revised 25 April 2016
Accepted 26 June 2016

**KEYWORDS**
Lexical development; cross-linguistic comparison; basic word classes; word comprehension; word production

**CONTACT** Ewa Haman ewa.haman@psych.uw.edu.pl Faculty of Psychology, University of Warsaw, Stawki 5/7, 00-183 Warsaw, Poland.

Most research on bi- and multilingual children's lexical development rests on an implicit assumption that vocabulary development is similar across languages. However, it is not clear to what extent this assumption is valid. Cross-linguistic data collected from monolingual children who were tested on lexical tasks designed to be uniform across languages could be useful in suggesting what differences may be expected across languages, independently of the bi- or multilingual status of the speakers. The novel assessment tool Cross-linguistic Lexical Tasks (LITMUS-CLT, henceforth CLT; Haman, Łuniewska, & Pomiechowska, 2015) was designed within COST Action IS0804 as a response to the need for cross-linguistically and cross-culturally comparable lexical assessment tools for children. CLT is a part of the Language Impairment Testing in Multilingual Settings (LITMUS) battery (Armon-Lotem et al., 2015).

This article presents a large-scale cross-linguistic study of expressive and receptive word knowledge in monolingual children assessed using this new tool. The goal of this study was twofold: (1) to compare lexical development across languages and cultures and (2) to evaluate the assessment tool itself.

Importantly, analyses of cross-linguistic data from monolingual children may contribute crucial information for the cross-linguistic assessment of bilinguals, who should be assessed in both of their languages (Bedore & Peña, 2008). If the timing and pace of monolingual language development is not the same across languages due to the factors intrinsic to linguistic or culture characteristics, then it should not be expected that the languages of a bilingual child will develop in a fully balanced way, even if input and other external factors are levelled out. Thus, any cross-linguistic differences found with regard to the timing and pace of lexical development could be used, for instance, to inform clinical practice about whether similar levels of lexical knowledge should be expected in the different languages of bilingual children who are diagnosed with a language disorder.

In this article, we analyse data from monolingual preschool children across 17 languages. These analyses provide a background for the other studies presented in this issue that refer to specific languages, language pairs or language problems (Altman, Goldstein & Armon-Lotem, 2017; Gatt, Attard, Łuniewska & Haman, 2017; Hansen, Simonsen, Łuniewska & Haman, 2017; Kapalková & Slančová, 2017; Khoury Aouad Saliby, dos Santos, Kouba-Hreich & Messarra, 2017) as well as the recently published (Potgieter & Southwood 2016). Furthermore, we investigate similarities in measurements of children's lexical knowledge across languages. We explore whether the assumptions underlying CLT affect the scores, as well as to what extent our results correspond to previous cross-linguistic research on the lexical skills of children.

## Why do we need cross-linguistic lexical tasks?

Many studies on the lexicons of bilingual children have analysed word knowledge by considering language scores from only one of the children's languages (e.g. Bialystok, Luk, Peets, & Yang, 2010; Pearson, 2010; Umbel, Pearson, Fernández, & Oller, 1992). Studies analysing both languages of bilingual children have typically focused on one of a limited number of specific language pairs. The most commonly investigated language pair is Spanish and English, and in most of the studies, English was the children's second language (L2). A comprehensive list of previously studied language pairs is presented in Appendix 1.

Several studies of young children below the age of 3 years (Armon-Lotem & Ohana, 2017; Conboy & Thal, 2006; De Houwer, Bornstein & Putnick, 2014; Gatt, 2017; Miękisz et al., 2017; O'Toole & Hickey, 2017; O'Toole et al., 2017) have combined two language adaptations of MacArthur-Bates Communicative Development Inventories (MB-CDI) to assess children's lexical development across their languages. This has been possible due to the large number of available adaptations for this inventory: 61 language versions were mentioned in a review by Dale and Penfold (2011), and this number is increasing as new language versions are developed (e.g. Baal & Bentzen, 2014; Dar, Anwaar, Vihman, & Keren-Portnoy, 2015). Although MB-CDI is potentially a useful tool for providing a comparable assessment of both languages spoken by young bilingual children (Law & Roy, 2008), this instrument was originally developed for a monolingual context. Therefore, MB-CDIs need to be used with caution in clinical practice or research in multilingual contexts (Gatt, O'Toole, & Haman, 2015). Furthermore, MB-CDIs are designed for children aged 8–30 months (or in some cases up to 36 months) and do not cover the full preschool age range.

Instruments designed for bilingual children older than 3 years are scarce and have typically been developed for one specific population only, such as the Bilingual English Spanish Assessment (BESA) for Spanish–English-speaking Americans (Peña, Gutierrez-Clellen, Iglesias, Goldstein, & Bedore, 2014), Sprachstandstest Russisch für mehrsprachige Kinder [Russian language proficiency test for multilingual children] for Russian–German children (Gagarina, Klassert, & Topaj, 2010), the Prawf Geirfa Cymraeg [Welsh Vocabulary Test] for bilingual Welsh–English children with different home-language backgrounds (Gathercole, Thomas, & Hughes, 2008), and the Bilingual Verbal Ability Test for children acquiring American English along with one of 17 minority languages (Muñoz-Sandoval, Cummins, Alvarado, & Ruef, 2005). Going beyond specific language pairs with preschool children has been a challenge due to the lack of comparable measures. Given the variety of language combinations in bilingual and multilingual populations within Europe, Working Group 3 of the recent COST Action IS0804 (Bi-SLI; http://bi-sli.org/) aimed to construct a set of quasi-universal lexical tasks that could be freely paired within an extensive list of languages. The CLT is thus a first attempt to design such a uniform tool across languages. CLT is in the process of being normed on mono- and bilingual children, and will subsequently be applied in individual diagnosis. Preparing a normed instrument for clinical practice is a lengthy and expensive process that should be preceded by extensive research on the tool's characteristics. Here we present one of the initial steps in research on designing a tool to assess word knowledge.

## Cross-linguistic comparisons of monolingual development

Cross-linguistic comparisons of monolingual children's lexical development are important for two reasons. First, such a comparison could shed light on the tool's cross-linguistic comparability. If the CLT reveals similar results across languages for monolinguals of equal age and socio-economic background (SES), we could assume that the CLT is cross-linguistically comparable not only in terms of design, but also in terms of its relative difficulty. As a consequence, we could assume that bilingual children who scored equally in both of their languages are balanced bilinguals in terms of their lexical knowledge. If the CLT is not directly cross-linguistically comparable (does not reveal similar difficulty across languages), the tool may still be useful, but analyses across languages would then need to rely on comparisons with language-specific norms. In any case, before the CLT is used in diagnosis of both monolingual and bilingual children, norming studies for specific populations are needed.

Second, cross-linguistic comparisons of monolingual children's lexical development can be used to investigate cross-linguistic variation per se. There is strikingly little research on cross-linguistic differences in lexical development in terms of first-words onset, word-learning rate or vocabulary size. We need cross-linguistic tools to assess potential cross-linguistic variability in lexical development. However, any variability found between languages using such tools could come from either cross-linguistic variability or inherent differences in the tools. The present study contributes to a deeper understanding of these issues.

Lexical development attracts the most research attention at its earliest stages: i.e. from the first words uttered or comprehended (Bergelson & Swingley, 2012; Fenson, Dale, Reznick, & Thal, 1993). A few publications have aimed to analyse cross-linguistic similarities and differences in the composition of early lexicons across several languages (Bornstein et al., 2004; Caselli et al., 1995; Conboy & Thal, 2006; Mayor & Plunkett, 2014). However, none of these studies have directly addressed the issue of cross-linguistic differences in the exact age of use of first words. The exception is the meta-analysis by Bleses et al. (2008) involving a comparison of the use of words and vocabulary size in 18 languages assessed using MB-CDIs.

Regarding age of the use of first words, a longitudinal multi-case study of spontaneous speech from English, French, Japanese and Swedish infants ($N = 20$) showed that Japanese children ($N = 5$) produced their first word and reached the 4 and 25 word milestones about 2 months later than the other infants (de Boysson-Bardies & Vihman, 1991), while the onset times for English, French and Swedish were very similar to each other.

Results by Bornstein et al. (2004) suggest that although the composition of the vocabularies of 20-month-old children was similar across seven languages (two Germanic: Dutch, English; three Romance: French, Italian, and Spanish; and two from other families: Hebrew and Korean) in terms of the prevalence of nouns over other word classes, vocabulary size varied across the languages, although the researchers did not directly comment on this. The data they provided regarding the average scores on ELI, an earlier version of MB-CDI (Borstein et al., 2004, Table 2, page 1124), showed that Korean children had the smallest vocabularies and the smallest variation between participants, while Hebrew children had the largest vocabularies. There were no significant differences among speakers of the Romance and Germanic languages. It seems that differences are more pronounced between language families than within a family. Note that the study by Bornstein et al. involved a significantly larger sample than de Boysson-Bardies and Vihman's (1991) study ($N = 269$, ranging from $N = 28$ to 51 per language).

The only large cross-linguistic study tapping directly into data on the age of use of first words and vocabulary size in early language development was carried out by Bleses et al. (2008); it involved 14 languages,[1] based on data from over 26,000 children (with a median

[1]Existing MB-CDI data were compared for Basque, Mandarin Chinese, Croatian, Danish, Dutch, American English, British English, Finnish, French, Galician, German, Hebrew, Icelandic, Italian, European Spanish, Mexican Spanish and Swedish.

sample size of 864 children per language, and sample sizes ranging from 30 children for Chinese to 6112 for Danish). The study showed that Danish children in the age span 8–30 months knew fewer words than children acquiring most of the other languages. Bleses et al. (2008) argue that this relative lag for Danish speakers is caused by specific phonological features (phonological reductions as compared to the closely related languages Norwegian and Swedish), which renders Danish words less phonologically transparent and harder to perceive. Similarly, de Boysson-Bardies and Vihman (1991) argued that the reason for the delay in Japanese vocabulary acquisition was related to specific features of word onsets that affected the articulatory process.

One piece of research that could shed light on potential cross-linguistic differences in lexical development is Bornstein and Hendricks (2012), which assessed children's language comprehension and production in 16 countries.[2] This study used extremely short parental reports (two simple yes/no questions about whether the child could understand talk directed to her/him and whether he/she could speak at all), which was filled in by the parents of over 100,000 children aged 2–9 years. Scores for rates of production by children aged 2–5 years varied from .84 (for Sierra Leone) to .99 (for Uzbekistan). The two countries with the lowest scores also have the lowest ratings on the Human Development Index (HDI), an indicator of life chances that – according to the authors – may influence language development. However, the study did not gather any detailed information on what language(s) the children had acquired; most of the countries were multilingual, and those with the lowest HDI were highly multilingual.

So far, we have considered studies that discuss cross-linguistic similarities and differences in vocabulary size or the pace of vocabulary acquisition in monolingual children. Studies like Bleses et al. (2008) have not yet been replicated with older children (above the age of 3 years), at least partly because of the lack of adequate tools. No instrument exists to directly measure lexical knowledge in a comparable way for older children across a similar range of languages. The development of the CLT was intended to fill this gap, tapping directly into both expressive and receptive vocabulary knowledge in children above the age of 3 years.

## CLT – construction and design

The CLT consists of picture-identification and picture-naming tasks aimed at assessing the comprehension and production of nouns and verbs via four subtasks, each consisting of 32 items. Each CLT language version was developed according to the same set of criteria. Target words were selected from a common set of 299 candidate words comprised of 158 nouns and 141 verbs. The list of candidate words was drawn up on the basis of a cross-linguistic picture-naming study conducted in 34 languages with adult native speakers (Haman et al., 2015). For a word to be included in the candidate set its meaning had to be shared in most of the 34 languages. The target word selection process takes into consideration two main factors which are assumed to contribute to the difficulty of word learning and processing for children and adults: the age of acquisition of the words (AoA) (D'Amico, Devescovi, & Bates, 2001; Ellis & Morrison, 1998; Juhasz, 2005) and a

[2]Albania, Bangladesh, Belize, Bosnia and Herzegovina, Central African Republic, Ghana, Iraq, Jamaica, Macedonia, Mongolia, Montenegro, Serbia, Sierra Leone, Thailand, Uzbekistan and Yemen.

complexity index (CI) which mainly takes into account the phonological (Morrison, Ellis, & Quinlan, 1992) and morphological (Baayen, Feldman, & Schreuder, 2006) characteristics of the target words. The AoA ratings for the words were obtained through a separate study (Łuniewska & et al., 2015).[3] The CI is based on a set of linguistic features: the number of phonemes in the word, morphological features (the number of roots for compound words, whether it is a derived word, plus the number of suffixes and prefixes), phonological features (the presence of initial fricatives, an initial consonant cluster or an internal consonant cluster), whether it is a recent loanword and the subjective frequency of exposure to the word, all as judged by linguists (one expert per language), who filled in a multipart form which contained questions about all the features for individual words. The exact formula used for calculating the CI can be found in Haman et al. (2015). In the construction of the CLT for each language, for both AoA and CI two-level categories were used (for AoA: early and late; for CI: low and high).

The production subtasks contain one picture for each target word. Each item in the comprehension subtask consists of a four-picture board containing one picture for the target word and three distractor pictures; one distractor was a picture used in the production task, while the other two were selected from words that matched the comprehension targets in AoA and CI. The target words for the two subtasks were different but had been carefully matched for their AoA and CI. Thus, across the four subtasks, children were presented with pictures of the comprehension targets only once. However, pictures of the production targets were presented twice: once in the production subtask and once in the comprehension subtask (as distractors).[4] The other distractor pictures never occur twice within the tasks.

All pictures were designed exclusively for the CLT. Some appeared in several versions to take into account cultural differences. In particular, pictures for actions involving people depicted different races and genders.

Although both the AoA and CI indices reflect word characteristics which were assumed to have an effect on word learning and processing, their impact on the accuracy of performance on the CLT tasks has not previously been directly assessed. This study is the first to analyse the interaction of the AoA and CI indices in 17 languages with the scores obtained by monolingual children; Hansen at al. (2017) do this in more detail using the same set of monolingual data for Polish and Norwegian, alongside bilingual Polish–Norwegian data. Similar analysis for Hebrew using monolingual Hebrew and bilingual Russian–Hebrew data can be found in Altman et al. (2017).

Target word selection for the CLT followed the same principles across languages, but the final list of 128 target words was specific for each language version. None of the 299 candidate words was selected as a target (either for the production or comprehension subtask) in more than 14 of the 17 languages, and there were no candidate words that were never used as a target. Figure 1 shows the number of times each candidate word occurred as a target across the

[3]Norwegian AoA ratings were obtained through a connected but distinct study, as described and discussed in Lind, Simonsen, Hansen, Holm, and Mevik (2015).

[4]Note that pictures for the production subtasks are never named by the researcher during the testing procedure. Pictures for the comprehension target words are named by the researcher once in a comprehension prompt (see next section). This asymmetry was inevitable in the construction of the CLT due to the limited number of candidate words. A total of 128 pictures/words were needed in each language, chosen from a set of 299, with strict matching criteria for distractors, which made the selection quite challenging.

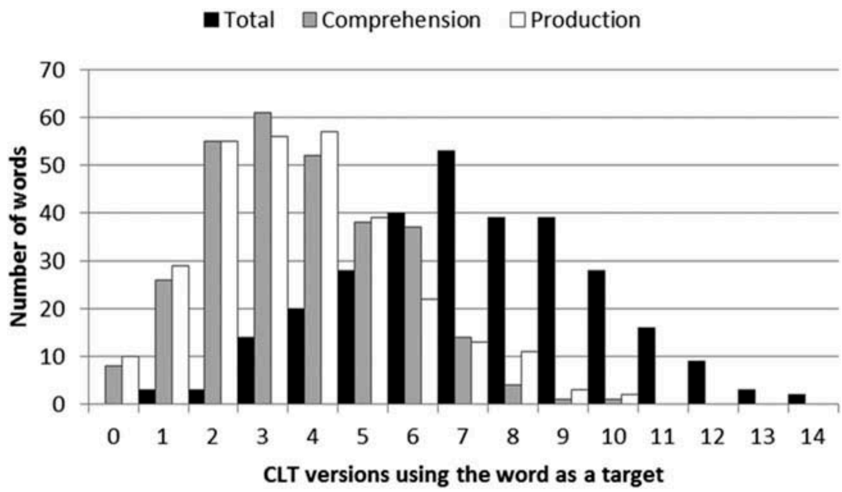


**Figure 1.** Distribution of frequency of word choice across 17 CLT language versions.

17 languages. This distribution is close to what we would expect if the selection of target words was random. Thus, all 299 words proved useful for this range of languages, which is important, as no constraints were imposed on the semantics of candidate words during the selection process. However, the AoA study conducted on the same word list showed that all these words are on average acquired between the ages of 2 and 8 years in the 25 languages studied (see Łuniewska et al., 2015). According to the estimated AoA, most of the CLT candidate words can be assumed to be acquired before the age of five, and only a few after the age of six. Thus, the CLT may be assumed to be a sensitive measure of lexical development in children within the age range involved in this study (i.e. 3–6 years), as it potentially contains target words that vary in difficulty for this age range.

## The current study

In this study, we address the issue of potential word-learning differences by children in terms of vocabulary size, lexicon composition (proportion of nouns and verbs) and receptive vs. expressive word knowledge across 17 languages from 8 language families: Baltic (Lithuanian), Bantu (isiXhosa), Finnic (Finnish), Germanic (Afrikaans, British English, South African English, German, Luxembourgish, Norwegian, Swedish), Romance (Catalan, Italian), Semitic (Hebrew), Slavic (Polish, Serbian, Slovak) and Turkic (Turkish). In view of previously published research findings, we expected participants to achieve higher accuracy on the lexical tasks for nouns than for verbs (Gentner, 1982; Gentner & Boroditsky, 2001; Tomasello & Merriman, 1995) and higher accuracy for comprehension than production (Bates & Goodman, 1999; Benedict, 1979; Clark, 2009; Fenson et al., 1994, Goldfield, 2000; Harris, Yeeles, Chasin, & Oakley, 1995; Reznick & Goldfield, 1992) across all languages studied. We also expected that the assessment would be sensitive to the participants' age, showing an increase in accuracy with age. To investigate the comparability of the different CLT versions, we also examined the potential impact on results of the language-specific background variables used in constructing the CLT, namely the AoA and CI.

We did not formulate specific hypotheses regarding overall differences in vocabulary size among the languages, since previous studies present ambiguous and incomplete results about this issue. Thus, cross-linguistic analyses concerning vocabulary size are exploratory in nature here.

The sample of languages shows some imbalance. Indo-European languages dominate the sample (13 out of the 17 languages), with half of the Indo-European group consisting of Germanic languages. Only four languages (Hebrew, Finnish, Turkish and isiXhosa) represent non-Indo-European language families. This reflects the fact that our data are drawn from the networking programme of COST Action IS0804, which focuses on languages of the European Union, rather than from a systematically constructed research project. It was only possible to add languages spoken in non-EU countries when COST awarded the country special status to be included in the Action. Additionally, although the CLT is now available for 25 languages (http://psychologia.pl/clts/), as can be seen in other articles in this issue, collecting monolingual data for some of them was not possible since there are no monolingual speakers of these languages (e.g. for Maltese, Gatt et al., 2017; and Lebanese, Khoury Aouad Saliby et al., 2017). Thus, for this study we analysed data from monolingual children speaking one of 15 mostly European languages.

## Method

### Participants

The participants consisted of 639 monolingual children (52% female) within an age range of 3;0–6;11 years. The distribution of participants was not equal across age groups (given in one-year intervals): the largest age group comprised 5-year-olds (46% of all children), followed by 4-year-olds (23%) and 6-year-olds (21%). Table 1 presents the number of participants by average age for each language group. Participants were recruited through preschools and schools, under the inclusion criteria that they were typically developing children with no previous diagnosis of language or cognitive problems. For 11 of the

**Table 1.** Number of participants per age and language group.

| Language\age group | 3 | 4 | 5 | 6 | Total per language | Mean age per language |
|---|---|---|---|---|---|---|
| Afrikaans | 1 | 20 | – | – | 21 | 4;5 |
| Catalan | 20 | 20 | 20 | – | 60 | 4;7 |
| English (British) | – | – | 8 | 9 | 17 | 5;11 |
| English (South African) | – | 10 | 18 | 1 | 29 | 5;2 |
| Finnish | 11 | 10 | 15 | 11 | 47 | 5;0 |
| German | – | – | 33 | 3 | 36 | 5;6 |
| Hebrew | – | – | 11 | 4 | 15 | 5;8 |
| IsiXhosa | – | 10 | – | – | 10 | 4;6 |
| Italian | – | – | 10 | 15 | 25 | 6;2 |
| Lithuanian | 3 | 9 | 14 | 16 | 42 | 5;6 |
| Luxembourgish | – | 17 | 38 | 34 | 89 | 5;8 |
| Norwegian | 6 | 9 | 11 | – | 26 | 4;8 |
| Polish | – | 11 | 38 | 15 | 64 | 5;6 |
| Serbian | – | 1 | 13 | 6 | 20 | 5;10 |
| Slovak | 18 | 18 | 22 | 15 | 73 | 5;0 |
| Swedish | – | 7 | 24 | 1 | 32 | 5;4 |
| Turkish | – | 7 | 20 | 6 | 33 | 5;5 |
| Total | 59 | 149 | 295 | 136 | 639 | 5;4 |

languages (Afrikaans, British English, South African English, Finnish, Hebrew, isiXhosa, Norwegian, Polish, Serbian, Slovak, Swedish), participants' basic SES data were available, which confirmed that most of the participants came from a mid-to-high SES. The exceptions were participants from South Africa, for which the SES was carefully ascertained and used in separate analyses (Potgieter & Southwood, 2016); half of the speakers of Afrikaans and South African English, and all the speakers of isiXhosa, came from a low SES background. For the remaining six languages (Catalan, German, Italian, Lithuanian, Luxemburgish and Turkish), no SES data were available for the individual child participants; however, their place of recruitment (e.g. school and type of neighbourhood) reflected a mid-to-high SES environment.

## Procedure

To assess children's lexical knowledge, we used the CLT in their respective languages. The children were assessed in their preschools or schools in a quiet setting (such as a separate room). They were acquainted with the experimenter prior to testing. For most of the languages tested, paper CLT versions were used: for the comprehension subtasks, one target picture and three distractors were presented per page, printed in colour in A4 format (landscape). The production subtasks contained a single-coloured picture per page, printed in A5 format (landscape). This ensured that the pictures were of a similar size across the subtasks. For three of the languages (Norwegian, Polish and Slovak), e-versions of the task were used, with the pictures presented on a computer touch-screen, and the prompts for target words were pre-recorded. For German, a PowerPoint version was used, with pictures presented on the computer screen and pre-recorded prompts, but without the automatic saving of responses (no touch-screen was available). Otherwise, the procedure was as similar as possible to the paper version described above. The differences in task delivery reflected specific research goals of the various language teams which went beyond the aims of the current analyses.[5] We consider the various versions of the CLTs to be equivalent, since the administration procedure was the same, with the introductory instructions always provided by the experimenter; the only difference was whether or not item prompts (questions) were pre-recorded. In both cases, children were asked to point to or name the picture which corresponded to the prompt. Considering the rapidly rising access of very young children in the mid-to-high SES groups to electronic and mobile devices which mostly use touch-screens (Holloway, Green, & Livingstone, 2013), we did not expect that the difference in picture presentation would affect the results of our study.

At the beginning of the assessment, the children were told, using simple wording, that they were going to view a series of pictures, and that the researcher would ask them about the pictures. They were informed that there would be one question per page, and that pointing to one picture or giving a one-word answer would be sufficient. The original introductory instructions were written in English and subsequently translated into the other languages, with the recommendation that the wording should be natural and play-like, using simple vocabulary appropriate for young children. The form of the prompts in the comprehension subtasks were: 'Where is the [x, target noun]?' (e.g. *squirrel*), 'Who is

[5]Specifically, the teams using e-versions were interested in word-processing speed for the two word classes assessed in the comprehension and production tasks. Reaction time measurement was not possible with the printed version.

**Table 2.** Order of CLT delivery.

| | First subtask | Second subtask | Third subtask | Forth subtask |
|---|---|---|---|---|
| ORDER 1 | Verb comp | Noun comp | Verb prod | Noun prod |
| ORDER 2 | Noun comp | Verb comp | Noun prod | Verb prod |
| ORDER 3 | Noun prod | Verb prod | Noun comp | Verb comp |
| ORDER 4 | Verb prod | Noun prod | Verb comp | Noun comp |

[x-ing, target verb]?' (for agentive verbs, e.g. *singing*) and 'Where is it [x-ing, target verb]?' (for stative verbs, e.g. *raining*). The form of the prompts in the production subtasks were: 'What/who is this?' for nouns, 'What is he/she doing?' for agentive verbs and 'What is happening here?' for stative verbs (e.g. *boiling*).

The order for administering the four subtasks was balanced so that nearly equal numbers of participants received each of the four possible orders, as shown in Table 2. A short break could be taken between the subtasks if needed. Once all the subtasks were completed, the children were thanked for their participation.

## Results

### Preliminary data analysis

#### Items removed from analysis

As mentioned above, there were 32 items in each subtask of the CLT for all language versions. The complete set of items is analysed here for 14 of the 17 languages. For three language versions, some items were removed from the analysis.

For British English, we have used results from the pilot version of the CLT. In the analysis here, we only include items that were used in both the pilot version and the final version of the tasks (28 items for both noun production and comprehension; 26 items for verb production; 25 items for verb comprehension).

Due to an error in constructing the Afrikaans version, two items were repeated in the production and comprehension tasks (*helikopter* in the noun subtasks, and *brei* 'to knit' in the verb subtasks). We dealt with this by counting *helikopter* as a target word for comprehension but not for production, and *brei* as a target word for production but not for comprehension; these are the subtasks where these items occur in the final corrected version of the Afrikaans CLT. Thus, for Afrikaans we have analysed 31 items for noun production and verb comprehension, and 32 items for the two other subtasks.

For isiXhosa, an error in constructing this version of the CLT led to most items in the verb comprehension subtask not being the right ones. As only six items were correct, we omit the isiXhosa verb comprehension subtask in the analysis.

#### Item and subtask difficulty

In order to assess the influence of AoA and CI, the language-specific variables used in constructing the CLTs, we analysed the effects of these factors on item difficulty, as measured by the percentage of children who responded correctly to a particular item in a given language. We calculated the Spearman $\rho$ correlations for AoA and CI with item difficulty in each subtask for each language version.

To analyse the accuracy of the monolingual children's performance on the CLTs, we calculated the mean percentage of correct responses for each of the four subtasks in each language. The percentage score was used instead of raw scores, as some items were excluded from the analyses, as discussed above.

Below, we report first on results concerning the evaluation of the CLT background variables, and then proceed to analyses that are linked to our expectations regarding higher accuracy for older participants, for nouns vs. verbs, and for comprehension vs. production, including also exploratory analyses regarding potential differences across languages.

## Effects of AoA and CI on item difficulty

For the AoA, we found a pattern of significant moderate-to-strong negative correlations with item difficulty (Table 3) in over 72% of cases (all the subtasks in all languages). The number of languages in which significant correlations were found differed across subtasks: 10 for noun comprehension, 11 for noun production, 12 for verb comprehension and 16 for verb production. The average Spearman $\rho$ of all significant coefficients for a subtask ranged from −.49 for noun comprehension to −.59 for noun production. In general, the correlation was stronger for verbs than for nouns.

This pattern of correlations was not repeated for the CI, where significant low-to-moderate negative correlations were found for only 13% of the subtasks, mostly verb production (Table 4). In 10 languages there was no effect of the CI at all.

## Accuracy

The mean accuracy rating for each subtask in each of the 17 languages is given in Figure 2. Accuracy ranged from 72% to 100% for noun comprehension (Mdn = 98%); from 80% to 98% for verb comprehension (Mdn = 92%), from 41% to 93% for noun production (Mdn = 82%), and from 28% to 85% for verb production (Mdn = 66%).

**Table 3.** Correlations between item difficulty and AoA in 17 languages (Spearman $\rho$ coefficients).

| | Comprehension | | Production | | | | |
|---|---|---|---|---|---|---|---|
| Language/task | Nouns | Verbs | Nouns | Verbs | *n* | Range: years | Range |
| Afrikaans | −0.41* | −0.61** | ns | −0.68*** | 21 | 0.98 | 3;11–4;11 |
| Catalan | −0.45** | −0.53*** | −0.46** | −0.49** | 60 | 2.59 | 3;4–5;11 |
| English (British) | NA | ns | ns | −0.53*** | 17 | 1.58 | 5;2–6;9 |
| English (South African) | −0.65*** | −0.62*** | −0.56** | −0.65*** | 29 | 2.18 | 4;0–6;2 |
| Finnish | ns | −0.67*** | −0.60*** | −0.43* | 47 | 3.92 | 3;0–6;11 |
| German | ns | −0.49** | ns | −0.42* | 36 | 1.22 | 5;0–6;3 |
| Hebrew | −0.49*** | −0.56*** | ns | −0.67*** | 15 | 1.29 | 5;0–6;3 |
| IsiXhosa | −0.63*** | NA | −0.79*** | −0.42* | 10 | 0.75 | 4;0–4;10 |
| Italian | ns | ns | −0.46** | −0.70*** | 25 | 1.65 | 5;3–6;11 |
| Lithuanian | ns | −0.64** | ns | ns | 42 | 3.5 | 3;5–6;11 |
| Luxembourgish | −0.47** | −0.74*** | −0.65*** | −0.58*** | 89 | 2.18 | 4;7–6;10 |
| Norwegian | −0.56** | −0.36* | −0.58*** | −0.69*** | 26 | 2.42 | 3;6–5;11 |
| Polish | −0.42* | −0.56** | −0.68*** | −0.64*** | 64 | 2.81 | 4;1–6;11 |
| Serbian | −0.30* | ns | ns | −0.53** | 20 | 1.59 | 4;11–6;6 |
| Slovak | ns | −0.70*** | −0.72*** | −0.55** | 73 | 3.66 | 3;4–6;11 |
| Swedish | −0.48** | ns | −0.53** | −0.36* | 32 | 1.63 | 4;4–6;0 |
| Turkish | ns | −0.42* | −0.44** | −0.59*** | 33 | 2.63 | 4;1–6;10 |

*Note*. English noun comprehension: no variance. *** means significance at p ≤ 0.001; ** means significance at p ≤ 0.01; * means significance at p ≤ 0.05; ns means non-significant result.

**Table 4.** Correlations between item difficulty and CI in 17 languages (Spearman ρ coefficients).

| | Comprehension | | Production | | | | |
|---|---|---|---|---|---|---|---|
| Language /task | Nouns | Verbs | Nouns | Verbs | *n* | Range: years | Range |
| Afrikaans | ns | ns | ns | −0,44* | 21 | 0,98 | 3;11–4;11 |
| Catalan | ns | ns | ns | ns | 60 | 2,59 | 3;4–5;11 |
| English | NA | ns | ns | ns | 17 | 1,58 | 5;2–6;9 |
| English (South African) | −0,37* | ns | −0,50** | ns | 29 | 2,18 | 4;0–6;2 |
| Finnish | ns | ns | −0,38* | −0,42* | 47 | 3.92 | 3;0–6;11 |
| German | ns | ns | ns | ns | 36 | 1,22 | 5;0–6;3 |
| Hebrew | ns | ns | ns | ns | 15 | 1,29 | 5;0–6;3 |
| IsiXhosa | ns | NA | ns | ns | 10 | 0,75 | 4;0–4;10 |
| Italian | ns | ns | ns | ns | 25 | 1,65 | 5;3–6;11 |
| Lithuanian | ns | −0,38* | ns | ns | 42 | 3,5 | 3;5–6;11 |
| Luxembourgish | ns | ns | ns | ns | 89 | 2,18 | 4;7–6;10 |
| Norwegian | ns | ns | ns | ns | 26 | 2,42 | 3;6–5;11 |
| Polish | ns | ns | −0,41* | ns | 64 | 2,81 | 4;1–6;11 |
| Serbian | ns | ns | ns | −0,39* | 20 | 1,59 | 4;11–6;6 |
| Slovak | ns | ns | ns | ns | 73 | 3,66 | 3;4–6;11 |
| Swedish | ns | ns | ns | ns | 32 | 1,63 | 4;4–6;0 |
| Turkish | ns | ns | ns | −0,53** | 33 | 2,63 | 4;1–6;10 |

*Note*. English noun comprehension: no variance. *** means significance at p ≤ 0.001; ** means significance at p ≤ 0.01; * means significance at p ≤ 0.05; ns means non-significant result.

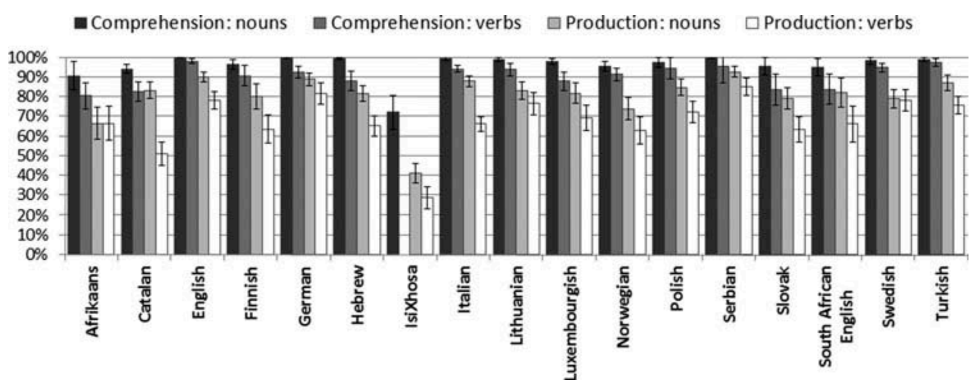


**Figure 2.** CLT accuracy across 17 languages.

*Note*. Verb comprehension data in isiXhosa were not included in the analysis. Error bars represent 1/2 SD. Tables – Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (CLTs)

### *Participants' age*

Analysis of the results for each of the 639 participants showed a significant positive correlation between overall accuracy (percentage of correct answers in all subtasks for each child) and the participants' age (in months) for the languages taken together ($\rho = .61$; $p < .001$), as well as for 11 individual languages (see Table 5). The Spearman ρ coefficients for the subtasks ranged from .26 (noun comprehension in Polish) to .82 (verb production in Norwegian) and, in eight of the 17 languages, the correlation was significant for at least three of the four subtasks.

### *Language, subtask and word class*

We first ran a multivariate analysis of variance (MANOVA) to explore the differences between the results in isiXhosa and the other languages. The dependent variables were the

**Table 5.** Correlations of the CLTs results with the participants' age (Spearman ρ coefficients).

| | Comprehension | | Production | | | | |
|---|---|---|---|---|---|---|---|
| Language/task | Nouns | Verbs | Nouns | Verbs | Total | *n* | Age range |
| Total | 0.49*** | 0.50*** | 0.52*** | 0.51*** | 0.61*** | 639 | 3;0–6;11 |
| Afrikaans | ns | ns | ns | ns | ns | 21 | 3;11–4;11 |
| Catalan | 0.52*** | 0.75*** | 0.61*** | 0.73*** | 0.81*** | 60 | 3;4–5;11 |
| English | NA | ns | ns | ns | ns | 17 | 5;2–6;9 |
| English (South African) | 0.59** | 0.62*** | 0.67*** | 0.72*** | 0.66*** | 29 | 4;0–6;2 |
| Finnish | 0.57*** | 0.69*** | 0.66*** | 0.68*** | 0.73*** | 47 | 3;0–6;11 |
| German | ns | ns | ns | ns | ns | 36 | 5;0–6;3 |
| Hebrew | ns | 0.55* | ns | ns | ns | 15 | 5;0–6;3 |
| IsiXhosa | ns | NA | 0.73* | ns | ns | 10 | 4;0–4;10 |
| Italian | ns | ns | ns | 0.40* | 0.46* | 25 | 5;3–6;11 |
| Lithuanian | 0.49** | 0.47** | 0.43** | 0.62*** | 0.61*** | 42 | 3;5–6;11 |
| Luxembourgish | 0.36*** | ns | 0.31** | ns | 0.25* | 89 | 4;7–6;10 |
| Norwegian | ns | 0.68*** | 0.74*** | 0.82*** | 0.80*** | 26 | 3;6–5;11 |
| Polish | 0.26* | 0.36** | 0.38** | 0.48*** | 0.50*** | 64 | 4;1–6;11 |
| Serbian | ns | ns | ns | 0.71** | 0.57** | 20 | 4;11–6;6 |
| Slovak | 0.39* | 0.67*** | 0.47*** | 0.52*** | 0.64*** | 73 | 3;4–6;11 |
| Swedish | 0.53** | 0.36* | 0.46* | 0.51** | 0.58** | 32 | 4;4–6;0 |
| Turkish | ns | 0.39** | ns | ns | 0.40* | 33 | 4;1–6;10 |

*Note*. There was no variance for the comprehension of nouns in the British English.
Verb comprehension data in isiXhosa were not included in the analysis.
***, ** and * indicate significance levels at $p \leq 0.001$, $p \leq 0.01$ and $p \leq 0.05$, respectively; ns – non-significant result.

scores on the three subtasks for which isiXhosa data were available (noun comprehension, noun production, verb production), and the independent variable was language. There was a significant effect of language ($F(48,1866) = 15.2$, $p < .001$). We then ran a Dunett *t* post-hoc test to determine whether the isiXhosa results differed from those in the other languages. All pairwise comparisons were significant ($p < .001$ in all 48 cases), revealing that the isiXhosa results were lower than results in all the other languages for all three subtasks. Because of this, the isiXhosa data were omitted from further analyses of the effects of language, subtask and word class.

We ran a repeated-measure ANCOVA using within-subject factors (type of task: comprehension vs production; and word class: noun vs. verb), a between-subject factor (language), and a covariate (age). This analysis revealed significant main effects of language, participants' age, subtask and word class (see Table 6). As the main effect of language was weak (partial $\eta^2 = .16$) and the main effects of subtask (partial $\eta^2 = .28$) and word category (partial $\eta^2 = .25$) were stronger, we ran partial comparisons of estimated marginal means for the latter two factors.

**Table 6.** Within-subject and between-subject effects in the ANCOVA.

| | | df | *F* | *p* | Partial $\eta^2$ |
|---|---|---|---|---|---|
| Between-subject | Intercept | 1 | 859.42 | <0.001 | 0.58 |
| | Age | 1 | 260.84 | <0.001 | 0.30 |
| | Language | 15 | 7.78 | <0.001 | 0.16 |
| | Error df | 612 | | | |
| Within-subject | Subtask | 1 | 234.42 | <0.001 | 0.28 |
| | Subtask * age | 1 | 56.75 | <0.001 | 0.09 |
| | Subtask * language | 15 | 5.20 | <0.001 | 0.11 |
| | Word category | 1 | 204.30 | <0.001 | 0.25 |
| | Word category * age | 1 | 82.52 | <0.001 | 0.12 |
| | Word category * language | 15 | 22.11 | <0.001 | 0.35 |
| | Task * word category | 1 | 0.78 | 0.38 | 0.00 |

### *Subtask*

A comparison of the marginal means with a Bonferroni correction for confidence intervals showed significant effects of subtask for all 16 languages: there were higher scores for the comprehension tasks than for the production tasks across all languages (Table 7).

### *Word class*

A comparison of the marginal means with a Bonferroni correction for confidence intervals showed significant effects for word class in 13 of the 16 languages, the exceptions being Afrikaans, Norwegian and Swedish. In those 13 languages, the scores were higher for the noun tasks than for the verb tasks. For the other three languages, the direction of difference was the same but was not significant. Table 8 presents the exact values of marginal means for all the languages.

**Table 7.** Marginal means of the subtask results across languages, with a Bonferroni correction for the confidence intervals.

| | | | | 95% Confidence interval | |
|---|---|---|---|---|---|
| Language | Subtask | Mean | SE | Lower bound | Upper bound |
| Afrikaans*** | Comprehension | 0.89 | 0.01 | 0.86 | 0.92 |
| | Production | 0.73 | 0.02 | 0.69 | 0.77 |
| Catalan*** | Comprehension | 0.91 | 0.01 | 0.90 | 0.93 |
| | Production | 0.72 | 0.01 | 0.70 | 0.74 |
| English (British)*** | Comprehension | 0.96 | 0.02 | 0.93 | 0.99 |
| | Production | 0.79 | 0.02 | 0.75 | 0.83 |
| English (South African)*** | Comprehension | 0.90 | 0.01 | 0.87 | 0.92 |
| | Production | 0.75 | 0.02 | 0.71 | 0.78 |
| Finnish*** | Comprehension | 0.95 | 0.01 | 0.93 | 0.97 |
| | Production | 0.74 | 0.01 | 0.71 | 0.76 |
| German*** | Comprehension | 0.95 | 0.01 | 0.93 | 0.97 |
| | Production | 0.83 | 0.01 | 0.80 | 0.86 |
| Hebrew*** | Comprehension | 0.92 | 0.02 | 0.89 | 0.95 |
| | Production | 0.71 | 0.02 | 0.66 | 0.75 |
| Italian*** | Comprehension | 0.93 | 0.01 | 0.90 | 0.95 |
| | Production | 0.71 | 0.02 | 0.67 | 0.74 |
| Lithuanian*** | Comprehension | 0.95 | 0.01 | 0.93 | 0.97 |
| | Production | 0.78 | 0.01 | 0.76 | 0.81 |
| Luxembourgish*** | Comprehension | 0.91 | 0.01 | 0.91 | 0.93 |
| | Production | 0.73 | 0.01 | 0.71 | 0.75 |
| Norwegian*** | Comprehension | 0.96 | 0.01 | 0.93 | 0.98 |
| | Production | 0.73 | 0.02 | 0.69 | 0.76 |
| Polish*** | Comprehension | 0.95 | 0.01 | 0.93 | 0.97 |
| | Production | 0.77 | 0.01 | 0.75 | 0.79 |
| Serbian** | Comprehension | 0.95 | 0.02 | 0.92 | 0.98 |
| | Production | 0.85 | 0.02 | 0.81 | 0.89 |
| Slovak*** | Comprehension | 0.91 | 0.01 | 0.89 | 0.92 |
| | Production | 0.74 | 0.01 | 0.72 | 0.76 |
| Swedish*** | Comprehension | 0.96 | 0.01 | 0.94 | 0.99 |
| | Production | 0.78 | 0.02 | 0.75 | 0.81 |
| Turkish*** | Comprehension | 0.98 | 0.01 | 0.96 | 1.00 |
| | Production | 0.81 | 0.01 | 0.78 | 0.84 |

*Note.* *** and ** indicate significance levels at $p \leq 0.001$ and $p \leq 0.01$, respectively.

**Table 8.** Marginal means of the word categories results across languages, with a Bonferroni correction for the confidence intervals.

| | | | | 95% Confidence interval | |
|---|---|---|---|---|---|
| Language | Subtask | Mean | SE | Lower bound | Upper bound |
| Afrikaans$^{ns}$ | Nouns | 0.82 | 0.01 | 0.80 | 0.85 |
| | Verbs | 0.80 | 0.02 | 0.76 | 0.84 |
| Catalan*** | Nouns | 0.92 | 0.01 | 0.90 | 0.93 |
| | Verbs | 0.72 | 0.01 | 0.70 | 0.74 |
| English (British)* | Nouns | 0.92 | 0.01 | 0.89 | 0.95 |
| | Verbs | 0.83 | 0.02 | 0.79 | 0.87 |
| English (South African)*** | Nouns | 0.89 | 0.01 | 0.87 | 0.91 |
| | Verbs | 0.75 | 0.02 | 0.72 | 0.78 |
| Finnish*** | Nouns | 0.90 | 0.01 | 0.88 | 0.91 |
| | Verbs | 0.79 | 0.01 | 0.77 | 0.82 |
| German*** | Nouns | 0.93 | 0.01 | 0.91 | 0.95 |
| | Verbs | 0.85 | 0.01 | 0.82 | 0.88 |
| Hebrew*** | Nouns | 0.89 | 0.02 | 0.86 | 0.92 |
| | Verbs | 0.74 | 0.02 | 0.69 | 0.78 |
| Italian*** | Nouns | 0.90 | 0.01 | 0.87 | 0.92 |
| | Verbs | 0.74 | 0.02 | 0.70 | 0.77 |
| Lithuanian** | Nouns | 0.90 | 0.01 | 0.88 | 0.92 |
| | Verbs | 0.84 | 0.01 | 0.81 | 0.86 |
| Luxembourgish*** | Nouns | 0.88 | 0.01 | 0.87 | 0.90 |
| | Verbs | 0.76 | 0.01 | 0.74 | 0.78 |
| Norwegian$^{ns}$ | Nouns | 0.87 | 0.01 | 0.85 | 0.89 |
| | Verbs | 0.81 | 0.02 | 0.78 | 0.85 |
| Polish*** | Nouns | 0.90 | 0.01 | 0.89 | 0.92 |
| | Verbs | 0.82 | 0.01 | 0.80 | 0.84 |
| Serbian* | Nouns | 0.94 | 0.01 | 0.91 | 0.97 |
| | Verbs | 0.86 | 0.02 | 0.82 | 0.90 |
| Slovak*** | Nouns | 0.89 | 0.01 | 0.87 | 0.90 |
| | Verbs | 0.76 | 0.01 | 0.74 | 0.78 |
| Swedish$^{ns}$ | Nouns | 0.88 | 0.01 | 0.86 | 0.90 |
| | Verbs | 0.86 | 0.02 | 0.83 | 0.89 |
| Turkish** | Nouns | 0.93 | 0.01 | 0.91 | 0.95 |
| | Verbs | 0.86 | 0.02 | 0.83 | 0.88 |

*Note.* ***, ** and * indicate significance levels at $p \leq 0.001$, $p \leq 0.01$ and $p \leq 0.05$, respectively; ns – non-significant result.

## Discussion

### *The impact of background variables on CLT results*

First, we evaluated the impact of the two background variables that were used to select target words for the assessment tasks – AoA and CI. As expected, for AoA, the correlations were negative for all languages (the higher a word's AoA value, i.e. the later a word is acquired, the lower its item accuracy). For 12 of the languages, correlations were significant for at least three of the four subtasks (Table 3). For British English, German, Italian and Lithuanian, the lack of significant correlations for noun subtasks can be attributed to a ceiling effect in the CLT results. It is harder to explain the lack of a significant correlation for Serbian noun production and Lithuanian verb production, the latter being the only non-significant result among all the verb production subtasks.

The second factor used for target word selection was the CI. Contrary to our predictions, there was no significant correlation between the CI and item accuracy for most languages and subtasks (Table 4). This may be due to the compositionality of the CI, which was meant to account for various word characteristics: phonological, morphological, whether it is a loanword, and children's exposure to the object or action depicted by the word (Haman et al., 2015). Hansen et al. (2017) discuss possible reasons underlying the absence of correlation between the CI and item accuracy. They suggest that including characteristics from several different domains in one composite score may lead to an inconsistent measure, as particular components may give contradictory values. Thus, the resulting average score (a word's CI) may fail to reflect the actual difficulty of each component. However, when these word characteristics were analysed separately, the only component that had some impact was exposure (whether children had frequent and easy access to the object or action depicted by the word). The expected correlation with phonology, morphology or borrowings was not found. It is not clear, however, whether the target words used in the Polish and Norwegian versions of the CLT presented enough variability in each of these domains to reveal significant effects, as target word selection for these CLTs was based on only two levels of the composite CI score: low and high (i.e. under or above the mean for each language). Another possibility is that the complexity measures used for the different domains did not capture actual word complexity for all languages. It is also possible that word complexity has more influence on word learning at earlier ages, while participants in this study were mostly 4–6-year olds. Phonology has indeed been shown to have an influence on word learning in several cross-linguistic comparison tasks for children below the age of 3 years, as discussed in Hansen et al. (this issue). It is possible that once the phonological system of a language is mastered, phonology exerts less influence on lexical development, at least as long as the words comply with the phonological characteristics of the language.

It is also possible that word complexity in terms of phonology and morphology influences word learning in different ways in each language. Calculating the CI in the same way for every language might thus be inadequate. For example, languages differ when it comes to typical word length (Garmann, Hansen, Simonsen, & Kristoffersen, in press). This suggests that the relationship between word length and word difficulty is not linear but depends on which phonological patterns are typical in the language. For instance, English and Danish children tend to prefer monosyllabic words in their production (Garmann et al., in press), while children acquiring Italian produce very few monosyllables and tend to acquire di- and polysyllabic words first (Caselli et al., 1995). When it comes to morphology, both inflectional and derivational morphology are mastered earlier in morphologically rich languages than in morphologically poor ones (Clark, 2001), and thus morphology might not have an impact on word difficulty for children of the age range under scrutiny in this study. Thus, the CI requires much more detailed investigation for individual languages, using new data possibly from younger children, before the hypothesis that it plays a role in word learning is rejected.

### *Effect of participants' age*

Next, we evaluated the CLT's ability to reflect the expected increase in vocabulary size with age. We found strong significant positive correlations between overall CLT scores and the participants' age across all samples. This result holds for all four subtasks, and for 11 out of 17 languages (see Table 5).

No age effect was found for four languages (Afrikaans, British English, Hebrew and isiXhosa). This may be due to the small sample sizes ($N \leq 21$) and narrow age ranges (< 1.58 years; see Table 1). For German, the sample size was moderate ($N = 36$), but the age range was low (1.22), and as the participants were all at or above the age of 5;0, there were ceiling effects in the CLT results (especially for noun comprehension, where there was no variation at all, see Figure 2). This meant that there was a lower correlation with age for the German group. When the sample size and age range were adequate, the CLT was sensitive to participants' age, which confirms that the measure reflects expected developmental changes in vocabulary size for children over the age of three.

## *IsiXhosa as an exception*

In the overall analysis of CLT performance, we found that one language was an outlier in comparison to all the other languages. The scores for the isiXhosa CLT were significantly lower than for all the other languages on all three subtasks analysed for isiXhosa. This could be due to the small sample size ($N$ = 10) and relatively low age of the isiXhosa participants (average age 4;6), but there is some evidence that this finding may reflect real differences in lexical development. First, Potgieter and Southwood (2016) report that the isiXhosa vocabulary size of monolingual isiXhosa-speaking children (the group which is reported on here) does not differ from that of South African trilinguals. They also state that monolingual isiXhosa children came from a low SES. This was unavoidable, as in the South African society, parents with a higher socio-economic status who speak isiXhosa have a strong tendency to raise their children bilingually, perceiving the ability to speak South African English as providing an opportunity for social advancement. Second, in the study of subjective age of acquisition in 25 languages carried out by Łuniewska et al. (2015), adult native speakers of isiXhosa generally rated the words as being acquired at a later age than speakers of other languages did. Note that isiXhosa, the only example of a Bantu language in our sample, is a seriously understudied language in terms of language acquisition. It is thus possible that the rate of isiXhosa lexical acquisition does indeed differ significantly from that of the other languages included here, but it is not clear whether this is due to linguistic properties of isiXhosa, the social status of isiXhosa speakers in South African society, or other factors.

## *Word class*

Our results confirmed that overall nouns are learned earlier than verbs across all languages apart from Afrikaans, Swedish and Norwegian. The lack of a word class effect for these three languages was due to high variability in accuracy on the verb tasks (Table 8), although the difference was in the expected direction (see Figure 2).

For the 14 other languages, there was a consistent pattern of better performance on the noun subtasks, which supports claims that nouns take precedence in lexical development (Black & Chiat, 2003; Bornstein et al., 2004; Gentner, 2006). Our findings may reflect a greater conceptual saliency of nouns over verbs, and a greater reliance of verbs on linguistic structure (Gentner & Boroditsky, 2001). It is, however, important to note that the fact that the CLT is based on pictures may in itself give an advantage to nouns, as the picture prompts for nouns are perceptually simpler than those for verbs. The nouns are typically represented by single objects or people, whereas most of the verbs are depicted by human characters performing an action. In many cases, the pictures for verbs include instruments required for carrying out the actions or clues about the environment in which the action is typically performed. For example, a picture prompt for a saw only requires the object itself, while a picture prompt for the verb 'to saw' requires an agent and an instrument. This difference was unavoidable and is also discussed in previous studies comparing noun and verb production through picture tasks (Bello et al., 2012; Kauschke, Lee, & Pae, 2007; Masterson, Druks, & Gallienne, 2008; Mätzig, Druks, Masterson, & Vigliocco, 2009). However, this effect (better performance on nouns than verbs) also holds across a wide variety of languages when other measures are used, e.g. in studies using MB-CDI or similar parental checklists (Bornstein et al., 2004; Caselli et al., 1995).

### Comprehension vs. production

As predicted, lexical comprehension was more advanced than production for participants from each of the languages studied (Table 7). Findings showing the primacy of comprehension over production (Benedict, 1979; Bornstein & Hendricks, 2012; Clark, 2012; Goldfield, 2000; Goldin-Meadow, Seligman, & Gelman, 1976; Harris et al., 1995; Reznick & Goldfield, 1992) are robust, and our data add to this body of research. Production is claimed to be more demanding in terms of long-term memory and lexical access. Clark (1993, 2009) suggests that children have separate word representations for comprehension and production, and that this allows them to operate on word meanings before mastering their articulatory form and actually enables them to come up with the adult form by comparing the non-mastered child form with the adult version, which is represented initially only at a comprehension level. According to Clark, the discrepancy between comprehension and production holds from early childhood to adulthood.

It should also be noted that in our study, this effect in part reflects the specificity of each subtask. For the comprehension tasks, there were only four possible answers, as the child was asked to choose among four pictures. In contrast, the production task required selecting a word from an open set limited only by the child's overall lexical repertoire (Markman, 1989; Quine, 1960). Although children may overcome this problem either by applying constraints in selecting possible meanings for a word (Clark, 1995; Markman, 1991) or by using more general conceptual strategies (Bloom, 2000), they may still provide a variety of adequate labels for a picture in a naming task. For this article, the only production responses considered to be correct were those involving the target word (allowing only for mispronunciations or unexpected inflections, as long as the target word stem was present), as discussed by Kapalková and Slančová (2017). This is because the AoA and CI, which underlie the development of the tasks, were available only for target words and not for other potential answers children may provide). When CLT is used to investigate differences between groups, synonyms typically count as correct responses, according to guidelines agreed upon in the Bi-SLI Working Group 3, which is discussed in detail by Kapalková and Slančová (2017).

### Cross-linguistic differences

Overall, we found only a small effect (partial $\eta^2$ = .16) of language on the CLT results when the outlier isiXhosa was removed. Thus, there were only small differences in vocabulary size, as measured by CLT, between the 16 other languages. However, as there were differences in the age range of the participants for the languages, we cannot draw any strong inferences about this factor. To gain a deeper insight into potential cross-linguistic differences in word learning, we need to systematically assess younger children in order to bypass ceiling effects which were evident in many of the languages for nouns and sometimes even for verbs. The current results obtained from monolingual children suggest, however, that for this age range the CLT may be an adequate screening tool for assessing lexical knowledge, as it is sensitive to the participants' age and differentiates well between comprehension and production, as well as between nouns and verbs, in line with previous research. Thus, we could expect that the CLT will also adequately assess word knowledge in bilingual and multilingual children. This has already been partially demonstrated for a limited number of language pairs: Polish–Norwegian (Hansen et al., 2017); Maltese–English (Gatt et al., 2017), and one trilingual configuration: South African English, Afrikaans and isiXhosa (Potgieter & Southwood, 2016). Further studies should confirm whether the CLT may also be used in assessing the severity of specific language impairment (SLI) in the lexical domain or for establishing the type of SLI (Friedmann & Novogrodsky, 2008). To date, a study of monolingual Slovak children has shown the CLT's sensitivity at the group level in differentiating between children with language impairment and their typically developing peers. The same study provided insights into different error patterns on naming subtasks between impaired children and typically developing peers who were matched on comprehension results (Kapalková & Slančová, 2017). Similarly, a study of bilingual children acquiring Lebanese as one of their languages showed differences in the Lebanese vocabulary size (as assessed by CLT) between children with and without a language impairment (Khoury Aouad Saliby et al., 2017).

### Limitations of this study

One limitation of this study is that it does not account in detail for potential differences in vocabulary caused by the socio-economic status of participants. The influence of this factor on children's lexical knowledge is now well attested (Hart & Risley, 1995; Hoff, 2003). The fact that isiXhosa was the only outlier language in our study may at least partially be explained by these participants' low socio-economic status, as discussed in more depth in Potgieter and Southwood (2016).

Although our overall sample is quite large, it should be noted that some language subsamples had only a small number of participants. It is also clear that the CLT has limits to its sensitivity in discerning age differences. In particular, a ceiling effect for noun comprehension, observable for some languages, indicates that for typical populations the CLT may not be sensitive enough above the age of 5–6 years. However, this does not exclude its usefulness as a potential screening tool for children who are suspected of having a language disorder or who have limited language input (e.g. successive bilinguals) at the age of 6 years or even beyond.

More studies need to be conducted using the CLT, and directed specifically at the problems highlighted above, before any claims can be made about its ability to account for e.g. socio-economically-driven differences in children's lexical knowledge. In addition, more research on impaired populations is needed to confirm the CLT's validity in assessing vocabulary development in a clinical context. At this point, we assume that the CLT is a sensitive measure of children's receptive and expressive lexicon, suitable for use with typically developing monolingual children above the age of 3 years, with the potential for it to become a wider assessment tool when new data (or norming studies) become available for particular languages or language pairs for specific populations (e.g. monolingual or bilingual children).

## CLT: Future directions

This article, along with the other articles in this issue, present a first cross-linguistic comparison of results obtained using the CLT for word knowledge in preschool children. We are still at the initial stages of applying the CLT for diagnostic and clinical purposes. To render the CLT useful in clinical settings, norming studies are needed, controlling for participants' SES (Hart & Risley, 1995; Roy & Chiat, 2012), as well as quantifying language input for bilingual and multilingual populations (Gathercole et al., 2008).

Revisions to the CLT should also be considered, particularly if it is meant to be used as an effective diagnostic tool. In the current study, all the participants took part in a relatively long testing session – they performed all subtasks on all items. Once more data are available, important modifications to the procedure can be envisaged, namely adaptive testing: adjusting the test difficulty to the child's performance during testing.

However, implementation of adaptive testing would only be possible with fully computerized versions of the CLT, which is another potential future direction for the development of the CLT. Fully computerized versions would also enable inclusion of additional variables, such as processing speed, which may prove to be of importance for clinical assessment, since children with SLI are sometimes claimed to have lower processing abilities when compared to typically developing children (Lahey & Edwards, 1996; Lahey, Edwards, & Munson, 2001).

For future studies, the CLT is freely available to all interested researchers (http://psychologia.pl/clts/). New language versions are under development (e.g. American English, Armenian and Malay), and new collaborators who are willing to develop the CLT for other languages are welcome.

## Acknowledgments

This study is a result of collaboration that took place within the COST Action IS0804 'Language Impairment in a Multilingual Society: Linguistics Patterns and the Road to Assessment' (www.bi-sli.org).

We would like to express our gratitude to all children who participated in the study as well as their parents and teachers. Our appreciation for help in data collection goes to Rachel Haas (British English), Eeva-Maria Kovalainen, Suvi Sirviö, Heta Toivola and Marika Pirnes (Finnish), and Annegret Klassert, Vanessa Borchers, Sabrina Klostermeier and Henriette Stoll (German).

## Declaration of interest

The authors report no conflict of interest.

## Funding

The research presented here was partially supported by Berliner Senate and Bundesministerium für Bildung und Forschung (BMBF), Germany [grant number 01UG1411]; the Research Council of Norway through its Centres of Excellence funding scheme [grant number 223265]; the National Science Centre, Poland [grant number 809/N-COST/2010/0]; Polish Ministry of Science and Higher

Education [grant number 0046/DIA/2013/42]; Faculty of Psychology, University of Warsaw [grant number BST2015-1744/4]; Ministry for Education, Science and Technological Development of Republic of Serbia [grant number 175012]; the Slovak Research and Development Agency [grant number APVV-0410-11]; Harry Crossley Foundation and National Research Foundation of South Africa [grant number 88631]; Ministerio de Economía y Competitividad, Spain [grant number FFI2014-56968-C4-1-P]; and Åke Wiberg Foundation [grant number H14-0104], Birgit and Gad Rausing Foundation (Sweden) [grant number S14-14]. We also acknowledge support from International Visegrad Fund partially enhancing collaboration reported in this article [grant number 21420015]. All pictures used for CLTs are subject of copyright of University of Warsaw (Poland).

## ORCID

Pernille Hansen http://orcid.org/0000-0003-3785-0132

## References

Allman, B. (2005). Vocabulary size and accuracy of monolingual and bilingual preschool children. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism* (vol. 5, pp. 58–77). Somerville, MA: Cascadilla Press.

Altman, C., Goldstein, T. & Armon-Lotem, S. (2017). Quantitative and qualitative differences in the lexical knowledge of monolingual and bilingual children on the LITMUS-CLT task. *Clinical Linguistics and Phonetics*, *31*(11–12), 931–954.

Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children. Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters.

Armon-Lotem, S., & Ohana, O. (2017). A CDI study of bilingual english-hebrew children – frequency of exposure as a major source of variation. *International Journal of Bilingual Education and Bilingualism*, *20* (2), 201–217.

Baal, B. A., & Bentzen, K. (2014, August 16–18). *Adapting the MacArthur-Bates CDI to North Sámi – minority language issues*. Paper presented at NorPhlex final seminar, Tromsø, Norway.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *55* (2), 290–313.

Barnett, W. S., & Lamy, C. (2006). Estimated impacts of number of years of preschool attendance on vocabulary, literacy and math skills at kindergarten entry. New Brunswick, NJ: National Institute for Early Education Research.

Bates, E., Goodman, J. C. (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (Ed.). *The emergence of language* (pp. 29–79). Mahwah, NJ: Erlbaum.

Bedore, L., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, *11* (1), 1–29.

Bello, A., Giannantoni, P., Pettenati, P., Stefanini, S., & Caselli, M. C. (2012). Assessing lexicon: Validation and developmental data of the Picture Naming Game (PiNG), a new picture naming task for toddlers. *International Journal of Language & Communication Disorders*, *47* (5), 589–602.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, *6* (02), 183–200.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109* (9), 3253–3258.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13* (04), 525–531.

Black, M., & Chiat, S. (2003). Noun-verb dissociations: a multi-faceted phenomenon. *Journal of Neurolinguistics*, *16* (2–3), 231–250.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, *35* (03), 619–650.
Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: The MIT Press.
Bohman, T. M., Bedore, L. M., Pena, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish English bilinguals. *International Journal of Bilingual Education and Bilingualism*, *13* (3), 325–344.
Bornstein, M. H., & Hendricks, C. (2012). Basic language comprehension and production in >100,000 young children from sixteen developing nations. *Journal of Child Language*, *39* (4), 899–918.
Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., Pêcheux, M. G., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75* (4), 1115–1139.
Caselli, C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10* (2), 159–200.
Chiang, P. Y., & Rvachew, S. (2007). English-French bilingual children's phonological awareness and vocabulary skills. *Canadian Journal of Applied Linguistics (CJAL)/Revue canadienne de linguistique appliquée (RCLA)*, *10* (3), 293–308.
Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
Clark, E. V. (1995). Language acquisition: The lexicon and syntax. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language and communication* (vol. 11, pp. 303–337). San Diego: Academic Press.
Clark, E. V. (2001). Morphology in Language Acquisition. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology*. Blackwell. Reference Online. 28 December 2007.
Clark, E. V. (2009). *First language acquisition* (2nd. ed.). Cambridge: Cambridge University Press.
Clark, E. V. (2012). First Language Vocabulary Acquisition. In C. A. Chapelle (Ed.), *The Encyclopaedia of Applied Linguistics*. Oxford: Blackwell Publishers.
Conboy, B. T., & Thal, D. J. (2006). Ties Between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development*, *77* (3), 712–735.
Dale, P. S., & Penfold, M. (2011). *Adaptations of the MacArthur-bates CDI into non-U.S. English languages*. Retrieved from http://www.sci.sdsu.edu/cdi/documents/AdaptationsSurvey7-5-11Web.pdf
D'Amico, S., Devescovi, A., & Bates, E. (2001). Picture naming and lexical access in Italian children and adults. *Journal of Cognition and Development*, *2* (1), 71–105.
Dar, M., Anwaar, H., Vihman, M. & Keren-Portnoy, T. (2015). Developing an Urdu CDI for early language acquisition in Pakistan. *York Papers in Linguistics, Series* 2 (*14*), 1–14.
de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, *67* (2), 297–319.
De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, *35* (06), 1189–1211.
Dixon, L. Q. (2011). The role of home and school factors in predicting English vocabulary among bilingual kindergarten children in Singapore. *Applied Psycholinguistics*, *32* (01), 141–168.
Duursma, E., Romero-Contreras, S., Szuber, A., Proctor, P., Snow, C., August, D., & Calderon, M. (2007). The role of home literacy and language environment on bilinguals' English and Spanish vocabulary development. *Applied Psycholinguistics*, *28* (01), 171–190.
Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24* (2), 515–523.
Engel de Abreu, P. M. J., Baldassi, M., Puglisi, M. L., & Befi-Lopes, D. M. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: Testing language-minority children with an immigrant background. *Journal of Speech, Language, and Hearing Research*, *56*(2), 630–642.
Engel de Abreu, P. M. J., Cruz-Santos, A., & Puglisi, M. L. (2014). Specific language impairment in language-minority children from low-income families. *International Journal of Language & Communication Disorders*, *49* (6), 736–747.
Engel de Abreu, P. M. J., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology*, *104* (4), 974.

Fenson, L., Dale, P. S., Reznick, J. S., & Thal, D. (1993). *The MacArthur-Bates Communicative Development Inventories (CDI I): Words and Gestures*. Baltimore, MD: Paul Brookes Publishing.

Friedmann, N., & Novogrodsky, R. (2008). Subtypes of SLI: SySLI, PhoSLI, LeSLI and PraSLI. W A. Gavarró & M. J. Freitas (Eds.), *Language acquisition and development* (pp. 205–217). Newcastle: Cambridge Scholars Publishing.

Gagarina, N. V., Klassert, A., & Topaj, N. (2010). Sprachstandstest Russisch für mehrsprachige Kinder [Russian language proficiency test for multilingual children]. *ZAS Papers in Linguistics*, *54*, Berlin: ZAS.

Garmann, N. G., Hansen, P., Simonsen, H. G., & Kristoffersen, K. E. (in press). Phonological characteristics of children's first words. In F. Gayraud, S. Kern & F. Chenu (Eds.). *Proceedings from the ELA 2012 conference*. Cambridge Scholar Publishing.

Gathercole, V. C. M., Thomas, E. M., & Hughes, E. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism*, *11* (6), 678–720.

Gatt, D. (2017). Bilingual vocabulary production in young children receiving Maltese-dominant exposure: individual differences and the influence of demographic and language exposure factors. *International Journal of Bilingual Education and Bilingualism*, *20* (2), 163–182.

Gatt, D., Attard, D., Łuniewska, M., Haman, E. (2017). The effects of bilingual status on lexical comprehension and production in Maltese five-year-old children: a LITMUS-CLT study. *Clinical Linguistics and Phonetics*, *31*(11–12), 844–873.

Gatt, D., O'Toole, C., & Haman, E. (2015). Using parental report to assess early lexical production in children exposed to more than one language. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 151–195). Bristol: Multilingual Matters.

Gentner, D. (1982). *Why nouns are learned before verbs: Linguistic relativity versus natural partitioning* (Technical Report No. 257). Cambridge, MA: Bolt Beranek and Newman Inc. Retrieved from http://files.eric.ed.gov/fulltext/ED219724.pdf

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544–564). Oxford: Oxford University Press.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge: Cambridge University Press.

Goldfield, B. A. (2000). Nouns before verbs in comprehension vs. production: the view from pragmatics. *Journal of Child Language*, *27* (03), 501–520.

Goldin-Meadow, S., Seligman, M. E. P., & Gelman, R. (1976). Language in the 2-year-old. *Cognition*, *5*, 189–202.

Gorman, B. K. (2012). Relationships between vocabulary size, working memory, and phonological awareness in Spanish-speaking English language learners. *American Journal of Speech-Language Pathology*, *21* (2), 109–123.

Haman, E., Łuniewska, M. & Pomiechowska, B. (2015). Designing cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In Armon-Lotem, S., Jong, J. d. & Meir, N. (Eds.), *Assessing multilingual children: Disentangling bilingualism from Language Impairment* (pp. 196–240). Bristol, UK: Multilingual Matters.

Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., & Goldstein, B. (2012). Predicting Spanish-English bilingual children's language abilities. *Journal of Speech, Language and Hearing Research*, *55* (5), 1251–1264.

Hansen, P., Simonsen, H. G., Łuniewska, M., Haman, E. (2017). Validating the psycholinguistic aspects of LITMUS-CLT: Evidence from Polish and Norwegian. *Clinical Linguistics and Phonetics*, *31*(11–12), 910–930.

Harris, M., Yeeles, C., Chasin, J., & Oakley, Y. (1995). Symmetries and asymmetries in early lexical comprehension and production. *Journal of Child Language*, *22* (01), 1–18.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Chicago: Paul H Brookes Publishing.

Hemsley, G., Holm, A., & Dodd, B. (2013). Conceptual distance and word learning: Patterns of acquisition in Samoan–English bilingual children. *Journal of Child Language*, *40* (04), 799–820.
Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74* (5), 1368–1378.
Holloway, D., Green, L., & Livingstone, S. (2013). *Zero to eight: Young children and their internet use*. LSE, London: EU Kids Online.
Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131* (05), 684–712.
Kan, P. F., & Kohnert, K. (2005). Preschoolers learning Hmong and English lexical-semantic skills in L1 and L2. *Journal of Speech, Language, and Hearing Research*, *48* (02), 372–383.
Kapalková. S., and Slančová, D. (2017). The vocabulary profile of Slovak children with primary language impairment compared to typically developing Slovak children measured by LITMUS-CLT. *Clinical Linguistics and Phonetics*, *31*(11–12), 893–909.
Kauschke, C., Lee, H. W., & Pae, S. (2007). Similarities and variation in noun and verb acquisition: A cross-linguistic study of children learning German, Korean, and Turkish. *Language and Cognitive Processes*, *22* (7), 1045–1072.
Khoury Aouad Saliby, C., dosSantos, C., Kouba-Hreich, E., Messarra, C. (2017). Assessing Lebanese bilingual children: the use of Cross-Linguistic Lexical Tasks in Lebanese Arabic language. *Clinical Linguistics and Phonetics*, *31*(11–12), 874–892.
Klassert, A., Gagarina, N., & Kauschke, C. (2014). Object and action naming in Russian- and German-speaking monolingual and bilingual children. *Bilingualism: Language and Cognition*, *17* (01), 73–88.
Kohnert, K. J., Bates, E., & Hernandez, A. E. (1999). Balancing bilinguals: Lexical-semantic production and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, and Hearing Research*, *42* (6), 1400–1412.
Kohnert, K., Kan, P. F., & Conboy, B. T. (2010). Lexical and grammatical associations in sequential bilingual preschoolers. *Journal of Speech, Language, and Hearing Research*, *53* (3), 684–698.
Lahey, M., & Edwards, J. (1996). Why do children with specific language impairment name pictures more slowly than their peers? *Journal of Speech, Language and Hearing Research*, *39* (5), 1081–1098.
Lahey, M., Edwards, J., & Munson, B. (2001). Is processing speed related to severity of language impairment? *Journal of Speech, Language, and Hearing Research*, *44* (6), 1354–1361.
Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Development Inventories. *Child and Adolescent Mental Health*, *13* (4), 198–206.
Lind, M., Simonsen, H. G., Hansen, P., Holm, E., & Mevik, B.-H. (2015). Norwegian Words: A lexical database for clinicians and researchers. *Clinical Linguistics & Phonetics*, *29* (4), 276–290.
Loizou, M., & Stuart, M. (2003). Phonological awareness in monolingual and bilingual English and Greek five-year-olds. *Journal of Research in Reading*, *26* (1), 3–18.
Łuniewska, M., Haman, E., Armon-Lotem, S., Etenkowski, B., Southwood, F., Anđelković, D., … & Ünal-Logacev, Ö. (2016). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words? *Behavior Research Methods*, *48*(3), 1154–1177.
Mancilla-Martinez, J., & Lesaux, N. K. (2011). Early home language use and later vocabulary development. *Journal of educational psychology*, *103* (3), 535–546.
Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: The MIT Press.
Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 72–106). Cambridge: Cambridge University Press.
Masterson, J., Druks, J., & Gallienne, D. (2008). Object and action picture naming in three- and five-year-old children. *Journal of Child Language*, *35* (02), 373–402.
Mätzig, S., Druks, J., Masterson, J., & Vigliocco, G. (2009). Noun and verb differences in picture naming: Past studies and new evidence. *Cortex*, *45* (6), 738–758.

Mayor, J., & Plunkett, K. (2014). Shared understanding and idiosyncratic expression in early vocabularies. *Developmental Science*, *17* (3), 412–423.

Messer, M. H. (2010). *Verbal short-term memory and vocabulary development in monolingual Dutch and bilingual Turkish-Dutch preschoolers* (Unpublished PhD dissertation). Utrecht University.

Miękisz, A., Haman, E., Łuniewska, M., Kuś, K., O'Toole, C. & Katsos, N. (2017). Impact of first generation immigrant environment on the heritage language: Productive vocabulary of polish toddlers living in the UK and Ireland. *International Journal of Bilingual Education and Bilingualism*, *20* (2), 183–200.

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, *20* (6), 705–714.

Muñoz-Sandoval, A. F., Cummins, J., Alvarado, C. G., & Ruef, M. L. (2005). *Bilingual verbal ability tests. Normative update*. Boston, MA: Houghton Mifflin Harcourt.

O'Toole, C., & Fletcher, P. (2010). Validity of a parent report instrument for Irish-speaking toddlers. *First language*, *30* (2), 199–217.

O'Toole, C. & Hickey, T.M. (2017). Bilingual language acquisition in a minority context: Using the Irish-English Communicative Development Inventory to track acquisition of an endangered language. *International Journal of Bilingual Education and Bilingualism*, *20* (2), 146–162.

O'Toole, C., Gatt, D., Hickey, T.M., Miękisz, A., Haman, E., Armon-Lotem, S., Rinker, T., Ohana, O., dos Santos, C., & Kern, S. (2017). Parent report of early lexical production in bilingual children across varied contexts: A CDI study. *International Journal of Bilingual Education and Bilingualism*, *20* (2), 124–145.

Pearson, B. Z. (2010). We can no longer afford a monolingual norm. *Applied Psycholinguistics*, *31* (02), 339–343.

Peña, E. D., Bedore, L., & Kester, E. S. (2015). Discriminant accuracy of a semantics measure with Latino English-speaking, Spanish-speaking, and English–Spanish bilingual children. *Journal of Communication Disorders*, *53*, 30–41.

Peña, E. D., Gutierrez-Clellen, V., Iglesias, A., Goldstein, B., & Bedore, L. (2014). *Bilingual English-Spanish assessment*. San Rafael, CA: AR-Clinical Publications.

Pham, G., & Kohnert, K. J. (2014). A longitudinal study of lexical development in children learning Vietnamese and English. *Child Development*, *85* (2), 767–782.

Potgieter, A., and Southwood, F. (2016). A comparison of proficiency levels in four-year-old monolingual and trilingual speakers of Afrikaans, isiXhosa and South African English across SES boundaries, using LITMUS-CLTs. *Clinical Linguistics & Phonetics*, *30* (2), 87–100.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: The MIT Press.

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology*, *28* (3), 406–413.

Rinker, T., Budde-Spengler, N., & Sachse, S. (2017). The relationship between L1 and L2 lexical development in young Turkish-German children. *International Journal of Bilingual Education and Bilingualism*, *20* (2), 218–233.

Roy, P., & Chiat, S. (2012). Teasing apart disadvantage from disorder. In C. R. Marshall (Ed.), *Current issues in developmental disorders* (pp. 125–150). New York: Psychology Press.

Scheele, A. (2010). *Home language and mono- and bilingual children's emergent academic language. A longitudinal study of Dutch, Moroccan-Dutch, and Turkish-Dutch 3- to 6-olds* (Unpublished PhD dissertation). Utrecht University.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15* (4), 426–445.

Tomasello, M., Merriman, W. (1995). *Beyond names for things: Young children's acquisition of verbs*. Mahwah, NJ: Erlbaum.

Uchikoshi, Y. (2006). English vocabulary development in bilingual kindergarteners: What are the best predictors? *Bilingualism, Language and Cognition*, *9* (1), 33–49.

Umbel, V. M., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development*, *63* (4), 1012–1020.

Van Tuijl, C., Leseman, P. P. M., & Rispens, J. (2001). Efficacy of an intensive home-based educational intervention programme for 4-to 6-year-old ethnic minority children in the Netherlands. *International Journal of Behavioral Development*, *25*(2), 148–159.

## Appendix 1. Studies analysing bilingual lexical development in various language pairs

| Language pairs | Tasks used | References |
|---|---|---|
| English–Spanish | CDI | Conboy and Thal (2006) |
| | Comprehension (PPVT, TVIP), naming (Expressive One Word Picture Vocabulary Test, EOWPVT – Spanish and English) | Allman (2005) |
| | Comprehension (PPVT + TVIP) | Barnett and Lamy (2006) |
| | BESOS | Bohman, Bedore, Peña, Mendez-Perez and Gillam (2010) and Peña, Bedore, and Kester (2015) |
| | Woodcock Language Proficiency Battery – Revised for English and Spanish | Duursma et al. (2007) and Mancilla-Martinez and Lesaux (2011) |
| | Woodcock Language Proficiency Battery – Revised for English and Spanish + PPVT+TVIP | Uchikoshi (2006) |
| | Naming (modified Receptive One-Word Picture Vocabulary Test) | Gorman (2012) |
| | Comprehension (PPVT + TVIP), naming (TELD-3 + PLS-3) | Hammer et al. (2012) |
| | Naming task prepared for this study | Kohnert, Bates, and Hernandez (1999) |
| English–French | Comprehension (PPVT + EVIP), naming (EOWPVT + French version of EOWPVT) | Chiang and Rvachew (2007) |
| | Comprehension (PPVT + EVIP – Canadian French PPVT), naming (Expressive Vocabulary subtest of CELF, expression-vocabuilare subtest of N-EEL) | Thordardottir (2011) |
| English–Greek | Comprehension & naming (Wechlser Preschool Primary Scale of Intelligence - Revised) | Loizou and Stuart (2003) |
| English–Hebrew | CDI | Armon-Lotem and Ohana (2017) |
| English–Hmong | Receptive + picture naming tasks developed for this study | Kan and Kohnert (2005) and Kohnert, Kan and Conboy (2010) |
| English–Irish | CDI | O'Toole and Fletcher (2010) and O'Toole and Hickey (2017) |
| English–Maltese | CDI | Gatt (2017) |
| English–Mandarin | PPVT translated from English | Dixon (2011) |
| English–Polish | CDI | Miękisz et al. (2017) |
| English–Vietnamese | ROWPVT + EOWPVT translated from English | Pham and Kohnert (2014) |
| English–Samoan | Receptive + picture naming tasks developed for this study | Hemsley, Holm and Dodd (2013) |
| Dutch–Arabic and Dutch–Turkish | Part of the Test for Bilingualism (Toets Tweetaligheid) (Verhoeven, Narain, Extra, Konak, & Zerrouk, 1995) | Messer (2010) |
| | Comprehension (Diagnostic Test of Bilingualism) + instrument specifically developed for research with bilingual immigrant children | Scheele (2010) |
| | Comprehension + naming (parts of the Diagnostic Test of Bilingual Development) | Van Tuijl, Leseman, and Rispens (2001) |
| Dutch–French | CDI | De Houwer et al. (2014) |
| German–Turkish | CDI | Rinker, Budde-Spengler, and Sachse (2017) |
| Luxembourgish–Portuguese | EOWPVT and BPVS translated from English | Engel de Abreu, Baldassi, Puglisi and Befi-Lopes (2013) |
| | EOWPVT and PPVT translated from English | Engel de Abreu, Cruz-Santos and Puglisi (2014) |
| Luxembourgish–German–French | EOWPVT translated from English | Engel de Abreu and Gathercole (2012) |
| Russian–German | Picture-naming task comprising pictures of objects and actions | Klassert, Gagarina, and Kauschke (2014) |