Running head: Word-related variables in vocabulary learning.

Title: Learning English through out-of-school exposure: how do word-related variables and proficiency influence receptive vocabulary learning?

Author 1: Vanessa De Wilde
Affiliations Author 1: Department of Translation, Interpreting and Communication, Ghent University; Artevelde University College Ghent
Author 2: Marc Brysbaert
Affiliation Author 2: Department of Experimental Psychology, Ghent University
Author 3: June Eyckmans
Affiliation Author 3: Department of Translation, Interpreting and Communication, Ghent University

Address for correspondence
Vanessa De Wilde
Groot-Brittanniëlaan 45
9000 Ghent
Belgium
vanessa.dewilde@ugent.be

Acknowledgements: We thank Geert De Meyer, Philip O'Neill, our students and colleagues for help with testing. We thank the schools, the teachers, the pupils and their parents for participating in the project. We thank the anonymous reviewers and the associate editor, Scott Crossley, for their suggestions and comments on previous versions of this manuscript.

Funding: This research was supported by Artevelde University College Ghent in the context of the PWO-project False Beginners.

To be published in Language Learning

Abstract

This study aims to investigate which word-related variables play a role in Dutch-speaking children's L2 word learning through out-of-school exposure prior to classroom instruction in the foreign language. Different measures were used to investigate the role of frequency, concreteness, cognateness, and age of acquisition in receptive vocabulary learning. The results show that cognateness and L1 age of acquisition are important predictors for receptive word knowledge for all the children. The findings confirm the importance of cognateness in vocabulary learning and show that less proficient learners tend to guess the meaning of words based on their L1. The results also show significant interactions between proficiency and cognateness, frequency, age of acquisition and concreteness, indicating that more proficient L2 learners are open to L2-related variables, such as L2 word frequency. This shows that word-related variables contribute in different ways according to learners' proficiency levels.

Keywords: incidental vocabulary learning, contextual word learning, cognates, lexical frequency, age of acquisition, concreteness

1. Contextual language learning through out-of-school exposure

Formal classroom teaching alone does not suffice to become proficient in a language. In order to acquire a large number of words, learners need more exposure to the language than what is traditionally offered in foreign language classrooms and thus formal learning should go hand in hand with learning in informal settings (Bybee & Hopper, 2001; Ellis, 2002; Ellis & Wulff, 2014). Recent research has looked at the effects of out-of-school English exposure on language learning. Studies investigating English language learning in young learners (Kuppens, 2010; Sylvén & Sundqvist, 2012; Lindgren &Muñoz, 2013; Jensen, 2017; De

Wilde, Brysbaert & Eyckmans, 2019) have shown that children are exposed to the English language through watching television, playing videogames and using social media; activities these children take part in, many of them on a daily basis. Some of these studies have demonstrated that contextual[1] language learning through out-of-school exposure takes place in children even before the start of formal instruction (Kuppens, 2010; Lefever, 2010; De Wilde et. al., 2019). These studies investigated children's language skills and vocabulary knowledge and show that some children are quite proficient in English. De Wilde et. al. (2019) studied contextual language learning through media exposure in Dutch-speaking Belgian 10-to-12 year old children and found that 25% of the participants have an A2-level (CEFR) for listening, 14% have an A2-level for speaking and 10% obtain an A2-level for reading and writing prior to any form of classroom instruction, which means that these children are able to deal with simple, straightforward information and can express themselves in familiar contexts (Council of Europe, 2001). However, the same study showed considerable differences between the children's language skills. The findings are in line with Lefever (2010) who found similar results in Icelandic eight-year-olds.

In the present study we will investigate contextual vocabulary learning in children in an informal setting where there has been no classroom instruction but abundant out-of-school exposure to L2.

2. Word-related variables involved in word learning.

Lexical proficiency is not an easy concept to define as there are many layers of knowledge. A distinction is often made between the breadth of vocabulary knowledge (i.e. the size of one's vocabulary) and the depth of vocabulary knowledge (how well different aspects of a word are known) as acquiring full knowledge of a word is a gradual process and it is possible to have

only partial knowledge of a word. However, independent of how word knowledge is defined, research has shown that a look at the (psycholinguistic) characteristics of words can help to better understand how learners process and produce a second language (Crossley, Salsbury, McNamara and Jarvis, 2011).

Traditionally, word frequency is thought to be the most important variable in vocabulary acquisition and researchers have linked language proficiency to the knowledge of less frequent words (e.g. the Lexical Frequency Profile by Laufer & Nation, 1995). Other variables investigated are concreteness, imageability, age of acquisition and cognateness, among others. A study by Crossley et. al. (2011), for instance, found that a variety of lexical indices predicted 44% of the variance in beginner to advanced level L2 writing tasks. Another study (Kim, Crossley & Shalicky, 2018) looked at the effects of word-related variables in reading comprehension. The results showed that word frequency, word concreteness and orthographic neighbourhood density predicted word processing times, together with text-level variables and the person variable 'L2 reading proficiency'. A third study, by Crossley and Salsbury (2010), illustrated the link between word-related variables and lexical production. The results showed that more frequent, more meaningful and more familiar words are produced more often. Again, this study illustrated the importance of different word-related variables in word learning.

Goriot, van Hout, Broersma, Lobo, McQueen and Unsworth (2018) explored the link between L1-related word variables and the results on an L2 vocabulary test. The authors tested Dutch learners of English in six different age groups and found that pupils scored higher on test items which were similar to their Dutch translation. They also found that, especially for younger learners, L1 frequency was predictive for the test scores whereas older learners' scores were predicted by L2 frequency. This is in line with findings in studies concerning receptive multilingualism, which is the ability of people to communicate in their

own language with people who speak a closely related but unknown language (Gooskens, van Heuven, Golubović, Schüppert, Swarte, & Voigt, 2018). One of the strategies used in receptive multilingualism is cognate guessing, which is guessing the meaning of words based on similarities with known cognates (Vanhove & Berthele, 2015).

In our study we wish to investigate the role of word-related variables in contextual word learning, prior to classroom instruction. Four variables, which are relevant for our study, will be discussed in more detail: frequency, cognateness, concreteness, and age of acquisition.

2.1. Frequency

One of the most widely investigated variables related to L1 word learning is word frequency. It has been observed that people know more high frequency words than low frequency words (see Brysbaert, Mandera & Keuleers, 2018, for a review of the word frequency effect).

Monaghan, Chang, Welbourne & Brysbaert (2017) showed that there are individual differences in the word frequency effect which are related to differences in the (amount of) language exposure. This means that language learners do not show the same frequency effect but that the effect depends on how much exposure learners have had to words. Learners with low exposure show a strong frequency effect for high-frequency words but not for low-frequency words (which they do not know). In contrast, speakers with high exposure also show a word frequency effect for low-frequency words, in line with the finding that beginners recognize and produce more high-frequency words than more advanced learners, who know more infrequent words (Laufer & Nation, 1995; Ellis, 2002; Crossley et. al., 2011).

In studies concerning contextual word learning in a second language, frequency of occurrence has also been proposed as one of the major factors contributing to vocabulary learning. Words are typically picked up when they have been encountered eight times or more (Elgort, Brysbaert, Stevens, & Van Assche, 2018; Gullberg, Roberts & Dimroth, 2012;

Pellicer-Sanchez & Schmitt, 2010). However, the amount of necessary encounters seems to depend on the input learners receive and studies by Webb and Chang (2017) and Bisson, Van Heuven, Conklin and Tunney (2014) suggest that learning from multimodal input already happens with fewer than eight exposures.

2.2. Cognateness

Cognates are words that have high orthographic and/or phonological overlap with their translation equivalent and in addition are very alike in meaning. The words can be identical (e.g. Dutch *bus* – English *bus*) or similar (e.g. Dutch *kalender* – English *calendar*).

Already in 1985, Palmberg wrote about the facilitative effect of cognates in contextual foreign language learning. In his study, 74 elementary school pupils (age 10), who had not received any English lessons prior to the study, took a translation test in which they were asked to write down Swedish equivalents of English words and expressions. The results showed that English words sharing formal similarities with Swedish words were more easily understood. Studies by Brenders, van Hell and Dijkstra (2011) and Poarch and van Hell (2012) also showed a cognate facilitation effect in word processing for children learning an L2.

A cognate effect was further observed in a study on contextual word learning from TV viewing (Peters & Webb, 2018) with 63 adult EFL learners who were in their first or second year at university. Chances of recognising the translation equivalent for a cognate were 2.5 times higher than those for non-cognates. This finding was also reported in earlier research (Lindgren & Muñoz, 2013; Vidal, 2011). Peters and Webb (2018) further ventured that cognateness could be more important for aural input than for written input, because learners cannot go back to spoken words and thus may have to rely more on L1 similarity.

Lindgren andnMuñoz (2013) ran a study with young learners from different language backgrounds to look into the influence of out-of-school exposure (watching, playing, listening), parents (educational level and use of the foreign language in the workplace) and cognate linguistic distance on children's performance on L2 listening and reading tests. Cognate linguistic distance is a distance measure based on the proportion of cognates in two given languages (Van der Slik, 2010). The results showed that cognate linguistic distance was the strongest predictor of the children's performance on both tests. Cognate linguistic distance has also been reported as a strong predictor of success in L2 learning (of Dutch) in Van der Slik's study (2010) of 5763 learners with 11 different language backgrounds. Furthermore, a study by Schepens, van der Slik and van Hout (2016) with multilinguals learning an L3 shows that cognate linguistic distances to both L1 and L2 play a role in L3 learnability.

2.3. Concreteness

A third factor that can facilitate word learning is concreteness: concrete words seem to be easier to learn than abstract words. A study by de Groot and Keijzer (2000), in which experienced foreign language learners were taught a number of pseudowords, showed that concrete words were easier to learn both receptively and productively. Concreteness is also linked to lexical proficiency in the sense that more proficient L2 learners tend to use more abstract words than concrete words (e.g in writing, Crossley, Salsbury, McNamara & Jarvis, 2010). This concreteness effect was discussed by Ellis and Beaton (1993) and has been linked to the dual-coding theory (Paivio, 1969) which states that concrete words are stored in memory both as a perceptual and a verbal representation, and therefore are easier to access than abstract words, which only have a verbal representation.

In a more recent study (Elgort & Warren, 2014), the concreteness effect was attested in contextual word learning. Participants read an authentic text in which some words were

replaced by pseudowords (5-6 letter made-up words similar to real English words). The concreteness values were derived from the hypernymy indices in Coh-Metrix (the hypernymy index depends on the number of levels in a conceptual taxonomic hierarchy above the word). Words with many hypernym levels are usually more concrete, whereas those with few hypernym levels are more abstract (Graesser, McNamara, Louwerse & Cai, 2004). The explicit knowledge of the pseudowords was tested after reading. Results of this study revealed that concrete words were easier to learn than abstract words.

2.4. Age of Acquisition

Items which have been learned earlier in life have a stronger memory representation and can therefore be accessed more easily than those which have been acquired later in life. This is referred to as the age of acquisition (AoA) effect and it has been observed in many studies investigating several aspects of L1 word learning (see Johnston & Barry (2006) and Juhasz (2005) for reviews).

Only a few studies have looked at AoA effects in L2 learning. Izura and Ellis (2002) found that words acquired at an earlier age in L2 were named faster in L2 picture naming and recognised faster in lexical decision. The authors did not find an effect of L1 AoA on reaction times in L2 tasks. Dirix and Duyck (2017) investigated how L2 AoA and AoA of the L1 translation equivalent influenced speed of reading in L2 (which was measured through eye-tracking). Different from the findings by Izura and Ellis (2002), they found that both L2 and L1 AoA influenced reading times of L2 words in a meaningful text. If a word was learned earlier in L2 or L1, it was read faster. In a study by Crossley and McNamara (2009) investigating whether lexical indices (related to cohesion and connectionist models) can distinguish L1 from L2 writing tasks, it was found that L2 writers (whose L1 was Spanish) used words with higher AoA values than L1 writers. This might be a result of the fact that L2

learners rely more on written texts than spoken texts and thus know words which are typically acquired later in spoken language (as early AoA-ratings are mainly based on spoken language).

The role of age of acquisition has not been investigated in contextual word learning yet, but, as it plays a role in several aspects related to word learning (such as object naming, word recognition, word reading and written proficiency), it could very well be a valuable characteristic.

3. Aims and research questions

In the present study we investigate how word learning is influenced by word-related variables in a context in which learners have not received any L2 instruction and in which their word knowledge stems from contextual L2 learning in informal settings.

Previous research (Laufer & Nation, 1995; Ellis, 2002; Crossley et al, 2011; Kim et al, 2018) has shown that the role of word-related variables is important when assessing someone's proficiency. Other studies (De Wilde et. al., 2019; Lefever, 2010) have revealed that learning takes place in informal settings but that there are considerable individual differences between learners. Therefore, the second aim of this study is to investigate whether the role of the respective word-related variables changes according to the learners' proficiency.

The research questions of this study thus are:

(1) Which of the following word-related variables contribute to contextual word learning in informal settings: L1 and L2 frequency, cognateness, concreteness, L1 and L2 age of acquisition?

(2) Does the influence of these variables differ according to learners' proficiency levels (from lower level to more advanced learners)?

We assume that all the item-related variables mentioned in research question 1 will play a role in word learning. We expect cognates to facilitate word learning as similarity to L1 words is the first thing children with low proficiency in L2 can rely on and given the observation in previous research that cognates facilitate vocabulary learning (Palmberg, 1985; De Groot & Keijzer, 2000; Vanhove & Berthele, 2015; Goriot et. al., 2018).

We expect L2 word frequency to also play a role since many studies in vocabulary learning have shown its importance (cf. 2.1). Quite a few children are exposed to English through different media each day (De Wilde et. al., 2019), which means that L2 frequency could be an important factor in contextual word learning prior to instruction. The study by Goriot et. al. (2018) showed that L1 frequency might play a role as well, particularly at the initial stages of L2 learning. Concreteness has been shown to be another significant variable in contextual word learning (Elgort & Warren, 2014) and we expect concrete words to be learned more easily than abstract words. Our final variable, age of acquisition, has been widely investigated in studies about word learning and word processing and has been linked to lexical proficiency but has not yet been addressed in the field of contextual word learning. Based on findings from other domains (Dirix & Duyck, 2017), we hypothesize that items with a translation equivalent that has been acquired earlier in the children's L1 will be picked up more easily.

Regarding the second research question, previous studies investigating lexical proficiency found that the role of different item-related variables such as frequency and concreteness, depends on the learners' proficiency (Laufer & Nation, 1995; Crossley et al., 2010; Kim et al., 2018). Based on findings from receptive multilingualism and the study by

Goriot et al. (2018), we hypothesize that cognateness and L1 frequency will explain more variance within the group of lower level learners as they might rely more on what they know from their L1 than the more advanced learners who might rely more on their L2 knowledge.

4. Method: participants, instruments, procedure

4.1. Participants

Participants are 779 children from the last year of primary school in Flanders, Belgium. The children were between 10 and 13 years old with the vast majority being 11 years old (n = 584) and only four children aged 13. Eighty-one children were 10 years old and 94 children were 12 years old at the time of the study. Seven children did not report their age. The study started out with 867 children but the children who reported they had already had English lessons or who spoke English with at least one of their parents were left out of the analyses, leaving us with 779 participants. The children came from 38 schools (50 classes) which were selected through a stratified random sampling method, ensuring geographical diversity and diversity of different school types[2]. The language of instruction in all schools was Dutch. None of the schools offered English in their curriculum, because English as a foreign language is only taught in secondary schools in Flanders. A total of 567 children reported that they only spoke Dutch with their parents; 206 children reported they (also) spoke another language at home with at least one of their parents.

4.2. Instruments and procedure

Receptive vocabulary knowledge

The children's English L2 receptive vocabulary knowledge was tested with the Peabody Picture Vocabulary Test 4 (PPVT-4), form A (Dunn & Dunn, 2007). In this test, children listened to a recording of a word while they were shown four drawings, from which they had

to choose the drawing that corresponded to the meaning of the word. This test was selected as it was the only available picture-based receptive vocabulary test suitable for children at the time of testing. Since none of the children had had any English lessons, we expected that a picture-based test would give better insight in the children's English vocabulary knowledge rather than e.g. a test based on translations or a test based on written words.

The first 120 items were tested (10 sets of 12 items). The test was administered in the classroom. By using a recording, it was ensured that all children heard the words pronounced in the same way. The words were all pronounced by the same male voice. The speaker was a highly proficient L2 speaker of English. The words were pronounced with an American English accent, since the children are familiar with this variety of English (through media exposure). They were allowed to listen to the recording twice if necessary. All children were tested on all 120 items, which is different from the procedure recommended in the PPVT-4 manual which has been written for L1 speakers of English and prescribes that testing should continue until a test-taker makes more than 7 mistakes in one set of 12 items. In our study all 120 items were tested as L2 learning might happen in a different order than L1 learning and this is something we wanted to investigate.

Second language proficiency

To measure second language proficiency, the children in this study also did the Cambridge test for Young Learners: Flyers (Cambridge English Language Assessment, 2016), which consists of three parts assessing different language skills: listening, reading and writing (tested together), and speaking.

Word-related variables

In order to have an idea of the role of word frequency, we used frequency measures from different corpora as there can be considerable differences depending on which corpus is used (Brysbaert, Mandera & Keuleers, 2018). As the participants in this study were mainly exposed to English through different media, it seemed appropriate to choose a corpus based on subtitles. Therefore, we calculated frequencies for the different test items from both the Subtlex-UK corpus (Van Heuven, Mandera, Keuleers & Brysbaert, 2014) and the Subtlex-US corpus (Brysbaert & New, 2009). We also added the frequencies from the CBBC-corpus (which is part of the Subtlex-UK corpus) as this corpus consists of words from television programs aimed at 6- to 11-year-olds, which is the language our participants are most likely exposed to. As an extra measure we added frequencies from the BNC-corpus. However, as this corpus is based on words found in written materials, we expected this measure to be less useful for the present study. We used frequency measures from the Subtlex-NL corpus (Keuleers, Brysbaert & New, 2010) to calculate L1 frequency of the Dutch translations of the target items. All frequencies were expressed as Zipf-values (Van Heuven, Mandera, Keuleers & Brysbaert, 2014). The Zipf scale is a scale going from 1 (very low frequency words) to 6 (very high frequency content words) or 7 (a few function words, pronouns, and auxiliary verb forms like "have"). The calculation of Zipf values equals log10(frequency per billion words) or log10(frequency per million words) + 3. A Zipf value of 1 corresponds to words with frequencies of 1 per 100 million words, a Zipf value of 2 corresponds to words with frequencies of 1 per 10 million words, a Zipf-value of 3 corresponds to words with frequencies of 1 per million words, and so on.

Cognateness was operationalised by calculating the orthographic and phonological similarity between the English targets and their Dutch translations. Similarity was determined by measuring the normalized Levenshtein distance (Schepens, Dijkstra, & Grootjen, 2012). The Levenshtein distance calculates how many letter changes (deletions, additions,

substitutions) are needed to go from the L2 word to the L1 translation. In order to establish phonological similarity, we first transcribed the English items and their Dutch translations by using the phonetic transcriptions from Webcelex (Max Planck Institute for Psycholinguistics, 2001). Orthographic and phonological similarity were then calculated using the Levenshtein.distance-function of the vwr-package in R (Keuleers, 2013) and normalized according to the equation found in Schepens et al. (2012): normalized distance score = 1 – (distance/length), with distance representing the Levenshtein distance between the English word and its translation and length representing the number of letters or sounds of the longest word (either the English word or its Dutch translation equivalent). This normalization is needed to control for word length, because otherwise the distance is likely to be larger for long words than for short words. Overall, the 120 target items from the PPVT are rather short (max. word length = 10 letters, 28 words consist of 3 or 4 letters).

Concreteness ratings for all 120 English items were taken from Brysbaert, Warriner, & Kuperman (2014) who provide a word list with a concreteness score for each item ranging from 1 (abstract) to 5 (concrete).

AoA ratings for Dutch were taken from Brysbaert, Stevens, De Deyne, Voorspoels and Storms (2014). AoA ratings for L1 speakers of English were taken from Kuperman, Stadthagen-Gonzalez and Brysbaert (2012). L2 AoA ratings for the 120 items of the PPVT-4 were gathered using the method of Brysbaert et al. (2014): all items were presented to 33 L1-speakers of Dutch, who were 14-16 years old at the time of rating. The raters were either in the third of fourth year of secondary education and had received one or two years of English lessons prior to the test. The participants were asked at which age they first knew the word. If they did not know the word, they were asked to write this down instead of the AoA. Thirty-five items of the PPVT-4 were also present in the L2 AoA ratings gathered by Dirix and Duyck (2017). The correlation between our ratings and those gathered by Dirix and Duyck

was .69. Differences in the ratings could be due to the different ages of the participants who rated the items (as the mean age of the participants in the study of Dirix and Duyck was 19 years).

An overview of the L1 translation equivalents (used to calculate linguistic similarity and L1 frequency scores) and the different word-related measures per item can be found in the supplementary materials.

5. Results

5.1. Exploratory statistics

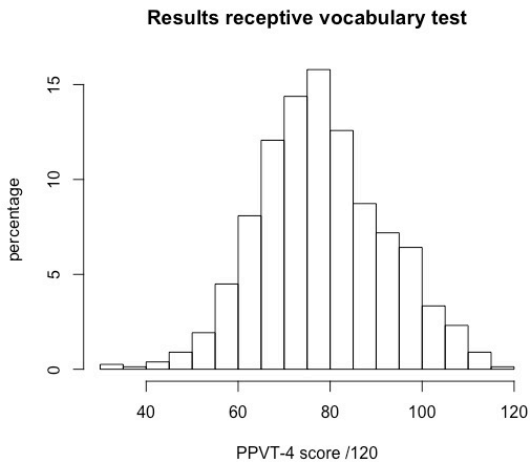
Results for the Cambridge proficiency test

765 children completed the Cambridge test for Young Learners: Flyers, which assessed written and spoken language perception and production at the A2-level, giving a maximum total score of 95. With a minimum score of 7/95, a maximum score of 95/95 and a standard deviation of 21.71, the test results showed a broad range in scores. The median was 38/95 and the mean 43/95. About one third of the children (N = 273) had a score above 50%.

Results for the Peabody receptive vocabulary test

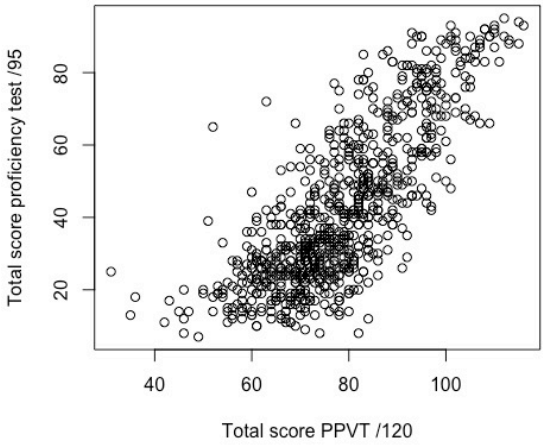
Figure 1 shows the overall score on the receptive Peabody Picture Vocabulary Test (first 120 items of PPVT- 4). The minimum score was 31/120 and the maximum score 116/120, pointing to a large standard deviation (SD =13.85). The median was 78/120 and the mean 78.6/120, indicating that the function is largely symmetric, as can also be seen in Figure 1.

Figure 1: Distribution of the results on the PPVT-4 vocabulary test



There was a high correlation between the Peabody receptive vocabulary test and the Cambridge proficiency test (r = 0.79) as can be seen in Figure 2.

Figure 2: Scatterplot showing the correlation between the score on the receptive vocabulary test and the results for second language proficiency.



Word-related variables

Frequency

As mentioned in paragraph 4.2, we looked at the frequency of the different items in different corpora (Subtlex-UK, Subtlex-US, BNC and CBBC-corpus). As expected, there are high correlations between the different frequency measures (cf. Table 2).

Table 1: Descriptive statistics of the different lexical measures for the target words of the PPVT

	Mean	SD	min	max
1. Subtlex-UK	4.07	0.65	2.35	5.56
2. Subtlex-US	3.98	0.64	2.30	5.52
3. BNC	3.99	0.65	2.17	5.62
4. CBBC-corpus	4.18	0.63	2.47	5.59
5. Subtlex-NL	3.89	0.74	2.04	6.00
6. Orthographic similarity	0.43	0.35	0.00	1.00
7. Phonological similarity	0.30	0.29	0.00	1.00
8. Concreteness	4.47	0.64	2.39	5.00
9. L1 AoA Dutch	6.95	1.97	3.46	12.76
10. L1 AoA English	5.72	1.83	2.39	11.56
11. L2 AoA English	9.60	1.57	4.94	13.20

Table 2: Correlations between the different lexical measures for the target words of the PPVT.

	1	2	3	4	5	6	7	8	9	10	11
1. Subtlex-UK	1	0.91***	0.91***	0.90***	0.66***	-0.01***	0.07***	-0.02***	-0.48***	-0.58***	-0.40***
2. Subtlex-US		1	0.86***	0.87***	0.69***	-0.04***	-0.02***	0.06***	-0.49***	-0.61***	-0.39***
3. BNC			1	0.76***	0.65***	-0.02***	0.03***	-0.06***	-0.40***	-0.44***	-0.26***
4. CBBC-corpus				1	0.64***	0.00	0.07***	0.04***	-0.55***	-0.68***	-0.44***
5. Subtlex-NL					1	-0.05***	-0.01**	-0.05***	-0.62***	-0.46***	-0.33***
6. Orthographic similarity						1	0.83***	0.17***	0.18***	0.05***	-0.48***
7. Phonological similarity							1	0.22***	0.07***	0.03***	-0.46***
8. Concreteness								1	-0.26***	-0.25***	-0.20***

9. L1 AoA Dutch	1	0.65***	0.36***
10. L1 AoA English		1	0.37***
11. L2 AoA English			1

*** p<.001, **p<.01, *p< .05

Correlations between the item scores in the receptive vocabulary test and the different word features show that the word frequency measure based on the CBBC corpus was the best measure for our data (cf. Table 3). This result could be expected as it is a corpus consisting of language targeted at children.

Table 3: Correlations between word-related variables and the results of the PPVT-4

	PPVT
1. Subtlex-UK	0.23***
2. Subtlex-US	0.20***
3. BNC	0.17***
4. CBBC-corpus	0.31***
5. Subtlex-NL	0.29***
6. Orthographic similarity	0.53***
7. Phonological similarity	0.56***
8. Concreteness	0.21***
9. L1 AoA Dutch	-0.32***
10. L1 AoA English	-0.24***
11. L2 AoA English	-0.63***

*** p<.001, **p<.01, *p< .05

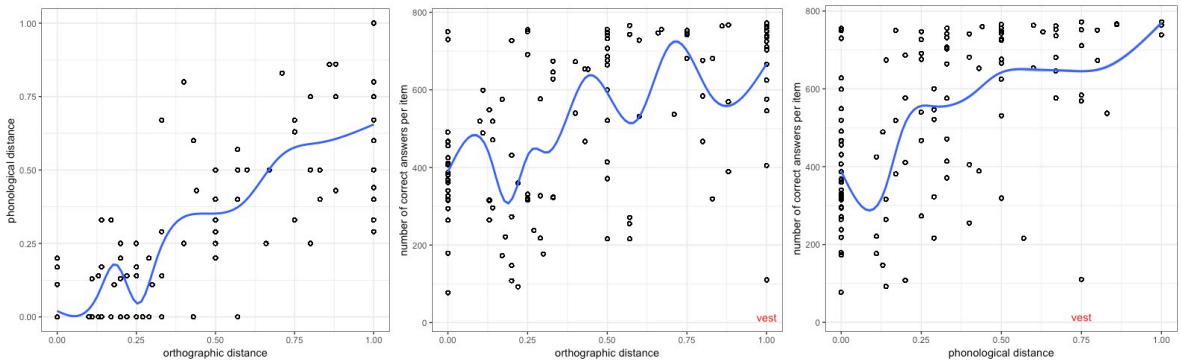
Linguistic similarity

Based on orthographic similarity, 54 words were Dutch (near-)cognates (normalized Levenshtein distance > .50). Based on phonological similarity, 35 words are Dutch (near-)

cognates. The difference between the number of orthographic and phonological cognates can be explained by the fact that Dutch and English have some vowels that are pronounced (slightly) differently but written the same. The correlation between both measures of linguistic similarity was .83. The correlations between orthographic and phonological similarity and the number of correct answers are nearly the same (cf. Table 3). The correlations with phonological similarity are slightly higher which is unsurprising as children typically have more exposure to spoken language than to written language (Lindren & Muñoz, 2013; De Wilde et. al., 2019). When looking at all the participants in the present study, the correlations between the test results and cognateness (orthographic and phonological similarity) are stronger than the correlations with frequency, concreteness and L1 AoA.

The scatterplots in Figure 3 show that items with a linguistic distance close to 1 are known to most of the children. The big exception is the word "vest", which is not a cognate but a homograph, a word written the same in two languages but with a different meaning ("vest" in Belgian Dutch means "coat", which was present as a distractor in the PPVT together with the picture of a waistcoat). The regression lines show the results predicted on the basis of a gam (generalized additive models) function, which takes into account non-linear relationships. They show that the main bulk of the covariations are captured by linear regressions.

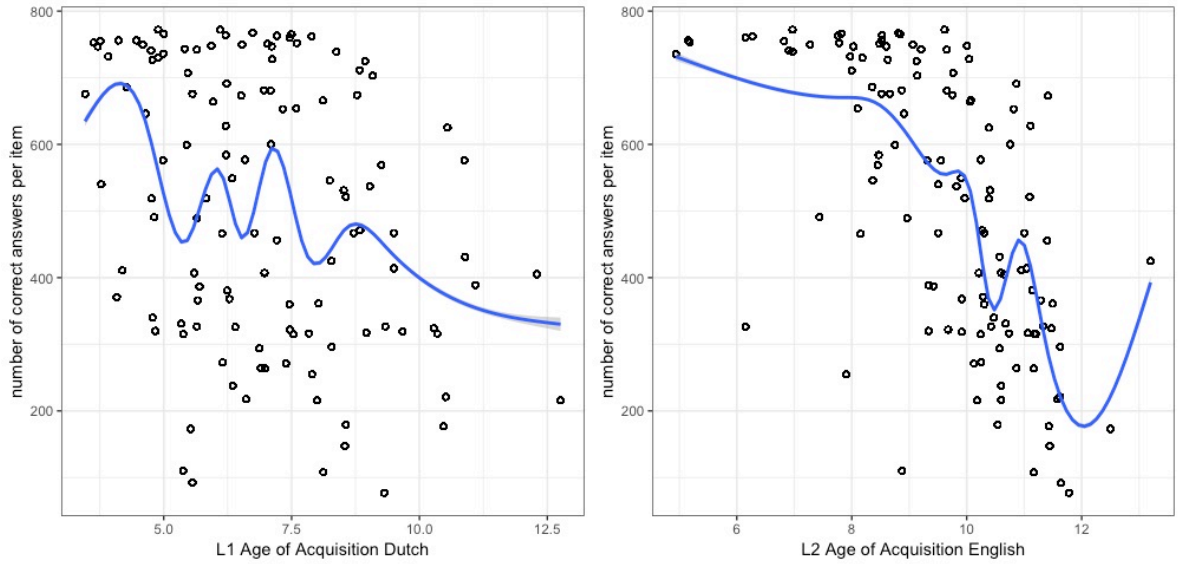
Figure 3: Scatterplots showing the correlations between different measures of linguistic similarity and between linguistic similarity and the score per item.



Age of Acquisition

AoA also seems to play a role in how easy it is to learn a word. Correlations between AoA of the Dutch translation equivalents and percentage of children who know the word can be found in Table 3 and Figure 4.

Figure 4: Scatterplots showing the correlations between the two measures of age of acquisition (Dutch L1 and English L2) and the scores per item.



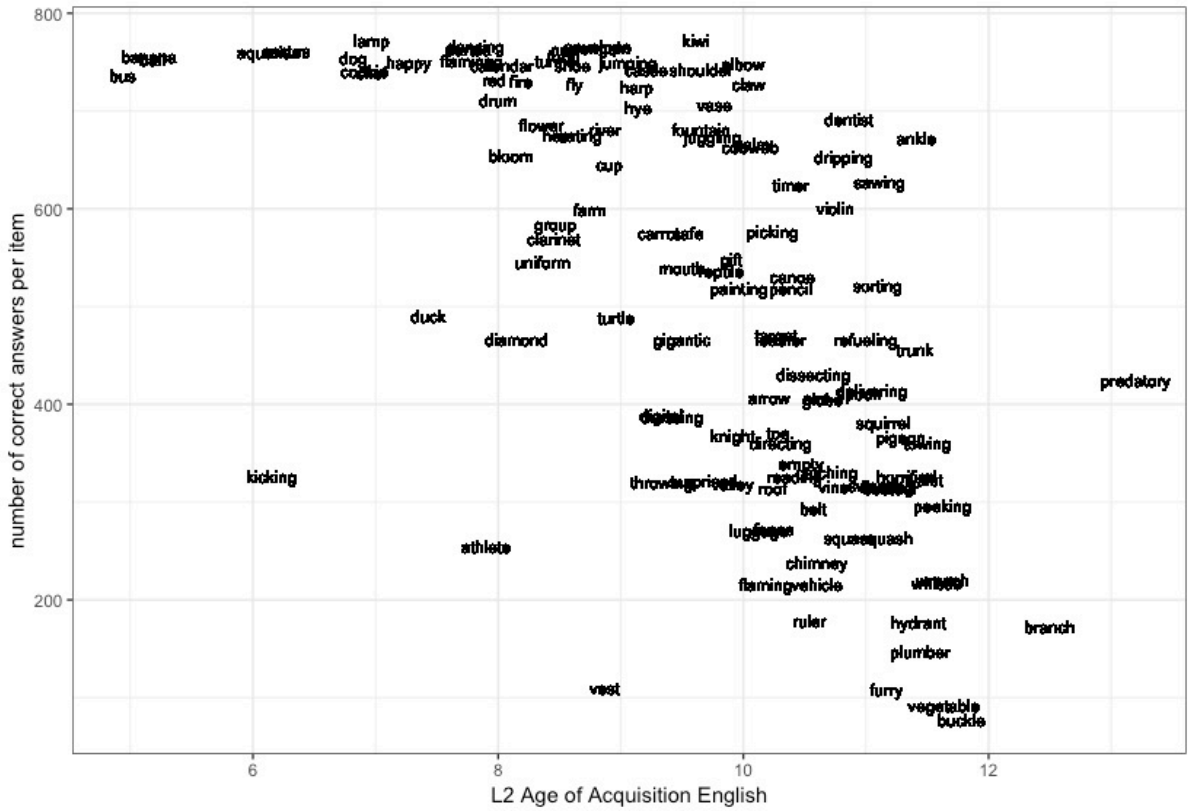
The correlation between Dutch AoA and English word knowledge shows that words learned early in Dutch are generally known better to the children in English, arguably because these words are easier to acquire or more important to know. Indeed, there are indications that

the words learned first in life form central nodes in the semantic system of an individual (Brysbaert & Ellis, 2016). In this respect we do not think it is a coincidence that the correlation of English word knowledge is higher with Dutch L1 AoA (r = -.32) than with English L1 AoA (r = - .24). Indeed, for Dutch-speaking learners of English, Dutch L1 AoA estimates represent the sequence of concept acquisition, and not English L1 AoA.

Figure 4 also shows the strong impact of English L2 AoA on word knowledge (with a correlation of -.63 and an indication of a ceiling effect for the first learned words). However, we do not think much theoretical impact can be attached to this correlation. Basically, it says that English words of which 16-year-olds indicate they acquired them early, are known to 11-year-olds in the PPVT. This is a tautology or at best an indication of the convergent validity of the measures. The correlation gives no further information about why some L2 words are acquired earlier than others. The high correlations between L2 AoA and orthographic and phonological similarity to L1 (.48 and .46 respectively) repeat the importance of cognateness, certainly given the much lower correlations between L1 AoA and orthographic similarity or phonological similarity. There may be other factors making some L2 words more difficult to learn than others, so that they take longer to be picked up even though they may be present in the environment (especially in informal settings, where there is no teacher to guide the learning process).

English L2 AoA has a low correlation with English L1 AoA (r = .37), further pointing to the fact that other factors may make words difficult to learn in L2 than in L1. Of course, there is also the difference in environment between Dutch-speaking bilinguals informally acquiring English (primary school children) and native speakers acquiring the language (infants). For those interested in the specific order of English L2 word acquisition by Dutch-English bilinguals in informal settings, Figure 5 repeats the correlation between the chances of a word known and English L2 AoA, but this time with the English words included.

Figure 5: Scatterplot showing the correlation between L2 age of acquisition and the score per item in text. (This is the same figure as 4b.)



Concreteness

The correlation between concreteness and the test results was rather lower. It should be noted, however, that most items in the test were concrete words, as it was a picture-based test. Only seven words had a concreteness score below 3 and more than 80 items had a concreteness score above 4.5. So, there was not much variation in concreteness in the words tested, which probably impacted the results.

5.2. Variance explained by proficiency and word-related variables in statistical models.

Method

In order to determine the percentage of variance in the test results that can be explained by the word-related variables, we analysed the data using generalized linear mixed modelling

(GLMM). In this kind of model both random and fixed effects can be included to account for individual variance among participants and items. We used the R-package lme4 (Bates, Mächler, Bolker, & Walker, 2015) to construct our model using the glmer-function, because the dependent variable was correct vs. wrong answer.

We first centred the variables on the mean. We then started with a basic model including three random effects (intercepts items, participants and schools) to account for individual variability among items, participants and schools. We next added L2 proficiency as a fixed effect. Because there is a high correlation between second language proficiency and the dependent variable (r = 0.79), we dichotomized the variable into two equal groups of learners with low and high proficiency (baseline = high, score on the proficiency test of 39/95 or higher) to avoid a singular fit. Otherwise, the model did not converge.

In a third step, we sequentially added the following item-related variables as fixed effects to predict the outcome variable: phonological similarity, L1 frequency, L2 frequency, L1 AoA Dutch, L1 AoA English, and concreteness. Because orthographic and phonological similarity had a high intercorrelation, we kept only phonological similarity as a predictor in the model to avoid a collinearity problem. As indicated above, most of the children's out-of-school exposure to English concerns aural rather than written input (De Wilde et. al., 2019; Kuppens, 2010) and so it makes sense to use the phonological similarity. Concerning English frequency, we used the CBBC-frequency measure, as this is the most indicated (and also provided the best model). To investigate the role of proficiency, interactions between proficiency and the item-related variables were added. We used anova(model1, model2) to assess the model fit and calculated marginal R^2, which measures the variance explained by the fixed effects only, and conditional R^2, which measures the variance explained by both the fixed effects and the random effects using the MuMin package (Nakagawa & Schilezeth, 2013) in R. Interactions were plotted using the sjPlots package (Lüdecke, 2018).

The model with all item-related variables (phonological similarity, L1 frequency, L2 frequency, L1 AoA Dutch, L1 AoA English, concreteness) and proficiency as fixed effects and the interactions between proficiency and five of these word-related variables (without the interaction with L1 AoA Dutch), turned out to be the best model. Model comparison using anova (model1, model2) shows that the maximal model is not significantly better than the model without the interaction between proficiency and L1 AoA ($\chi^2(1) = 0.70$, $p = .40$).

Results

The basic model (with only the three random effects) shows that there was little variance between the schools (variance = 0.03, SD = 0.17) relative to the variance between participants (variance = 0.60, SD = 0.77). The random effect with most of the variance was 'items' (variance = 3.24, SD = 1.80).

We then added the fixed effect 'proficiency' to the model. This variable was significant ($p < 0.001$) and explained 4 percent of the variance (marginal $R^2 = .04$, conditional $R^2 = .54$).

After adding the word-related variables, the results showed a significant positive main effect of phonological similarity and significant negative main effects of proficiency and L1 age of acquisition Dutch. Interestingly, the same main effects remained in a model without interactions.

The model with interactions showed five significant interactions between proficiency and item-related variables. There was an interaction between proficiency and phonological similarity ($p = 0.03$), between proficiency and English L1 AoA ($p < 0.001$), between proficiency and frequency in English, ($p < 0.001$), between proficiency and frequency in Dutch ($p < 0.001$), and between proficiency and concreteness ($p < 0.001$). The estimated standard errors and z-values can be found in Table 4. A visual representation of the interactions effects can be found in Figure 6. The model reported a marginal R^2 of .28 and a

conditional R^2 of .55, which shows that an extra 24% of the variance is explained by the word-related variables and their interactions.

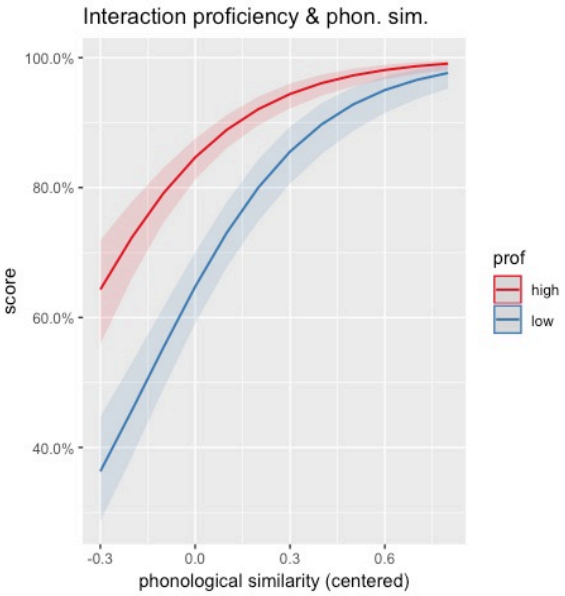
Table 4: GLME model predicting right or wrong answers on the vocabulary text.

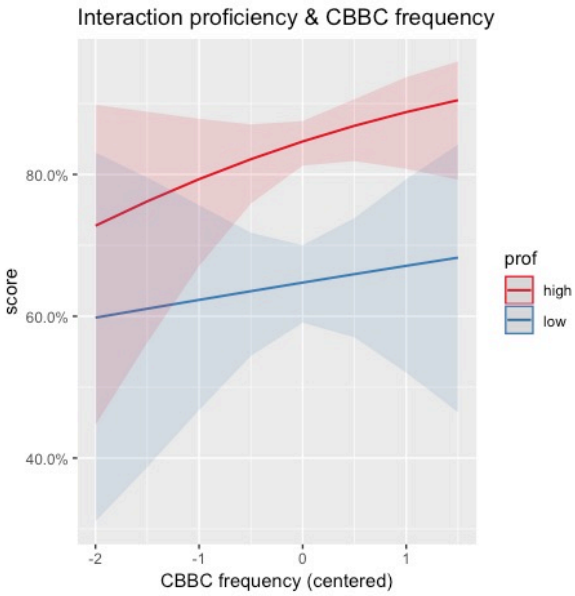
Random effects	Variance	SD
Participant	0.339	0.582
Item	1.650	1.285
School	0.007	0.085

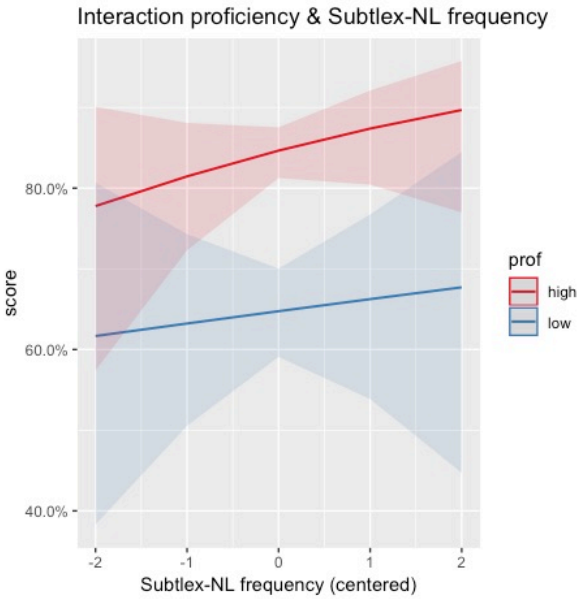
Fixed effects	Estimate	SE	Z-value	
(Intercept)	1.71	0.12	13.84	***
Phonological Similarity	3.73	0.44	8.50	***
Proficiency	-1.10	0.05	-23.20	***
L1 AoA Dutch	-0.29	0.09	-3.11	**
CBBC frequency	0.36	0.30	1.21	
Subtlex-NL frequency	0.23	0.24	0.97	
L1 AoA English	-0.03	0.10	-0.30	
Concreteness	0.03	0.21	0.16	
Phonological Similarity x proficiency	0.17	0.08	2.21	*
CBBC frequency x proficiency	-0.26	0.05	-5.65	***
Subtlex-NL frequency x proficiency	-0.16	0.03	-5.02	***
L1 AoA English x proficiency	0.10	0.01	7.20	***
Concreteness x proficiency	0.10	0.03	3.73	***

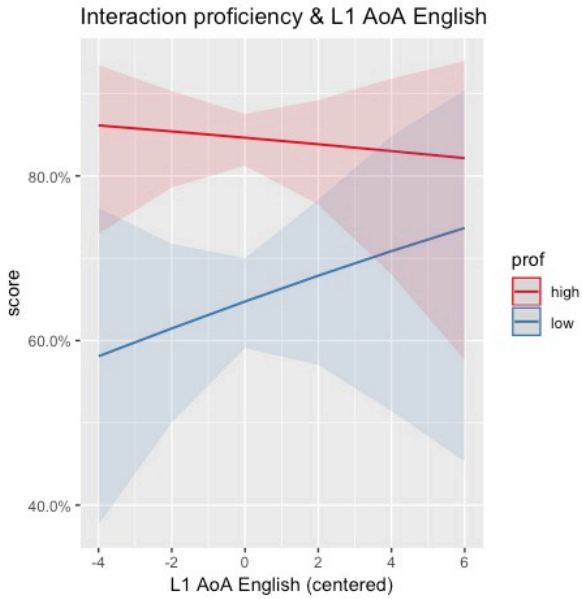
*** p<.001, **p<.01, *p< .05
Baseline for proficiency = high

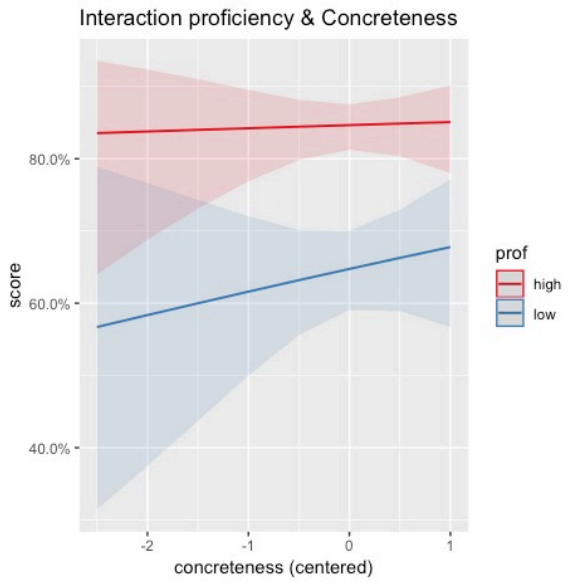
Figure 6: Predictions of the score per word-related variable according to proficiency (=prof).


Interaction proficiency & phon. sim.
100.0%
80.0%
60.0%
40.0%
score
-0.3
0.0
0.3
0.6
phonological similarity (centered)
prof
high
low


Interaction proficiency & CBBC frequency
80.0%
60.0%
40.0%
score
-2
-1
0
1
CBBC frequency (centered)
prof
high
low


Interaction proficiency & Subtlex-NL frequency
80.0%
60.0%
40.0%
score
-2
-1
0
1
2
Subtlex-NL frequency (centered)
prof
high
low


Interaction proficiency & L1 AoA English
80.0%
60.0%
40.0%
score
-4
-2
0
2
4
6
L1 AoA English (centered)
prof
high
low



6. Discussion

When looking at the random effects, we see that the effect with the most variance is 'item'. This means that learning difficulty is more determined by characteristics of the words than characteristics of the learner. There is also variance between learners but far less so. Schools do not seem to have an effect, which could be expected as these children learn English through exposure outside the school.

Regarding the word-related variables, we hypothesized that all variables considered in this study would influence the children's receptive vocabulary learning. This proves to be the case.

When looking at the results of the generalised mixed effects model, we see that there are three significant main effects in the model. The first significant main effect is proficiency, which is unsurprising as it is to be expected that more proficient English learners know more words. Regarding proficiency, we are mainly interested in the interactions between proficiency and the word-related variables. These interactions will be discussed below.

The second main effect which is significant, is phonological similarity (the normalised Levenshtein distance between the spoken English word and its spoken Dutch translation). The effect is positive, meaning that English words which are more similar to their Dutch

translation equivalent are easier to learn. Phonological similarity is the strongest predictor of word knowledge, which is in line with previous research showing the importance of cognates in language learning in general and in contextual language learning in particular (Goriot et al., 2018; Peters & Webb 2018; Keijzer & De Groot, 2000; Palmberg, 1985).

The third significant main effect is L1 AoA of the Dutch translation equivalents, which means that words learned early in life in L1 are also easier to learn in a foreign language. This is in line with the findings of Dirix and Duyck (2017) who reported that words acquired at an earlier age in L1 are processed faster in L2. A possible explanation is the semantic hypothesis about the AoA effect. This hypothesis assumes that words/concepts learned earlier in life are more central in the semantic network and are thus more easily accessible (Brysbaert & Ellis, 2016; Brysbaert, Van Wijnendaele & De Deyne, 2000; Steyvers & Tenenbaum, 2005). As a result, they are easier to learn in L2.

The model also showed significant interactions between a child's proficiency and five word-related variables (Table 4, Figure 6), indicating that the word-related variables play a different role at different stages in children's L2 learning.

The interaction between proficiency and phonological similarity shows that for low-proficiency children the effect of phonological similarity to L1 is stronger than for the more proficient children, indicating that beginning learners mainly rely on formal similarity with L1 during receptive word acquisition. These children seem to draw heavily on cognate guessing, which is guessing the meaning of words based on similarities with known cognates (Vanhove & Berthele, 2015). This finding is in line with the parasitic hypothesis (Hall, 2002) which states that, especially in the early stages of language learning, learners use their existing vocabularies to make meaning of the input they receive and link the forms they recognize to the knowledge they have about similar forms and in this way establish an initial

memory representation of the new L2 word. The study by Goriot et. al. (2018) also found that L1-L2 similarity influenced the scores on the PPVT-test.

The interaction between proficiency and L2 frequency is due to the fact that the L2 frequency effect is stronger for proficient learners than for less proficient learners. As L2 learners become more proficient, they tend to know more high-frequency English words than less frequent English words. Something similar can be observed for AoA in English. The distribution in the high proficiency group is as expected: words with an earlier AoA in English tend to be known better; for the lower proficiency group, scores are higher for words which are learned later in English, which is surprising and seems to suggest that L1 AoA English does not play a role in this group. The results for English frequency and English AoA point to L2 influences in the high proficiency group, whereas the lower proficiency group seemed to rely mainly on L1. Goriot et al. (2018) also found that L2 frequency is more important with older, more proficient learners than with younger learners.

The interaction between L1 frequency and proficiency shows that more proficient learners seem to score better on words that are more frequent in Dutch. This is somewhat surprising as for the other variables it seems that L1-related variables mainly benefit less proficient leraners. It is also different from what was found in Goriot et al. (2018), where younger learners, and thus less proficient learners, showed a stronger influence of Dutch frequency. We have to keep in mind that in our study even the high proficiency group is only proficient at the CEFR A2-level, which means they are basic users of English. Thus, one possible explanation might be that our very low proficiency group is strongly focused on formal similarities with L1 (hence the interaction with phonological similarity) whereas our high proficiency group - which seem to be the most efficient learners in informal learning conditions – have already become more active in searching and remembering translations for words they often use in L1. This hypothesis would have to be looked into in future research.

The final significant interaction is that between concreteness and proficiency. When we look at the plot in Figure 6, there does not seem to be an effect of concreteness in the high proficiency group but more concrete words seem to be easier to learn for the learners with low proficiency. As there was not much variation in concreteness in the items of the PPVT and most words are concrete words, the results of learners with a higher proficiency do not seem to be influenced by this variable. For the low proficiency group the most concrete words were easier to learn which is in line with previous research (de Groot & Keijzer, 2000; Elgort & Warren, 2014).

Overall, the results show that several word-related variables contribute to how difficult or easy it is to recognize a word in L2. The role of these variables is different according to the proficiency level of the speaker. Similar findings have been attested in previous research (Laufer & Nation, 1995; Crossley et al., 2011; Kyle & Crossley, 2015), but have now been extended to informal, out-of-school learning.

L2 AoA was not added to the GLME model, because this variable turned out to be less revealing than we thought when it comes to predicting word knowledge. Both L2 AoA ratings and percentages of L2 speakers who know a word largely measure the same construct. In contrast, L2 AoA ratings are likely to be an interesting variable with reference to the speed with which high-proficiency L2 speakers can recognize and produce L2 words (Dirix & Duyck, 2017; Izura & Ellis, 2002).

A limitation of this study is that the number of individual differences that have been included in the analyses is restricted. Gender was not included, although previous research (Sylvén & Sundqvist, 2012) has shown that gender and out-of-school exposure are linked (e.g. boys are more frequent gamers than girls, which may give them an intial advantage in informal language learning). Gender may also affect which words are learned as a result of diverging interests. Other variables worth looking at in future research may be similarlity

between L1 and L2, and language background. Our study counted both monolingual and multilingual learners. All participants followed monolingual Dutch education and were in the final year of regular primary education, so they all had rather extensive knowledge of Dutch, which is one of the languages closest to English. It would be very interesting to look into the influence of learners' background (different L1, different schooling language, influence from L1 and schooling language on L3, balanced/unbalanced bilinguals, …) in more detail.

Our study was also limited to vocabulary. Another question is how grammar is acquired out of school and how this is influenced by L1 (and the language of instruction if different).

7. Conclusion

The present study showed how young learners can build up receptive vocabulary knowledge in a foreign language through contextual word learning alone. The results indicate that children's vocabulary development is kickstarted by similarities between L1 and L2. The fact that Dutch and English share many cognates (Van der Slik, 2010) helps the children to independently learn English as an L2 without the assistance of a school teacher. The mixed effects model suggests that the low proficiency group might still be in the stadium of cognate guessing, whereas the results of the more proficient group show that word learning in this group is influenced by the characteristics of the English language itself.

Our findings are in line with those of Goriot et al. (2018) and add evidence to the fact that, when using the PPVT, the number of cognates can strongly influence the results. Therefore, researchers should be careful when using the PPVT to compare learners with different language backgrounds, and it is advisable to report the number of cognates in the PPVT for each language. This would give researchers an idea of how much learners can rely on cognate guessing.

Finally, the present results are relevant for educational practice. The study shows that word characteristics play an important role when it comes to learning difficulty. This has implications for teachers and coursebook writers. Our study may encourage these stakeholders to take lexical indices into account when designing vocabulary lessons or selecting texts. Selecting appropriate materials, i.e. taking into account word-related variables such as L2 frequency, may lead to larger learning gains and thus boost children's L2 vocabulary learning. The study also shows that teachers can expect learners to deal with cognates independently and they can adapt their lessons accordingly. For example, when engaged in vocabulary development or reading/listening activities teachers do not need to pre-teach transparent words, not even for learners at an elementary level.

Notes

[1] Contextual word learning is often referred to as incidental word learning, which is typically considered to be the opposite of deliberate learning. Learning from context is not necessarily accidental, which is why Elgort, Brysbaert, Stevens, & Van Assche (2018) proposed the intentionality-agnostic term 'contextual word learning', a term that will also be used throughout the present article.

[2] The educational system in Flanders is organized through three networks: state schools, subsidized public schools and subsidized free schools. All three networks are represented in the sample.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). The Role of Repeated Exposure to Multimodal Input in Incidental Acquisition of Foreign Language Vocabulary: Repeated Exposure in Incidental FL Acquisition. *Language Learning*, *64*(4), 855–877. https://doi.org/10.1111/lang.12085

Brysbaert, M. & Ellis, A.W. (2016). Aphasia and age-of-acquisition: Are early-learned words more resilient? *Aphasiology, 30*, 1240-1263. https://doi.org/10.1080/02687038.2015.1106439

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80–84. https://doi.org/10.1016/j.actpsy.2014.04.010

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Brysbaert, M., Wijnendaele, I. V., & Deyne, S. D. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*(2), 215–226. https://doi.org/10.1016/S0001-6918(00)00021-4

Bybee, J. L., & Hopper, P. J. (Eds.). (2001). Frequency and the Emergence of Linguistic Structure (Vol. 45). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45

Cambridge English Language Assessment. (2016). Cambridge English: Young Learners: Flyers. Retrieved from http://www.cambridgeenglish.org/exams/young-learners-english/

Council of Europe (Ed.). (2009). *Common European Framework of Reference for Languages: learning, teaching, assessment* (10. printing). Cambridge: Cambridge Univ. Press [u.a.].

Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, *18*(2), 119–135. https://doi.org/10.1016/j.jslw.2009.02.002

Crossley, S., & Salsbury, T. (2010). Using lexical indices to predict produced and not produced words in second language learners. *The Mental Lexicon*, *5*(1), 115–147. https://doi.org/10.1075/ml.5.1.05cro

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, *28*(4), 561–580. https://doi.org/10.1177/0265532210378031

De Groot, A. M. B., & Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Language Learning*, *50*(1), 1–56. https://doi.org/10.1111/0023-8333.00110

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, 1–15. https://doi.org/10.1017/S1366728918001062

Dirix, N., & Duyck, W. (2017). The first- and second-language age of acquisition effect in first- and second-language book reading. *Journal of Memory and Language*, *97*, 103–120. https://doi.org/10.1016/j.jml.2017.07.012

Dunn, L., & Dunn, L. (2007). *Peabody Picture Vocabulary Test (4th ed)*. Circle Pines, MN: American Guidance Service.

Elgort, I., & Warren, P. (2014). L2 Vocabulary Learning From Reading: Explicit and Tacit Lexical Knowledge and the Role of Learner and Item Variables: L2 Vocabulary Learning From Reading. *Language Learning*, *64*(2), 365–414. https://doi.org/10.1111/lang.12052

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: an eye-movement study. *Studies in Second Language Acquisition*, *40*(2), 341–366. https://doi.org/10.1017/S0272263117000109

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(02), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic Determinants of Foreign Language Vocabulary Learning. *Language Learning*, *43*(4), 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Ellis, N. C., & Wulff, S. (2014). Usage-based approaches to SLA. In VanPatten, B. & Williams, J. (Eds.), *Theories in second language acquisition: An introduction* (2nd Edition, pp. 75–93). New York: Routledge.

Gooskens, C., van Heuven, V. J., Golubović, J., Schüppert, A., Swarte, F., & Voigt, S. (2018). Mutual intelligibility between closely related languages in Europe. *International Journal of Multilingualism*, *15*(2), 169–193. https://doi.org/10.1080/14790718.2017.1350185

Goriot, C., van Hout, R., Broersma, M., Lobo, V., McQueen, J. M., & Unsworth, S. (2018). Using the peabody picture vocabulary test in L2 children and adolescents: effects of L1. *International Journal of Bilingual Education and Bilingualism*, 1–23. https://doi.org/10.1080/13670050.2018.1494131

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. https://doi.org/10.3758/BF03195564

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *International Review of Applied Linguistics in Language Teaching*, *4*(50), 239–276. https://doi.org/10.1515/iral-2012-0010

Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *International Review of Applied Linguistics in Language Teaching*, *40*(2). https://doi.org/10.1515/iral.2002.008

Izura, C., & Ellis, A. W. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Psicológica, 23*(2).

Jensen, S. (2017). Gaming as an English Language Learning Resource among Young Children in Denmark. *CALICO Journal*, *34*(1), 1–19. https://doi.org/10.1558/cj.29519

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13*(7–8), 789–845. https://doi.org/10.1080/13506280544000066

Juhasz, B. J. (2005). Age-of-Acquisition Effects in Word and Picture Identification. *Psychological Bulletin*, *131*(5), 684–712. https://doi.org/10.1037/0033-2909.131.5.684

Keuleers, E. (2013), vwr: Useful functions for visual word recognition research. R-package version 0.3.0. https://cran.r-project.org/web/packages/vwr/index.html

Keuleers, E., Brysbaert, M. & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods, 42*(3), 643-650. https://doi.org/10.3758/BRM.42.3.643

Kim, M., Crossley, S. A., & Skalicky, S. (2018). Effects of lexical features, textual properties, and individual differences on word processing times during second language reading

comprehension. *Reading and Writing*, *31*(5), 1155–1180. https://doi.org/10.1007/s11145-018-9833-x

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*(1), 65–85. https://doi.org/10.1080/17439880903561876

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Lefever, S. (2010). *English skills of young learners in Iceland: "I started talking English when I was 4 years old. It just bang… just fall into me."*. Presented at the Menntakvika Conference, Reijkjavik, Iceland.

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*(1), 105–129. https://doi.org/10.1080/14790718.2012.679275

Lüdecke, D. (2018, July 12). Sjplot - Data Visualization For Statistics In Social Science. Zenodo. https://doi.org/10.5281/zenodo.1310947

Max Planck Institute for Psycholinguistics. (2001). *Web-based CELEX*. Retrieved April 9, 2018, from http://celex.mpi.nl/

Monaghan, P., Chang, Y.-N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, *93*, 1–21. https://doi.org/10.1016/j.jml.2016.08.003

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*(3), 241–263. https://doi.org/10.1037/h0027272

Palmberg, Rolf. (1985). How Much English Vocabulary Do Swedish-Speaking Primary-School Pupils Know before Starting to Learn English at School? In Ringbom, Hakan (Ed.), *Foreign Language Learning and Bilingualism* (Vol. 105, pp. 89–97). Åbo: Publications of the Research Institute of the Åbo Akademi Foundation.

Pellicer-Sanchez, A., & Schmitt, N. (n.d.). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart. *Reading in a Foreign Language*, *22*(1), 31–55.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 1–27. https://doi.org/10.1017/S0272263117000407

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, *15*(01), 157–166. https://doi.org/10.1017/S1366728910000623

Schepens, J., van der Slik, F., & van Hout, R. (2016). L1 and L2 Distance Effects in Learning L3 Dutch: L1 and L2 Distance Effects in L3 Learning. *Language Learning*, *66*(1), 224–256. https://doi.org/10.1111/lang.12150

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, *29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, *24*(03), 302–321. https://doi.org/10.1017/S095834401200016X

Van der Slik, F. W. P. (2010). Acquisition of Dutch as a second language. *Studies in Second Language Acquisition*, *32*(03), 401–432. https://doi.org/10.1017/S0272263110000021

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A New and Improved Word Frequency Database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vanhove, J., & Berthele, R. (2015). The lifespan development of cognate guessing skills in an unknown related language. *International Review of Applied Linguistics in Language Teaching*, *53*(1), 1–38. https://doi.org/10.1515/iral-2015-0001

Vidal, K. (2011). A Comparison of the Effects of Reading and Listening on Incidental Vocabulary Acquisition: Incidental Vocabulary Acquisition. *Language Learning*, *61*(1), 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Webb, S., & Chang, A. (2017). *To what extent are L2 multiword combinations incidentally learned in different modes of input?* Presented at the EuroSLA27 Conference, Reading.