Productive vocabulary knowledge in L2 German : which word-related variables matter?

Productive vocabulary knowledge in L2 German: Which word-related variables matter?

Abstract

In vocabulary research, factors affecting second/foreign language (L2) receptive vocabulary have been investigated extensively. However, few studies have examined factors affecting productive knowledge (i.e., form recall), especially concerning Languages Other Than English (LOTEs). Moreover, not much is known about the influence of time on L2 learners' productive single word knowledge. Nevertheless, it is crucial to understand which variables affect productive vocabulary learning in order to optimize the learning process. A 3-year longitudinal study was conducted with Dutch-speaking university learners of German (N = 64) to investigate to what extent time, frequency, cognateness, concreteness, and age of acquisition (AoA) play a role in the development of productive single word knowledge in L2 German. Results revealed that time, frequency, cognateness, and AoA together explained 28% of the variance. L2 frequency was found to be the main predictor for L2 form recall. Cognateness also played a role, although the impact was lower compared to the findings of studies on receptive vocabulary knowledge. Concreteness did not affect L2 learners' form recall scores. Based on the results of the study, suggestions for materials and pedagogical implications will be discussed.

Keywords

productive vocabulary, form recall, L2 German, frequency, cognateness, concreteness, age of acquisition

1. Introduction

The success of learning a word in a second or foreign language (L2) depends on the learning burden of this particular word. According to Nation (2022, p. 49), "the learning burden of a word is the amount of effort required to learn it. Different words have different learning burdens for learners with different language backgrounds." Some words are thus easier to learn than others, depending on a number of word-related factors (Laufer, 1997). For example, de Groot & Keijzer (2000) demonstrated that it was easier to learn cognates and concrete L2 words, and that these items were less likely to be forgotten compared to non-cognates and abstract words.

In vocabulary research, many factors that can affect vocabulary acquisition have been identified. Peters (2020) gives an overview of factors that have been suggested to affect the learning of single word items, such as cognateness, similarity in form (e.g., synforms such as *adopt/adapt*, *historic/historical*, *price/prize* (Laufer, 1988)) or meaning (e.g., synonyms), word length, part of speech, concreteness and imageability, polysemy and homonymy, frequency of occurrence, and L1 or L2 frequency.

However, research in this domain has mainly focused on receptive vocabulary learning, so one of the remaining questions is whether productive learning of L2 single words is affected by the same factors and to the same extent as receptive learning (Peters, 2020). Besides, studies that focus on learners' L2 productive vocabulary knowledge are rather scarce, despite the repeated calls to examine this further (e.g., Fitzpatrick & Clenton, 2017; Laufer & Nation, 1999). Second, more longitudinal studies are needed to explore whether and how the effect of word-related variables on word learning changes over time. These kinds of insights can help teachers to decide what to focus on in the classroom depending on the learners' language level. Third, we believe that research is needed into a wider variety of languages and language pairs, to verify whether certain word-related variables affect

vocabulary learning in the same or in a different way in various languages. Most studies in the field have focused on L2 English, and there are only a few studies on vocabulary development with a focus on German (e.g., Knopp, 2022; Lindgren & Bohnacker, 2020). These studies, however, look at the factors affecting vocabulary development of bilingual children. Additional studies with L2/foreign language learners are thus needed, because novel insights into the variables that affect productive vocabulary learning can help teachers to identify words that are easy or hard to learn and to guide their learners to successful productive word learning.

In this article, we report on a longitudinal study that looks into how several word-related variables might influence L1 Dutch-speaking learners' productive knowledge of L2 German words in a university setting. The research questions are the following:

1) How do several word-related variables (i.e., frequency, cognateness, concreteness, and age of acquisition) influence L2 learners' productive vocabulary knowledge on a form recall level?
2) To what extent does the influence of these word-related variables change over time?

2. Literature review

2.1. Defining and measuring L2 productive vocabulary knowledge

Although L2 vocabulary knowledge is considered to be a complex construct, a well-known distinction in the literature on L2 vocabulary is the distinction between *receptive* (*passive*) and *productive* (*active*) vocabulary knowledge (e.g., Laufer, 1998; Nation, 2001; Webb, 2008). Receptive vocabulary knowledge refers to the ability to recognize and comprehend a word when it is encountered in listening or reading, whereas productive vocabulary knowledge means that we try to recall or produce language in speaking and writing (e.g., Meara & Miralpeix, 2021; Nation, 2001; 2022; Webb, 2008). In this regard, it

should be noted that many vocabulary researchers do not consider the difference between receptive and productive L2 vocabulary knowledge as a dichotomy, but rather as a continuum (e.g., Laufer & Goldstein, 2004; Schmitt, 2010). According to this view, receptive knowledge comprises the initial stage of L2 vocabulary acquisition and this receptive knowledge then eventually develops to the extent that it becomes productive (ibid.).

When it comes to testing learners' vocabulary knowledge, the terms *meaning recognition* and *form recognition* (for receptive/passive knowledge), and *meaning recall* and *form recall* (for productive/active knowledge) are often used (e.g., Schmitt, 2010). Laufer and Goldstein (2004) distinguish between four degrees of strength of knowledge, i.e., passive recognition, active recognition, passive recall and active recall. In their study, they suggest that there is a hierarchy of knowledge, and state that "the most advanced degree of knowledge is reflected in active recall and the least advanced knowledge is passive recognition" (Laufer & Goldstein, 2004, p. 408).

In the context of the present study, we are interested in learners' "most advanced degree of knowledge", and will assess learners' productive vocabulary on the form recall level. We intend to measure what is called *controlled productive vocabulary*, meaning that we will test word knowledge that learners have to produce in a prompted task, in which they are asked to provide a specific word with the first letter or letters given (Laufer, 1998; Laufer & Nation, 1999; Nation, 2001).

2.2. Word-related variables affecting L2 word knowledge

2.2.1. Frequency

A variable that has been widely studied and shown to influence L2 vocabulary learning is L2 frequency. L2 frequencies of words are usually derived from a corpus. These corpus frequencies are then used to estimate learners' real-world exposure to L2 input. It has

generally been accepted that high-frequency words are normally learned before low-frequency words, because high-frequency words are encountered more often (e.g., Ellis, 2002; Nation, 2022). Regarding receptive vocabulary knowledge in incidental/contextual learning of young learners (children between 10 and 12/13 years old), several studies indicate that learners' receptive vocabulary was indeed influenced by word frequency (De Wilde et al., 2020; Puimège & Peters, 2019a). Regarding productive knowledge, the number of studies is smaller. With younger learners, Segura et al. (2022) used a pre-test post-test design and detected a frequency effect in pre-primary bilingual EFL learners (aged 4 and 5) through a soft CLIL approach for productive vocabulary. When it comes to young adult L2 learners (aged between 19 and 21), Puimège and Peters (2019b) investigated the incidental learning of single words and formulaic sequences by watching English language television, also using a pre-test post-test design. One of the item-related factors that affected the learning in a form recall test was corpus frequency. Both studies are intervention studies, so the role of L2 frequency in a less controlled setting should be further explored.

Not only L2 frequency can be a factor in the learning difficulty of a L2 word, L1 frequency might also play a role. According to Peters (2020), the frequency of the corresponding L1 word is an underinvestigated factor that might influence L2 vocabulary learning. Some studies indeed suggest a positive influence of L1 frequency. De Groot et al. (2000) investigated the effect of L1 frequency and other word-related variables on paired-associate word learning and observed a small effect of L1 frequency. Goriot et al. (2021) found that L1 lexical frequency and cognate similarity influenced L1 Dutch pupils' performance on a receptive L2 English vocabulary test. Paquot (2017) analyzed lexical bundles in the argumentative essays of French and Spanish EFL learners and found a significant influence of L1 frequency. However, to the best of our knowledge, no studies have investigated how L1 frequency affects L2 single word productive vocabulary knowledge.

2.2.2. Cognateness

It is generally accepted that learners often do not start from scratch when they learn a foreign language. Cognates – words with a high similarity in form and meaning between L1 and L2 (e.g., *green*, *groen*, and *grün* in English, Dutch and German respectively) – are said to facilitate foreign language learning (e.g., de Groot & Keijzer, 2000; Elgort, 2013; Goriot et al., 2021; Lemhöfer et al., 2004). Several studies on incidental and contextual vocabulary learning have demonstrated that cognates are easier to learn compared to non-cognates. Peters and Webb (2018), for example, explored incidental vocabulary learning through viewing L2 television. Their study showed a facilitative effect for cognates. In different studies with young L2 learners – children between 7 and 12 years old –, an overall finding was that cognateness was an influencing factor on L2 meaning recognition (e.g., De Wilde et al., 2020, 2022; Bosma & Nota, 2020; Muñoz et al., 2018; Puimège & Peters, 2019a). Regarding the learnability of cognates, there are some studies in which paired-associate learning was used. Learners had to learn a list of cognates and a list of non-cognates, and form recall scores demonstrated that learners performed better for cognates than for non-cognates (e.g., de Groot & Keijzer, 2000; Rogers et al., 2015). On the other hand, a study by Otwinowska et al. (2020) suggests that the advantage of cognates does not hold in all learning contexts. The study revealed that Polish learners of English knew more cognates than other word types before the longitudinal quasi-experimental classroom study, but that during the experiment, they acquired cognates at the same rate as other word types. It is therefore necessary to further explore the role of cognateness in a non-experimental context, where learners learn new words both inside and outside the classroom.

Genetically related languages such as Dutch and German, both being Germanic languages, share more cognates than non-related languages. Van der Slik (2010) calculated

the cognate linguistic distance – a measure based on the number of cognate words in two different languages – between Dutch and 11 other European languages. The cognate linguistic distance between Dutch and German is .838, indicating that these languages share many cognates. It has been shown that two determining factors of cross-linguistic intelligibility are the percentage of cognate words in the two related languages and the formal similarity between the cognates (e.g., Tang & van Heuven, 2015). Language pairs that are so closely related can thus follow the principle of receptive multilingualism, which holds that "the speakers are able to communicate each using their own language without prior language instruction" (Gooskens et al., 2018, p. 170). L1 users can rely on the strategy of cognate guessing for understanding the L2, which implies that they try to guess the meaning of a word in the L2 based on similarities with words in their L1 (Vanhove & Berthele, 2015). For a closely related language pair such as Dutch-German, it has been shown that there is a cognate-learning-advantage for receptive vocabulary (e.g., Gooskens et al., 2011). When it comes to speaking proficiency, Schepens et al. (2013) demonstrated that linguistic distance predicted general difficulty of learning Dutch as a second language. L1 German learners thus obtained better scores compared to learners with a L1 with a larger linguistic distance. In their 2016 study, Schepens et al. found that larger linguistic distances from the L1 to the L3 and from the L2 to the L3 correlate with lower degrees of L3 learnability, based on speaking proficiency test scores of L3 Dutch learners. The question remains, however, if the cognate-learning-advantage applies for productive vocabulary knowledge at the form recall level too.

2.2.3. Concreteness

Concrete words can be defined as words that are easy to imagine, whereas abstract words are words that are more difficult to imagine (e.g., Peters, 2020). In the literature, the terms concreteness and imageability are often used interchangeably because of the high

correlation between the two (e.g., Brysbaert et al., 2014). A learning advantage in both L1 and L2 of concrete (or highly imageable) words over abstract words, the so-called concreteness effect, has been reported in a large number of studies (e.g., de Groot & Keijzer, 2000; Ding et al., 2017; Palmer et al., 2013; Puimège & Peters, 2019b). In L1 novel word learning, for example, Palmer et al. (2013) found that L1 speakers – who had to learn novel words in English – gave faster responses to the concrete novel words compared to the abstract words in a semantic categorization task and lexical decision task. Ding et al. (2017) observed that L1 Chinese learners, after a short learning phase, acquired both concrete and abstract L1 novel words, although the concrete words were learned better. In this context, the studies of de Groot and Keijzer (2000) and Puimège and Peters (2019b) should be mentioned too, since those studies used both receptive and productive testing, and showed that concrete words were easier to learn compared to abstract words. However, those results were observed in an experimental context. It would therefore be interesting to investigate whether the results would be the same in a non-experimental context.

2.2.4. Age of acquisition

Age of acquisition (AoA) refers to the age at which a concept or a skill is learned (Hernandez & Li, 2007). Early acquired words have an advantage over words acquired at a later age in both the L1 and the L2 (Brysbaert et al., 2000; DeKeyser, 2013; Izura & Ellis, 2002; Johnston & Barry, 2006; Llanes & Muñoz, 2013; Saito, 2013). Izura and Ellis (2002) investigated AoA effects in L1 and L2 word recognition and production in dominant Spanish-English bilinguals and conducted four experiments. In the first experiment, they detected an AoA effect (both for L1 and L2) in a productive picture naming task. The second, third and fourth experiment all employed a lexical decision task. In the second experiment, AoA effects were found for L1 and L2, whereas in the third and fourth experiment, there was no L1 AoA

effect in the L2 lexical decision task. They suggest that the AoA effect is related to the learning of word forms rather than meanings. In contrast, De Wilde et al. (2020) revealed that AoA of the L2 target words and AoA of the L1 translation equivalent both had an impact on young learners' L2 English receptive word knowledge. The authors found an overall effect of AoA of the L1 translation equivalent, indicating that words which were acquired at an earlier age in the L1 were easier to learn in the L2. For L2 English learners who were absolute beginners there was no effect of English AoA but for learners who were more proficient, it was shown that L2 English receptive vocabulary knowledge was also impacted by English AoA. Regarding L2 reading, Dirix and Duyck (2017) examined in a sample of 14 undergraduate students how L2 AoA and L1 AoA affected L2 reading speed through eye-tracking. It was shown that L2 reading speed was influenced by both L2 and L1 AoA. Since not much is known about L1 and L2 AoA effects on L2 productive vocabulary knowledge, it would be interesting to find out the effect of AoA on L2 learners' form recall knowledge.

3. Method

3.1. Participants

The participants in this study were 64 undergraduate language students with Dutch as L1, majoring in German and an additional foreign language at a university in Flanders, the northern Dutch-speaking part of Belgium. Their 3-year bachelor's program consisted of an in-depth study of Dutch and two foreign languages. They were all exposed to the same formal classroom instruction in German at university (190 contact hours in the first year, 215 in the second, and 140 in the third year). No prior knowledge of German is required for the study program. The targeted level for graduating is a B2/C1 level (upper-intermediate for speaking and writing; advanced for listening and reading) according to the Common European Framework of Reference (Council of Europe, 2001).

Participants' productive vocabulary knowledge was tested once a year, three times in total. Only students who participated in at least two of the three vocabulary tests were selected as study participants. As a result, there were 64 participants in this study. 40 of them took part in all three tests. Seven students completed test 1 and test 2, three students test 1 and test 3, and 14 students participated in test 2 and 3. There were thus 50 participants for the first test, 61 for the second and 57 for the third. Participants took part on a voluntary basis and provided informed consent.

3.2. Materials

3.2.1. Productive vocabulary test

To assess students' L2 productive word knowledge on a form recall level, a productive vocabulary cloze test for German was administered. This test, modelled after Nation's Productive Vocabulary Levels Test (PVLT) for English (Nation, 1990) is a widely used measure (e.g., Eguchi et al., 2022; Kormos et al., 2022; Sonbul et al., 2023) and aims to assess participants' vocabulary knowledge on levels of 1000 words each. For the German version, these levels are based on the frequency lists developed from the Herder/BYU-corpus (Jones et al., 2006). The test was developed by the Institute for Test Research and Test Development in Leipzig, Germany. It consists of 90 cloze items (18 items per frequency level), which measure students' productive vocabulary on the levels 1000, 2000, 3000, 4000, and 5000. Each target word is surrounded by one or two sentences and the first letter or letters of the word are provided to disambiguate the item. For example: *In der Klasse gibt es zehn Jungen und zwölf Mä*_____. („In the class, there are 10 boys and 12 gi____.")

Each year, a different version of the test – containing other items per frequency level – was administered. One version is freely available on the internet (see http://www.itt-leipzig.de/static/vltgerman_01p/index.html), and has proven to be a highly reliable instrument

for assessing productive vocabulary knowledge of L2 learners (Tschirner, 2021). A second and third version of the test (in paper and pencil format) were made available by the Institute for Test Research and Test Development for the purpose of this study. Since each test has 90 items, the dataset of this study contained 270 items in total.

3.2.2. Word-related variables in the study

Frequency: Although the test is based on the frequency lists of the Herder/BYU-corpus (Jones et al., 2006), the decision was made to use the word frequencies from the SUBTLEX-DE corpus (Brysbaert et al., 2011) for this study. The SUBTLEX-DE corpus consists of 25.4 million German words from popular films and television series. It was shown that these subtitle-based frequencies outperform frequencies based on written sources for predicting both word processing efficiency in psychological experiments (Brysbaert et al., 2011) and L2 vocabulary learning (De Wilde et al., 2020). To verify the effect of L1 frequency, we also added the frequencies of the SUBTLEX-NL corpus, consisting of 44 million Dutch words based on film and television subtitles (Keuleers et al., 2010). The SUBTLEX Zipf scale (van Heuven et al., 2014) was used. All frequency counts were log transformed with the formula log10 (frequency per million words)+3. The advantage of this scale is that it is logarithmic and that the values are easy to interpret (ibid.).

Cognateness: To identify translation equivalents that are cognates, the cross-language orthographic similarity of each word pair was assessed. To this end, Levenshtein distance between the German items and their Dutch translation equivalents was calculated using the Levenshtein.distance function from the "vwr" package in R (Keuleers, 2013). The Levenshtein distance is the minimum number of insertions, deletions or substitutions required to change one word into another. To make sure that orthographic similarity is comparable

between word pairs of different length, the Levenshtein distance was normalized, as suggested by Schepens et al. (2012). The normalized distance score = 1–(distance/length), with distance representing the Levenshtein distance between both words, and length representing the number of letters of the longest word (either the German item or its Dutch translation equivalent). This results in a score between 0 and 1. When there is no overlap between the words, the normalized Levenshtein distance is 0, whereas identical words show a value of 1.

Concreteness: To determine the target items' concreteness values, the dataset of Charbonnier and Wartena (2021) was used. For German, various word norms with concreteness values have been created, but most of them are quite small. Charbonnier and Wartena created a merged dataset, containing 4181 words. 133 German target items of the present study appeared in the dataset. For the remaining 137 items, no concreteness value was found. Therefore, we used Brysbaert et al.'s (2014) concreteness values to determine the concreteness of the Dutch translation equivalent of the German target items. The advantage of the database of Brysbaert et al. (2014) is that the concreteness norms are available for 30,000 Dutch words.

AoA: Although a large number of AoA ratings exist in many languages, there are not so many available for German. To explore AoA effects, the database of Birchenough et al. (2017) was used. This German AoA database contains subjective AoA ratings for 3,259 German words, including 2,363 nouns and 473 verbs. To the best of our knowledge, the database of Birchenough et al. (2017) is the most extensive one for German until today. AoA ratings for the 270 German target items of this study were searched for in this downloadable database, and 94 target items were found, meaning that there were no AoA ratings for German

for the remaining 176 target items. AoA norms for the Dutch translation equivalent of the 270 German target items were taken from Brysbaert et al. (2014), a database with AoA ratings for 30,000 Dutch words.

3.3. Procedure

Participants were tracked for three years. Once a year, a different version of the productive vocabulary test was administered in class as a paper-and-pencil test. There was a time limit of 30 minutes to complete the test, and students completed it individually without using a dictionary or other resources. In total, there were three test sessions. The first test was administered at the beginning of students' first year of university. The second one was taken in the second semester of the second year, the third one in the second semester of the third year.

All tests were corrected manually by the first author of this study by using an answer key that had been provided by the test developers. Part of the data were checked by the second author of the study, and no mistakes were found. Each test item was scored either one point for a completely correct answer (e.g., *erstaunlich* for "astonishing") or zero points for an incorrect answer (*erstaunend or erstaunt* for "astonishing"), following the *all-or-nothing approach* (Schmitt, 2010). Spelling had to be correct, because the missing of an umlaut (i.e., two dots on the vowels a, o, u, changing them into ä, ö, ü), for example, can completely alter the meaning of a word in German (e.g., *drucken* "to print" is not the same as *drücken* "to push"). A binary score for each item was given, which was then used in the analyses.

Regarding the word-related variables, a Dutch translation equivalent was needed to be able to calculate Levenshtein distance, and to correlate the German concreteness and AoA norms with the Dutch norms. Because some words have more than one translation equivalent, the decision was made to use the equivalent that fitted best into the context sentence in which

the word was embedded. An example: the German word *Hals* (“throat” or “neck”) can be translated in Dutch with (“keel” or “nek”). The sentence in the test was: *Ich habe Schmerzen im Ha_____. Ich kann kaum sprechen.* („I have a sore throat. I can hardly speak.“), so the Dutch equivalent that was used here was *keel* (“throat”).

3.4. Analyses

All analyses were carried out using the R software environment (version 4.1.2.; R Core Team, 2022). We first calculated descriptive statistics for the overall score on the vocabulary test and for the various word-related variables. Then, we calculated Pearson correlations between the different word-related variables and between the word-related variables and the score on the different items of the vocabulary test. In order to investigate the impact of various word-related variables on the difficulty of the vocabulary items, we built a generalized linear mixed model (since the score was a binary score, 0 or 1) using the lme4-package in R (Bates et al., 2015). We first centered the word-related variables around the mean. Then we built a basic model with only the random effects, random intercepts for items and participants and a random slope for time. We then consecutively added the fixed effects (time and various word-related variables). To answer the second research question, we added interactions between time and the word-related variables. To assess model fit, we applied the procedure anova (model1, model2) (Brysbaert, 2020). 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, were calculated using the MuMin package (Nakagawa & Schielzeth, 2013) in R. All data and code can be found on https://osf.io/jh7gf/?view_only=33b525599113433e900b0342a26e3fff.

4. Results

First, we calculated descriptive statistics for the total score on the vocabulary tests at different times (T1, T2 and T3; cf. Table 1 and Figure 1). The results show a large increase in the score between times 1 and 2 and a decrease in the score between times 2 and 3. T-tests showed the difference in the score was significant between times 1 and 2 with a large effect size (d = -3.37) and between times 2 and 3 with a medium effect size (d = 0.56). Tables 2 and Table 3 show the descriptive statistics for the various word-related variables overall and per test.

Table 1. Descriptive statistics for the vocabulary tests at test times 1, 2 and 3 (max score = 90)

	Min	Max	Median	Mean	SD
1. Time 1	15	58	38	37.82	9.12
2. Time 2	48	82	66	65.46	7.37
3. Time 3	45	80	61	61.19	7.81

Note: At times 1, 2 and 3 there were, respectively, 50, 61 and 57 study participants.

Figure 1. Boxplot showing the vocabulary test scores at times 1, 2 and 3.

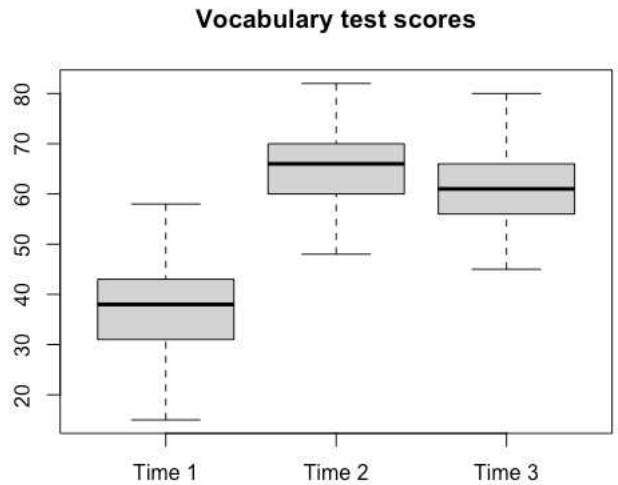

Table 2. Descriptive statistics for the word-related variables

	Min	Max	Median	Mean	SD
1. Subtlex-DE German	0.30	4.47	2.52	2.42	0.79
2. Subtlex-NL Dutch	1.66	6.20	4.11	4.00	0.90
3. Concreteness German	2.48	6.69	4.64	4.80	1.13
4. Concreteness Dutch	1.13	5.00	2.80	2.96	0.97
5. Age of Acquisition German	2.67	12.15	5.60	5.73	2.08
6. Age of Acquisition Dutch	3.76	13.72	8.11	8.04	2.19
7. Cognateness	0.00	1.00	0.45	0.48	0.26

Table 3. Mean and standard deviation for the word-related variables per test

	Test 1 M(SD)	Test 2 M(SD)	Test3 M(SD)
1. Subtlex-DE German	2.49 (0.80)	2.43 (0.79)	2.34 (0.75)
2. Subtlex-NL Dutch	4.06 (0.90)	4.01 (0.94)	3.94 (0.87)
3. Concreteness German	4.67 (1.20)	4.73 (1.06)	4.99 (1.29)
4. Concreteness Dutch	2.90 (0.96)	2.92 (0.94)	3.05 (1.01)
5. Age of Acquisition German	5.77 (2.12)	5.93 (2.19)	5.42 (1.81)
6. Age of Acquisition Dutch	8.04 (2.19)	8.09 (2.08)	7.97 (2.29)
7. Cognateness	0.48 (0.27)	0.48 (0.25)	0.47 (0.28)

In order to get a better view on the relationship between the various word-related variables and the word-related variables and the score on the various items from the vocabulary tests, we calculated Pearson correlations (cf. Tables 4 and 5). The results show

large positive correlations between concreteness and age-of-acquisition norms for the German items and their Dutch translation equivalents. Because these variables correlate strongly and could possibly lead to multicollinearity and because the norms for both concreteness and age of acquisition in German are incomplete, it was decided to use the Dutch norms in our further (generalized mixed model) analyses. Furthermore, as German is a foreign language which is typically learnt in a formal context from adolescence onwards rather than in a naturalistic setting, one might expect a larger impact from L1 Dutch AoA on word learning. There are moderate positive correlations between time and frequency and the item score, suggesting that more words become known over time and more frequent words are easier to learn. There is a moderate negative correlation between age of acquisition and item score, which might suggest that words which are learnt earlier in life by L1 German speakers and L1 Dutch translations that are acquired earlier in life are easier to learn for L2 German learners. Correlations between the item score and concreteness and cognateness are still significant but are somewhat lower.

Table 4. *Summary of correlations (Pearson's r) between time and the various word-related variables*

	1	2	3	4	5	6	7	8
1. Time		-.07***	-.05***	.11***	.06***	-.06***	-.01	-.02*
2. Subtlex-DE Ge			.77***	.33***	.21***	-.62***	-.59***	.08***
3. Subtlex-NL Du				.33***	.23***	-.36***	-.70***	-.03***
4. Concreteness Ge					.80***	-.51***	-.60***	-.02
5. Concreteness Du						-.40***	-.59***	.02
6. Age of Acquisition Ge							.78***	-.25***

7.	Age of Acquisition	-.02*
	Du	
8.	Cognateness	

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

Table 5. *Correlations (Pearson's r) between the various word-related variables and the vocabulary item score*

	Item score
1. Time	.35***
2. Subtlex-DE German	.45***
3. Subtlex-NL Dutch	.34***
4. Concreteness German	.27***
5. Concreteness Dutch	.19***
6. Age of Acquisition German	-.37***
7. Age of Acquisition Dutch	-.38***
8. Cognateness	.11***

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

To answer the first research question – namely, how several word-related variables (i.e., frequency, cognateness, concreteness, and age of acquisition) influence L2 learners' productive vocabulary knowledge on a form recall level – we built a generalized linear mixed effects model. The basic model included random intercepts for participants and items and a random slope for time: glmer(score ~ 1|item + time|id). It showed that most of the variation was explained by the variable 'item' (variance = 6.40, *SD* = 2.53); the variable 'participant' explained far less variation (variance = 0.85, *SD* = 0.92). The model with a random slope for

time (variance = 0.11, SD = 0.34) was better than the model without the random slope. We then added time and the various word-related variables as fixed effects. The model containing fixed effects for time, German frequency, age of acquisition (Dutch) and cognateness was the best model (cf. Table 6). The model explained 28% of the variance in the scores.

Table 6. GLME Model Predicting Right or Wrong Answers on the Vocabulary Tests

Random effects		Variance	SD
Item	(Intercept)	3.59	1.90
Participant	(Intercept)	0.80	0.89
Time		0.10	0.31

Fixed effects	Estimate	SE	Z-value	P
(Intercept)	-1.62	0.34	-4.71	<.001 ***
Time	1.22	0.16	7.84	<.001 ***
Subtlex-DE German	1.46	0.20	7.42	<.001 ***
Dutch AoA	-0.16	0.07	-2.40	.01*
Cognateness	1.24	0.46	2.67	.008 **

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

As we also wanted to gauge how the impact of various word-related variables could differ at the various time points, we built another model in which we added interactions between time and the word-related variables which showed to have an impact in the best model without interactions (cf. Table 5). None of the interactions were significant and the model without interactions proved a better model than the model with interactions. This

indicates that the impact of several word-related variables on word learning does not change over time.

5. Discussion

This study aimed to investigate the role of several word-related variables (i.e., frequency, cognateness, concreteness, and age of acquisition) in the development of productive single word knowledge in L2 German. To our knowledge, it is the first study to examine the effect of those variables regarding L2 single word productive knowledge on a form recall level, and to evaluate their effect over time.

5.1. How do several word-related variables influence L2 learners' productive vocabulary knowledge?

Frequency: The present study found that L2 German word frequency – based on the SUBTLEX-DE frequencies – had a statistically significant positive effect on participants' productive L2 vocabulary knowledge and was the main predictor for participants' form recall scores. High-frequency words in the L2 were thus recalled better than low-frequency ones by our participants, which is in line with the frequency-based model of vocabulary learning (Ellis, 2002). This frequency effect has been found in several studies on word recognition (e.g., De Wilde et al., 2020; Puimège & Peters, 2019a), and in intervention studies on word production (e.g., Segura et al., 2022; Puimège & Peters, 2019b). The present study illustrates that L2 frequency also affects word form recall in a non-experimental language learning setting.

Because L1 word frequency has been suggested to influence receptive vocabulary learning (e.g., Goriot et al., 2021), we also added the SUBTLEX-NL frequencies in the analyses. The best final model, however, did not include this variable. The fact that our findings differ from Goriot et al. (2021) might be due to the fact that our participants were

young adults learning L2 German at university, whereas the participants in Goriot et al.'s study were children and adolescents learning L2 English. In Goriot's study, it was found that younger and unexperienced L2 learners' scores were predicted by L1 frequency, whereas L2 frequency predicted older and more experienced learners' scores. The university learners in the present study can be considered more experienced learners, so it is likely that they – in contrast to children – relied less on L1 frequency. Another explanation might be that the effect for L1 frequency was not there for Dutch-speaking learners, because of the high linguistic similarity between German and Dutch, or that the effect was so small it could not be detected through the current sample size. Our findings also contrast with those of Paquot (2017), who found strong positive correlations between the frequency of a lexical bundle in French and Spanish EFL learners' essays and the frequency of its equivalent from L1 French and Spanish corpora. However, writing an argumentative essays is very different compared to completing a productive vocabulary cloze test, so it is difficult to compare the findings. In any case, further research is warranted to clarify the impact of L1 word frequency on productive vocabulary learning.

Cognateness: Another variable in the final model that predicted learners' word recall scores was cognateness, i.e., the cross-language orthographic similarity between the German items and their Dutch translation equivalents. This means that our participants were better in recalling L2 German words that were orthographically similar in L1 and L2. The effect for receptive vocabulary knowledge has been shown repeatedly (e.g., De Wilde et al., 2020, 2022; Muñoz et al., 2018; Puimège & Peters, 2019a) and it is in a way unsurprising as the cognate word can be heard or read and linked to its translation equivalent. The present study suggests that the effect is also present – although to a smaller extent – in a form recall task, in which neither word (not the L2 word nor the translation equivalent) can be heard or seen but still seems to be easier to recall when there is more similarity between the L1 and L2 word. It

should be noted, however, that the test format of the productive vocabulary test, in which the first letter or letters of a word were given, might have helped students to recall the word. Especially words with a large orthographic similarity might have been easier to recall. For example, for the sentence *Du sollst deinem Gefühl fo_____ und dich dann entscheiden.* ("You should fo_____ your intuition and then decide."), the target word in German is *folgen*, in English *follow*, in Dutch it is *volgen*, which is very similar in form. In any case, the finding on cognateness in our study corroborates the results of the study of de Groot and Keijzer (2000), in which it was shown that learners' scores were positively influenced by cognateness, both for receptive and productive testing. Since our study was conducted in another context, it adds to the existing literature by showing that the advantage of cognates also holds in a non-experimental language learning context.

Concreteness: Although many other studies on L2 vocabulary learning have shown a learning advantage of concrete words over abstract words (e.g., de Groot & Keijzer, 2000; Puimège & Peters, 2019b), we did not find a facilitative effect for concreteness in the present study. When participants had to recall the form of a specific word, the concreteness of the word did not seem to have an impact on their scoring. This contrasts with the findings of Puimège and Peters (2019b), who found that concreteness did affect the incidental learning of single words from watching L2 television. Although concrete words might be easier to learn incidentally, this might be different in another learning context, in which vocabulary is learned inside and outside the classroom. In a longitudinal study into L2 learners' productive collocation knowledge (Boone et al., 2023), which was conducted in a similar university learning context, no effect for imageability was found either. It thus seems that the effect for concreteness might differ depending on the kind of vocabulary knowledge (receptive vs. productive) tested or depending on the learning context.

Age of acquisition: Regarding the variable AoA, the results indicated a statistically significant negative effect of the L1 (Dutch) AoA ratings on L2 learners' productive vocabulary knowledge, meaning that L2 German words of which the L1 equivalent was learned at an earlier age were associated with a higher score on the form recall test. The finding that early-learned words have an advantage over later-learned words is not new, but has been shown, for example, in studies on lexical processing or receptive knowledge (e.g., Brysbaert, 2000; De Wilde et al., 2020). The AoA effect of the present study, in which form recall was measured, is in line with the study of Izura and Ellis (2002), who suggest that AoA affects the acquisition of word forms rather than meanings. However, in contrast to the findings of Izura and Ellis, we found that L1 AoA influenced learners' L2 vocabulary scores. It should be noted that in the present study, we only took into account the AoA-norms of the L1 translation equivalent, because of the large correlations between the AoA-norms of the German items and their Dutch translation equivalents, because the norms for AoA in German were incomplete, and because our learners did not follow a naturalistic learning trajectory. In any case, the L1 AoA effect found here aligns with the results of De Wilde et al. (2020) and Dirix and Duyck (2017).

5.2. To what extent does the influence of these word-related variables change over time?

The results did not indicate any statistically significant interaction between time and the significant variables (i.e., frequency, cognateness, and L1 AoA), meaning that the effect of those word-related variables on productive word learning did not change over time. Thus, frequent words, cognate words, and L1 words that were acquired early in life, had a positive effect on L2 students' productive vocabulary knowledge at T1, T2 and T3. To the best of our knowledge, this is the first longitudinal study which investigates the effects of various word-related variables on form recall, so more research is needed to be able to generalise the

findings. In the study of Boone et al. (2023) on productive collocation knowledge, for example, it was found that the impact of congruency (i.e., the presence of a L1-L2 literal translation equivalent) on learners' productive collocation knowledge did change over time. In that study, students' knowledge of congruent collocations remained relatively stable over time, whereas their knowledge of incongruent items significantly increased. Therefore, more longitudinal research is needed to gauge the effect of several word-related variables on vocabulary development, both for single words and multi-word units.

In the final model of the present study, time emerged as significant predictor, which is not surprising, since our participants were learning German at university and are thus supposed to acquire more vocabulary as time goes by. However, here it is interesting to note that there is a big leap between T1 and T2, and a slight decrease in vocabulary scores between T2 and T3. This might be explained by the formal instruction context students were in and the exposure to German at university, where students got 190 contact hours in the first year, 215 in the second, and 140 in the third year. When the first test was administered, students were still at the very beginning of their learning trajectory and had had almost no contact hours. The second test was taken in the second semester of the second year, meaning that they had had two years of intensive classroom instruction, with an explicit focus on vocabulary and grammar. In the third year, the focus in their program was more on language practice, and the contact hours were limited to 140. It can be observed (see Table 1) that at this point, their productive vocabulary knowledge slightly decreases (the mean score of 65.46 becomes 61.19 out of a maximum of 90). In our opinion, this reduction in L2 input at university may explain the decrease in vocabulary knowledge between T2 and T3 as "input never ceases to play a role in an instructed language learning setting" (Muñoz, 2011, p. 113). This finding shows some resemblance to the results obtained by Akbarian et al. (2020), who found that the amount of instructional exposure explained the variation in learners' lexical production. In

this regard, it is important to note that in L2 lexical acquisition, there seems to be a strong association between input and output (e.g., Crossley et al., 2016). Another explanation might be that the three tests contained different items. Although there were no significant differences between test 1, 2 and 3 regarding the word-related variables investigated in the present study, it is not impossible that the second test – compared to the third test – contained some more words students were familiar with, maybe because they had encountered some of them. Finally, our findings suggest that it is necessary to actively engage with vocabulary in the classroom, even with more intermediate/advanced learners. Especially on a productive level, active engagement is needed, since active recall is the most advanced degree of knowledge and the most difficult to acquire (Laufer & Goldstein, 2004).

5.3. Pedagogical implications

Although our study has focused on only one level of L2 vocabulary knowledge and on one language pair (Dutch-German), we believe that it has important implications for both language teachers and textbook writers. As we have shown, learners' form recall scores were influenced positively by L2 frequency, cognateness and L1 AoA. This means that less frequent words, non-cognate words and L1 words that were learned later in life were those words that were difficult for learners – in this case students in a university language program – to recall correctly. In order to design effective vocabulary learning material or exercises, teachers and textbook writers should thus take into account those word-related variables and cross-check the selected materials against existing lexical databases (e.g., with frequencies or AoA ratings). A special focus in the materials or in the classroom on low frequent words and words that were learned later in life, for example, may boost L2 learners' vocabulary knowledge. Based on learners' higher recall scores for cognate words, our study also suggests a cognate learning advantage for productive vocabulary knowledge. Paying special attention

to non-cognate words in the classroom, especially for a language pair with a small cognate distance such as Dutch-German, might prove useful here. Since the effect of the word-related variables did not change over time, the implications here seem to be relevant to the levels from beginner to upper intermediate. However, the findings of the present study should be interpreted with caution and future research replicating these findings is necessary. Although the results of our study are in line with the existing literature on L2 receptive vocabulary and on other language pairs (e.g., English-Dutch), more studies on productive vocabulary and word-related variables influencing this kind of knowledge should be done before firm conclusions can be drawn.

5.4. Limitations and future directions

The findings of this study have to be seen in light of some limitations that could drive future research. First, the sample of participants is relatively small. However, it has to be noted that we investigated a LOTE, and university learners who choose to study German are not as numerous as L2 English learners. It should also be remarked that for a 3-year longitudinal study, the sample of participants is reasonable. In future studies, it would be interesting, however, to conduct a study with more participants – also non-language specialists – and with other language backgrounds (e.g., French, Spanish, Chinese, …) to see whether the same findings would be obtained. Second, the number of word-related variables investigated here is restricted. Future research could thus investigate the role of other influencing variables, such as polysemy or homonymy, part of speech, and others (for an overview, see Peters, 2020). Although in our study, most of the variance was explained by the variable 'item', also learner-related variables could be added in future studies to explore the extent to which they contribute to L2 learners' productive single word learning.

Conclusion

The present study investigated longitudinally to what extent time, frequency, cognateness, concreteness and AoA affect productive single word knowledge in L2 German. Students' L2 productive word knowledge on a form recall level was measured three times, using three different versions of the German Productive Vocabulary Levels Test. Overall, this study contributed to the existing literature on vocabulary research by showing that several word-related variables (i.e., L2 frequency, cognateness and L1 AoA) that have been suggested to influence L2 receptive word knowledge also seem to have an impact on productive word knowledge. On the other hand, the present study did not find a concreteness effect, which tentatively suggests that the role specific word-related variables play in learning new vocabulary may depend on the learning context and the type of vocabulary knowledge. Finally, it is interesting to note that the effect of the investigated variables on learners' scores of L2 form recall did not change over time.

References

Akbarian, I., Farajollahi, F., & Jiménez Catalán, R. M. (2020). EFL learners' lexical availability: Exploring frequency, exposure, and vocabulary level. *System*, *91*, 102261. https://doi.org/10.1016/j.system.2020.102261

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1).

Birchenough, J. M. H., Davies, R., & Connelly, V. (2017). Rated age-of-acquisition norms for over 3,200 German words. *Behavior Research Methods*, *49*(2), 484–501. https://doi.org/10.3758/s13428-016-0718-0

Boone, G., De Wilde, V., & Eyckmans, J. (2023). A longitudinal study into learners' productive collocation knowledge in L2 German and factors affecting the learning. *Studies in Second Language Acquisition*, *45*(2), 503–525. https://doi.org/10.1017/S0272263122000377

Bosma, E., & Nota, N. (2020). Cognate facilitation in Frisian–Dutch bilingual children's sentence reading: An eye-tracking study. *Journal of Experimental Child Psychology*, *189*, 104699. https://doi.org/10.1016/j.jecp.2019.104699

Brysbaert, M. (2020). *Basic statistics for psychologists (2nd edition)*. Macmilan International.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

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., 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

Charbonnier, Jean, & Wartena, Christian. (2021). *Predicting the Concreteness of German Words*. In Proceedings of the 5th Swiss Text Analytics Conference (SwissText) & 16th Conference on Natural Language Processing (KONVENS). CEUR-WS. Vol-2624. https://doi.org/10.25968/OPUS-2075

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Crossley, S., Kyle, K., & Salsbury, T. (2016). A Usage-Based Investigation of L2 Lexical Acquisition: The Role of Input and Output. *The Modern Language Journal*, *100*(3), 702–715. https://doi.org/10.1111/modl.12344

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

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, *63*, 52–67. https://doi.org/10.1111/j.1467-9922.2012.00737.x

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English Through Out-of-School Exposure: How Do Word-Related Variables and Proficiency Influence Receptive Vocabulary Learning? *Language Learning*, *70*(2), 349–381. https://doi.org/10.1111/lang.12380

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2022). Formal versus informal L2 learning: How do individual differences and word-related variables influence French and English L2 vocabulary learning in Dutch-speaking children? *Studies in Second Language Acquisition*, *44*(1), 87–111. https://doi.org/10.1017/S0272263121000097

Ding, J., Liu, W., & Yang, Y. (2017). The Influence of Concreteness of Concepts on the Integration of Novel Words into the Semantic Network. *Frontiers in Psychology*, *8*, 2111. https://doi.org/10.3389/fpsyg.2017.02111

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

Eguchi, M., Suzuki, S., & Suzuki, Y. (2022). Lexical competence underlying second language word association tasks: Examining the construct validity of response type and response time

measures. *Studies in Second Language Acquisition*, *44*(1), 112–142. https://doi.org/10.1017/S0272263121000164

Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing*, *30*(2), 253–272. https://doi.org/10.1177/0265532212459028

Ellis, N. C. (2002). Frequency effects in language processing: A Review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Fitzpatrick, T., & Clenton, J. (2017). Making Sense of Learner Performance on Tests of Productive Vocabulary Knowledge. *TESOL Quarterly*, *51*(4), 844–867. https://doi.org/10.1002/tesq.356

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

Jones, R. L., Tschirner, E. P., Goldhahn, A., Buchwald, I., & Ittner, A. (2006). *A frequency dictionary of German: Core vocabulary for learners*. Routledge.

Gooskens, C., Kürschner, S. & van Bezooijen, R. (2011). Intelligibility of standard German and low German to speakers of Dutch. *Dialectologia 7,* 35–63.

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. (2021). Using the peabody picture vocabulary test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism*, *24*(4), 546–568. https://doi.org/10.1080/13670050.2018.1494131

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, *133*(4), 638–650. https://doi.org/10.1037/0033-2909.133.4.638

Izura, C., & Ellis, A. W. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Psicológica, 23*(2), 245–281.

Keuleers, E. (2013). vwr: Useful functions for visual word recognition research. Retrieved from https://cran.r-project.org/package=vwr

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

Knopp, E. (2022). Effects of Dominance in Exposure and Linguistic Distance in the Bilingual Proficiency Profiles of Dutch-German and Greek-German School-Age Bilinguals. *Journal of Home Language Research*, *5*(1), 2. https://doi.org/10.16993/jhlr.42

Kormos, J., Suzuki, S., & Eguchi, M. (2022). The role of input modality and vocabulary knowledge in alignment in reading-to-speaking tasks. *System*, *108*, 102854. https://doi.org/10.1016/j.system.2022.102854

Laufer, B. (1988). The concept of 'synforms' (similar lexical forms) in vocabulary acquisition. *Language and Education*, *2*(2), 113–132.

Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 140–155). Cambridge: Cambridge University Press.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, *19*(2), 255–271. https://doi.org/10.1093/applin/19.2.255

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. https://doi.org/10.1177/026553229901600103

Lemhöfer, K., Dijkstra, T., & Michel, M. (2004). Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language and Cognitive Processes*, *19*(5), 585–611. https://doi.org/10.1080/01690960444000007

Lindgren, J., & Bohnacker, U. (2020). Vocabulary development in closely-related languages: Age, word type and cognate facilitation effects in bilingual Swedish-German preschool children. *Linguistic Approaches to Bilingualism*, *10*(5), 587–622. https://doi.org/10.1075/lab.18041.lin

Llanes, À., & Muñoz, C. (2013). Age effects in a study abroad context: Children and adults studying abroad and at home. *Language Learning*, *63*(1), 63–90. https://doi.org/10.1111/j.1467-9922.2012.00731.x

Łuniewska, M., Haman, E., Armon-Lotem, S., *et al.* (2016). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words? *Behavior Research Methods*, *48*(3), 1154–1177. https://doi.org/10.3758/s13428-015-0636-6

Meara, P., & Miralpeix, I. (2021). Bayesian vocabulary tests. *Vigo International Journal of Applied Linguistics*, *18*, 177–205. https://doi.org/10.35869/vial.v0i18.3370

Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *IRAL - International Review of Applied Linguistics in Language Teaching*, *49*(2). https://doi.org/10.1515/iral.2011.006

Muñoz, C., Cadierno, T., & Casas, I. (2018). Different starting points for English language learning: A comparative study of Danish and Spanish young learners. *Language Learning*, *68*(4), 1076–1109. https://doi.org/10.1111/lang.12309

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.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Newbury House Publishers.

Nation, I. S. P. (2022). *Learning vocabulary in another language* (Third edition). Cambridge University Press.

Otwinowska, A., Foryś-Nogala, M., Kobosko, W., & Szewczyk, J. (2020). Learning orthographic cognates and non-cognates in the classroom: Does awareness of cross-linguistic similarity matter? *Language Learning*, *70*(3), 685–731. https://doi.org/10.1111/lang.12390

Palmer, S. D., MacGregor, L. J., & Havelka, J. (2013). Concreteness effects in single-meaning, multi-meaning and newly acquired words. *Brain Research*, *1538*, 135–150. https://doi.org/10.1016/j.brainres.2013.09.015

Paquot, M. (2017). L1 frequency in foreign language acquisition: Recurrent word combinations in French and Spanish EFL learner writing. *Second Language Research*, *33*(1), 13–32. https://doi.org/10.1177/0267658315620265

Peters, E. (2020). Factors affecting the learning of single-word items. In: S. Webb (Eds.), *The Routledge Handbook of Vocabulary Studies*, pp. 125–142. Routledge.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Puimège, E., & Peters, E. (2019a). Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning*, *69*(4), 943–977. https://doi.org/10.1111/lang.12364

Puimège, E., & Peters, E. (2019b). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*(4), 424–438. https://doi.org/10.1080/09571736.2019.1638630

R Core Team (2018). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org

Rogers, J., Webb, S., & Nakata, T. (2015). Do the cognacy characteristics of loanwords make them more easily learned than noncognates? *Language Teaching Research*, *19*(1), 9–27. https://doi.org/10.1177/1362168814541752

Saito, K. (2013). Age effects on late bilingualism: The production development of /ɹ/ by high-proficiency Japanese learners of English. *Journal of Memory and Language*, *69*(4), 546–562. https://doi.org/10.1016/j.jml.2013.07.003

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, *15*(1), 157–166. https://doi.org/10.1017/S1366728910000623

Schepens, J., van der Slik, F., & van Hout, R. (2013). Learning complex features: A morphological account of L2 learnability. *Language Dynamics and Change*, *3*(2), 218-244.

Schepens, J. 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

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2010). *Researching Vocabulary*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230293977

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913–951. https://doi.org/10.1111/lang.12077

Sonbul, S., El-Dakhs, D. A. S., & Masrai, A. (2023). Second language productive knowledge of collocations: Does knowledge of individual words matter? *Studies in Second Language Acquisition*, *45*(2), 480–502. https://doi.org/10.1017/S0272263122000341

Tang, C., & van Heuven, V. J. (2015). Predicting mutual intelligibility of Chinese dialects from multiple objective linguistic distance measures. *Linguistics*, *53*(2). https://doi.org/10.1515/ling-2015-0005

Tschirner, Erwin (2021): *Examining the validity and reliability of the ITT vocabulary size tests*. Leipzig (Research papers in assessment 3).

Van der Slik, F. W. P. (2010). Acquisition of Dutch as a second language: The explanative power of cognate and genetic linguistic distance measures for 11 West European first languages. *Studies in Second Language Acquisition*, *32*(3), 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

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, *30*(01). https://doi.org/10.1017/S0272263108080042