

A longitudinal study into learners' productive collocation knowledge in L2 German and factors affecting the learning

Abstract

This longitudinal study explored the roles of item- and learner-related variables in L2 learners' development of productive collocation knowledge (L1 = Dutch; L2 = German; $N_{\text{Learners}}= 50$). Learners' form recall knowledge of 35 target collocations was measured three times over a 3-year period. The item-related variables investigated were L1-L2 congruency, corpus frequency, association strength, and imageability. We also explored the learner-related variables L2 prior productive vocabulary knowledge and L2 immersion. Mixed-effects regression modeling indicated a significant effect of time, congruency and prior productive vocabulary knowledge on learners' collocation learning. While learners' knowledge of congruent collocations remained relatively stable after year one, knowledge of incongruent collocations increased significantly. Learners' prior productive vocabulary knowledge was clearly associated with growth of productive collocation knowledge, but besides overall growth there were instances of attrition.

Introduction

Various types of formulaic language, including collocations, are widespread in language and are of great importance in learning an additional language (e.g., Erman & Warren, 2000; Wray, 2002). Especially when it comes to the fluent, accurate and idiomatic production of both spoken and written language, collocation knowledge plays a crucial role (e.g., Sinclair, 1991). However, despite the large and growing number of pertinent studies, and the fact that collocations have been shown to be quite challenging even for advanced L2 learners (e.g.,

Boers et al., 2014; Laufer & Waldman, 2011; Nesselhauf, 2003), relatively few studies have addressed the longitudinal development of learners' productive collocation knowledge. The few studies that have addressed this subject have mainly focused on learners' use of collocations in (academic) writing (e.g., Edmonds & Gudmestad, 2021; Li & Schmitt, 2010; Siyanova-Chanturia, 2015; Siyanova-Chanturia & Spina, 2020) or they were intervention studies designed to test the effect of a specific type of input on the incidental acquisition of collocations (e.g., Vu & Peters, 2021). These longitudinal studies have undoubtedly provided valuable insights into learners' collocation development. Even so, much remains to be learned about how learners' L2 collocation knowledge develops over a longer time span as well as about the factors that affect this knowledge.

In a foreign language learning context, vocabulary and collocation learning can occur both intentionally and incidentally (e.g., Hulstijn, 2003). The term "intentional" has commonly been used in collocation research when investigating the effectiveness of different forms of classroom-based teaching, and "incidental" when examining learners' collocational gains after they engaged in meaning-focused communicative activities such as reading or viewing television (see, for a review, Szudarski, 2017). However, language learners – especially university foreign language majors who wish to develop their language skills for professional or personal reasons – might also have the intention to learn when reading or listening in the L2, as "coming across unfamiliar words during reading may trigger different kinds of processes, from basic visual intake and semantic integration to deliberate attempts to encode form and derive meaning" (Elgort et al., 2018, p. 363). Therefore, we believe that the term "contextual learning" from Elgort et al. (2018) is the most suitable term to refer to collocation learning in this study, in which we aim to investigate university students' productive form recall knowledge of L2 German collocations by focusing on 35 specific target collocations and by examining a range of item-related and learner-related variables.

Background

Identifying collocations

In L2 research, there is a firm consensus that collocations (e.g., *pay attention*) are one type of formulaic language, alongside other types of multiword expressions such as idioms (e.g., *it's a piece of cake*), conventional situational expressions (e.g., *nice to meet you*), and lexical bundles (e.g., *I don't know if*). In a broad sense, the concept “collocation basically refers to a syntagmatic relationship among words which co-occur” (Wood, 2019, p. 31). However, precise definitions have varied according to researchers' analytical approach. Two approaches, the phraseological and the frequency-based, have been predominant in past research (e.g., Granger & Paquot, 2008). The former approach identifies a collocation as a type of restricted word combination based on the semantic and/or syntactic relationship between two (or more) words (e.g., Howarth, 1998). The frequency-based approach sees collocations as sets of words which have a high statistical probability of appearing together in natural language (e.g., Firth, 1957; Sinclair, 1991). Granger and Paquot (2008) argued for a definition that takes account of frequency, semantics, and syntax all together. In their hybrid view collocations are seen as habitually co-occurring lexical partnerships that have relatively transparent meanings (e.g., *make a mistake*, *strong coffee*) unlike, say, idioms. We adopt this hybrid view. Specifically, we take a collocation to be a word combination which (1) represents a specific syntactic pattern (e.g., adjective + noun, verb + noun), (2) occurs within a given word span in a corpus, and (3) has a relatively transparent meaning.

Item-related and learner-related variables affecting L2 collocation knowledge

Previous research on the processing and use of collocations indicates that L2 collocation development is influenced by item-related and learner-related variables. Item-related variables reported to influence the acquisition and processing of L2 collocations include L1-L2

congruency (e.g., Ding & Reynolds, 2019; Vu & Peters, 2021; Wolter & Gyllstad 2011, 2013), collocation frequency (e.g., Durrant, 2014; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018), and frequency of node words (e.g., Nguyen & Webb, 2017). Important learner-related variables include learners' knowledge of single word L2 vocabulary (e.g., Gyllstad, 2009; Nguyen & Webb, 2017; Vilkaitė, 2017; Vu & Peters, 2021) and L2 immersion (e.g., Edmonds & Gudmestad, 2021; Siyanova & Schmitt, 2008). Although these variables have been shown to have a (positive) influence on L2 collocation learning, they are rarely studied together in a single study. With respect to incidental collocation learning, Vu and Peters (2021) explored the effect of three different modes of reading (reading-only, reading-with-listening, and reading with textual input enhancement), prior vocabulary knowledge, and five item-related variables (congruency, frequency of occurrence, MI score, corpus frequency, and type of collocation). They found that all three modes of reading resulted in learning gains, with a superior effect for reading with textual input enhancement. Learners' prior vocabulary knowledge and congruency were found to be significant predictors of the incidental learning. This study was carried out over a 9-week-period, but to the best of our knowledge, no study to date has investigated the aforementioned variables over a longer time span.

Item-related variables

Congruency

It has been widely observed that even advanced L2 learners produce unconventional and perhaps odd L2 collocations owing to over-reliance on word-for-word translation from their L1 (e.g., Laufer & Waldman, 2011; Nesselhauf, 2003). In most studies on L1 influence, the present study included, collocations are considered “congruent” in L1 and L2 if there is a word-for-word translation of the L1 expression for the concept that the learner has in mind,

and “incongruent” if there is no such translation equivalent (e.g., Nesselhauf, 2005; Wolter & Gyllstad, 2013). An example of a congruent German collocation for Dutch learners of German is *eine Rolle spielen*^G – *een rol spelen*^D (‘play a role’). An incongruent collocation is *in Ruhe lassen*^G – *met rust laten*^D (‘leave alone’), because the word-for-word translation would be **mit Ruhe lassen* (*‘leave with silence’). Note that only one word (the preposition) of *in Ruhe lassen* is incongruent according to the above definition. However, there are collocations that have more than one incongruent part, for example *Wert legen (auf)*^G – *belang hechten (aan)*^D (‘attach importance [to]’). In our study, a collocation is considered incongruent when there is no literal translation equivalent of at least one of the constituent parts.

The effect of L1-L2 congruency on the learning and processing of L2 collocations has been well-documented in the SLA literature. For example, Peters (2016) and Vu and Peters (2021) reported evidence that learners’ ability to recall forms is generally better for congruent collocations than for incongruent. Additionally, it has been found that congruent L2 collocations are processed more quickly and accurately than incongruent ones (e.g., Ding & Reynolds, 2019; Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2018; Yamashita & Jiang, 2010). The fact that this effect is observed even in very advanced learners suggests that there might be a continuing influence of the L1 (Wolter & Gyllstad, 2013). To the best of our knowledge, no study to date has examined the influence of congruency over multiple years to see how, or even whether, the effect of L1-L2 congruency effect changes.

Frequency and association strength

Usage-based theories hold that language learning is experience-driven and that an extremely important fact of this experience is that different vocabulary items occur in input with different frequencies (Ellis, 2002). In SLA research, corpus frequencies are used to estimate real world input frequencies. There is much evidence that learners tend to acquire high-

frequency words before low-frequency words because high-frequency words are encountered more often (e.g., Ellis, 2002; Nation, 2001). Some researchers suggest that frequency matters in this way for collocations as well (e.g., Durrant, 2014). It has been reported that the learnability of a collocation is not only influenced by the frequency of the collocation as a whole, but also by the frequencies of its constituent words (e.g., Nguyen & Webb, 2017; Wolter & Yamashita, 2018). However, Vu and Peters (2021) found that corpus frequency of the target collocations (consisting of high-frequency words) was not a significant predictor of students' learning gains. Given these mixed findings, more research is needed into the effect of corpus frequency and L2 collocation development.

Inextricably linked to corpus frequency is interword association strength, which is often measured by t-scores or Mutual Information (MI) scores. Rankings based on t-scores tend to highlight very frequent word combinations (e.g., *good example*), whereas high MI scores tend to highlight relatively infrequent combinations made up of words that are strongly associated (e.g., *tectonic plates*) (Durrant & Schmitt, 2009). The t-score is computed as an adjusted value of collocation frequency based on the raw frequency minus random co-occurrence frequency divided by the square root of the raw frequency (Gablasova et al., 2017). The MI score compares the probability of observing the two words of the collocation to the probabilities of observing the words independently (Church & Hanks, 1990). Both measures are widely used in learner corpus studies on L2 writing, in which collocations are usually extracted from learners' written productions. The few longitudinal studies on this topic, with participants in an immersion context, present differing results in terms of change over time. Yoon (2016) found no significant changes in MI scores when comparing the essays written at the start and the end of one semester. Li and Schmitt (2010) found a moderate increase in the t-score, whereas the MI scores remained relatively stable, although they also found considerable variation between the individual students. In a large-scale learner corpus

study Siyanova-Chanturia and Spina (2020) did not find that MI scores underwent a statistically significant change over time. However, in an earlier study, Siyanova-Chanturia (2015) found that learners' writings at the end contained not only more higher-frequency combinations but also more collocations with relatively high MI scores than did the writings at the beginning. Edmonds and Gudmestad (2021) also found a change in MI scores in learners' output 8 months after these learners' time abroad but no change in the overall frequencies of collocations. A possible explanation for these mixed findings relates to the immersion experience itself, in which the degree of language acquisition might not only depend on the amount of exposure, but also on the comparatively active engagement with the L2 in social interactions (e.g., González Fernández & Schmitt, 2015).

Studies of L2 collocation knowledge in a non-immersion context seem to point to a lack of sensitivity of L2 learners to the association strength between words. For example, using a prompted productive collocation test, González Fernández and Schmitt (2015) measured 108 Spanish learners' productive form recall knowledge of 50 English collocations which vary widely in respect of corpus frequency, t-score and MI score and observed the following Pearson's correlations with learners' collocation scores: Raw corpus frequency: .45; t-score: .41; MI score: -.16. The authors concluded that "increasing the 'tightness' of the combinational bonding does not seem related to collocation learning" (p. 107). Weak mean correlations between MI scores and learners' collocation knowledge were found in a meta-analysis carried out by Durrant (2014). He argued that L2 learners, unlike L1 speakers, may notice only whole collocation frequency and not association strength between the constituent parts, concluding that "L1 learners notice both collocations and their components, while L2 learners focus only on the whole collocation" (Durrant, 2014, p. 472). In contrast, Wray (2002) suggested that L2 learners tend to focus on individual words. Again, findings have been mixed and further research is needed.

Imageability

Imageability, defined as a lexeme's "capacity to evoke a mental image" (Steinel et al., 2007, p. 449) and concreteness, defined as "the degree to which the concept denoted by a word refers to a perceptible entity" (Brysbaert et al., 2014a, p. 904) are often used interchangeably because of the typically high correlation between both measures (e.g., Brysbaert et al., 2014b). Both imageability and concreteness are known to be potent facilitators of L2 word learning (e.g., De Groot and Keijzer, 2000; Ding et al., 2017; Ellis and Beaton, 1993). According to Steinel and colleagues (2007), imageability may also facilitate the learning of L2 idioms. Because imageability has not been investigated yet as a possible variable in collocation learning, our analysis also took account of this variable.

Learner-related variables

Prior L2 vocabulary knowledge

A learner-related variable thought to be especially important for L2 collocation learning is learners' prior L2 vocabulary size, that is, the number of known words, operationalized as "knowledge of the form–meaning connection" (Schmitt, 2014, p. 915). Vilkaitė (2017), who investigated the effects of adjacency and prior vocabulary knowledge on the incidental acquisition of L2 collocations, found that learners' prior receptive vocabulary knowledge – as measured by the Vocabulary Levels Test (VLT) (Nation, 2001; Schmitt et al., 2001) – had a positive effect on the learning of collocations. Specifically, with an increase of one point in a learner's VLT score, the predicted probability of learning a collocation increased by 10% in the immediate post-test and by 13% in the delayed post-test. Vu and Peters (2021) found that learners with a higher score on the VLT had a better chance of learning the form of the collocation: with an increase of one unit in the VLT score, the odds of a correct response increased by 2.2%. However, in the study of Toomer and Elgort (2019), participants' L2

vocabulary knowledge – also measured by the VLT – seemed not to affect L2 learners' gains of collocations. What should be remarked though, is that the VLT measures form recognition. In general, researchers agree that learners' mastery of form (and meaning) recognition lags behind their mastery of form (and meaning) recall (e.g., Schmitt, 2014) and that productive tasks are often more demanding than receptive ones (e.g., Webb, 2008). Thus, if receptive vocabulary size is an influential factor in collocation learning, then it may be assumed that learners with a larger productive vocabulary size tend to be especially able to acquire productive collocation knowledge. Therefore, it seemed worth investigating whether learners' prior productive vocabulary knowledge – as measured by a German version of the Productive VLT (PVLT) – influences the acquisition of collocations over time.

L2 immersion

Another factor that has been shown to influence the learnability of collocations is learners' exposure to the L2. Usage-based theories predict that extensive exposure is needed for language learning in general (e.g., Ellis, 2002) and for collocation learning in particular (e.g., Durrant & Schmitt, 2010). However, González Fernández and Schmitt (2015) have argued that “it may not be exposure per se that is important, but the kind of high-quality engagement with language that presumably occurs in a socially-integrated environment, where learners wish to use the L2 for meaningful and pleasurable communication” (p. 101). Thus, it would be reasonable to expect that spending time in a L2 environment may facilitate collocation learning. Some studies have indeed observed a positive effect of L2 immersion abroad on learners' collocation knowledge (González Fernández & Schmitt, 2015; Groom, 2009; Macis & Schmitt, 2017; Siyanova & Schmitt, 2008). In other studies though it was found that a stay in the target language country did not lead to an appreciably higher or more accurate use of L2 collocations (Boone, 2021; Li & Schmitt, 2010; Nesselhauf, 2005). Because findings have been mixed, L2 immersion will be added as a variable in our analysis.

The present study

Although the studies reviewed above have undoubtedly contributed to a deeper understanding of L2 collocation development, more longitudinal studies are needed (e.g., Siyanova-Chanturia & Spina, 2020). First, the longitudinal studies outlined above ranged in duration from 9 weeks (Vu & Peters, 2021) to 21 months (Edmonds & Gudmestad, 2021). Examining development over a longer period may contribute to a better understanding of the language learning process, which is often dynamic, complex and long-ongoing (e.g., Larsen-Freeman, 1997). Second, most longitudinal studies of collocation development have been corpus studies focusing on L2 learner output in writing, elicited by means of writing assignments, from which collocations are extracted. Longitudinal studies of L2 productive collocation knowledge (i.e., form recall) of specific target items, rather than collocation use, are scarce, but might provide additional insights on the learnability of specific collocations. Third, with the exception of Vu and Peters (2021), previous relevant longitudinal studies did not take into account a broad range of item- and learner-related variables. Examining item-related variables may help to identify characteristics that make collocations comparatively easy or hard to learn, and whether the effect of these variables changes during the learning process. Additionally, it is crucial to take account of individual learner profiles (e.g., Boers, 2020) to see how these learner-related variables influence collocation development. Fourth, although other L2 languages than English are starting to be explored (e.g., Edmonds & Gudmestad, 2021; Siyanova-Chanturia, 2015), the majority of studies in the field so far have focused on use or acquisition of collocations in L2 English. In sum, our study aimed to add to the existing body of research on collocation development by: (1) adopting a 3-year longitudinal design, (2) testing learners' productive collocation knowledge by focusing on their correct or incorrect (written) production of a given collocation in a form recall test format, (3) taking

into account several item- and learner-related variables, and (4) exploring an underrepresented L2, namely German.

The research questions are the following:

- 1) How does learners' collocation knowledge develop over time?
- 2) How do several item-related variables (congruency, corpus frequency, association strength, and imageability) influence this development?
- 3) How do two learner-related variables (prior productive vocabulary knowledge and L2 immersion) influence this development?
- 4) Does the influence of these item-related and learner-related variables change over time?

Methodology

Participants

The participants in this study were 50 L1 Dutch undergraduate students (nine male, 41 female), majoring in German and an additional foreign language at a Belgian university. Twenty-one of them were studying French, 16 English, nine Spanish, two Italian, one Russian and one Turkish as their extra foreign language. 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 the third year). Their bachelor's program consists of an in-depth study of Dutch and two foreign languages and has an explicit focus on grammar and vocabulary in the first year, whereas there is more language practice (e.g., within translation, speaking and writing courses) in the second and third year. No prior knowledge of German is required for the program, and 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). As a curriculum requirement,

during the third academic year the students were expected to participate in a compulsory 5-month exchange program abroad. Of the 43 students who participated in the third collocation test, twenty-four of them went to a German-speaking country and 19 spent the semester in a non-German-speaking country. All students continued to study German at the host universities.

Data collection started at the beginning of students' university program and ended after 3 years. All 50 students participated in the PVLТ and in at least two data collection points of the collocation test. Participants took part on a voluntary basis and provided informed consent.

Target collocations

We wanted to develop a sample of targets representative of collocations that students might encounter during a learning trajectory aiming for a B2/C1 level. Identification of representative collocations for this level is complicated, because of the variety and sheer number of collocations, and because there is no published, validated list of collocations to work from. However, there is an official German B1 word list (Glaboniat et al., 2013), which comprises about 2400 lexical items that learners should know at this level. A number of collocations can be found in the example sentences next to the lexical items in this list (for example, *packen* – *Ich muss noch meinen Koffer packen*, 'I still have to pack my suitcase'). For this study, we selected only adjective-noun, noun-verb and preposition-noun-verb collocations. The German collocation dictionaries of Quasthoff (2011) and Häcki Buhofer et al. (2014) were then consulted, and candidate collocations appearing in at least one of these dictionaries were selected. The result was a pool of 55 collocations. To make sure there was sufficient variety in terms of frequency and association strength, these collocations were cross-checked with the German Web Corpus 2013 (deTenTen), a corpus of 16.5 billion

words, made up of texts collected from the internet using the concordance tool in SketchEngine (<https://www.sketchengine.eu/>). We also checked that the targets did not appear in the vocabulary lists of students' course textbooks and had not been addressed explicitly in vocabulary class, which was confirmed by the teachers.

Instruments

Productive collocation test

To measure students' productive collocation knowledge of the target collocations, a productive collocation form recall test was developed, which took the form of a gap-fill translation test. Students had to complete the German sentences by adding the appropriate German collocation, as indicated by a L1 (Dutch) translation provided in parenthesis. For example: Zwischen Gesundheit und Armut besteht ein (nauw verband) _____ . ('There is a *close link* between health and poverty.')

We ran a pilot study, in which we administered the collocation test to 77 first-year students of German. The aim was to test the internal consistency of the 55 items and to identify and omit ambiguous candidate collocations. Internal consistency of the items was measured using Cronbach's alpha and was found to be high ($\alpha = .92$). However, the pilot showed that several items were not suitable for the purposes of our study. For example, the Dutch collocation *moeite doen* ('make an effort') can be translated in German with the collocation *sich Mühe geben*, but also with the reflexive verb *sich anstrengen* (which is not a collocation). Another issue was that there were collocations that have multiple correct translations in German. An example is *Nebel* ('fog') which occurs in *dichter Nebel*, *dicker Nebel*, *starker Nebel* – all meaning 'dense fog' (Dutch *dichte mist*). After exclusion of the collocations deemed problematic for this reason, 35 collocations remained (13 adjective-noun, 15 noun-verb, and 7 preposition-noun-verb). Cronbach's alpha showed a good internal

consistency of the 35 items in the task ($\alpha = .87$). The same 35 items were used in the same test format each year, but the items were put in randomized order, and students did not receive feedback on their performance. The 35 target collocations can be found in Appendix A, the collocation test in Appendix B (Supplementary Material).

Since we did not want to attract participants' attention to the targets already at the very beginning of their learning trajectory, we did not administer a pretest (e.g., Toomer & Elgort, 2019). Instead, to estimate baseline knowledge of the target collocations we collected proxy pretest scores from a very similar sample of learners: 32 Dutch-speaking undergraduate students of German at the beginning of their first year of university. The test was the same as the year-end tests in the main study. The proxy pretest scores show that only one congruent collocation, *Ziel erreichen* ('achieve a goal') had a mean score of 0.47 (SD = 0.50). The mean score for the remaining 16 congruent collocations was 0.09, the mean score for the 18 incongruent collocations was 0.01. These scores show that there was negligible productive knowledge of the target collocations in this group.

Item-related and learner-related variables in the study

As part of the study, we collected measures for several potential predictors of learners' L2 collocation knowledge. Appendix C (Supplementary Material) gives the values for all variables.

Congruency: To determine the effect of L1 congruency, the target items were labelled as congruent (1) or incongruent (0), based on the ratings of 11 university lecturers who were asked to decide whether the target item has a literal L1 translation equivalent in the L2 ("+congruent") or not ("–incongruent"). These university lecturers had between 4 and 30 years of experience in teaching German. To estimate the reliability of the ratings for the 35

targets the intra-class correlation coefficient (ICC) was computed. The relevant version of the ICC as a measure of consistency is “2-way average random raters”. We used the *psych* package in R (Revelle, 2020) and found $ICC = 0.95$, 95% confidence interval (CI) [.93, .97], which indicates excellent reliability (Koo & Li, 2016). The target items were assigned the congruency values 1 or 0 depending on the rating of the majority of the lecturers. The result was that 17 collocations were signaled as congruent (1) and 18 signaled as incongruent (0). Of the 18 incongruent collocations, 13 contain one word that does not translate literally from Dutch, four of them two words, and one three (including the preposition). Congruency was included as a dichotomous variable in the analysis.

Frequency and association strength: For all 35 targets, raw corpus frequency values (for the entire collocation and for the noun), t-score and MI score were obtained from the German Web Corpus 2013. We used the SUBTLEX Zipf scale (Van Heuven et al., 2014) to log transform all frequency counts with the formula $\log_{10}(\text{frequency per million words})+3$. The advantage of this scale is that it is logarithmic and that the values are easy to interpret (Van Heuven et al., 2014).

Imageability: To determine imageability, we collected subjective ratings of imageability on a 7-point Likert scale from 17 very advanced L2 speakers of German (all holding a Master’s degree in German language/literature and using German regularly in their jobs or daily lives) and seven L1 German speakers for a list of 66 collocations, including the 35 target collocations and 31 non-target collocations. One purpose of the added 31 items was to allow our new imageability ratings to be validated (i.e., compared with a previously published set of collocation ratings). Seven of the 31 non-target collocations functioned as list-initial “calibrator” items intended to serve as examples of the various levels of the rating scale. All

31 non-target collocations were selected from the database compiled by Citron et al. (2016) which gives concreteness ratings of 619 German phrases. Given the typically strong correlations between ratings of concreteness and imageability and because no collections of imageability ratings for German are yet available, we used the database of Citron et al. (2016) to be able to validate our ratings.

The randomized list of to-be-rated collocations was presented to the raters. To increase the reliability of the ratings, raters were invited to rate the collocations twice, with a pause before the second round. For the second round, the collocations were presented in a new randomized order. Sixteen of the raters rated the collocations twice and eight rated them once. The two sets of ratings from the raters who completed the ratings twice were averaged to yield a single set of mean ratings for that person. Finally, a mean rating across all raters was calculated for each collocation (Appendix D).

To validate our ratings, we calculated correlations between the existing concreteness ratings (Citron et al., 2016) and our new imageability ratings for the 31 non-target collocations, finding $r = .79$, CI [.61, .89], which is very similar to the range reported for correlations between imageability and concreteness in the literature (e.g., Brysbaert et al., 2014b). To estimate the reliability of the ratings for the 35 target collocations we calculated the appropriate version of the ICC = 0.92, CI [.88, .95]. This indicates excellent reliability (Koo & Li, 2016).

Prior productive vocabulary size: To assess students' prior vocabulary, we administered the PVLТ for German developed by the German Institute for Test Research and Test Development in Leipzig and modeled after Nation's PVLТ for English (Nation, 2001). The test contains five subtests, which measure learners' vocabulary knowledge on the vocabulary levels of 1000, 2000, 3000, 4000, and 5000 words respectively. These levels are based on the

frequency lists derived from the Herder/BYU-German corpus (Jones et al., 2006). There are 18 cloze items per subtest (i.e., per frequency level). Each target word is embedded in one or two sentences. To disambiguate the target items, the first letter (or letters) of a targeted word is provided. For example: In dem Dorf steht eine alte Ki_____. ('In the village, there is an old ch_____'.')

L2 immersion: All students participated in a compulsory 5-month exchange program in a country in which one of their languages of study is spoken. In our analysis, L2 immersion or study abroad (SA) was coded as SA_TL if the participant spent the semester abroad in a German-speaking country ($n = 24$), or SA_nonTL if otherwise ($n = 19$). It should be remarked, however, that for Time 1 and Time 2 L2 immersion was not coded, because students only went abroad between Time 2 and Time 3.

Procedure

Participants were tracked for 3 academic years. Both the PVLТ and the collocation test were administered in class and students could not use a dictionary. In total, there were four test sessions. The PVLТ was administered at the beginning of students' first year of university, as a paper-and-pencil test. The time needed for completion was 30 minutes. The first collocation test was administered at the end of students' first year of university, the second at the end of the second year, and the third at the end of the third year. The first two versions of the collocation test were taken as paper-and-pencil tests, but due to the COVID-19 pandemic, the third version had to be administered online. The time needed to complete this test was 25 minutes.

Scoring and analyses

The PVLТ and the collocation tests were corrected manually, and each test item was scored either one point for a (completely) correct answer (e.g., *Ziel erreichen* for ‘achieve a goal’) or zero points for an incorrect or incomplete answer (e.g., ... *erreichen* or *Zweck erzielen*). For the collocation test, a binary score for each collocation was given, which was used in the analyses. For the PVLТ, learners’ mean percentage score across all subtests (measuring learners’ vocabulary knowledge on the vocabulary levels of 1000, 2000, 3000, 4000, and 5000 words respectively) was used in subsequent regression modeling. All analyses were carried out using the R software environment (version 4.1.2, R Core Team, 2021).

To visualize how learners’ collocation knowledge develops over time (RQ1), we calculated descriptive statistics for the proxy baseline test and the three collocation tests. The *ggplot2* package (version 3.3.5; Wickham, 2016) was used to create line plots. To explore the influence of the learner- and item-related variables on students’ collocation score (RQ2 and 3) and the influence of time (RQ4), a generalized linear mixed model was used since the outcome variable (collocation score) is binary. A linear mixed model is an extension of a simple linear model to allow both random and fixed effects that account for individual variation between items and participants. The model was constructed using the *gmler*-function from the package *lme4* (version 1.1.26; Bates et al., 2021).

First, the continuous variables were centered on the mean. Then, a basic model was built with only random effects: items and learners. Next, the fixed effect “time” and the learner-related fixed effect “baseline productive vocabulary” were added. To be able to integrate the other learner-related variable – L2 immersion – a separate model was built, because students went abroad during their third year and consequently, only the results of the final test might have been influenced by this L2 immersion experience. Also the item-related fixed effects congruency, collocation frequency, MI and imageability were added. To avoid a

collinearity problem, noun frequency and t-score were not included (see Table 2 for a correlation matrix). Interactions between time and the other fixed effects were added. Finally, variables and interactions were omitted until the best fit was identified. Models were fit using a maximum likelihood technique (Laplace Approximation) technique. Model fit was assessed using the anova-function in R. Marginal R^2 was calculated, 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 performance package (version 0.7.2, Lüdtke et al., 2021) in R.

Results

The study collected binary (correct vs incorrect) learner responses to 35 German collocations at three times. Since 50 learners were enrolled in the study, the potential number of binary scores for tests 1 to 3 was 5250. However, owing to learner absences the actual total was 4235. Baseline knowledge of the 35 collocations was estimated by testing 32 learners similar to the ones participating in our study. The test-to-test correlations between the by-item scores on tests of productive collocation knowledge, with bootstrapped 95% CIs, are as follows: Proxy test to Test 1: $r = .71$ [.48, .84], Test 1 to Test 2: $r = .77$ [.59, .89], Test 2 to Test 3: $r = .90$ [.76, .96].

To visualize the results for RQ1 (How does learners' collocation knowledge develop over time), descriptive statistics (mean, median, standard deviation and range) of the by-item scores are given in Table 1. As can be seen, there is general progress from time 0 to time 3.

Table 1

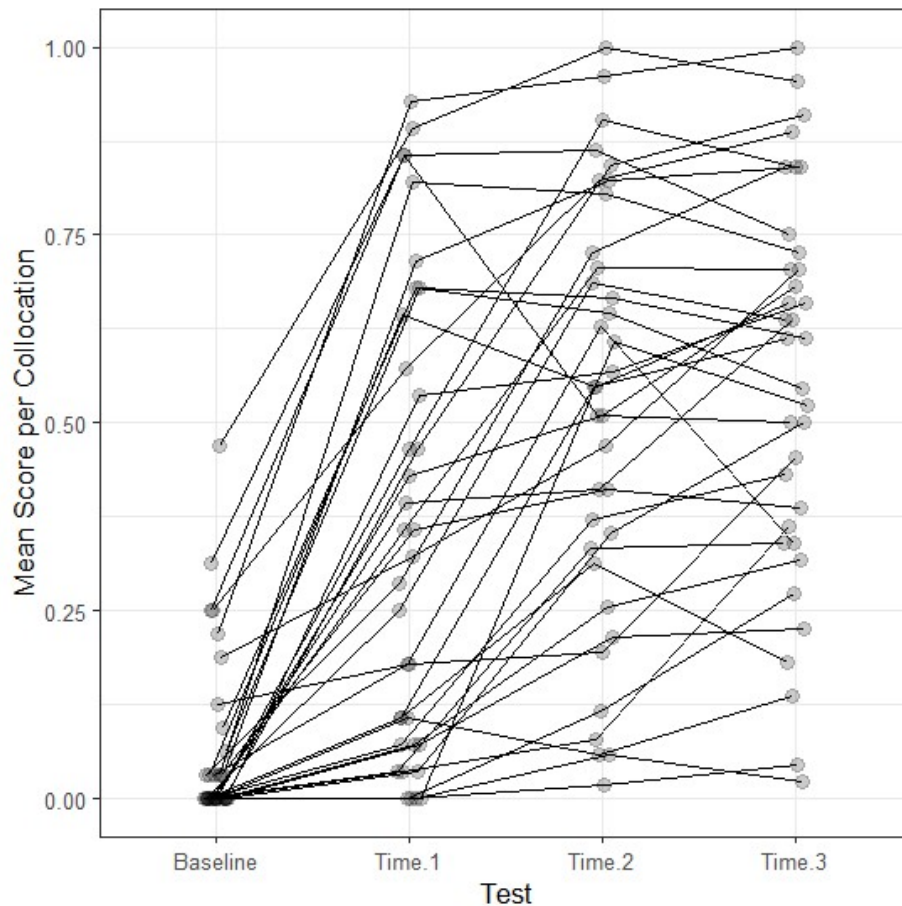
Descriptive statistics of the by-item scores (as proportions of the maximum) on the collocation tests

| | Mean | SD | Median | Range |
|------------|------|-----|--------|------------|
| Proxy test | .06 | .11 | .00 | .00 — .47 |
| Test 1 | .36 | .31 | .32 | .00 — .93 |
| Test 2 | .51 | .28 | .55 | .02 — 1.00 |
| Test 3 | .55 | .26 | .61 | .02 — 1.00 |

Figure 1 represents the learning trend of the 35 target collocations. In each time point represented in the figure there would be 35 dots (one per collocation) if no collocation had the same mean score as any other. Although all the dots have been randomly “jittered” to minimize complete overlaps when multiple collocations have the same mean score, it is still the case that some dots are not visible. A dot that is especially dark corresponds to more than one collocation. Dots in two columns that relate to the same collocation are connected by a line. To sum up, Figure 1 shows that most baseline scores were at or near zero. General progress in collocation learning is indicated by the fact that most of the lines slope upward from left to right.

Figure 1

The trend of collocation learning from baseline to Time 3

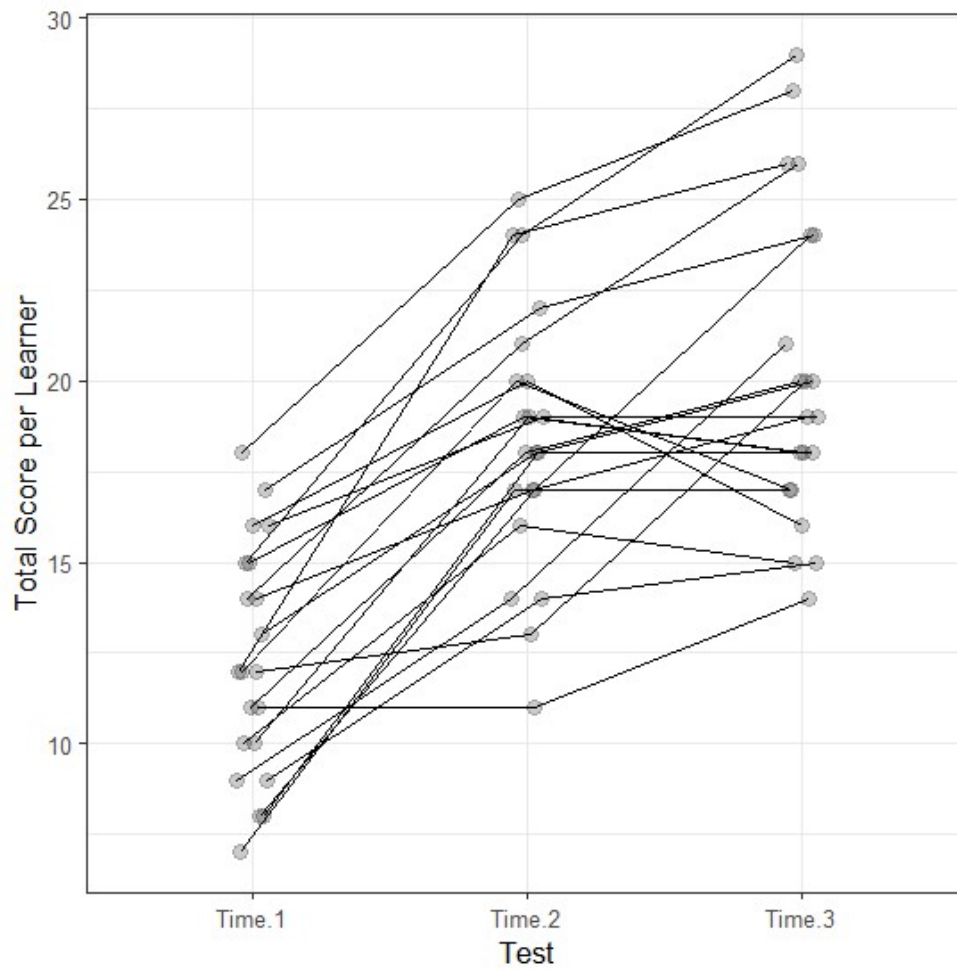


Note. The baseline scores come from 32 similar learners. At times 1 to 3 there were, respectively, 28, 50, and 43 study participants per collocation.

Twenty-one learners took all three year-end collocation tests. For these learners the total per-collocation test scores correlate fairly strongly from test to test: Test 1 to Test 2, $r = .60$; Test 2 to Test 3, $r = .71$. Figure 2 shows the trend of collocation learning for the 21 learners who took all three year-end tests of productive collocation knowledge. Overall progress is indicated by the fact that the great majority of the lines slope upward from test to test. It is plain, however, that there was some forgetting, especially during the final year.

Figure 2

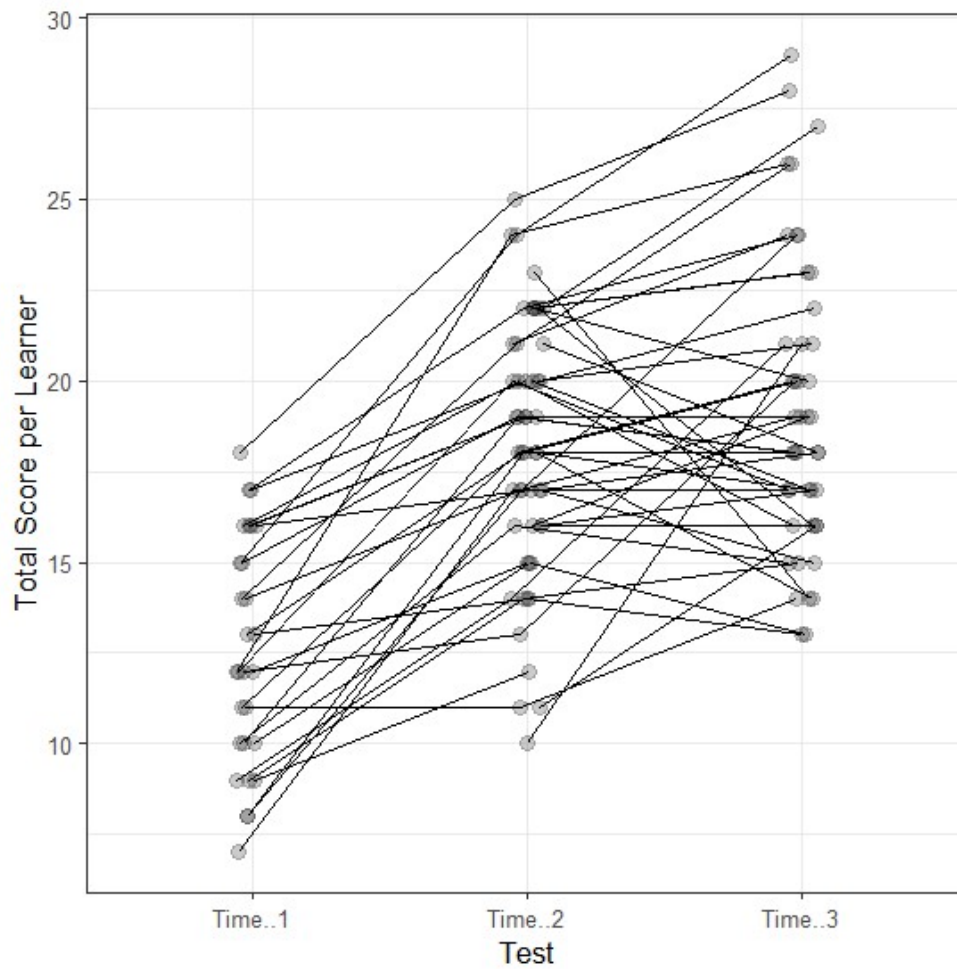
The trend of collocation learning for the 21 learners who took all three year-end tests



Lastly, Figure 3 gives an overview of the collocation learning of all learners who were present for at least two consecutive year-end tests. Again, there was general progress but also some forgetting.

Figure 3

The trend of collocation learning for all learners taking at least two sequential year-end tests.



To explore which variables contributed to learners' collocation development (RQ2 and 3), two mixed effects logistic regression models were built. Table 2 provides descriptive statistics and a correlation matrix for the continuous item-related variables.

Table 2

Descriptive statistics and Spearman's correlations for the continuous item-related variables

| Item-related variable | Mean | SD | Min | Max | r_s | r_s | r_s | r_s | r_s |
|---------------------------------------|------|------|------|------|--------|-------|-------|-------|-------|
| ($N_{\text{collocations}} = 35$) | | | | | 1 | 2 | 3 | 4 | 5 |
| 1. collocation frequency ^a | 2.82 | 0.71 | 1.30 | 3.78 | — | | | | |
| 2. noun frequency ^a | 4.82 | 0.48 | 3.82 | 5.64 | .543** | — | | | |

| | | | | | | | | |
|-----------------|--------|--------|-------|--------|--------|--------|-------|-------|
| 3. t-score | 151.20 | 116.65 | 19.37 | 533.04 | .946** | .557** | — | |
| 4. MI | 7.98 | 2.58 | 2.83 | 12.28 | .007 | -.236 | .033 | — |
| 5. imageability | 4.83 | 1.12 | 3.04 | 6.88 | -.115 | -.078 | -.103 | .414* |

^a Zipf transformed values.

* $p < .05$. ** $p < .01$.

First, models were run without the variable L2 immersion, since this variable could only affect the results at Time 3. The basic generalized linear mixed effects model included only the random effects of “learner” and “item”, and showed that the variable “item” explained most of the variation (variance = 2.21, SD = 1.49, ICC = .39). Far less variation was explained by the variable “learner” (variance = 0.21, SD = 0.46, ICC = .04). Then, two basic models were compared using the anova-function in the lme4 package in R, which gives a chi-square test of the relative fit of two embedded regression models (Brysbaert, 2020). Adding the random intercept for “learner” contributed significantly to improving the model fit $\chi^2(1) = 82.25, p < .001$. The best model to answer our research questions included random intercepts for item and learner and three significant fixed effects (time, learners’ productive vocabulary knowledge and congruency). Table 3 shows the final model, which has a marginal R^2 of .20 and a conditional R^2 of .45. This means that the fixed effects in the model explain 20% of the variance, and that an extra 25% of the variance was explained by the random effects. The odd coefficient for productive vocabulary was 1.03, $CI_{95\%} [1.01, 1.04]$, meaning that a one unit positive difference in participant’s mean percentage score on the PVLТ corresponds to a 3% positive difference in the probability of having productive knowledge of a collocation.

Table 3

Mixed-effects logistic regression model predicting right or wrong answers on the three productive collocation tests.

| Random effects | | Variance | SD | | |
|-----------------------|-------------|----------|-----------|----------|------------|
| Item | (Intercept) | 1.412 | 1.188 | | |
| Participant | (Intercept) | 0.117 | 0.342 | | |
| Fixed effects | | <i>B</i> | <i>SE</i> | <i>Z</i> | <i>P</i> |
| Intercept | | -6.567 | 1.380 | -4.759 | < 0.001*** |
| Time | | 0.817 | 0.077 | 10.612 | < 0.001*** |
| Productive vocabulary | | 0.025 | 0.006 | 4.193 | < 0.001*** |
| Congruency | | 2.507 | 0.494 | 5.075 | < 0.001*** |
| Collocation frequency | | 0.538 | 0.312 | 1.725 | 0.085 |
| Imageability | | 0.318 | 0.217 | 1.461 | 0.144 |
| MI | | 0.001 | 0.091 | 0.006 | 0.995 |
| Time x Congruency | | -0.558 | 0.101 | -5.446 | < 0.001*** |

Note. Baseline for congruency = incongruent.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Then, another model was built with only the data of Time 3, including the variable L2 immersion (1505 observations, 43 learners). This model also showed a significant effect of learners' prior vocabulary knowledge and congruency. Having studied in the target language country did not affect the results on the collocation test significantly. The results of the final model can be found in Table 4. For this model, we found a marginal R^2 of .13 and a conditional R^2 of .42.

Table 4

Mixed-effects logistic regression model predicting right or wrong answers on the last productive collocation test (Time 3).

| Random effects | | Variance | SD | | |
|--------------------------------|-------------|----------|-----------|----------|------------|
| Item | (Intercept) | 1.476 | 1.215 | | |
| Participant | (Intercept) | 0.168 | 0.410 | | |
| Fixed effects | | <i>B</i> | <i>SE</i> | <i>Z</i> | <i>p</i> |
| Intercept | | -4.441 | 1.467 | -3.028 | 0.002** |
| Productive vocabulary | | 0.035 | 0.009 | 3.973 | < 0.001*** |
| Congruency | | 1.037 | 0.464 | 2.232 | 0.025* |
| Collocation frequency | | 0.517 | 0.328 | 1.572 | 0.116 |
| Imageability | | 0.272 | 0.228 | 1.189 | 0.234 |
| MI | | -0.002 | 0.096 | -0.025 | 0.980 |
| L2 immersion (Germany/Austria) | | 0.073 | 0.179 | 0.410 | 0.682 |

Note. Baseline for congruency = incongruent.

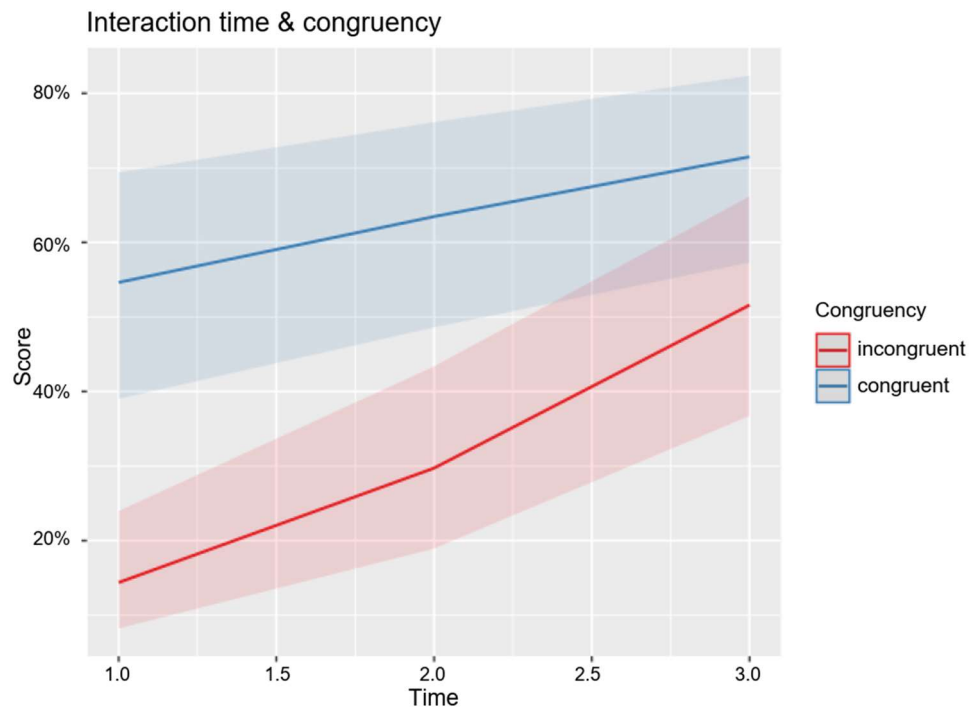
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

To answer RQ4, which was to determine whether the influence of these item-related and learner-related variables changes over time, we explored interactions between time and the significant predictors (productive vocabulary and congruency). A significant interaction between time and congruency was found. In order to make it easier to interpret this interaction effect, Figure 4 was added. It shows a rising trend for the knowledge of congruent collocations, although the curve is not very steep, rather gradual. In contrast, the predicted probability that learners will know an incongruent collocation clearly rises. Specifically, at Time 1 the two contrasted probabilities are far apart. At Times 2 and 3 they are markedly less far apart, showing that the effect of time diminishes for congruent versus incongruent collocations. The learning curve for incongruent collocations is steeper compared to the curve

of the congruent collocations, even though more congruent collocations are still known compared to incongruent collocations at Time 3.

Figure 4

Interaction between time and congruency with the predicted probabilities of score



Discussion

How does learners' collocation knowledge develop over time?

Our results show that there was a general increase in collocation knowledge (i.e., form recall knowledge) after 3 years of studying German. This seems unsurprising, since our participants were motivated language specialists who engaged with German almost daily at university in classes that include language production. However, if we look at the trend of collocation learning (Figures 1, 2 and 3), we see that at Time 3, not one learner was able to produce all 35 collocations correctly. Figures 2 and 3 show that total per learner scores range from 13 to 29. Although our learners were German majors, some of them still had rather limited knowledge

of the target collocations after 3 years. This finding is in line with previous studies which indicate that the acquisition of collocations is slow and challenging even for advanced learners (e.g., Boers et al.; 2014, Nesselhauf, 2003; Laufer & Waldman, 2011). It also appears to confirm the evidence in vocabulary research that “form recall is the most difficult degree of mastery of the form-meaning link” (Schmitt, 2014, p. 929). In addition, it seems that the learning process was rather nonlinear, both with respect to the items (Figure 1) and the individual learners (Figures 2 and 3). These outcomes seem to be in line with the dynamic systems approach to language learning, in which language development is expected to be a nonlinear, chaotic, and highly individual process, with a learning curve “filled with peaks and valleys, progress and backsliding” (Larsen-Freeman, 1997, p. 151). For collocation development, this type of process was already illustrated in the longitudinal study of Li and Schmitt (2010), who reported the variation in collocation development of four learners followed over one year. This is confirmed in our study, in which there is considerable variation both in how well individual collocations were learned and in how well individual learners learned collocations. For the majority of the learners, there is a clear upward trend in the learning curve, but for some learners, some attrition from Time 2 to 3 was observed. This attrition might be explained by the fact that some of these individuals might have had less input to L2 German (e.g., through out-of-class activities such as reading books or articles, watching television, listening to music or using social media in the L2), which also means less opportunities for contextual vocabulary learning. To get more insight into the causes hereof, qualitative interview data could prove useful.

The per learner scores at Time 3 do not only show that collocation learning is slow, but also raise the question of the effectiveness of contextual (incidental) learning. Research has shown that L2 collocations can be acquired both incidentally and intentionally, and that intentional learning results in greater gains (Szudarski, 2017). It is likely that the long-term

retention of the 35 collocations would have been better if they had been used as targets in a study on intentional learning. Explicit collocation instruction is definitely needed, but since only a small number of collocations can be taught in the classroom, it is important to know which variables affect learning in order to make informed choices about which collocations should be selected for classroom learning, and how to deal with individual differences.

How do several item-related variables (i.e., congruency, corpus frequency, association strength, and imageability) influence L2 collocation development?

This study found that congruency had a statistically significant positive effect on learners' productive collocation knowledge. Students' better knowledge of congruent collocations at time 1, 2 and 3 compared to their knowledge of incongruent collocations might be explained by the fact that German and Dutch are highly related Germanic languages. On the other hand, this positive congruency effect has been shown for other language pairs too: for less highly related Germanic language pairs like English-Dutch (Peters, 2016), English-German (Nesselhauf, 2003, 2005), and English-Swedish (Wolter & Gyllstad, 2011, 2013) and also for much less related language pairs like English-Chinese (Ding & Reynolds, 2019), English-Japanese (Yamashita & Jiang, 2010), and English-Vietnamese (Vu & Peters, 2021). Our results are thus in line with previous studies and indicate that 1) students often tend to rely on word-for-word translation when producing L2 collocations (Laufer & Waldman, 2011) and that 2) L1-L2 congruency is an important factor in the processing and use of collocations, which should be taken into account in teaching.

With respect to the other item-related variables (i.e., collocation frequency, MI, imageability), the results were non-significant in the final model. Although some studies did find that collocation frequency related to some degree to L2 collocation knowledge, they also point out that corpus frequency is only one factor of influence (e.g., Durrant, 2014; González Fernández & Schmitt, 2015). Vu and Peters (2021), who included a larger number of factors

in their study, found no significant effect for corpus frequency. It is thus clear that the relationship between corpus frequency and collocation knowledge is not straightforward, and that a study's findings may also depend on the corpus and the target items used (e.g., collocations consisting of infrequent words may yield different results compared to collocations consisting of high-frequency words or delexical verbs). The non-significant effect of MI in this study seems to be in line with other findings on L2 collocation knowledge in a non-immersion context, in which the strength of association between the words of a collocation does not seem related to L2 collocation learning (Durrant, 2014; González Fernández & Schmitt, 2015). Regarding imageability, it has been shown that it facilitates L2 word learning (e.g., De Groot and Keijzer, 2000), and that it may also facilitate the learning of L2 idioms (Steinel et al., 2007). Our study could not confirm a facilitating effect for L2 collocations, although we think that due to the limited number of items, further research into the influence of imageability on L2 collocation learning is needed.

How do the learner-related variables prior vocabulary knowledge and L2 immersion influence L2 collocation development?

Learners' baseline productive vocabulary emerged as a statistically significant predictor of productive collocation knowledge at the form-recall level. This finding seems to support our hypothesis that, if receptive vocabulary knowledge predicts collocation knowledge, both on a receptive level (e.g., Gyllstad, 2009; Nguyen & Webb, 2017; Vilkaitė, 2017) and on a form-recall level (e.g., Peters, 2016; Vu & Peters, 2021), productive vocabulary will do so too. In our study, we found that with an increase of one point in the mean percentage PVL score, the odds of learning a collocation increased by 3%. These results extend the evidence of the "the-rich-get-richer" phenomenon in vocabulary learning, whereby larger vocabulary sizes, receptive or productive, are associated with better learning outcomes (e.g., James et al., 2017). The findings might also point to the fact that in our study, two widely recognized constructs

of vocabulary knowledge, namely vocabulary size (i.e., knowledge of the form–meaning connection) and vocabulary depth (e.g., collocation knowledge) (Schmitt, 2014) appear to be related, since students’ productive vocabulary size predicted their development of productive collocation knowledge.

Although some studies indicate that L2 immersion plays a role in collocation knowledge, we did not find a significant effect of a stay abroad in a German-speaking country on learners’ collocation knowledge at Time 3. This might be explained by the fact that 5 months might be quite short for collocation development, or by the fact that also the students going to a non-German-speaking country continued studying German at the universities abroad. It might also depend on students’ active engagement with the L2 in social interaction abroad (e.g., González Fernández & Schmitt, 2015), which might take place outside the target language country too (e.g., Boone, 2021). Also here, qualitative interview data, for example on students’ L2 exposure and use or their L2 learning experience during SA, could yield relevant information.

Does the influence of these item-related and learner-related variables change over time?

A significant interaction effect between time and congruency was found in this study, indicating that the influence of congruency may change over time. Specifically, we found that students’ knowledge of congruent collocations remained comparatively stable from Time 1 to Time 3, whereas their knowledge of incongruent collocations significantly rose with time (Figure 4). However, it should be noted that for learners who already produced some of the congruent collocations correctly at Time 1, it was mathematically impossible to make as much progress as was the case with respect to the incongruent collocations.

Because of the important impact of the L1 on the processing of L2 collocations even at advanced levels of proficiency, Wolter and Gyllstad (2013) assumed a persisting congruency effect in L2 collocation processing. In our study, we see that the probability of knowing a

congruent collocation compared to an incongruent collocation is higher at Time 3, but our study also suggests that as learners' proficiency level rises, learners may have increasing success in acquiring incongruent collocations. In short, the results of our study are consistent with previous findings of a fairly general positive effect of congruency but also show that the substantive importance of the effect dwindles as learning progresses.

Interestingly, the two collocations with the lowest score at Time 3 are incongruent collocations with more than one incongruent constituent part, which points at a possibility that the level of congruency may play a role in collocation learning. If this finding were borne out by further research, there would be pedagogical implications for the foreign language classroom. Teachers could devote extra attention to incongruent collocations, especially to the 'very incongruent' ones (i.e., with both constituent parts being incongruent), since those are likely to cause problems for learners (e.g., Nesselhauf, 2003). A contrasting L1-L2 approach, making students aware of L1-L2 differences, can be recommended. Although there was one maximum by item score (at Time 3 for *eine Rolle spielen* ('play a role')), we believe that congruent collocations need attention too, as Wolter and Gyllstad (2011) have already pointed out. We suggest that especially in the beginning of a learning trajectory, when the gap between learners' knowledge of congruent and incongruent collocations is large, teachers should give attention to incongruent collocations by setting exercises with known potential to enhance learners' collocation knowledge (Boers & Lindstromberg, 2012; Szudarski, 2017). However, since the learning curve of congruent collocations hardly changes from Time 1 to Time 3, it may be useful in a later phase to give extra attention to congruent collocations that seem likely to be relatively hard to learn, because they are infrequent or contain low-frequency words, for example.

Limitations and suggestions for future research

Our findings have to be seen in light of several limitations. First, the number of both targets and learners was fairly low. Here, it is relevant that L2 learners of German are not as numerous as L2 English learners. However, smaller numbers of learners should not dissuade researchers from investigating other languages, since each language has its own characteristics and deserves its place in the field of applied linguistics. Additionally, compared to other longitudinal studies, the sample size is reasonable. On the other hand, we think it is recommendable to adopt a mixed approach in future studies, where quantitative results of smaller samples are complemented with qualitative results to see how individual learners deal with the challenges of learning L2 collocations and to get more context for the findings (e.g., on the attrition or on the L2 immersion experience). Qualitative insights are important to get the complete picture, because, as Henriksen (2013) puts it: “It is more than likely that collocational acquisition is much more idiosyncratic in nature and dependent on specific language use situations than single-word acquisition” (Henriksen, 2013, p. 48-49). A second methodological issue is the use of multiple tests to measure learners’ development. In this study, a positive testing effect cannot be entirely ruled out, i.e. some learning might have happened during test taking. However, we tried to reduce this effect by leaving a gap of a year between the completion of the tests. Third, the number of variables influencing collocation knowledge is undoubtedly much higher than the number investigated in this study. The predictors investigated here explain about 20% of the variation in test scores, which provides an opportunity for further studies to identify other factors involved. Fourth, it is not impossible that participants were able to translate some congruent collocations correctly even if they had never encountered these collocations before. For cognates (i.e., words with a similar form and meaning in the L1 and the L2), it has been shown that they “can grant learners access to a reservoir of potential target language vocabulary without explicit

instruction” (Vanhove & Berthele, 2015, p. 2). It is likely that the same applies for congruency. Using a literal L1 equivalent works perfectly for congruent collocations, but not in case of incongruent collocations. However, this kind of ‘guessing effect’ is difficult to avoid, and might also be an indication of how students produce language. In relation to this, also the other languages known by the participants may have had an effect on students’ collocation scores. An incongruent collocation targeted in our study, for example, might have been congruent in another foreign language our participants were familiar with. It is not impossible that students’ additional languages served as a bridge to translate the L2 target collocations. We did not investigate this in the present study, but it might be interesting to explore in future research. Finally, since most of our incongruent targets contained only one word without a literal translation equivalent, future studies should look at learners’ acquisition of incongruent collocations of different incongruency levels.

Conclusion

The goal of this study was to investigate L2 learners’ productive collocation development in German and to examine the effect of several item- and learner-related variables. The results indicate that there was general progress, despite some forgetting. The variation in both per learner and per collocation scores shows that collocation learning is influenced by multiple variables.

As to item-related variables, our results corroborate previous findings that L1-L2 congruency is an important predictor of collocation knowledge. What is more, the congruency effect was found to persist throughout learners’ 3-year trajectory. To maximize collocation learning, we recommend that teachers and materials creators direct learners’ attention toward both congruent and incongruent collocations, with special attention to incongruent collocations at the beginning of the learning trajectory. As to learner-related variables, we

found that learners with a comparatively large productive vocabulary at the beginning of the learning trajectory were more likely to produce correct L2 German collocations, which shows the importance of increasing one's vocabulary as much as possible even in the early stages of learning an additional language.

Supplementary material

Supplementary material is available at

https://osf.io/yp2j4/?view_only=b0a7c06c30904072a6240f86a4ff1ff2

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2021). R package *lme4*, version 1.1-26: *Linear mixed-effects models using 'Eigen' and S4*. <https://github.com/lme4/lme4/>
- Boers, F. (2020). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies*. Routledge.
- Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83–110.
<https://doi.org/10.1017/S0267190512000050>
- Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). Some explanations for the slow acquisition of L2 collocations. *Vial-Vigo International Journal of Applied Linguistics*, 11, 41–62.
- Boone, G. (2021). How social interaction affects students' formulaic development in L2 German in a multilingual SA context : four case studies. In R. Mitchell & H. Tyne (Eds.), *Language, mobility and study abroad in the contemporary European context* (pp. 159–170). Routledge.
<https://doi.org/10.4324/9781003087953-11>
- Brysbaert, M. (2020). *Basic statistics for psychologists*. Macmillan International.

- Brysbaert, M., Warriner, A., & Kuperman, V. (2014a). Concreteness ratings for 40,000 generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911.
<https://doi.org/10.3758/s13428-013-0403-5>
- Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014b). 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>
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, *16*(1), 22–29.
- Citron, F., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., & Jacobs, A. (2016). When emotions are expressed figuratively: Psycholinguistic and affective norms of 619 idioms for German (PANIG). *Behavior Research Methods*, *48*, 91–111. <https://doi.org/10.3758/s13428-015-0581-4>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- De Groot, A., & 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*, 1–56. <https://doi.org/10.1111/0023-8333.00110>
- Ding, C., & Reynolds, B. (2019). The effects of L1 congruency, L2 proficiency, and the collocatenode relationship on the processing of L2 English collocations by L1-Chinese EFL learners. *Review of Cognitive Linguistics*, *17*, 331–357. <https://doi.org/10.1075/rcl.00038.din>
- 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>

- Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, 19, 443–477.
<https://doi.org/10.1075/ijcl.19.4.01dur>
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL - International Review of Applied Linguistics in Language Teaching*, 47.
<https://doi.org/10.1515/iral.2009.007>
- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26, 163–188. <https://doi.org/10.1177/0267658309349431>
- Edmonds, A., & Gudmestad, A. (2021). Collocational development during a stay abroad. *Languages*, 6, 12. <https://doi.org/10.3390/languages6010012>
- 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. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, N., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, 43, 559–617. <https://doi.org/10.1111/j.1467-1770.1993.tb00627.x>
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text - Interdisciplinary Journal for the Study of Discourse*, 20.
<https://doi.org/10.1515/text.1.2000.20.1.29>
- Firth, J. (Ed.). (1957). *Papers in linguistics 1934–1951*. Oxford University Press.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence: Collocations in

Corpus-Based Language Learning Research. *Language Learning*, 67(S1), 155–179.

<https://doi.org/10.1111/lang.12225>

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *ITL - International Journal of Applied Linguistics*, 166, 94–126. <https://doi.org/10.1075/itl.166.1.03fer>

Glaboniat, M., Perlmann-Balme, M., & Studer, T. (2013). *Zertifikat B1: Deutschprüfung für Jugendliche und Erwachsene: Prüfungsziele, Testbeschreibung*. Hueber Verlag.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology* (pp. 27–49). John Benjamins.

<https://doi.org/10.1075/z.139.07gra>

Groom, N. (2009). Effects of second language immersion on second language collocational development. In A. Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language* (pp. 21–33). Palgrave Macmillan. https://doi.org/10.1057/9780230245327_2

Gyllstad, H. (2009). Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 153–170). Palgrave Macmillan.

Häcki Buhofer, A., Dräger, M., Meier, S., & Roth, T. (2014). *Feste Wortverbindungen des Deutschen: Kollokationenwörterbuch für den Alltag*. Francke.

Henriksen, B. (2013). Research on L2 learners' collocational competence and development: A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge, and use* (pp. 29-56). EuroSLA Monograph Series 2. EuroSLA.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19, 24–44. <https://doi.org/10.1093/applin/19.1.24>

Hulstijn, J. (2003). Incidental and intentional learning. In Doughty, C., & Long, M. (Eds.), *The Handbook of Second Language Acquisition* (pp. 349–381). Blackwell Publishing Ltd.

- James, E., Gaskell, G., Weighall, A. & Henderson, L. (2017). Consolidation of vocabulary during sleep: The rich get richer? *Neuroscience and Biobehavioral Reviews*, 77, 1–13.
[10.1016/j.neubiorev.2017.01.054](https://doi.org/10.1016/j.neubiorev.2017.01.054)
- Jones, R., Tschirner, E. P., Goldhahn, A., Buchwald, I., & Ittner, A. (2006). *A frequency dictionary of German: Core vocabulary for learners*. Routledge.
- Koo, T., & Li, M. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163.
<https://doi.org/10.1016/j.jcm.2016.02.012>
- Larsen-Freeman, D. (1997). Chaos/Complexity science and second language acquisition. *Applied Linguistics*, 18, 141–165. <https://doi.org/10.1093/applin/18.2.141>
- Laufer, B., & Waldman, T. (2011). Verb-Noun collocations in second language writing: A corpus analysis of learners' English verb-noun collocations in L2 writing. *Language Learning*, 61, 647–672. <https://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Li, J., & Schmitt, N. (2010). The development of collocations use in academic texts by advanced L2 learners: a multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 23-46). Continuum.
- Lüdecke, D. (2021). Performance: Assessment of regression models performance, Version 0.7.2. Computer freeware. <https://cran.r-project.org/web/packages/performance/performance.pdf>
- Macis, M., & Schmitt, N. (2017). Not just 'small potatoes': Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, 21, 321–340.
<https://doi.org/10.1177/1362168816645957>
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24, 223–242.
<https://doi.org/10.1093/applin/24.2.223>

- Nesselhauf, N. (2005). *Collocations in a Learner Corpus* (Vol. 14). John Benjamins.
<https://doi.org/10.1075/scl.14>
- Nguyen, T., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21, 298–320.
<https://doi.org/10.1177/1362168816639619>
- Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, 20, 113–138.
<https://doi.org/10.1177/1362168814568131>
- Quasthoff, U. (2011). *Wörterbuch der Kollokationen im Deutschen*. De Gruyter.
- R Core Team (2021). R: A language and environment for statistical computing (Version 4.1.2). R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revelle, W. (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.0.12, <https://CRAN.R-project.org/package=psych>.
- Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows: Size and Depth of Vocabulary Knowledge. *Language Learning*, 64(4), 913–951.
<https://doi.org/10.1111/lang.12077>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55–88.
<https://doi.org/10.1177/026553220101800103>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System*, 53, 148–160. <https://doi.org/10.1016/j.system.2015.07.003>

- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64, 429–458.
<https://doi.org/10.3138/cmlr.64.3.429>
- Siyanova-Chanturia, A., & Spina, S. (2020). Multi-Word Expressions in Second Language Writing: A Large-Scale Longitudinal Learner Corpus Study. *Language Learning*, 70(2), 420–463. <https://doi.org/10.1111/lang.12383>
- Steinel, M., Hulstijn, J., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition*, 29, 449–484.
<https://doi.org/10.1017/S0272263107070271>
- Szudarski, P. (2017). Learning and teaching L2 collocations: Insights from research. *TESL Canada Journal*, 34, 205–216. <https://doi.org/10.18806/tesl.v34i3.1280>
- Toomer, M., & Elgort, I. (2019). The Development of Implicit and Explicit Knowledge of Collocations: A Conceptual Replication and Extension of Sonbul and Schmitt (2013). *Language Learning*, 69(2), 405–439. <https://doi.org/10.1111/lang.12335>
- Van Heuven, W., 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.
- 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>
- Vilkaitė, L. (2017). Incidental acquisition of collocations in L2: Effects of adjacency and prior vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, 168, 248–277.
<https://doi.org/10.1075/itl.17005.vil>

- Vu, D.V., & Peters, E. (2021). Incidental learning of collocations from meaningful input: a longitudinal study into three reading modes and factors that affect learning. *Studies in Second Language Acquisition*, 1-23. <https://doi.org/10.1017/S0272263121000462>
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(01). <https://doi.org/10.1017/S0272263108080042>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.
- Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, 32, 430–449. <https://doi.org/10.1093/applin/amr011>
- Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, 35, 451–482. <https://doi.org/10.1017/S0272263113000107>
- Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, 40, 395–416. <https://doi.org/10.1017/S0272263117000237>
- Wood, D. (2019). Classifying and Identifying Formulaic Language. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 30–45). Routledge. <https://doi.org/10.4324/9780429291586-3>
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, 44, 647–668. <https://doi.org/10.5054/tq.2010.235998>
- Yoon, H.-J. (2016). Association strength of verb-noun combinations in experienced NS and less experienced NNS writing: Longitudinal and cross-sectional findings. *Journal of Second Language Writing*, 34, 42–57. <https://doi.org/10.1016/j.jslw.2016.11.001>