VALIDATING THE PSYCHOLINGUISTIC ASPECTS OF CLT: EVIDENCE FROM POLISH AND NORWEGIAN

CORRESPONDING AUTHOR:
Pernille Hansen, University of Oslo: pernille.hansen@iln.uio.no, (0047) 997 74 337

AUTHORS:
Pernille Hansen (pernille.hansen@iln.uio.no)
University of Oslo

Hanne Gram Simonsen (h.g.simonsen@iln.uio.no)
University of Oslo

Magdalena Łuniewska (magdalena.luniewska@psych.uw.edu.pl)
University of Warsaw

Ewa Haman (ewa.haman@psych.uw.edu.pl)
University of Warsaw

ABSTRACT

The novel assessment tool Cross-Linguistic Lexical Tasks (CLT) aims for comparable cross-linguistic assessment of multilingual children’s lexical skills by basing each language version on two language-specific variables: age of acquisition (AoA) and complexity index (CI), a novel measure related to phonology, morphology, exposure and etymology. This paper investigates the validity of this methodology, asking whether the underlying language-specific properties are robust predictors of children’s performance. The Polish and Norwegian CLTs were used to assess lexical skills in 32 bilingual Polish-Norwegian, 34 monolingual Norwegian and 36 monolingual Polish children. The effects of AoA and CI were investigated through regression analyses and contrasted with frequency in child directed speech (CDS) and imageability, two known predictors of lexical development. AoA was a reliable predictor of performance within all parts of CLT, in contrast to CI. Apart from AoA, only exposure and CDS frequency had a significant effect within both monolinguals and bilinguals. These results indicate that CLT assesses lexical skills in a cross-linguistically comparable manner, but suggest a revision of the CI measure.

Keywords: lexicon, minority language, bilingual assessment, CLT, age of acquisition, imageability
All words have inherent properties linked to form, meaning and usage pattern affecting when they are acquired and how easily they are processed in language production and comprehension. Due to cultural and linguistic differences, words for the same concept may be acquired at a different age in different languages, and they could be more easily retrieved from memory in some languages than in others.

Tools for lexical assessment of multilingual pre-schoolers are rare, and those that exist have so far focused on specific language pairs (for a review, see Haman et al., this volume). To assess multilingual children beyond these language pairs, researchers and clinicians have been compelled to use tools constructed for monolingual populations, such as the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 2007). However, as PPVT is translated directly between languages without fully accounting for cross-linguistic and cross-cultural differences, the test may be linguistically or culturally biased (Haitana, Pitama, & Rucklidge, 2010; Peña, 2007). Combining two language versions of PPVT to assess bilingual children has nevertheless been a common procedure for decades (Bialystok, 1988; Davidson, Jergovic, Imami, & Theodos, 1997; Davidson, Raschke, & Pervez, 2010; Diaz, 1985).

In contrast, the novel lexical assessment tool CLT aims for full cross-linguistic comparability, in order to aid identification of language impairment in the multilingual population. To achieve this, the tool is constructed individually for each language based on two inherent, language-specific properties: an estimation of age of acquisition (AoA) of the target words, and a composite measure of the target words’ complexity (Haman, Łuniewska, & Pomiechowska, 2015). This paper investigates the validity of this methodology through analyses of CLT results from Polish-Norwegian bilinguals and Polish and Norwegian monolinguals. We ask whether the language-specific properties underlying CLT are robust predictors of children’s performance, and compare these properties to two factors demonstrated to affect language acquisition and processing: frequency in child directed speech (CDS) (Goodman, Dale, & Li, 2008), and imageability (Ma, Golinkoff, Hirsh-Pasek, McDonough, & Tardif, 2009; McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2011).

The properties underlying CLT

Every language version of CLT consists of four task sets with 32 items in each: comprehension and production of nouns and verbs. Haman et al. (2015) present several reasons for this setup: For instance, nouns tend to precede verbs in lexical development (Bates et al., 1994; Caselli et al., 1995; Gentner, 1982; Hansen, 2016; Hao et al., 2015), but this may not hold for every language (Tardif, 1996; Tardif, Shatz, & Naigles, 1997), and a cross-linguistic lexical assessment tool should thus not be limited to nouns. Regarding the two task types, lexical comprehension is viewed as a more accurate measure (Clark, 2009), but production could differentiate better between children with and without a language impairment (Messer & Dockrell, 2006).
The CLT target words for the four task sets are selected from a list of 158 nouns and 142 verbs chosen for their consistency in meaning across languages and cultures (see Haman et al., 2015). CLT is based on two language-specific measures: subjective AoA and complexity index (CI). Measures based on child language data or CDS were considered in the development of CLT (Haman, 2010, 2011; Haman, Szewczyk, Łuniewska, & Pomiechowska, 2011), but eventually abandoned because comparable data could not easily be obtained across the more than 30 languages involved. Notably, no semantic measure was included in the construction procedure.

**Subjective age of acquisition**

The construction of the tasks is based on subjective AoA. For each language version of the tasks, at least 20 native speakers of the language have estimated how old they were when they could understand each CLT word, on a scale from 0 (i.e. during the first year of life) to 18 (i.e., at age 18 or later) (Łuniewska et al., 2016). Comparing results from 25 different languages, Łuniewska et al. (2016) report the method to be both reliable, as judgments corresponded across languages, and valid, as the AoA judgments correlated with CDI data on early lexical development within six different languages, including Polish. The correlation with CDI data has also been confirmed for Norwegian (Lind, Simonsen, Hansen, Holm, & Mevik, 2015). AoA was included in the CLT construction procedure as a substitute of direct data on child language or CDS, and it is thus interesting to see how it compares to CDS frequency in Polish and Norwegian.

**Complexity**

To account for cross-linguistic differences in word complexity, a composite measure, the so-called Complexity Index (CI) was devised specifically for the CLT construction (Haman et al., 2015, pp. 212-214). Table 1 lists the variables included in CI, and their individual contribution to the index.

As indicated in table 1, the first four factors pertain to the words’ phonology. Word length in phonemes has the greatest impact on the score: word length in phonemes is normalised within word class by subtracting the mean word length and dividing by the standard deviation. This result value is doubled ‘in order to emphasize the impact of this factor on overall complexity’ (Haman et al., 2015, p. 226).

The next three factors relate to morphology, followed by two factors concerning children’s exposure to the referent. The last criterion included in the CI is whether the word in question is a recent loanword (Haman et al., 2015). Loanwords may stand out when it comes to phonology and morphology, and lexical borrowings are more common within some semantic domains than others (Haspelmath, 2008).

To exemplify the CI calculation, the Norwegian noun *løk* ‘onion’ has a CI of -0.56: With only three phonemes, it has a normalised word length of -1.03. There are no consonant clusters or initial fricatives, and the word is

---

1 For Norwegian, the instructions were essentially the same, but the design was slightly different, as the study was a part of the preparations for the database *Norwegian Words*. The procedure is presented in Lind, Simonsen, Hansen, Holm, and Mevik (2015).
not a derivation, has no prefixes or suffixes, and consists of one stem only (1 point). Onions were judged as available to direct experience and quite often accessible for preschool children (0.5 points). Finally, it is not a recent loanword.

\[CI_{\text{onion}} = 2 \times (-1.03 + 1 + 0.5 + 0) = -0.56\]

On the other hand, the Polish verb *telefonować* ‘talk on the phone’ has a CI of 7.44: The word is relatively long with eleven phonemes, amounting to a normalised word length of 1.72. It has no consonant clusters or initial fricatives and consists of only one stem (1 point), but is a derivation (1 point) with a suffix (1 point). The action of talking in the phone is judged as available to direct experience and very often accessible for preschoolers. Finally, *telefonować* is a recent loanword (1 point).

\[CI_{\text{telefonować}} = 2 \times 1.72 + 3 + 0 + 1 = 7.44\]

**AoA and CI in the CLT construction**

The 300 potential CLT target words are divided into two levels of AoA and CI, depending on whether they are below or above the average for their word class in a given language. The CLT construction ensures that target words for each of the four task sets are balanced between the CI and AoA levels. Table 2 shows the

<table>
<thead>
<tr>
<th>Table 1. Variables included in the complexity index (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Phonology</strong></td>
</tr>
<tr>
<td>Word length in phonemes</td>
</tr>
<tr>
<td>Word-initial fricative or affricate?</td>
</tr>
<tr>
<td>Word-initial consonant cluster?</td>
</tr>
<tr>
<td>Word-medial consonant clusters?</td>
</tr>
<tr>
<td><strong>Morphology</strong></td>
</tr>
<tr>
<td>How many stems?</td>
</tr>
<tr>
<td>Is the word a derivation?</td>
</tr>
<tr>
<td>Is there a prefix or suffix added to the stem?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Exposure</strong></td>
</tr>
<tr>
<td>Is the object/action available to direct experience in your country?</td>
</tr>
<tr>
<td>How often would preschool children in your country have access to the object/activity?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Is the word a loanword?</td>
</tr>
</tbody>
</table>
Hansen, Simonsen, Łuniewska and Haman

Table 2. Distribution of low and high CI and early and late AoA in Polish and Norwegian CLTs

<table>
<thead>
<tr>
<th></th>
<th>Polish targets</th>
<th>Norwegian targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early AoA</td>
<td>60</td>
<td>68</td>
</tr>
<tr>
<td>Late AoA</td>
<td>68</td>
<td>60</td>
</tr>
<tr>
<td>Low CI</td>
<td>68</td>
<td>70</td>
</tr>
<tr>
<td>High CI</td>
<td>60</td>
<td>58</td>
</tr>
</tbody>
</table>

distribution of low and high CI and early and late AoA target words in the Polish CLT (Haman, Łuniewska, Pomiechowska, Szewczyk, & Wodniecka, 2012) and the Norwegian CLT (Simonsen, Hansen, & Łuniewska, 2012). There is no significant difference between the two languages in the distribution of early and late AoA ($\chi^2(1, n=256)=0.77, p=.38$) or low and high CI ($\chi^2(1, n=256)=0.02, p=.90$) among the target words. There is also no language difference in the values of AoA (median_{PL} 3.90, median_{NO} 3.85, W=8376, $p=.76$) or CI (mean_{PL} 2.73, mean_{NO} 2.66, $t=1.95, p=.85$).

CDS frequency

Word frequency affects both lexical processing (Bates et al., 2003; Strain & Herdman, 1999) and children’s lexical development (Goodman et al., 2008; Hansen, 2016). Concerning processing, Bates et al. (2003), who performed a picture naming task across seven different languages (English, German, Spanish, Italian, Bulgarian, Hungarian and Mandarin), reported a highly significant frequency effect on naming response time within all languages in adult native speakers. When it comes to lexical development, Goodman et al. (2008), who investigated the lexical development of children acquiring American English, demonstrated that within a word class, young children acquire words earlier the more frequent they are. They found CDS frequency as a far better predictor of lexical development than frequency in speech between adults (Goodman et al., 2008); these findings were confirmed for Norwegian by Hansen (2016).

As discussed, CDS frequency was initially considered as a background factor for CLT, but rejected because it could not be obtained for all the languages involved. However, there are available frequency lists for both languages investigated here, based on eight corpora and about 800,000 words for Polish (Haman, Etenkowski, et al., 2011) and two corpora and about 67,000 words for Norwegian (Hansen, 2016).

Imageability

A word’s imageability is defined as the degree to which it arouses a mental image or a sensory experience (Paivio, Yuille, & Madigan, 1968), and previous research has demonstrated that words are acquired earlier (Bird, Franklin, & Howard, 2001; Ma et al., 2009; McDonough et al., 2011) and processed faster (Strain & Herdman, 1999) the more imageable they are. Imageability is similar to concreteness, but also depends on
experience: Words denoting feelings are typically abstract, often experienced, and quite imageable (e.g., *anger*), whereas words denoting rarely experienced objects may be concrete, but hard to imagine as they are seldom experienced (e.g., *armadillo*) (Bird et al., 2001; Paivio et al., 1968).

Imageability is an interesting measure in the analysis of CLT results as no semantic measure was included in the CLT construction. The relationship between imageability and experience is also interesting, as the CI measure does have an experience related component. Imageability ratings are currently not available for Polish. The analyses of imageability effects are hence only carried out on the Norwegian data.

**Research questions**

In this paper, we ask the following three research questions: First, are nouns easier than verbs and comprehension tasks easier than production tasks for both monolinguals and bilinguals, in line with previous research? Second, are the language-specific properties underlying CLT robust predictors of children’s performance in Polish and Norwegian? Third, how do these properties compare to other, well-known predictors of lexical development available for the two languages? Concerning the second question, we have the following two hypotheses:

1. We expect that words with an early AoA are easier than words with a late AoA
2. We expect that words with a low CI are easier than words with a high CI

If the second hypothesis does not hold, some of the CI components could still have an impact. We will analyse the possible effects of the components related to phonology, morphology, exposure and loanword status (see table 1) to investigate this. Regarding the third question, we are interested in how well these two measures can account for children’s performance, as well as in how they compare to the properties underlying the CLT. We have the following hypotheses:

3. We expect highly frequent words to be easier than words with a low frequency
4. We expect words to be easier the more imageable they are

We will analyse results from monolinguals as well as bilinguals. Monolingual results from carefully age-matched groups are suitable to investigate whether CLT has reached the aim of cross-linguistic comparability, but at the same time, the tool is created for cross-linguistic assessment of multilingual children; taking a closer look at how the various factors affect bilingual children’s performance may thus indicate whether CLT is suitable for its target population.
METHODS

Participants

The data set in this study consists of Norwegian and Polish CLT results from 32 children acquiring Polish as a first language and Norwegian as an early second language, as well as 34 monolingual Norwegian and 36 monolingual Polish peers. The gender distribution and age span are given in table 3. Concerning gender, the monolingual groups are quite balanced, whereas boys are overrepresented in the bilingual group – since it is the smallest group, we still chose to include all the boys in the study.

The children were recruited via day-care centres, the bilinguals and the Norwegian monolinguals in the Oslo region in eastern Norway and in Hordaland on the Norwegian western coast, and the Polish monolinguals in Warsaw, central Poland. Norwegian and Polish pilot versions of the background questionnaire Parents of Bilingual children Questionnaire (PABIQ—(COST Action IS0804, 2011; Kuś, Otwinowska, Banasik, & Kiebzak-Mandera, 2012; Simonsen & Hansen, 2012; Tuller, 2015) were used to assess the children’s developmental history and linguistic and socio-economic background. This background information was used to ensure that none of the children were at risk of language impairment, and that the children fulfilled the inclusion criteria regarding language exposure presented below. For a discussion on the bilingual children’s lexical skills in relation to their linguistic background, see (Hansen et al., 2016).

For the bilingual group we set as inclusion criteria that the children had been exposed to Polish from birth, that at least one parent spoke Polish at home, and that the children were exposed to Norwegian in their day-care centre. These are quite loose criteria, and as a result, the group is quite diverse: Although they all predominantly heard Polish at home, two of the children heard another language from their father (Turkish and Urdu respectively), and one is reported to have heard English in addition to Polish from both parents. According to the parents’ judgments of the children’s language skills and language use, most of the bilinguals were dominant in Polish, and they were thus expected to score higher on CLT in Polish than in Norwegian.

Concerning the monolingual Norwegian group, we set as inclusion criteria that Norwegian was the L1 of both parents and that the child was predominantly exposed to Norwegian. However, Norway is a multilingual society, and this is evident in our data. Two of the 34 monolingual Norwegian participants had heard both Norwegian and English from birth, and for ten of the children, parents reported some exposure to one or more languages other than Norwegian.

Table 3. Participants

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>F:M</th>
<th>Mean age (SD)</th>
<th>Age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilingual Polish-Norwegian</td>
<td>32</td>
<td>9:23</td>
<td>4.9 (0.9)</td>
<td>3.5 – 5.9</td>
</tr>
<tr>
<td>Monolinguals Norwegian</td>
<td>34</td>
<td>17:17</td>
<td>4.9 (0.9)</td>
<td>3.6 – 5.11</td>
</tr>
<tr>
<td>Monolingual Polish</td>
<td>36</td>
<td>18:18</td>
<td>4.10 (0.6)</td>
<td>4.0 – 5.10</td>
</tr>
</tbody>
</table>

PABIQ is in part based on the ALEQ (Paradis, 2011) and the ALDeQ (Paradis, Emmerzael, & Duncan, 2010).
languages in addition to Norwegian. Nevertheless, all 34 children were primarily exposed to Norwegian both in day-care and at home.

Whereas Norway has undergone a positive net migration the last 50 years, with more people entering the country than leaving it, Poland is in the opposite situation (United Nations, 2013). Although other languages than Polish are spoken within the country’s border, the Polish society is far less multilingual than Norway. Hence, the 36 Polish monolingual participants are more homogenous: they were all exposed to Polish only, apart from some English classes in the day-care centres.

**Procedure**

All children were tested individually in their own day-care centre in a separate, quiet room. The order of each item in each task set was fixed, but the order of the four task sets (comprehension and production of nouns and verbs) was counterbalanced across children. Prior to testing, the children were given some time to get to know the experimenter. We took breaks between each task set, and the whole session lasted about 15 minutes. For the bilingual group, the two languages were assessed separately, about a week from each other, with the order counterbalanced across the children. The Norwegian assessment was carried out by a native speaker who did not know Polish, whereas a majority of the experimenters carrying out the bilinguals’ Polish assessment were L2 learners of Norwegian. However, the language of instruction and conversation was Polish. In the comprehension tasks, the child is asked a question (e.g. ‘Where is the zebra?’ or ‘Who is drinking?’) and presented with a picture board of four pictures. In the production tasks, the child is presented with only one picture and asked what the object is or what is happening in the picture. We used the computer version of CLT for both languages, where these questions are pre-recorded and presented in a sequence controlled by the computer program: Each question is played 150 ms prior to the corresponding picture or picture board.

**Ethical considerations**

The assessment of Polish monolinguals in Poland was approved by the Committee for Research Ethics at the Faculty of Psychology, University of Warsaw. The assessment of bilinguals and Norwegian monolinguals in Norway was approved by the Norwegian Social Science Data Services. Informed consent was obtained from

<table>
<thead>
<tr>
<th>Table 4. Number of children included for each task set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant group</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Bilinguals: Polish</td>
</tr>
<tr>
<td>Bilinguals: Norwegian</td>
</tr>
<tr>
<td>Norwegian monolinguals</td>
</tr>
<tr>
<td>Polish monolinguals</td>
</tr>
</tbody>
</table>
the parents, and the children were informed that they could stop the assessment at any point. Some chose to do so, and in addition, some of the bilinguals refused the whole assessment in one of their languages. In this paper, children from whom we do not have full results are not automatically excluded, but we did exclude results from a whole task set if a child refused to finish it. Table 4 shows the number of children from each group who are included for the four task sets.

Assessment and scoring

In the computer version of CLT, comprehension responses are scored automatically. The production responses are recorded and transcribed and scored manually according to a set of criteria set by members of COST Action IS0804 involved in the development and early piloting of the CLT. For analyses of word related factors such as in this paper, only responses including the stem(s) of the target word count as correct answers. Any phonological errors, unexpected or wrong inflections and derivations (within or across word class) are disregarded. Similarly, novel words and responses including words or morphemes from another language are scored as correct as long as they contain the stem of the target word in the language assessed. The word related variables investigated here are related to the exact word forms selected as targets in the tasks, and hence, synonyms and regional variants are scored as incorrect.3

Research assistants, native speakers of the respective language transcribed and coded all the production responses. The first author (a native Norwegian speaker) and the third author (a native Polish speaker) listened through and checked the coding of all production responses from bilinguals together to uncover and correctly score any response involving both languages.

Word-related properties and cross-linguistic comparability

As discussed in the introduction, CLT is constructed based on a strict procedure designed to ensure cross-linguistic comparability. The targets are balanced between early and late AoA and high and low CI in both Polish and Norwegian, with no significant differences between the two languages. In contrast, the construction procedure does not balance the targets for neither CDS frequency nor the individual CI components. Several of these factors are biased towards low values; the reason could be that the target words have been selected among words assumed to be acquired by the age of six. Regarding morphological complexity, there is a difference between the two languages, possibly reflecting a more complex morphology in Polish than in Norwegian. There are also significant dependencies between some of the factors. In this section, we describe the measures we have taken to account for these conditions in our statistical analyses.

3 The line between correct and erroneous responses in the production tasks is slightly different in analyses of individual children’s performance on the tasks. Here, synonyms and regional variants may count as correct responses.
Concerning the phonological CI component, the main contributor to the score is word length in phonemes, which is standardised within each word class for each language. Although the Polish targets are slightly longer than the Norwegian targets (median\text{PL} 6, median\text{NO} 5, W=6060.5, p<.001), there is no cross-linguistic difference in the phonological CI component (mean\text{PL} 0.85, mean\text{NO} 1.11, t=-1.025, p=.31), and the values are fairly evenly distributed (see figure 1).

Both languages have a strong bias towards morphologically simple words and words with a high exposure. The morphological CI component ranges from 1 to 3, and the exposure related CI component ranges from 0 to 2, but for both components, the majority of the target words have the lowest value. We remedy this skewness by dividing morphology and exposure into two levels: Any word with a morphology complexity of 1 is categorised as simple, and words with higher values are categorised as complex. For exposure, words with a score of 0 are categorised as high exposure words, and other words are categorised as low exposure words. The distributions of these categories are given in table 5, along with the distributions of loanword status.

As is evident in the table, there is a strong bias within both languages towards morphologically simple, high exposure and non-loanwords. This bias is equally strong in both languages for exposure ($\chi^2(1, n=256)=0.07, p=.79$) and loanword status ($\chi^2(1, n=256)=0.77, p=.38$), but for the morphological component, the bias towards simple words is stronger for Norwegian than for Polish ($\chi^2(1, n=256)=5.11, p=.02$). The difference in morphology is not unexpected, as Polish has a rich morphology (Laskowski, 1998) and Norwegian is a morphologically simple language (although slightly more complex than English) (Ragnarsdóttir, Simonsen, & Plunkett, 1999).

Figure 1. Distribution of phonological CI in Norwegian (left) and Polish (right)
Both CDS frequency lists (Haman, Etenkowski, et al., 2011; Hansen, 2016) are skewed towards few occurrences, as frequencies tend to be, and hence logarithmically transformed (on the formula ln(x+1)). The Polish and Norwegian frequency lists are very different in size and hence not directly comparable. To analyse the frequency effects across languages, the (transformed) frequencies are centred on zero within each language through robust standardisation (based on medians and median absolute deviations). There is no significant difference in the standardised scores between the two languages ($W=8055, p=.817$).

**Imageability**

This property is only available for Norwegian, through the database *Norwegian Words* (Lind et al., 2015). As discussed by Simonsen, Lind, Hansen, Holm, and Mevik (2013), the imageability ratings available in the database are skewed towards high values. This is also true for the CLT target words (6.44 SD 0.44), and the imageability ratings were hence transformed on an exponential scale for the analyses here. According to a Shapiro-Wilk test, the distribution of the transformed ratings (680 SD 254) does not depart significantly from a normal distribution ($W=0.99, p=.34$).

**Interactions between variables**

To assess potential interactions between the word-related variables, we investigated dependencies between categorical variables (e.g. AoA level) with Pearson’s Chi-squared test, the correlation continuous variables (e.g. frequency) with Spearman’s rank sum correlation, and the relationship continuous and categorical variables with two-sample t-tests or Wilcoxon rank-sum tests, depending on whether the subsets were normally distributed. We investigated each language separately, running 18 pairwise tests for Polish and 26 for Norwegian (the same as for Polish, plus eight involving imageability). To control for the false discovery rate, that is, the expected proportion of false positives when carrying out multiple pairwise tests, p-values were adjusted with the Benjamini-Yekutieli procedure (2001).
Significant interactions were found in seven cases for each language (see table 6): In both languages, early-AoA words tend to be high in frequency and exposure, and high exposure words tend to be highly frequent. Among the Polish target words, late-AoA words tend to be loanwords, and loanwords tend to be nouns. The nouns are more phonologically complex (M=1.61) than the verbs (M=0.09), and have a higher CI level. Among the Norwegian target words, morphologically complex words have a higher frequency than morphologically simple words, and there is an intermediate, negative correlation between frequency and the phonological CI component. Regarding imageability, the Norwegian nouns are more imageable (M=856) than the verbs (M=505), in accordance with previous cross-linguistic research (Bird, Ralph, Patterson, & Hodges, 2000; Cortese & Fugett, 2004; Ma et al., 2009; Masterson & Druks, 1998; Simonsen et al., 2013), and high-exposure words (M=732) are more imageable than low-exposure words (M=579).

Analysis

This paper investigates monolingual and bilingual results separately. Within each group of children, we calculate the proportion of correct answers for each word according to the criteria discussed above. There is a ceiling effect in the data, especially within monolingual children (see figure 2), but there are at the same time some outliers in the lower end of the scale: Not a single child in neither group responds correctly on the Norwegian verb production target haike ‘hitchhike’, and none of the bilinguals and only eight of 35
monolinguals respond correctly to the Polish noun production target *rakieta* ‘racket’. To avoid undue influences from these items without excluding them altogether (Wilcox, 2012; Wilcox & Keselman, 2012), the data are analysed through robust regression with MM-type estimators, using backwards stepwise selection, first on a model including task type, word class, AoA, CI and any two-factor interaction and a model where CI is replaced by its four components. We then turn to frequency and imageability, including each of these to a model including the significant predictors of the first analyses. Statistics are performed with R 3.2.2 (R Core Team, 2015), using the robust (Wang et al., 2014) and robust HD (Alfons, 2014) packages. To compare the relative predictive power of each variable, standardised beta (β) coefficients are calculated by running regression models with all continuous factors centred through robust standardisation.

**RESULTS**

Our first research question regarded the descriptive statistics: We asked whether the nouns were easier than the verbs and the comprehension tasks were easier than the production tasks, as expected from the literature. Assuming that the tasks are cross-linguistically comparable, there should be no difference in performance between Polish monolinguals and their Norwegian peers. For the bilinguals, who had all heard Polish from birth and acquired Norwegian as an early second language, we expected higher scores in Polish.

The results are in accordance with these expectations (see figure 2). For the monolinguals, both task type ($\beta=-0.66$, $t(252)=-4.22$, $p<.001$) and word class ($\beta=-0.59$, $t(252)=-3.76$, $p<.001$) are significant predictors of the children’s performance. For the bilinguals, we only found a significant main effect of word class ($\beta=-0.29$, $t(252)=-2.78$, $p=.006$).

![Figure 2. Distribution of word scores by task set, language and participant group](image-url)
There is also a significant effect of the interaction between the two (\(\beta=-0.40, t(244)=-2.38, p=.018\)). That is, the difference between nouns and verbs is larger within the production tasks. There was no language effect on the monolingual results, whereas the bilinguals scored significantly lower on the Norwegian tasks (\(\beta=-0.52, t(244)=-3.31, p=.001\)). The language also influenced the effect of task type (\(\beta=-0.44, t(244)=-2.63, p=.009\)): The difference between production and comprehension is larger in Norwegian than in Polish.

The language-specific CLT properties: AoA and CI

We found a main effect of AoA within both monolinguals (\(\beta=-0.65, t(252)=-4.06, p<.001\)) and bilinguals (\(\beta=-0.30, t(244)=-2.03, p<.044\)). For the bilinguals, there is a significant interaction with task type (\(\beta=-0.35, t(244)=-2.10, p=.037\)) and language (\(\beta=-0.44, t(244)=-2.63, p=.009\)): The AoA effect is largest within production tasks, and larger for Norwegian than for Polish. This may be attributed to the overall patterns presented in figure 2; the interactions are significant where there is more variation to account for.

In contrast, there is no main effect of CI within either group. There is however one significant interaction: Relative to the overall language score, the children perform slightly higher on high-CI words in Norwegian than in Polish (\(\beta=0.35, t(244)=2.06, p=.04\)). As shown in figure 3, this difference is related to a weak tendency towards a higher performance on low CI words in Polish, whereas there is no difference in Norwegian.

Replacing the CI measure with its four components: (phonology, morphology, exposure and loanword status), we found no significant main effect, but some significant interactions for morphology and exposure: For the monolinguals, a complex morphology added to the difficulty of the production items (\(\beta=-1.46, t(227)=-2.02\),
and Haman (2015), the tendency is that the recent loanwords are somewhat easier than non-loanwords. A plausible explanation for this effect is that many of these recent loanwords are also cognates (e.g. Norwegian tiger and Polish tygrys, both meaning ‘tiger’).

### CDS frequency

To compare frequency to the effects of AoA and the exposure related CI, we performed backwards stepwise selection on a robust regression model including word class, task type, AoA, exposure, frequency and any two-factor interactions. We found no main effect of frequency, but a significant interaction with task type: Words in the production tasks were easier the higher their frequency for monolinguals ($\beta=0.54$, $t(237)=2.59$, $p=.010$) as well as bilinguals ($\beta=0.15$, $t(242)=1.97$, $p=.050$). In addition, frequency was significant in interaction with language for the bilinguals ($\beta=0.25$, $t(242)=3.02$, $p=.003$), with a larger effect within Norwegian than within Polish. This difference may be related to the larger variation in their Norwegian results.

Although the lemma frequencies have been standardised within language, the size difference between Polish and the Norwegian frequency lists may mask potential effects. We thus performed the same analyses within each language, finding significant positive effects for Polish, but not for Norwegian: Here, there was still no main effect, and frequency still had a positive effect within production tasks for both monolinguals ($\beta=0.51$, $t(108)=2.55$, $p=.012$) and bilinguals ($\beta=0.19$, $t(116)=2.70$, $p=.008$). In addition, frequency had a positive effect within verbs for the Polish monolinguals ($\beta=0.41$, $t(108)=2.18$, $p=.031$), and within late-AoA words for the bilinguals’ Polish assessment ($\beta=0.19$, $t(116)=2.55$, $p=.012$).

### Imageability

As mentioned, imageability ratings are only available for Norwegian, and hence only the results from the Norwegian CLT are analysed here. We performed backwards selection on a regression model including the same factors as for the frequency analysis presented above, including imageability instead of frequency. We found no effect of imageability on the bilingual results, but for the monolinguals, words in the production tasks ($\beta=0.62$, $t(115)=2.38$, $p=.019$), verbs ($\beta=0.81$, $t(115)=2.18$, $p=.031$) and late-AoA words ($\beta=3.33$, $t(115)=6.08$, $p<.001$) were easier the higher their imageability.

### DISCUSSION

This paper investigates the cross-linguistic comparability of the new lexical assessment tool CLT, asking whether the two language-specific properties underlying the tool are robust predictors of mono- and bilingual
children’s performance based on the example of Polish and Norwegian. We contrasted these properties, AoA and CI, to frequency in CDS and imageability.

Following Haman et al. (2015), we hypothesised that words with an early AoA should be easier than words with a late AoA, and that words with a low CI should be easier than words with a high CI. Our analyses of CLT results from Polish and Norwegian monolinguals and Polish-Norwegian bilinguals confirmed the hypothesis for AoA within both participant groups, whereas there was no main effect of CI as a compound measure.

The CI measure constitutes four components related to phonology, morphology, exposure and loanword status. For the phonological and morphological components, Haman et al. (2015) expected words to be harder the higher their complexity. A high exposure was expected to make words easier, and recent loanwords were assumed to be harder than non-loanwords. We found some support for two of these expectations: The morphological CI component added to the difficulty within monolinguals’ production results, and exposure had the expected effect within verbs for the monolinguals, and within production items for the bilinguals. Notably, ceiling effects could mask parallel effects within nouns and comprehension tasks. Overall, AoA was still a far more robust predictor than any of the CI components, especially when it comes to the bilinguals’ performance.

We expected words with a high frequency in CDS to be easier than words with a low frequency. This expectation was confirmed within the production tasks, and within the Polish results also within verbs for the monolinguals and within late-AoA words for the bilinguals. The effect of frequency is thus similar to the exposure related CI component. This is not unexpected, as we also found a significant difference in frequency between low- and high-exposure words within both Polish and Norwegian. Again, as in the case of CI components, ceiling effects could mask parallel effects within nouns and comprehension tasks.

For imageability, we expected words to be easier the more imageable they are. We only found support for this hypothesis among the (Norwegian) monolinguals, and only within the harder items; that is, within production tasks, for verbs and within late-AoA words. Ceiling effects may be the reason for the lack of an imageability effect in the other task sets.

Our findings raise two questions: First, why is AoA such a superior predictor of children’s CLT results? Second, why is there no effect of CI?

Why does AoA work so well?

There are at least two possible explanations to the AoA effect: First, AoA might be a good measure of which of the words the children have acquired. As discussed in the introduction, both Łuniewska et al. (2016) and Lind et al. (2015) demonstrate a strong correlation between this measure and parental reports on young
children’s lexical development. In our data, a word’s AoA was tightly connected to the exposure related component.

Second, AoA could affect how words are encoded in the lexicon or how easily they are retrieved. A vast amount of studies have found a robust effect of AoA on lexical processing for both children and adults in a variety of test designs, and the effect has often equalled or surpassed the effects of frequency and imageability (for a review, see Juhasz, 2005). Several different theories have been proposed to explain this effect. According to the cumulative frequency hypothesis (H. D. Ellis & Young, 1977), words that are acquired early are easier to retrieve because we have encountered them more times than words acquired later in life. If AoA measures the exposure to a word, and the exposure CI component measures the exposure to the denoted object or action, this could explain why exposure accounts for some of the same variation as AoA in our data.

According to Bates, Burani, D’Amico, and Barca (2001), the AoA effect is related to both frequency and semantic-conceptual properties. Assessing Italian speaking adults, they found AoA effects in both word reading tasks and picture naming tasks. They demonstrated that the effects on reading could be accounted for as frequency effects, while the effects on picture naming seemed to depend on both frequency effects and words’ semantic-conceptual properties (Bates et al., 2001).

Juhasz (2005) suggests that the AoA effect on semantic processing could be caused by differences in encoding or differences in retrieval between early and late acquired words: According to the network plasticity hypothesis (A. W. Ellis & Lambon Ralph, 2000), patterns that are entered into a network early affect the way the network structures itself, whereas some of this plasticity is lost when later patterns are entered. This suggests that the way our semantic networks are encoded may be more adapted to the first words we acquire than to the words acquired later. According to the semantic locus hypothesis (Brysbaert, Van Wijnendaele, & De Deyne, 2000; Van Loon-Vervoon, 1989), concepts that are new to us will be defined in terms of already known concepts. This way, words that are acquired early will often reside in dense semantic networks facilitating retrieval, while words acquired late will be connected to relatively few words and hence be harder to retrieve.

**Why is there no effect of complexity?**

We can think of three possible culprits for this lack of predictive power for CI. First, there could be a CI effect that is not revealed because the CLT target words are too similar in complexity. Second, some factors could work as expected, but be neutralised by other factors turning out to have an opposite effect of what was expected by Haman, Łuniewska, and Pomiechowska (2015). Third, the properties included in the compound measure may simply not play a decisive role in children’s lexical skills, at least not in the two languages investigated here.
Concerning the first hypothesis, the list of 300 possible target words for all language versions of CLT consist of words children are expected to know by the age of six, and they could all be too simple to trigger a CI effect. The distribution of morphological and exposure related complexity among the target words may be taken as support for this explanation: As discussed in the methods section, there is a bias towards simple morphology and high exposure. In order to test this hypothesis, we would need data on children’s performance on a set of words more complex than those included in the CLT, or to assess younger children.

As for the second hypothesis, there is some support for this explanation in the data. We did indeed find effects of exposure and morphology in interaction with other factors, and loanword status had a close to significant opposite effect on bilinguals than predicted by Haman, Łuniewska, and Pomiechowska (2015). However, this effect was limited to Norwegian, and no such effect was found on the monolingual performance.

The phonological CI component did not have any effect on the results, supporting the third hypothesis: the component does not seem to play a decisive role in children’s lexical skills, at least not at the age studied here. Below, we will discuss possible reasons for the lack of effect from the phonological CI and the limited effect from the morphological component, as it was only significant for the monolinguals’ performance.

The phonological CI component

The largest impact on overall phonological complexity comes from word length; the number of phonemes is doubled to emphasize the importance of this measure (Haman, Łuniewska, & Pomiechowska, 2015, p. 226). However, cross-linguistic research indicates that although longer words may be harder to acquire (Hansen, 2016) and process in naming tasks (Morrison, Ellis, & Quinlan, 1992) than short words, lexical acquisition may be more complex: Disyllables, not monosyllables, seem to be the most common word form in young children’s early vocabularies in most languages (Vihman & Croft, 2007). Furthermore, Garmann, Hansen, Simonsen, and Kristofferson (in press) report cross-linguistic differences in the word length of children’s early words: Young Italian children prefer disyllabic and polysyllabic words, whereas Norwegian, English and Swedish children are balanced between mono- and disyllabic words and Danish children prefer monosyllables. These differences may stem from differences in the children’s input: For Italian, Norwegian, English and Swedish, corresponding proportions were found in text samples.

These findings indicate that a measure that aims to account for phonological effects on lexical acquisition may need to take cross-linguistic differences into account, at least as far as word length is concerned. This could prove to be a challenging task, as comparable data on language-specific phonological patterns may not be available across the wide variety of languages for which CLT has been constructed so far. Garmann et al. (in press) used phonetic transcriptions of the Aesop’s fable *The north wind and the sun* to compare Italian, Norwegian, English, Swedish and Danish; such transcriptions are available in many languages and could potentially be used as a basis for a more refined measure of phonological CI component.
The morphological CI component

The morphological complexity consists of three questions: Whether the word is a compound, whether it is a derivation and whether it has any prefixes or suffixes. However, the number of morphemes in itself may not be vital for when children acquire words. Evidence from Turkish suggests that children can acquire a complex morphology very early as long as it is transparent (Aksu-Koç & Ketrez, 2003). Ragnarsdóttir et al. (1999) investigate the acquisition of past tense in Norwegian, a morphologically simple language, and Icelandic, which has a much more complex morphology. They report that although Icelandic children lag behind their Norwegian peers at age four, they have caught up with them by age six, concluding that morphological complexity may be most important at the early stages of development. Phonological properties may account for the cross-linguistic differences; for instance, vowel patterning in the present tense may function as a clue for Icelandic children, but not for Norwegian children (Ragnarsdóttir et al., 1999). Type frequency also played an important role for the acquisition of morphology within both languages.

These findings indicate that a measure that successfully accounts for morphological complexity may need to take properties such as transparency and type frequency into account. Again, this may prove challenging, given the vast variation in the morphology of the languages involved.

CONCLUSION

Should CI be revised or removed in subsequent construction of new CLT versions? The loanword related component could be removed, and a revision could potentially remedy the phonological and the morphological component. However, finding ways to account for the phonological and morphological properties across such a wide variety of languages as involved in the overall CLT design may prove to be a difficult task. Another option would be to keep only exposure in the CI, as this was the only factor that had the expected effect within both monolinguals and bilinguals. However, the strong correspondence between exposure and the more robust predictor AoA indicates that this measure may be redundant.

Alternatively, the CI could be removed entirely from the CLT construction. The aim for the CLT was a tool that would help in differentiating multilingual five-year-olds with SLI from their typically developing peers by assessing vocabulary in all of a child’s languages in a comparable manner (Haman et al., 2015). The question of how well CLT differentiates between children with and without SLI is addressed elsewhere in this volume (Kapalková & Slančová, this volume; Khoury Aouad Saliby, dos Santos, Kouba-Hreich, & Messarra, this volume). As for cross-linguistic comparability, we found no significant effect of language on the Polish and Norwegian monolinguals’ performance, indicating that for these two languages, the goal has been met.³

³ The cross-linguistic comparability of CLT results across a wide variety of languages is discussed by Haman, Łuniewska, Hansen, et al. (this volume).
This suggests that the CLT construction procedure has succeeded in balancing the linguistic and cultural differences between Polish and Norwegian.

It would be impossible to ensure comparable results across a wide variety of languages if neither AoA nor CI had any predictive power, but one may be enough: As AoA was a robust and significant predictor of CLT results from both monolinguals and bilinguals and CI had little effect, the lack of difference between the Polish and Norwegian CLT results could be a result of AoA alone. To make these conclusions more universal across languages, similar analyses involving the effects of AoA and CI should also be performed for other languages or language pairs.

ACKNOWLEDGEMENTS

We wish to thank all the participants, their parents and day-care centre employees for taking part in the study; Elisabeth Holm, Katarzyna Chyl, Ingeborg Ribu and Stefan Markiewicz for data collection and coding; Ewa Wapinska and Emilia Dymarczyk for their contributions to the data collection from bilinguals; members of COST Action IS0804 for discussions on the development of CLT and preliminary results from this study; Yeşim Sevinç and two anonymous reviewers for valuable comments on the manuscript.

This study was partially supported by the following institutions: National Science Centre (Poland; grant no. 809/N-COST/2010/0); Polish Ministry of Science and Higher Education (contract no. 0046/DIA/2013/42); Council of Norway through its Centres of Excellence funding scheme (project number 223265); and the European Cooperation in Science and Technology (COST) through its funding of Short-Term Scientific Missions (ECOST-STSM-IS0804-121112-023461). All pictures used for CLTs are subject of copyright of University of Warsaw (Poland).

DECLARATION OF INTEREST

The authors report no conflicts of interest
REFERENCES


Hansen, Simonsen, Łuniewska and Haman


Hansen, P. (2016). What makes a word easy to acquire? The effects of word class, frequency, imageability and phonological neighbourhood density on lexical development. First Language, Advanced online publication. DOI: 10.1177/0142723716679956.


Hansen, Simonsen, Łuniewska and Haman


