ACCEPTED VERSION

See https://doi.org/10.1177/13621688211020412 for published version.

Finding the sweet spot: Learners' productive knowledge of mid-frequency lexical items

Ben Naismith and Alan Juffs

University of Pittsburgh

Abstract

Research into vocabulary knowledge often differentiates between breadth (how many words a person knows) and depth (how well the words are known). Both theoretical categories are essential for understanding language learners' lexical development but how the different aspects of vocabulary knowledge interconnect has not received the same attention as each individual dimension (Haomin & Bilü, 2017), especially in terms of productive knowledge (Mantyla & Huhta, 2014).

This study analyzes lexis from mid-frequency lemmas in the K3-K9 frequency bands from the learner corpus PELIC (Juffs et al., 2020). Critically for learners, mastery of lexis in this frequency range is essential for achieving the English proficiency required for university study. From these mid-frequency items, a dataset of 7,554 tokens were collected from word families with multiple derivations and manually annotated. The findings showed high rates of collocational and derivational accuracy for the forms learners opted to use. However, compared to expert speaker texts in the Corpus of Contemporary American English (COCA; Davies, 2008-), learners overused the verb forms and underused the noun forms of these lexical items. These patterns provide evidence of the interplay between breadth and depth in learners' productive vocabulary usage, suggesting that increased lexical depth will naturally lead to greater lexical breadth and vice versa. Pedagogical implications reaffirm the importance of developing learners' explicit morphological awareness (Ishikawa, 2019) and collocational accuracy (Crossley et al., 2015). Suggestions for mid-frequency lexical items to prioritize in language learning are also provided, with a view to helping learners achieve academic readiness.

Key words: collocation, derivation, learner corpus, lexical breadth, lexical depth, lexical knowledge, vocabulary knowledge

Finding the sweet spot: Learners' productive knowledge of mid-frequency lexical items

I Introduction

Research into lexical knowledge[1] often differentiates between breadth of knowledge and depth of knowledge, a distinction first attributed to Anderson and Freebody (1981). Breadth refers to the number of words a person knows, whereas depth describes how well the words are known. Of the two, lexical depth is more challenging to define and measure due to its multi-dimensional nature (Haomin & Bilü, 2017; Schmitt, 2010) and because breadth and depth grow in relation to one another (Milton, 2009).

To address this issue, various dimensions of depth have been isolated and analyzed, including collocational knowledge and derivational knowledge, with each of these dimensions demonstrating unique patterns of development (Chen & Truscott, 2010). Approaching these constructs from another angle, researchers have investigated the relationship between derivational awareness (depth) and overall vocabulary size (breadth) (Haomin & Bilü, 2017). Furthermore, lexical knowledge is commonly divided into *receptive* and *productive* knowledge, that is, words whose meaning can be retrieved when reading/listening, compared to words that express a desired meaning when writing/speaking (Lenko-Szymanska, 2019). In contrast, productive depth of knowledge features have been studied much less than breadth of knowledge features (Crossley et al., 2015).

To illustrate the interconnections between these various types of lexical knowledge, consider a learner's productive knowledge of the different lemmas *accept* (v)*, acceptance* (n)*, acceptable* (adj)*,* and *acceptably* (adv), all of which belong to the same word family. If the learner has form-meaning knowledge of these forms and can use them, they have the breadth of productive

knowledge of four unique lemmas. And, if the learner knows all four forms and recognizes their base+suffix patterns, then they have the derivational knowledge of the words as well. Logically, form-meaning knowledge usually precedes other aspects of vocabulary knowledge including derivational knowledge. Extending this example, if a learner can also use one collocation with each of these forms, for example, *gladly accept, growing acceptance, prove acceptable,* and *perform acceptably*, then they are demonstrating further depth of productive knowledge (the collocational dimension) and breadth of knowledge (eight unique lemmas).

One method to better understand these different aspects of learners' productive lexical knowledge is to compare learner production to expert speaker production.[2] This approach is based on the premise that second language (L2) production "directly taps into learners' lexical knowledge and is closely aligned with language proficiency and levels of vocabulary acquisition" (Crossley & Skalicky, 2019, p. 386). By comparing expert speaker and learner production, it is possible to see where learners are overusing or underusing language forms (breadth of productive knowledge), as well as with which language forms they are committing errors (depth of productive knowledge).[3] As we shall see in the current and previous studies, learners' production exhibits certain patterns which differ from those of expert speakers, for example an overreliance on one or two preferred forms (Schmitt & Zimmerman, 2002), indicating gaps in their knowledge of word families (Brown, 2013).

Ultimately, learners will be better served if we develop pedagogy and curricula that simultaneously target and develop multiple key aspects of productive lexical knowledge, which would improve accuracy and extend range. Clearly, this suggestion is not a novel one, as is the case with most 'new' ideas for teaching vocabulary proposed by researchers (see, e.g., Hatch & Brown, 1995). Rather, a more realistic and useful aim is to 'tweak' existing teaching practices so

as to maximize their efficacy. Take, for example, this extract from a learner's written text from our data:

(1) `Because the number of infections agents are increasing,`

In (1), teachers will easily recognize the error, *infections*, instead of *infectious*, unintentionally drawing the reader's attention away from the learner's message in an otherwise acceptable clause. However, how to address this issue is less clear-cut. Is it enough to point out that an adjective is needed instead of a noun? Should other forms in this word family be taught at this time, such as *infectious* (adj) or *infect* (v)? Should longer chunks be taught, and if so which ones, for example, *infectious agents* or *number of* ___ *agents*? Is there value in raising awareness of the word parts *infect* (stem) + *-ious* (affix to form adj from verb)? Or perhaps as the communicative intent is clear, the 'error' should not be highlighted at all.

The management of such cases in a principled manner requires a thorough understanding of which lexical forms, and which aspects of these forms, should be the focus of instruction. Equipped with this knowledge, teachers, materials writers, and other stakeholders can make the best possible use of the limited class time available to learners. To facilitate clarifying instructional practice, this study investigated a set of specific mid-frequency lexical items occurring in both a learner corpus and an expert-speaker corpus to answer three research questions:

1) Within mid-frequency word families, which forms do learners show a preference/dispreference for producing and how do these preferences compare to those of expert speakers?
2) Within mid-frequency word families, how often do learners at different proficiency levels make derivational and collocational errors?
3) How do the productive lexical depth dimensions of derivational accuracy and collocational accuracy interact in learners' writing?

By answering question 1, we aim to uncover which word forms would be the best candidates for instructional focus with the goal of improving productive lexical breadth. By answering questions 2 and 3, we hope to uncover the relative importance of different dimensions of productive lexical depth and to better understand their relationship to one another.

II Lexical knowledge

As noted previously, the concept of lexical knowledge is notoriously hard to pin down (Tan, Pandian, & Jaganathan, 2016). Nevertheless, Read (2004) provides an exhaustive definition, categorizing vocabulary knowledge into 'precision of meaning', 'comprehensive word knowledge', and 'network knowledge'. It is this second category which we consider here, "Knowledge of a word which includes not only its semantic features but also its orthographic, phonological, morphological, syntactic, collocational and pragmatic characteristics" (Read, 2004, p. 211). Under this umbrella category, numerous studies have confirmed the essential role of lexical knowledge for other skills in English including reading (e.g., Moghadam, Zaidah, & Ghaderpour, 2012), listening (e.g., Bonk, 2012), speaking (e.g., Milton, Wade, & Hopkins, 2010), and writing (e.g., Dabbagh & Enayat, 2019).

1 Lexical knowledge frameworks

To conceptualize the many dimensions of 'comprehensive word knowledge', various frameworks have been proposed in the L2 pedagogical literature in ESL (e.g., Qian 1999; Nation 2001; Richards, 1976). Most influential among these is Nation's (2001) comprehensive model that considers productive and receptive elements of form, meaning, and use. However, it does not provide any hierarchical information or links among the dimensions, thereby reducing its

pedagogical value (Schmitt, 2019). For the current study, the framework from Dóczi and Kormos' longitudinal study (Figure 1; 2016) was selected because it captures the notion that acquisition of one dimension is connected to acquisition of the others (Schmitt, 2000), such that categories higher in the pyramid rely on knowledge components lower down. The model also provides a clear suggested order of development (each row in the pyramid).

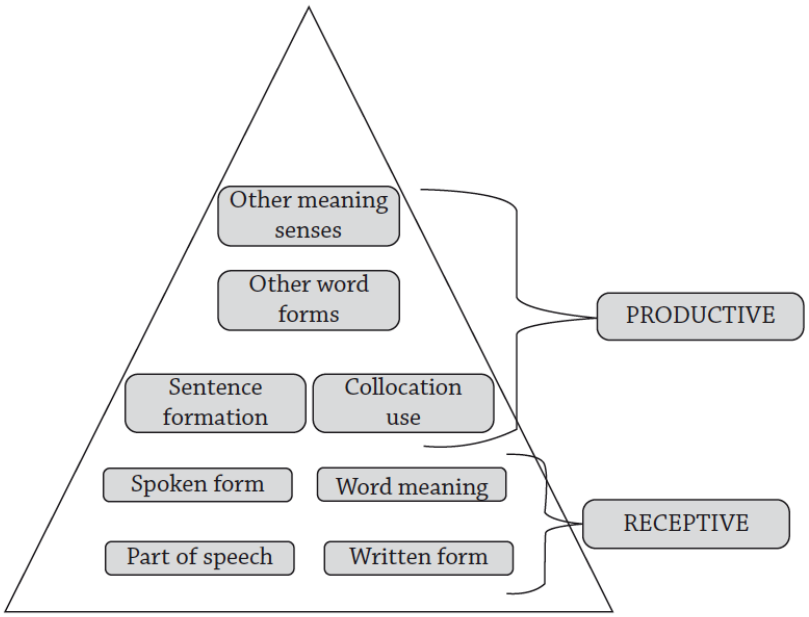


Figure 1. The order of development for word knowledge types (Dóczi & Kormos, 2016, p. 81)

2 Productive/Receptive lexical knowledge

Both the Nation and Dóczi and Kormos frameworks make a distinction between productive and receptive lexical knowledge. This division has ecological and statistical validity as it is a common classroom phenomenon for learners to demonstrate understanding of words which they cannot produce (Schmitt, 2014), and it is more challenging for learners to demonstrate productive than receptive knowledge (Laufer et al., 2004). To be clear, however, the Dóczi and Kormos (2016) framework is not suggesting that knowledge of polysemy is restricted to only productive knowledge. The pyramid depicts those dimensions of lexical knowledge measured in their study

rather than all possible dimensions, with productive dimensions showing later or more gradual patterns of development.

Overall in terms of breadth, learners know approximately twice as many items receptively as they know productively (Milton, 2009), though proficiency and word frequency impact this ratio (Waring, 1997). Typically, expert speakers and L2 speakers alike possess only receptive mastery of many low-frequency words (Schmitt & Meara, 1997). The reason appears to be that receptive knowledge takes less time to acquire and is more easily retained (Waring, 1997). In contrast, productive knowledge is more complicated (Nation, 2016), requiring a series of mental processes prior to the moment of actual production (Levelt, 1989, 2001). Productive knowledge is undeniably critical for both speaking and writing, yet it is often under-researched, with many lexical development studies occurring under controlled settings rather than reflecting 'natural' classroom use where students have access to mediational tools such as (on-line) dictionaries (Juffs, 2019). As a result, a gap remains in the field with respect to understanding productive lexical knowledge and how well lexis can be used (Schmitt, 2010).

III Lexical depth

There are two common approaches for measuring lexical depth, a developmental approach and a dimensional approach (Schmitt, 2010). In the developmental approach, depth is measured along a self-reported scale (e.g., Paribakht & Wesche, 1993; Tan et al., 2016). Such scales can be useful in that they provide a continuum rather than a binary outcome (Schmitt, 2014) and because they are practical to administer. Nevertheless, self-reporting brings its own inherent set of issues as learners may not be aware of what they know/don't know. As a result, depth scales can be "rather speculative" (Schmitt, 2010, p. 217).

In contrast, specific elements of depth are measured independently in the dimensions (or components) approach "to capture all the different aspects of knowing a word" (Pellicer-Sánchez, 2019, p. 512) (see Figure 1). To measure and compare dimensions, researchers have developed batteries of tests that target various aspects of depth of knowledge (e.g., Chen & Truscott, 2010; Ishii & Schmitt, 2009). Chen and Truscott focused on incidental vocabulary acquisition of L1 Mandarin university students. With regards to lexical depth, seven dimensions (three productive and four receptive) were tested in a 7-page vocabulary test. Their findings showed that the various dimensions of lexical knowledge had different patterns of development, with spelling benefitting most from the first few exposures, knowledge of meaning requiring between three to seven exposures, and knowledge of parts of speech developing smoothly across exposures. The Chen and Truscott study is representative of many studies adopting the dimensions approach in that a narrow number of lexical items are tested in depth, rather than a broader yet shallower view of a range of lexical items (e.g., Fitzpatrick, 2012; Haomin & Bilü, 2017; Read, 1998).

1 Derivations

Knowledge of derivations encompasses two intertwined elements: knowledge of different word forms within a word family and knowledge of affixation. Although it is beyond the current paper to tease apart how/when morphemes are stored separately from words (see Juffs [2020] for a review), both construction types are fundamental (Booij, 2017). With over sixty affixes, derivation is the most productive word formation process in English, used in 12.8% of words (Nation, 2001).[4] As such, it is an important aspect of lexical knowledge, and one which greatly affects depth and breadth (Mochizuki & Aizawa, 2000); it is estimated that as many as 5000 word families must be known in order to master receptive knowledge of most affixes (Milton, 2009). A word family

consists of "a base word, its inflected forms and a small number of reasonably regular derived forms" (Nation & Waring, 1997, p. 8; see Bauer & Nation [1993] for a thorough discussion). Therefore, not all derived forms of a base necessarily belong to the same family, for example, *hard* and *hardly.* Consequently, how to count word families, and by extension how to measure vocabulary size, is a challenge.

Unfortunately for learners, acquisition of derivational knowledge is a daunting task. For example, in Schmitt and Meara (1997), students' suffix knowledge over one year of study showed only weak improvement (5% productive, 4% receptive). Similarly, in Schmitt and Zimmerman (2002) learners were found to typically have productive knowledge of only two derivations of major word classes of the target lexis. More recently Dóczi and Kormos (2016) did report noticeable gains in *written* derivational knowledge development for a group of L1 Hungarian teens. Participants were assessed by means of three interviews across a 16-month period, focusing on 21 high-frequency words that had multiple potential derivations. Knowledge of nine different dimensions were assessed, including part-of-speech knowledge. Significant improvement was observed in seven of the nine dimensions, including written part-of-speech knowledge, but there were no significant gains in spoken part-of-speech knowledge.

In one of the few studies focusing explicitly on the relationship between dimensions of lexical depth and breadth, Haomin and Bilü (2017) explored morphological awareness and overall vocabulary size. In this study, 198 Chinese learners of English took four tests, two relating to morphological awareness and two to vocabulary size. Results of path analysis found that derivational knowledge was a significant contributor to lexical breadth and depth, in other words, when learners have greater awareness of the derivational properties of words, this knowledge can be used to establish connections with other lexical items.

2 Collocations

The definition of 'collocation' in vocabulary research differs subtly across studies (see Granger [2018] and Wray [2018] for a full discussion). For example, the focus may be on words which are immediately adjacent as in 'illegally parked' or 'speak English' (Bestgen & Granger, 2014) or include words which are separated, often by functional words, as in '**bread** and **butter**' or '**drink** a **beverage**' (Church & Hanks, 1990). In general, however, there are two key components: (1) the frequency with which the words occur together, and (2) the semantic link between the words. As a starting point, we adopt here Laufer and Waldman's (2011, p. 648) definition which incorporates these two elements:

> "[Collocations are] habitually occurring lexical combinations that are characterized by restricted co-occurrence of elements and relative transparency of meaning."

In this conceptualization of collocation, frequency (co-occurrence) is essential, and can be measured statistically. Raw frequency counts can be used, i.e., how often two words occur near each other within a selected span (or 'window'), but such metrics do not address the element of meaning seen in the Laufer and Waldman definition. For instance, sequences like 'I do not' and 'there is a' are highly frequent, but there are clearly no strong lexical bonds between the words.

In contrast, other measures have gained popularity in the collocation research community because they are believed to correlate with human intuitions. Among these, the two most popular measures are Mutual Information (MI) and t-score, both of which are association measures to describe collocational strength based on co-occurrence between words (but see Kang [2018] for a critique). For both metrics, a higher score indicates that the words are more likely to co-occur

compared to chance. Where the two measures differ is in the types of collocations they identify. MI prioritizes less common words that are typically found together, for example *furrowed brows*. A t-score gives greater weight to frequency, for example collocations composed of high-frequency words which may also be found in a wider range of contexts, such as *good work* (Durrant, 2018; Bestgen & Granger, 2014; Granger & Bestgen, 2014). A typical convention based on previous research is to consider words with an MI score over 3 or a t-score over 2 to be a collocation (Church & Hanks, 1990; Hunston, 2002; Stubbs, 1995).

In the current study, we consider collocations in windows of five words: a node word, plus two words on either side. This cut-off is slightly below the most common span size of 3-5 words each side (Evert, 2009; Sinclair, 1991) so the study could focus on more immediate collocations, which are more likely to be salient and used by learners. Importantly, in automated identification of collocations, the association is most commonly between two words only (but see Saito [2020] for an exception); three-word combinations would require a different minimum MI or t-score to be considered an acceptable collocation and thus cannot easily be compared. This study therefore adopts a contingency-based approach: human intuitions of collocation may include more than two words, whereas the automated measures are restricted to considering each two-word combination in a span. Considering (1) again, in the 5-word window, MI and t-scores would be calculated for *number-infections*, *of infections, infections agents,* and *infections are*. Human judgements of collocations would also consider 3-word combinations such as *number of infections* and *infections agents are*:

(1) Because the number of **infections** agents are increasing,

Like derivational knowledge, collocational knowledge grows gradually (Fitzpatrick, 2012), is acquired late, and is often not mastered by L2 language learners (Henriksen & Stenius Stæhr,

2009). There are many factors for these tendencies. For instance, collocates may not appear frequently enough in input (Gyllstad & Wolter, 2015), and even when they do appear, learners might struggle to recognize them as such (Lee, 2019). In learners' output, these issues are realized in different ways relating to over/under use and errors. Considering overuse, Liu and Shaw (2006) showed that learners overuse collocations with high-frequency verbs, a finding similarly reported by Laufer and Waldman (2011). Conversely, Granger (1998) and Lorenz (1999) noticed underuse of adverb-adjective combinations and restricted collocations. In general, it is apparent that learners use fewer 'native-like' collocations (Granger, 2018), limiting the lexical breadth of their production.

Compared to over/underuse, Laufer and Waldman (2011) deemed collocational errors to be a more serious problem, leading to depressed proficiency scores. Crossley et al.'s (2015) study seems to support Laufer and Waldman's claim. In their analysis of written and spoken samples, human ratings were compared to quantitative measures. They found that collocational accuracy was one of three dimensions most predictive of human judgements of overall lexical proficiency, accounting for much of the variance in both written (84%) and spoken (89%) samples. Identifying collocational errors is a challenge, however, and reported collocational error rates vary greatly across studies. For example, Nesselhauf (2005)'s analysis of collocations in the ICLE learner corpus showed that in 2000 collated verb-noun collocations, there was a 50% error rate as judged by human annotators. In contrast, Laufer and Waldman (2011) estimated that in the ILCoWE learner corpus there was an approximately 33% error rate.

IV Current study

In this study, we adopt the dimensions approach to investigating lexical depth, focusing on the two form-related dimensions reviewed earlier: derivational knowledge and collocational knowledge. Consider, for example, the following four samples from our data in which the collocation span is underlined. Extract (2) demonstrates accurate productive derivational and collocational knowledge of the key word *vary* (v), whereas (3) exemplifies the opposite in which both the derivation (*variable* [adj]) and the collocation (*support variable meals*) are inaccurate. Extracts (4) and (5), on the other hand, reveal how one of these dimensions may be accurate while the other is inaccurate. In (4), the verb form *vary* is acceptable, but the collocation should be *vary in accordance (with)*. Conversely, in rare cases like (5), the collocation *experiencing varied* is acceptable, though in the wider context of the concordance line we see that the derivation should be the adjective form *varying*.

(2) [parking machines . Parking fees **vary** depending on the area .]

(3) [best family restaurant should support **variable** meals which include enough nutrition]

(4) [The food the countries **vary** of accord a your location]

(5) [of the world are experiencing **varied** degrees of the temperature increase]

Using a corpus-based approach, a subset of two corpora are compared, a learner corpus and a general reference corpus of expert speakers (described in the next section). This data subset was extracted using a set of criteria relating to frequency and derivations, producing 7,554 tokens from the learner corpus embedded in concordance lines. The data in these corpora are free production, unlike many of the controlled experiments reviewed in previous sections. Rates of usage were compared using a variety of statistical tests, and errors were annotated manually before undergoing statistical analysis, focusing on the relationship between accurate production of collocations and correct choice of word form.

V Methodology

1 Dataset

The two corpora analyzed are COCA (The Corpus of Contemporary American English; Davies, 2008-, 2018 version) and the PELIC (The University of Pittsburgh English Language Institute Corpus; Juffs et al., 2020). Rather than querying through a web browser interface, all data files were stored and processed locally on a personal computer using Python and R programing languages in Jupyter notebooks.[5]

a COCA. COCA was selected as the expert-speaker reference corpus, using frequency information from the COCA list of top 100,000 words.[6] To match the L2 corpus information, only frequencies from the written domains were included in calculations. The written COCA data matches the L2 corpus in that American English is the norm and the target variety of EAP students in the U.S. The written subset of the corpus used in our study totals 419 million tokens equally distributed across domains and years, and can therefore be considered a "large, genre-balanced corpus of American English" (Davies, 2008-).

From the total frequency information, 'mid-frequency' lemmas in the K3-K9 frequency bands were specified. Items in this range are useful in terms of pedagogy (Schmitt, 2010) and deserve attention by learners once high-frequency words are known (Nation, 2016). Having established the mid-frequency lemmas, all the possible forms within each lemma's word family were extracted through *Compleat Lexical Tutor*'s 'familizer' function (Cobb, 2020), which builds

a list of headwords from an inputted text. For example, for the lemma *correct* (adj), the word family includes *correct* (v), *correction* (n), *correctly* (adv), etc.

Because the goal of the study was to compare the errors and distributions of different derivations within word (or lemma) families, it was necessary to narrow the analysis to those mid-frequency word families containing a sufficient range of possible lemmas for which students might (mis)use one form over another. As such, mid-frequency families were only selected if they had a minimum of four different mid-frequency derivations which were either nouns, verbs, adjectives, or adverbs. Furthermore, word families were excluded if they had a lemma in the top 100 most frequent COCA lemmas since such lemmas masked and skewed the overall data due to their drastically higher frequencies. Based on this restriction, the two families containing *time* (n) and *new* (adj) were removed. For the same reason, one outlier was also later removed; after compiling the concordancer, it was discovered that the lemma *reason* (n) accounted for 2,724 of the total 9,898 tokens under analysis, which was 27.5% of the total dataset. This preponderance of tokens was likely due to the high lemma rank of the noun form (413), its inclusion in task prompts (e.g., *Give reasons why...*), and its natural utility in the genre of argumentative essays. As a result of the above calculations, the final COCA dataset for analysis had the following characteristics: 27 word families (Appendix 1), 262 total lemma types, and 120 mid-frequency lemma types.

b PELIC. The selected L2 corpus is PELIC, a corpus consisting of spoken and written texts from learners in the Intensive English Program (IEP) at the University of Pittsburgh English Language Institute, collected from 2006-2012. PELIC was chosen as source of learner data because of its size, rich metadata, and open access. PELIC is also considered suitable for comparing to the American English of COCA because the data were collected in the United States.

In the current study, only the written texts were analyzed, totaling 4.2 million words across the three principal proficiency levels (intermediate, upper-intermediate, advanced). PELIC texts were collected in the five class types offered in the IEP – speaking, listening, reading, writing, and grammar – with the majority of tokens (63%) produced in the writing classes. This 'traditional' division of skills allowed for a split-level format in which students could be placed into different levels for different skills according to their needs (Juffs, 2020). There are 30 L1s represented in PELIC, though the five most common are Arabic (37.3%), Chinese (18.7%), Korean (18.2%), Japanese (5.7%), and Spanish (4.8%). Unlike most learner corpora which are cross-sectional (Callies, 2015), PELIC can be considered longitudinal in that learners often produced texts at different levels even though they did not always start at the beginning of the program or complete every level.

For this study, PELIC data were further narrowed by only including the first version of texts so as to eliminate near-identical drafts by students. Furthermore, only texts from 'Writing' classes were admitted resulting in a sub-corpus of extended written compositions, primarily essays and paragraphs. These texts averaged 172 words in length and were in response to a wide variety of 1355 prompts on a range of topics. From the remaining texts, a subset was collected of all texts containing any of the lemmas from the 'key families' in our COCA dataset. In sum, the learner data for analysis totaled 3,461 texts with 7,554 key family tokens. Of note, spelling mistakes in PELIC key family items were manually corrected to allow their inclusion in the analysis as the focus of this study was not the orthographic dimension of lexical depth.

2 Annotations

To prepare for human annotations of data, we programmed a concordancer for the key family tokens (sample in Appendix 2).[7] Each concordance line consisted of the node word and a span of five words on either side. Concordance lines were also randomized to reduce annotator bias and fatigue. The first 100 lines were set aside for annotator training purposes.

There were three expert human annotators consisting of a graduate linguistics student (Annotator 1), an undergraduate linguistics student (Annotator 2), and a linguistics faculty member (Annotator 3). Annotators 1 and 2 completed and compared answers on the training data before independently annotating the remaining 7,454 items. Annotator 3 acted as an adjudicator for the items for which there was disagreement between the original annotators. Using a small number of trained, independent annotators is considered superior to a larger number of untrained annotators (Bhardwaj & Ide, 2010), in-line with other fine-grained linguistic annotations and transcriptions (e.g., Crossley et al. 2015; Hanssen et al., 2015). To assess inter-annotator reliability, simple agreement rates (e.g., Hovy et al., 2006) were calculated rather than the commonly used Kappa. Although Kappa is designed to account for bias caused by chance agreement, it can be affected if the number of observed and chance matches coincide because it may underestimate the agreement of a rare category (Viera & Garrett, 2005). This is the case with our data, in which observed and chance error rates were very low. Both Kappa and agreement rates are frequently used in Natural Language Processing (Hovy & Lavid, 2010).

a Derivation annotations. Derivational errors were annotated manually, which is the norm in learner corpus research (Gries & Berez, 2017), due to the difficulties in accurately automating the process of error detection and correction. For each node word, the annotator decided whether its derivation was correct, and if not, decided the correct derivation and part of speech.

b Collocation annotations. Following an adapted version of Nesselhauf's (2005) procedure, collocations were assigned to one of four categories: *acceptable, unacceptable, questionable,* and *not applicable*. This four-way distinction provides greater flexibility than studies involving binary decisions of collocational acceptability (e.g., Laufer & Waldman, 2011). Acceptability judgments of this nature are to an extent inherently subjective but have a long history as a valid data type in linguistics (Schütze & Sprouse, 2013). The initial annotator training also served to increase reliability.

Critically, in contrast to collocation studies focusing on one specific construction such as verb-noun collocations (Laufer & Waldman, 2011; Nesselhauf, 2005), annotators gave an 'acceptable' rating if any acceptable collocation with the node word was present. Thus, [get some technical skills to **compete** on attractive positions] contains an acceptable collocation due to the noun-verb collocation *skills to compete* despite the infelicity of the verb-preposition collocation *compete on*, which should be *compete for* in this context. This methodological choice was made because our study is not investigating a specific collocational pattern, but rather taking a wider view of learners' overall collocational knowledge of our key families. Extracts (6) to (9) provide examples from the annotator training data of each category type. The span is underlined and the node word is in bold:

(6) [First, before he **opened** the restaurant, he did]

Acceptable: opened the restaurant is an acceptable collocation.

(7) [may grow up in the **varied** stages of social economy]

Questionable: varied stages should perhaps be *various stages*. No other collocations.

(8) [He could not watch **clearly.** Then he stopped the]

Unacceptable: watch clearly should be *see clearly.*

(9) [The unplanned budget, the **restriction** and the menu are three]

Not applicable: No collocation is expected as *the restriction* is an item in a list.

3 Automated measures

For comparison with the human annotations of collocation, four automated measures supplement the analysis. Two binary 'True/False' labels were assigned if any of the possible collocations in the collocation span appear in the Oxford Collocation Dictionary (McIntosh, Francis, & Poole, 2009) or in the top ten most frequent collocations in COCA. Two other associations measures, MI and t-scores, produced continuous numeric variables (described in Section 2). To calculate MI and t-scores, ngram frequency data from COCA were used[8] with the formulas from Davies (2008-) and Evert (2009) respectively.[9] To align the manner of human annotation, automated association measures were calculated for each two-word combination in the span with the key word, so that for (6), the possible combinations were *before opened, he opened, opened the,* and *opened restaurant.* The combinations with the highest MI and t-scores were kept as the 'best' potential collocations.

4 Data analysis

We report rates of use of items in the dataset by the two samples (expert speakers and learners) first, using descriptive statistics and linear mixed-effects models (LMMs). To compare how the two groups differed in their use of different parts of speech, relative percentages provide a broad overview. LMMs created in R (version 4.0.0; R Core Team, 2014) using the lme4 package (version 1.1-23; Bates et al., 2015) then present a more in-depth look at underlying factors and are

accompanied by proportions tables to describe significant effects. Similarly, descriptive statistics and logistic mixed-effects models are employed to examine derivation and collocation error rates in the annotated data. The use of mixed-effects regression models is considered appropriate for data containing groups of this nature and allow for sophisticated, realistic models in linguistics (Speelman, Heylen, & Geeraerts, 2018).

VI Findings

1 Use of 'key' word families

A total of 27 'key' word families which met the criteria outlined in Section V.1 were in the collected dataset, that is, families containing at least four mid-frequency lemmas in addition to other high- and low-frequency lemmas. Within these 27 families, learners in PELIC produced slightly more than half (51.9%) of all COCA lemmas, and an even greater percentage (85.8%) of the COCA mid-frequency lemmas that this study focused on (Table 1). As expected, rates of use increased across proficiency levels with advanced learners more likely to use a wider range of lexis. Table 2 provides further support for this fact; as the proficiency level increased, so too did the number of tokens per text from the 'key' families, indicating the importance of these lexical items for achieving advanced proficiency.

Table 1. Lemma forms of key families from COCA (total and mid-frequency)

	Total lemmas (type)	**Total lemmas (percentage)**	**'Key' family lemmas (type)**	**'Key' families lemmas (percentage)**
COCA	**262**	**100.0%**	**120**	**100.0%**
PELIC	**136**	**51.9%**	**103**	**85.8%**
Level 3 (Int)	79	30.2%	52	43.3%
Level 4 (Upper-Int)	124	47.3%	86	71.6%
Level 5 (Adv)	137	52.3%	100	83.3%

Table 2. Sources of tokens from key families in PELIC

Level	Students	Texts	Texts per student	'Key' tokens	'Key' tokens per text
Level 3 (Int)	269	574	2.13	865	1.51
Level 4 (Upper-Int)	592	1,678	2.82	3,551	2.12
Level 5 (Adv)	323	1,209	3.74	3,138	2.60

Patterns of usage also emerged when considering the parts of speech of key family lemmas. As Table 3 and Figure 2 reveal, expert speakers (COCA) and language learners (PELIC) were likely to use adjective and adverb forms attested in COCA with similar frequency, as compared with noun and verb forms. However, the ratios of noun and verb forms were nearly mirror images for the two corpora. Whereas in COCA the noun form was preferred over the verb form (33.4% to 23.9%), in PELIC this figure was reversed (21.0% to 31.9%), with the verb form occurring more frequently. A review of the individual lemma forms in Appendix 3 revealed following tendency: eight of the ten lemmas most used in COCA compared to PELIC were noun forms, whereas seven of the ten lemmas most used in PELIC compared to COCA were verb forms. For example, in COCA the lemma *confusion* (n) was produced 22.3% more often compared to other parts of speech than in PELIC. In contrast, the lemma *confuse* (v) was produced in PELIC 25.7% more often compared to other parts of speech than in COCA.

Table 3. Overview of part-of-speech distributions for PELIC and COCA

	Adjective	Adverb	Noun	Verb	**Total**
COCA	33.3% (232,067)	12.2% (85,172)	33.4% (232,896)	21.0% (146,321)	**100.0% (696,456)**
PELIC	30.9% (2,320)	13.3% (1000)	23.9% (1,792)	31.9% (2,396)	**100.0% (7,508)**

Note: 46 items were mis-tagged due to learner language as particles (40), prepositions (5), and pre-determiners (1) and were not included in this table.

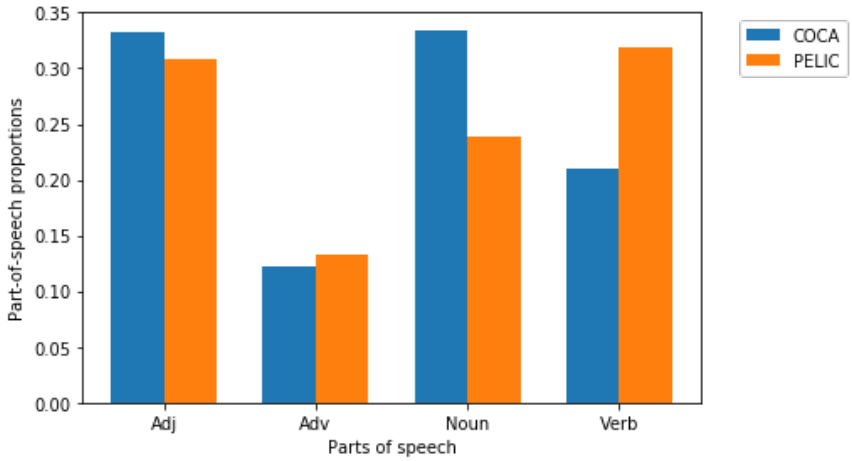


Figure 2. Part of speech proportions of key lemmas

To check the significance of these findings, four LMMs were run using the lmer() function. These regressions compared the parts of speech so that the outcome variable was the proportion of the time that one part of speech was used against the three others. In each one, the fixed factors were the corpus, the part of speech, the lemma frequency (log transformed and scaled), the interaction between corpus and frequency, and the interaction between corpus and part of speech. There was also a random effect for lemma to account for item-level variation. The fixed factors of corpus and frequency were significant for all LMMs ($t > 5$, $p < .001$), indicating that (1) a specific lemma was more likely to be produced if it had a higher raw frequency (regardless of the frequency of other lemmas in the same family), and (2) a lemma was more likely to be produced if it belonged to a part of speech which occurred more frequently in that corpus. Of greater relevance to our investigation, the interaction of corpus and part of speech was only significant in two of the four LMMs, the noun/non-noun and verb/non-verb LMMs (highlighted in bold), but not in the adjective/non-adjective and adverb/non-adverb LMMs. This factor from each of the four LMMs is compiled in Table 4. Here the baseline default is the interaction of COCA and the other three parts of speech, so that, for example, the interaction of the corpus and the noun forms significantly

decreased the intercept by -5.36 and the interaction of the corpus and the verb forms significantly *increased* the intercept by 6.41.

Table 4. Linear mixed effects models summary of interactions of part of speech and corpus

	Fixed effects		
Parameters	Estimate	*SE*	*t*
corpus PELIC:Adj	0.01	1.76	0.01
corpus PELIC:Adv	2.07	2.47	0.84
corpus PELIC:Noun	-5.36	1.73	**-3.10****
corpus PELIC:Verb	6.41	2.12	**3.03****

Note. *p<.05, **p<.01, ***p<.001
Models formula: lmer(proportion ~ corpus + POS + log(frequency+1) + corpus:POS + corpus:log(frequency+1) + (1 | lemma)).

2 Errors

a Collocation accuracy. Originally, the annotators made a three-way distinction between *Acceptable*, *Questionable*, and *Unacceptable* (as well as *Not applicable*). However, upon examination of the results (Table 5), only 1.2% of ratings were *Questionable*, so these were combined with the *Unacceptable* ratings under a broader *Not Acceptable* category. After collapsing the categories, the inter-annotator agreement rate was 91.3%. There were 90 *Not applicable* items which were excluded from further collocational analysis.

The descriptive statistics overview (Table 5) shows that the majority of human annotations (90.3%) considered there to be an acceptable collocation with the node words (the key words in the centre of the concordance lines). These judgements also appear to align with all three automated measures of collocation acceptability in that items with an acceptable rating were found more often in the dictionary and corpus, and they had higher MI and t-scores (Figure 3). A small number of outliers were removed when calculating the means of association scores, those items

more than 2.5 SD from the mean (1.9% of MI scores, 0.5% of t-scores). The 5356 'Acceptable' collocation MI scores (M = 2.26, SD = 2.83, 95% CI = 2.18, 2.33) compared to the 538 'Non-acceptable' collocation MI scores (M = 0.71, SD = 2.71, 95% CI = 0.48, 0.94) were significantly higher, t(5892) = 12.1, p =< .001, d = 0.56, as were the 5437 'Acceptable' t-scores (M = 18.6, SD = 30.1, 95% CI = 17.8, 19.4) compared to the 546 'Non-acceptable' t-scores (M = 9.8, SD = 32.0, 95% CI = 7.1, 12.5), t(5981) = 6.5, p =< .001, d = 0.14, though with a much smaller effect size. This finding means that the automated measures generally aligned with the pattern of human judgements of collocational acceptability, even though the mean MI of acceptable collocations was lower than the typical collocation threshold of 3. This discrepancy may in part be potentially due to grammatical collocations with high-frequency words (higher t-scores/lower MI), leniency of the part of the annotators, the acceptability of combinations of more than two words, or the annotators considering collocates beyond the target span.

Table 5. Overview of collocation acceptability measures

Human Rating	**Total N**	**Dictionary look up**	**Corpus look up**	**Mean MI**	**Mean t-score**
Acceptable	90.3% (6,647/7,364)	12.2% (704/6,647)	15.0% (994/6,647)	2.26	18.64
Not Acceptable	9.7% (717/7,364)	6.1% (44/717)	5.2% (37/717)	0.71	9.80
Total	**100% (7,364/7,364)**	**10.2% (748/7,364)**	**14.0% (1031/7,364)**	**2.10**	17.80

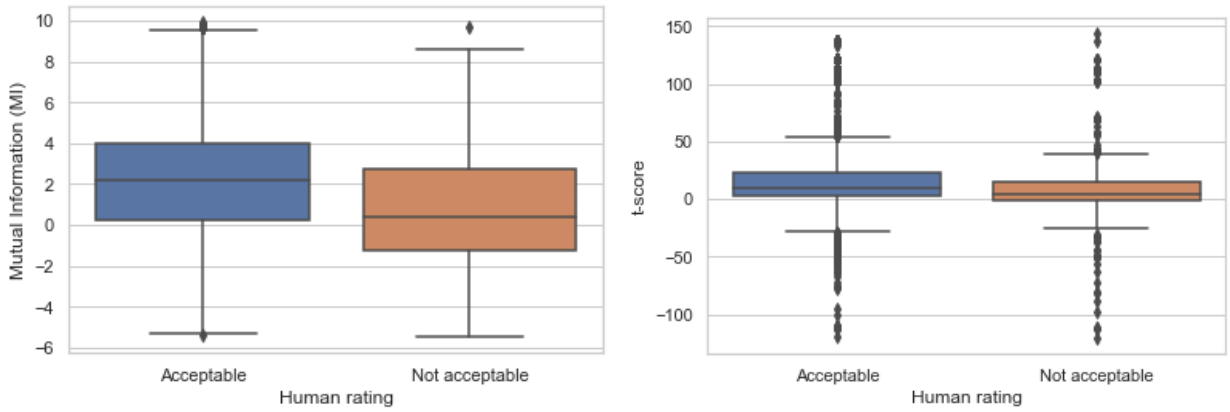


Figure 3. Collocation association scores and human ratings

Table 6 presents a sample of two-word combinations from the data with a range of association scores. At one extreme, combinations like *became confused* met the criteria of all the collocation acceptability measures, whereas combinations like *excited life* met none of them. In the middle, combinations like *compete for* or *be open* showed a more jagged profile, acceptable by some but not all of the collocation measures. In the case of `[being a open character person]`, there was no appropriate collocate with *open* recognized by the annotators, even though *be open* was listed in the collocation dictionary.

Table 6. Sample collocations and acceptability measures

Collocation	Human Rating	Dictionary look up	Corpus look up	MI	t-score
I became **confused** and at	Acceptable	True	True	10.84	9.32
use medium **heat** at this	Acceptable	True	True	10.33	53.54
individual to **compete** for him	Acceptable	True	False	3.69	70.30
being a **open** character person	Unacceptable	True	False	0.65	8.91
my case **specially** I want	Unacceptable	False	False	-2.14	-1.20
life to **excited** life	Unacceptable	False	False	-2.94	-4.03

Note. The key word in each concordance line is in bold, and the two-word combination with the highest MI in the span is underlined.

Delving deeper into the uniformity of these broad trends, a mixed-effect logistic regression (Table 7) was conducted to provide further insight into the relative significance of the various collocation

acceptability measures. The model was created using the glmer() function through a stepwise model selection (Baayen, 2008) in which the initial model was the dependent variable with a random intercept for participants. Independent variables and interactions were then added one-by-one to determine whether they were significant ($p<.05$) and improved model fit. In this model, the outcome variable was the binary human rating of *Acceptable/Not acceptable*, and the intercept was the likelihood (log odds) of an acceptable rating. The fixed factors were the three automated collocation acceptability measures (dummy coded), the learners' level, and the part of speech. Using a "maximal" random effects structure (Barr et al. 2013), crossed random effects for subject and items were fit for variables with sufficient data points. This process resulted in the inclusion of two random intercepts for subjects (students) and items ('key' families). No other main effects or interactions were significant. Of note, non-significant factors excluded from the model included t-scores and learners' L1s.

Table 7. Logistic mixed effects model for factors predicting 'acceptable' collocation rating

					Random effects	
	Fixed effects				By Subject	By Family
Parameters	Estimate	*SE*	*z value*	*Pr(>|z|)*	*SD*	*SD*
Intercept	2.33	0.16	14.71	**<0.001*** **	0.87	0.43
Level	0.32	0.12	2.61	**0.01** **		
POSAdv	-0.91	0.21	-4.41	**<0.001*** **		
POSNoun	-0.09	0.16	-0.55	0.59		
POSVerb	-0.28	0.15	-1.86	0.06		
Dictionary lookup	0.18	0.19	0.93	0.35		
Corpus lookup	0.21	0.22	0.95	0.34		
MI	0.22	0.02	10.74	**<0.001*** **		

Note. *$p<.05$, **$p<.01$, ***$p<.001$
Model formula: col_rating ~ level + POS + col_in_dict + col_in_COCA + MI + (1 | student_id) + (1 | family).

The results in Table 7 show that of the collocation measures, only MI was significant, providing a positive effect (estimate = 0.22, SE = 0.02, z value = 10.74). This finding is explored further in Figure 4 which shows that a higher MI score predicted a higher chance of a human judgement of *Acceptable*, taking into account all other factors (the small vertical lines indicate the density of observation). There was also a significant positive main effect for level (estimate = 0.32, SE = 0.12, z value = 2.61). Likewise, there was an effect for part of speech, but only when contrasting adjectives to adverbs; it seems adverbs were used less expertly, resulting in unacceptable collocations (estimate = -0.91, SE = 0.21, z value = -4.41). The overall conditional R^2 of the model was 0.31.

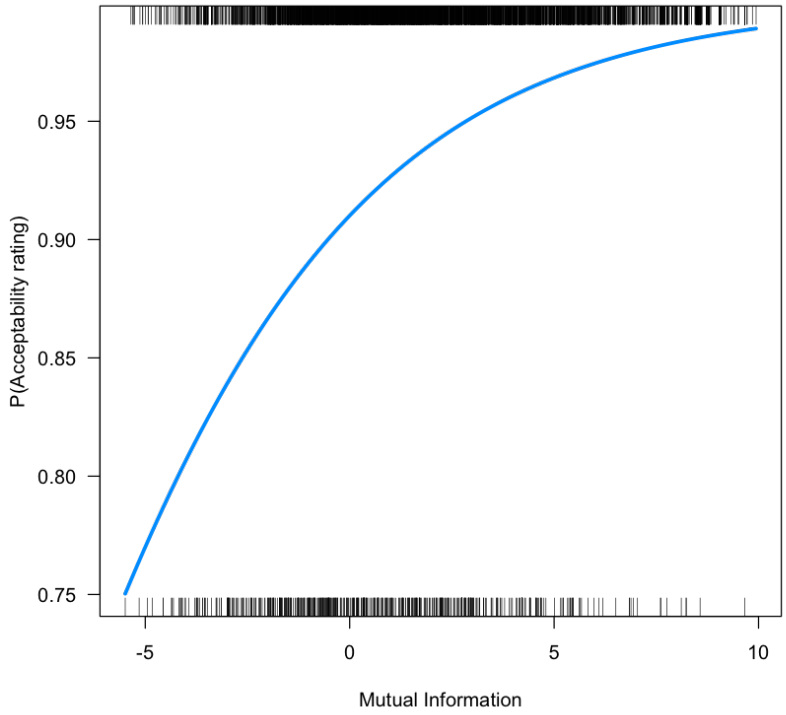


Figure 4. Probabilities of an 'Acceptable' collocation rating with respect to Mutual Information

b *Derivation accuracy.* For derivational accuracy there was a very high inter-annotator agreement of 97.7%, with the findings presenting a more straightforward picture than the

collocations. Overall, accuracy rates were very high, with negligible differences between proficiency levels (Table 8). When errors did occur, certain specific error types emerged:

- confusing *-ed* and *-ing* endings in adjective pairs like *embarrassed/embarrassing*:

  `[tutleneck . It was so **embarrassed** . Final reason is that]`

- using the wrong derivation with nearly identical orthography to the correct form (e.g., *especially/specially*):

  `[screen to see new movies **specially** comedian movies and to know]`

- confusing the adjective and adverb forms (e.g., *correct/correctly*):

  `[spelling , and choose the **correctly** words . After long time]`

- errors with compound words (e.g., *open mind/open-minded*):

  `[it requires you to be **open** mind and acquire the qualities]`

Table 8. Overview of derivation errors

Derivation accuracy	Level 3 (Int)	Level 4 (Upper-Int)	Level 5 (Adv)	**Total**
Accurate	98.8% (853/863)	98.0% (3,430/3,500)	98.5% (3,044/3,091)	**98.3% (7,327/7,454)**
Inaccurate	1.2% (10/863)	2.0% (70/3,500)	1.5% (47/3,044)	**1.7% (127/7,454)**

An LMM was conducted for derivational accuracy (Table 9), using the same methodology described for Table 7, but with 'Derivational accuracy' as the outcome variable. In this case, only one main effect was significant and improved the model, the human collocation rating level (estimate = 4.48, SE = 0.29, z value = 15.24). The same two random effects as in Table 7 were also significant, two random intercepts for subjects (students) and items ('key' families). Learners' L1, proficiency level, and part-of-speech were all non-significant. The conditional R^2 of the model was 0.47.

Table 9. Logistic mixed effects model for factors predicting derivation errors

					Random effects	
	Fixed effects				By Subject	By Family
Parameters	Estimate	*SE*	*z value*	*Pr(>|z|)*	*SD*	*SD*
Intercept	1.77	0.26	6.93	**<0.001*** **	0.57	0.92
Col rating	4.48	0.29	15.24	**<0.001*** **		

Note. *p<.05, **p<.01, ***p<.001
Model formula: Formula: derivation_accuracy ~ col_rating + (1 | student_id) + (1 | family).

3 Summary of findings

Our findings address two aspects of the learners in PELIC's productive lexical knowledge: the over/underuse of lemmas *within* the key families (indicative of a potential lack of lexical breadth), and the collocational and derivational accuracy of these lemmas in their writing (indicative of a potential lack of lexical depth). Of the two areas, the former showed the most striking results; in answer to our first research question, there was a clear difference between the use of key lemma forms by expert speakers and by learners, especially in terms of the part-of-speech preference. Whereas learners favored verb forms, expert speakers more often favored noun forms (Figure 2).

With respect to the findings relating to accuracy, our survey of collocation acceptability measures found MI and t-scores to correlate with human judgements, though human judgements were more lenient in terms of what constituted an acceptable collocation. In answer to our second research question, acceptable collocations were seen to increase across all three proficiency levels, though only significantly from levels 3 to 4. Interestingly, collocations were less acceptable when adverbs were involved (in comparison to adjectives). Compared to other collocation studies reviewed (Nesselhauf, 2005; Laufer & Waldman, 2011), our collocation error rates were much lower, likely due to the consideration of all collocations with the node word rather than assessing

only one type of collocation. Similarly, derivation accuracy rates were uniformly high, and increased only insignificantly across levels. We now turn to the implications of these findings, considering their relevance to language pedagogy.

VII Discussion

1 Pedagogical implications

Returning to Dóczi and Kormos's framework, the dimensions towards the top of the hierarchy must become the goal for learners striving to reach academic readiness for university study. As such, productive knowledge of collocations and different word forms would seem an appropriate target. It is true, as we reviewed, that there are many challenges to acquiring these types of knowledge, and in our study learners differed from expert speakers in their patterns of usage. However, we also observed that collocational accuracy improved across proficiency levels, and that collocational accuracy was interconnected with derivational accuracy. Consequently, there is the potential for explicit instruction which targets these lexical knowledge dimensions from multiple angles in a cohesive and principled manner to increase breadth and depth of knowledge.

The inclusion of explicit collocation instruction in ESL/EFL contexts is well-established (Nizonkiza & Vdpoel, 2019), often credited to the popularity of the Lexical Approach (Lewis, 1993) and authors such as Willis (1990) (cf. Ellis, 2001; Wray, 2002 for arguments in favor of *implicit* instruction). Though it inevitably requires additional time and attention, knowing collocational patterns can reduce overall learning burden (Nation, 2001) because these formulaic sequences provide scaffolding in relating meaning and form in the appropriate syntactic context. Scaffolding of this type is important for learners in specific discourse communities, like EAP

learners developing their writing, and is correlated to higher grades and perceived proficiency (Durrant, 2018).

However, though there is general unanimity about the need for learning collocations, the decision as to which collocations to teach appears to often be haphazard, with vocabulary selection continuing to be "unprincipled and opportunistic" (Macis & Schmitt, 2017, p. 334). Thus, there is still no consensus on this important pedagogical matter (Nizonkiza & Vdpoel, 2019). To be sure, there is value in teaching collocations as productive language needs emerge (Meddings & Thornbury, 2009), but principled lexical instruction remains essential (Cobb, 2016). What is known is that learners attend to frequent combinations like collocations (Siyanova-Chanturia, 2015), and the majority of collocates of high-frequency items are likely to be high-frequency themselves (Nizonkiza & Vdpoel, 2019). It would seem then that what is less likely to be acquired incidentally are collocations with items in the mid-frequency (K3-K9 bands) which occur less regularly in the input and may also collocate with less-frequent words. Furthermore, with respect to frequency, it is unclear the extent to which MI is being taken into consideration when selecting lexical items to teach, with raw frequencies and frequency bands being the predominant criterion. Since it has been documented that L2 university students lack high MI collocations in their writing (Ellis, Simpson-Vlach, & Maynard, 2008), and that use of higher MI collocations are a feature of more advanced proficiency (Granger & Bestgen, 2014; Paquot, 2017), a focus on this particular metric would seem prudent.

Considering derivations, Mantyla and Huhta (2014) have commented that knowledge of word parts is rarely taught, although some coursebook series, such as *Words for Students of English*, have made it a priority (Davis et al., 2015a; Davis et al., 2015b). Some evidence, however, indicates that such knowledge increases vocabulary size (i.e., lexical breadth) and reading speed

(Nation, 2001), and it can aid delayed retrieval of vocabulary (Zhang, 2002). To teach affixation, Mochizuki and Aizawa (2000) suggest an order for learning affixes based on level of difficulty, and Bauer and Nation (1993) propose orders based on various factors including frequency and regularity. However, another frequency-based option is to teach affixation in relation to frequency bands and word families, for example the affixes present in the 'key' families from this study (Appendix 1).

Returning to the error types found in our data, these errors seem to correspond to less salient morphemes, also a key factor in the natural order of inflectional L2 morpheme acquisition (Goldschneider & DeKeyser, 2001). Such morphemes may also have lower functional load (Bardovi-Harlig, 2000) because they do not contribute as much as other morphemes to the intended meaning of the sentence. For example, there is a weak vowel sound in ***e****specially* and *open-mind****ed*** which can alter the part of speech of the word. Similarly, the *-ed*/*-ing* endings are ambiguous in that they can be either inflectional or derivational, depending on the syntactic frame in which they occur. Other pairs of challenging lemmas include adjective/adverb pairs like *correct/correctly* which may be used interchangeably in less formal registers but not in academic writing. One implication is that without explicit instruction to promote noticing, learners may struggle to become aware of these types of inaccuracies in their own production. By teaching collocational patterns with these 'hard-to-notice' lexical items, collocational breadth and derivational accuracy can be promoted in tandem.

The importance of affix knowledge notwithstanding, we suggest that teaching affixation in isolation neglects other elements of derivational competence, as well as failing to acknowledge the individual differences of learners. After all, some learners may prefer learning whole words (i.e., bigger 'chunks' or constructions) and the complexity of the learners' L1 morphology may also

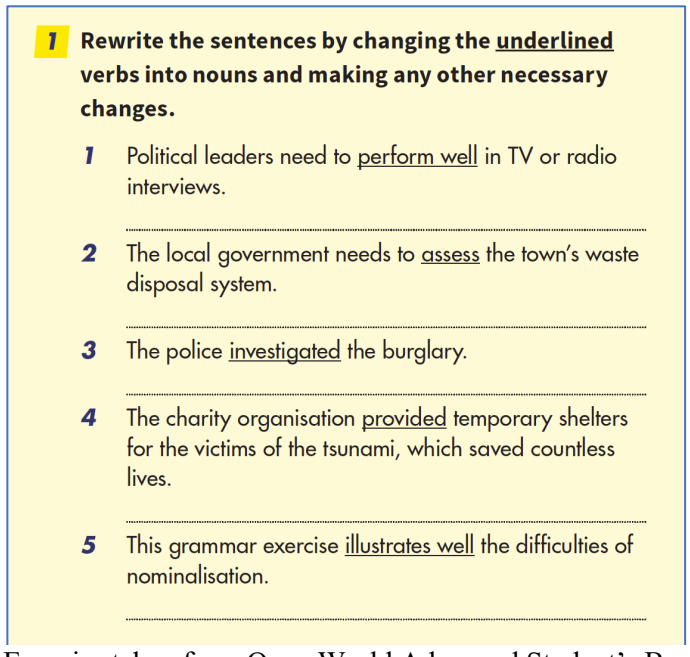
dictate to some extent whether 'affix-focused' or 'word-family focused' instruction is more effective (Wu & Juffs, 2021). What can be asserted with greater certainty is that most learners "are likely to have only mastery of a limited number of word knowledge categories for a large percentage of words in their lexicon" (Schmitt & Meara, 2007, p. 18), and that knowledge of one word form does not entail productive knowledge of others (Brown, 2013; Schmitt & Zimmermann, 2002). As a result, learners may rely on the well-known strategy of avoidance in their own writing.

As this pattern pertains to the current study, our data point to gaps in productive knowledge occurring more frequently with the noun forms within word families. These findings therefore support the research of Parent (2019), which found that for learners, complex words were more likely to be misused as nouns than verbs, and that some expected forms were not used by the learners at all. If true, then the impact on the perceived proficiency of learners' writing could be substantial. In an analysis of written complexity, Lu (2011) found evidence that complex nominal measures are an important index of syntactic complexity. Furthermore, modern academic writing relies heavily on nominalized structures (Biber & Gray, 2010). It therefore stands to reason that for learners to produce more complex, nominal phrases, they must first acquire productive knowledge of the noun forms which form the basis of such phrases.

To address this need for increased word family knowledge, particularly of noun forms, a common practice is to expose learners to input containing the target language, potentially through academic reading which contains a high proportion of derivates and complex words (Schmitt & Zimmerman, 2002). Nevertheless, productive practice is essential to increase productive skills proficiency, necessitating a need to carefully select which word families and forms to target. The following sample task illustrates some of the potential strengths and areas for improvement of current coursebook materials with respect to developing derivational and collocational knowledge.

2 A sample task sequence

The following exercise is from an advanced exam preparation coursebook (Figure 5; Cosgrove & Wijayatilake, Forthcoming, p. 223). The entire page focuses on nominalization, with this particular exercise providing controlled written practice through a transformation task. As it is currently constructed, the exercise has many merits, namely, it requires written production; it deals with the important academic language feature of nominalization (and by extension derivation); and it requires learners to draw on their existing lexical knowledge as the words required for the answers are not provided for them in the text or elsewhere. Accordingly, this exercise is an excellent springboard for developing lexical knowledge, both breadth and depth.



1 **Rewrite the sentences by changing the underlined verbs into nouns and making any other necessary changes.**

1 Political leaders need to perform well in TV or radio interviews.

2 The local government needs to assess the town's waste disposal system.

3 The police investigated the burglary.

4 The charity organisation provided temporary shelters for the victims of the tsunami, which saved countless lives.

5 This grammar exercise illustrates well the difficulties of nominalisation.

Exercise taken from Open World Advanced Student's Book, page 223, © Cambridge University Press and UCLES 2021. Reprinted with permission

Figure 5. Nominalization task (Cosgrove & Wijayatilake, Forthcoming, p.223)

Still, the items themselves suffer from a somewhat unprincipled approach to inclusion, common in textbooks (Koprowski, 2005), by focusing on lexical items from a range of frequency bands

(though elsewhere, dedicated vocabulary sections do closely adhere to CEFR levels). It is true that in some cases the choice of lexis may result from task or topic restrictions. However, with careful deliberation, there is usually sufficient topic-related lexis at a range of frequency bands to be able to satisfy task/topic and frequency considerations. In order of the answers, the key nouns and their COCA lemma frequencies are presented in Table 10. These figures show that three out of the five key nouns are high-frequency items (K1-K2), with two mid-frequency items, *provision* and *illustration* (K5), and no K3 or K4 items. It can be argued that by working with high-frequency items, learners are able to direct their attention to the grammar of nominalization. However, given the expectations on advanced learners, omitting greater consideration of mid-frequency lexis is perhaps a missed opportunity.

Table 10. Frequency of nouns in sample task

Task number	Possible noun in answer	COCA lemma frequency	COCA frequency band
1	performance	674	K1
2	assessment	1711	K2
3	investigation	1381	K2
4	provision	4867	K5
5	illustration	4313	K5

What we would propose is to further exploit a task such as this, first by adjusting the lexical items being practiced, and second by supplementing the exercise with other related extension tasks, either in the student's book or as extension tasks in the teacher's book. In terms of substitution, reworking of the text could allow for a focus on more mid-frequency lemmas, like *provision* and *illustration.* For example, the underlined words in item 1 could easily be reworded to *be exciting* to test *exciting* (adj) → *excitement* (n), one of the lemmas from our key families that is typically underused by learners, rather than *perform* (n). In doing so, typical errors, like those found in our concordances could be targeted (6-8):

(6) [use your imagination and feel **exciting** and horror . Above all]

(7) [could bring us not only **excitations** but also other direct experiences]

(8) [according to the level of **exciting** , I have three options]

Alternatively, this nominalization could also take the form of a Key Word Transformation exercise in which students are given key words and the beginning/ending of the sentence. For example, item 3 would appear as follows (answer: *conducted an investigation into*):

The police investigated the burglary.
conducted
The police _____ the burglary.

Next, awareness-raising tasks could cohesively promote noticing by students of the language being nominalized. For instance, there are opportunities for consideration of the affixes used to transform verbs into nouns (*excite/assess* + *-ment*; *investigate/provide/illustrate* + *-sion/tion*). Focusing on larger constructions, collocations with these nouns could then extend learners' breadth and depth of knowledge, for example, a task where students select which verbs do not collocate with the key nouns (Figure 6):

Cross out the verbs which do not collocate with the noun in bold.

1 create / build / ~~make~~ / show **excitement**

2 conduct / receive / complete / perform **an assessment**

3 conduct / launch / complete / perform **an investigation**

4 give / include / contain / allow **a provision**

5 provide / feature / reveal / depict **an illustration**

Figure 6. Sample collocation follow-up task

Of course, these few tasks are not only beneficial for the isolated dimensions of lexical knowledge. By recycling these same lexical items multiple times, each time highlighting a new aspect,

cognitively learners strengthen the connections between different elements of form, and the level of challenge remains sufficiently high. What is more, from a classroom and pedagogical viewpoint, lesson planning and delivery is both cohesive and efficient, fully exploiting the one short task to develop knowledge of both specific lemmas and more generalizable patterns. From the perspective of learner training, such knowledge may in turn also facilitate future noticing (Schmidt, 2001) of other linguistic patterns.

VIII Conclusions

The aim of this study was to uncover learners' patterns of usage with respect to mid-frequency lexical items in order to learn more about their productive derivational and collocational lexical knowledge. To do so, we conducted a corpus analysis of learners' writing and compared it to an expert-speaker target. The results suggest that there are significant differences between the two populations for the aspects under investigation, with learners underusing noun forms when verb forms are otherwise available, despite nominalization being a key attribute of academic prose (Biber & Gray, 2010; Wells, 1960). The findings also support the notion that learners overly rely on one or two forms within word families to the exclusion of others (Schmitt & Zimmerman, 2002), resulting in low collocational and derivational error rates.

As far as the limitations of this study are concerned, the extent to which the findings are corpus-specific must be considered, which is true of any corpus-based investigation. Although the patterns described in this study are informative and fit well within other research in the field, their generalizability to other EAP environments remains to be confirmed. For validation, similar or replication studies with other learner corpora would be conducted, including those which have sampled students with different or homogenous L1s. In addition, comparable studies utilizing

spoken data or data from other genres would afford further insights into how productive lexical knowledge is realized in different contexts. Another desideratum is a closer examination of how data such as ours might be applied in classroom environments through material design, curriculum design, and pedagogical choices. In Section VII.2, one possible application was reviewed, but this is clearly only a preliminary step. Classroom interventions, quantitative and qualitative research, and two-way discussion with the teaching community are necessary in order to determine the efficacy of prioritizing the type of lexical items investigated in this study.

As Roche and Harrington (2013) among others point out, English has become the dominant language of instruction in tertiary education. Consequently, this reality has required L2 English learners worldwide to meet the challenge of attaining advanced proficiency, of which one critical aspect is accurate, productive lexical knowledge. This study contributes to our growing understanding of the complex nature of this knowledge and raises important implications for future lexical teaching and research. By making concrete pedagogical choices to focus on key word families that match certain criteria, it is feasible to promote the parallel development of multiple aspects of lexical knowledge.

Acknowledgements

We express our gratitude to John Starr and Dr. Na-Rae Han for their data annotation, and to the journal editors and anonymous reviewers for their helpful suggestions. All errors are, of course, our own.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this work was supported by the Social Sciences and Humanities Research Council of Canada and is based in part upon work supported by the National Science Foundation [grant number SBE-0836012].

Notes

1. *Lexical knowledge* is the preferred term in this paper, though original terms have been maintained when discussing other authors' work (e.g., Chen & Truscott, 2010; Crossley & Skalicky, 2019). Such knowledge is called *word knowledge* (e.g., Dóczi & Kormos, 2016; Milton, 2009), or sometimes 'vocabulary' in general.

2. The term *native speaker* is often used when describing corpora but is ideologically problematic (Holliday, 2006). Instead, this paper uses Rampton's (1990) term *expert speaker*, which includes L1 English speakers, highly-proficient L2 English speakers, and balanced bilinguals.

3. Some authors argue that it is not necessarily best practice to compare expert speaker and learner production (e.g., Fitzpatrick, 2007; Zareva & Wolter, 2012) because differences from the expert-speaker target may not be due to a lack of proficiency. Other factors which could explain

differences in usage include cultural and educational backgrounds, language experiences, and the way words are conceptualized and associated in the lexicon.

4. Here we adopt the mostly uncontroversial position that learners are sensitive to L2 morphology and therefore notice affixation and not just whole words. However, morphological sensitivity is not a binary quality, and the extent of this sensitivity is highly dependent upon a number of factors including the learners' L1 (Diependaele et al., 2011; Rehak & Juffs, 2011), proficiency levels (Bosch et al., 2016; Beyersmann et al., 2015), and explicit morphological knowledge (Deng et al., 2016; Ishikawa, 2019).

5. Dataset and code available at https://github.com/ELI-Data-Mining-Group/Lexical_knowledge_Naismith_Juffs_2021

6. https://www.wordfrequency.info/purchase.asp

7. Other measures of lexical depth beyond the scope of this study were also annotated for: 'accuracy of inflection' and 'clarity of meaning'.

8. https://www.ngrams.info/purchase_iweb.asp

9. MI = log((AB * corpus size)/(A * B * span)) / log(2); t-score = (O − E) / √O

References

Anderson, R. C. & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthire (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Neward, DE: International Reading Association.

Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R.* Cambridge University Press.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use.* Oxford: Blackwell.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), Volume 68, Issue 3, 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bauer, L. & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*, 253-279.

Bestgen, Y. & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of second language writing, 26*, 28–41.

Beyersmann, E., Casalis, S., Ziegler, J., & Grainger, J. (2015). Language proficiency and morpho-orthographic segmentation. *Psychonomic Bulletin & Review, 22*(4), 1054-1061.

Bhardwaj, V & Ide, N. (2010). Anveshan: A framework for analysis of multiple annotators' labeling behavior. *Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010*, (pp. 47–55), Uppsala, Sweden, 15-16 July 2010.

Biber, D., & Gray, B. (2010). Nominalizing the verb phrase in academic science writing. *The verb phrase in English: Investigating recent language change with corpora*, 99-132.

Bonk, W. J. (2000). Second Language Lexical Knowledge and Listening Comprehension, *International Journal of Listening*, *14*(1), 14-31.

Booij, G. (2017). Construction morphology. In M. Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics.* Oxford: Oxford University Press.

Bosch, S., Krause, H., & Leminen, A. (2016). The time-course of morphosyntactic and semantic priming in late bilinguals: A study of German adjectives. *Bilingualism: Language and Cognition*, 1-22.

Brown, D. (2013). Types of words identified as unknown by L2 learners when reading. *System, 41*(4), 1043-1055.

Callies, M. (2015). Learner corpus methodology. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 35-56). Cambridge: Cambridge University Press.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693–713.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics, 16*(1), 22–29.

Cobb, T. *Familizer + Lemmatizer v.2.2* [computer program]. Available at https://lextutor.ca/familizer/ (accessed January 2020).

Cobb, T. (2016). Numbers or Numerology? A response to Nation (2014) and McQuillan (2016). *Reading in a Foreign Language, 28*(2), 299-304.

Cosgrove, A. & Wijayatilake, C. (Forthcoming). *Open World Advanced Student's Book without Answers with Online Practice*. Cambridge: Cambridge University Press.

Crossley, S. A., Salsbury, T., Mcnamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: a case for collocation accuracy. *Applied Linguistics, 36*(5), 570-590.

Crossley, S. & Skalicky, S. (2019). Examining lexical development in second language learners: An approximate replication of Salsbury, Crossley & McNamara (2011). *Language Teaching, 52*(3), 385-405.

Dabbagh, A. & Enayat, M. J. (2019). The role of vocabulary breadth and depth in predicting second language descriptive writing performance. *The Language Learning Journal, 47*(5), 575-590.

Davies, M. (2008-). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Available at https://corpus.byu.edu/coca/ (accessed October 2018).

Davis, B., Juffs, A., McCormick, D. E., Mizera, G., O'Neill, M. C., Slaathaug, M., & Smith, D. (2015). *Academic Vocabulary Building in English, Low Intermediate* (Vol. 1). Ann Arbor: University of Michigan Press.

Davis, B., Juffs, A., McCormick, D. E., Mizera, G., O'Neill, M. C., Slaathaug, M., & Smith, D. (2015). *Academic Vocabulary Building in English, Intermediate* (Vol. 2). Ann Arbor: University of Michigan Press.

Deng, T., Shi, J., Bi, H., Dunlap, S., & Chen, B. (2017). The relationship between the morphological knowledge and L2 online processing of derivational words. *International Journal of Bilingualism*, *21*(4), 402-418.

Diependaele, K., Duñabeitia, J.A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language, 64*: 2344-358.

Dóczi, B. & Kormos, J. (2016). *Longitudinal Developments in Vocabulary Knowledge and Lexical Organization*. Oxford: Oxford University Press.

Durrant, P. (2018). Formulaic language in English for academic purposes. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 211-227). Routledge.

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-68). Cambridge: Cambridge University Press.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic Language in native and second language speakers: psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*, 375-396.

Evert, S. (2009). Corpora and collocations. In A. Lüdeling and M. Kytö (Eds.), *Corpus Linguistics. An International Handbook* (pp. 1212-1248). Mouton de Gruyter.

Fitzpatrick , T. (2007), Word association patterns: unpacking the assumptions. *International Journal of Applied Linguistics, 17*, 319–331.

Fitzpatrick, T. (2012). Tracking the changes: Vocabulary acquisition in the study abroad context. *The Language Learning Journal,* 40(1), 81–98.

Goldschneider, J.M., & DeKeyser, R.M. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language Learning, 51,* 1–50.

Granger, S. (Ed.). (1998). *Learner English on computer*. London & New York: Addison Wesley Longman.

Granger, S. (2018). Formulaic sequences in learner corpora: Collocation and lexical bundles. In A. Siyanova-Chanturia & Pellicer-Sanchez, A. (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 228-246). Routledge.

Granger, S. & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics, 52*(3), 229-252.

Gries, S. Th. & Berez, A. L. (2017). Linguistic annotation in/for corpus linguistics. In N. Ide and J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation* (pp. 379-409). Dordrecht: Springer Science+Business Media.

Gyllstad, H. & Wolter, B. (2015). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning, 66,* 296-323.

Hanssen, E., Versloot, A., Hoekstra, E., Banga, A., Neijt, A., & Schreuder, R. (2015). Morphological variation in the speech of Frisian-Dutch bilinguals: (Dis)similarity of linking suffixes and plural endings. *Linguistic Approaches to Bilingualism, 5*(3), 356-378.

Haomin, Z. & Bilü, Z. (2017). Multi-faceted morphological awareness and vocabulary knowledge in English as a second language learners: A multivariate analysis. *Chinese Journal of Applied Linguistics, 40*(1), 42-55.

Hatch, E. & Brown, C. (1995). *Vocabulary, semantics, and language education*. Cambridge: Cambridge University Press.

Henriksen, B., & Stenis Stæhr, L. (2009). Processes in the development of L2 collocational knowledge — a challenge for language learners, researchers and teachers. In Barfield, A.,

& H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 224–231). Basingstoke: Palgrave Macmillan.

Holliday, A. (2006). Native-speakerism. *ELT Journal* 60(4), 385-387.

Hovy, E. & Lavid, J. (2010). Towards a 'science' of corpus annotation: A new methodological challenge for corpus linguistics. *International Journal of Translation, 22*(1), 13-36.

Hovy, E. H., Marcus, M., Palmer, M., Ramshaw, L. & Weischedel, R. (2006). OntoNotes: the 90% solution. In *Proceedings of theHLT- NAACL-2006 Conference*, 57-60.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Ishikawa, K. (2019). Incidental and explicit learning of L2 derivational morphology and the nature of acquired knowledge. *Applied Psycholinguistics, 40*(6), 1377-1404.

Ishii, T. & Schmitt, N. 2009. Developing an integrated diagnostic test of vocabulary size and depth. *RELC, 40*(1), 5-22.

Juffs, A. (2019). Lexical development in the writing of English language program students. In R. M. DeKeyser & G. P. Botana (Eds.), *Reconciling methodological demands with pedagogical applicability* (pp. 179-200). John Benjamins.

Juffs, A. (2020). *Aspects of Language Development in an Intensive English Program*. Routledge.

Juffs, A., Han, N-R., & Naismith, B. (2020). The University of Pittsburgh English Language Corpus (PELIC) [Data set]. http://doi.org/10.5281/zenodo.3991977

Kang, B. (2018). Collocation and word association: Comparing collocation measuring methods. *International Journal of Corpus Linguistics, 23*(1), 85-113.

Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal*, *59*(4), 322-332.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing, 21*, 202-226.

Laufer, B. & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647-672.

Lee, S. (2019). L1 transfer, proficiency, and the recognition of L2 verb-noun collocations: A perspective from three languages, *International Review of Applied Linguistics in Language Teaching,* (published online ahead of print).

Lenko-Szymanska, A. (2019). *Defining and Assessing Lexical Proficiency*. New York: Routledge.

Levelt, W. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. (2001). Spoken word production: a theory of lexical access. *Proceedings of the National Academy of Sciences, USA, 98*, 13464-13471.

Lewis, M. (1993). *The lexical approach: The state of ELT and the way forward.* Hove, UK: Language Teaching Publications.

Liu, E. & Shaw, P. M. (2006). Investigating learner vocabulary: A possible approach to looking at EFL/ESL learners' qualitative knowledge of the word, *IRAL – International Review of Applied Linguistics in Language Teaching*, *39*(3), 171-194.

Lorenz, G. R. (1999). *Adjective intensification – learners versus native speakers: A corpus study of argumentative writing*. Amsterdam; Atlanta, GA: Rodopi.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*(1), 36-62.

Macis, M. & Schmitt, N. (2017). Not just 'small potatoes': Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, *21*(3), 321-340.

Mantyla, K. & Huhta, A. (2014). Knowledge of word parts. In J. Milton, & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp. 45-59). UK: Palgrave Macmillan.

McIntosh, C., Francis, B., & Poole, R (Eds.). (2009). *Oxford Collocations Dictionary: For Students of English.* Oxford: Oxford University Press.

Meddings, L. & Thornbury, S. (2009). *Teaching Unplugged: Dogme in English Language Teaching*. Peaslake UK: Delta Publishing.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. UK: Short Run Press.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in English as a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, M Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83-98). Toronto: Multilingual Matters.

Mochizuki, M. & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, *28*, 291-304.

Moghadam, S., Zaidah, Z., & Ghaderpour, M. (2012). A review on the important role of vocabulary knowledge in reading comprehension performance. *Procedia - Social and Behavioral Sciences* 66, 555-563.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing.* John Benjamins.

Nation, P. & Waring, R. (1997). Vocabulary Learning Strategies. In N. Schmitt and M. McCarthy (Eds.) *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6-19). Cambridge University Press.

Nesselhauf, N. (2005). *Collocations in a learner corpus.* John Benjamins.

Nizonkiza, D & Vdpoel, K. (2019). Mind the gap: Towards determining which collocations to teach. *Stellenbosch Papers in Linguistics Plus, 56*. 13-30.

Paribakht, T. S. & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*(1), 9-29.

Pellicer-Sánchez, A. (2019). Examining second language vocabulary growth: Replications of Schmitt (1998) and Webb & Chang (2012). *Language Teaching, 52*(4), 512-523.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*, 282-308.

Paquot, M. (2017). The phraseological dimension in interlanguage complexity research. *Second Language Research,* 1-25.

Parent, K. (2019). Do language learners walk the walk? noun-verb converted forms in learner English. *언어학 연구, 50*(50), 215-244.

R Core Team. (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.Vienna, Austria. http://www.R-project.org/.

Rampton, M. B. H. (1990). Displacing the 'native speaker': expertise, affiliation, and inheritance. *ELT Journal, 44*(2), 97-101.

Read, J. (1998). Validating a Test to Measure Depth of Vocabulary Knowledge. In A. Kunnan (Ed.), *Validation in Language Assessment* (pp. 41–60). Routledge. https://doi.org/10.4324/9780203053768

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a second language*: *Selection, acquisition, and testing* (pp. 209-227). Amsterdam/Philadelphia, PA: John Benjamins.

Rehak, K.M., & Juffs, A. (2011). Native and non-native processing of morphologically complex English words: Testing the influence of derivational prefixes. In Granena, G., et al. (Eds.), Selected Proceedings of the 2010 Second Language Research Forum (pp. 125–142). Somerville, MA: Cascadilla Proceedings Project.

Richards, J. (1976). The Role of vocabulary teaching. *TESOL Quarterly, 10*(1), 77-89.

Roche, T. & Harrington, M. (2013). Recognition vocabulary knowledge as a predictor of academic performance. *Language Testing in Asia, 3*(1), 1-13.

Saito, K. (2020), Multi- or single-word units? The role of collocation use in comprehensible and contextually appropriate second language speech. *Language Learning, 70*, 548-588.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. UK: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 1-39.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*(2), 261-274.

Schmitt, N. & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *SSLA, 20*, 17-36.

Schmitt, N. & Zimmerman, C. B. 2002. Derivative word forms: What do learners know? *TESOL Quarterly, 36*(2), 145-171.

Schütze, C.T., & Sprouse, J. (2013). Judgment data. In Podesva, R.J., & D. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). New York: Cambridge University Press.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System, 53,* 148-160.

Speelman, D., Heylen, K., & Geeraerts, D. (Eds.). 2018. *Mixed-Effects Regression Models in Linguistics*. Springer.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative methods. *Functions of language 2*(1). 23-55.

Tan, D., Pandian, A., & Jaganathan, P. (2016). Lexical testing and the reliability of the modified vocabulary knowledge scale. *Advances in Language and Literary Studies, 7*(5), 91-96.

Viera, A. J. & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine*, *37*(5), 360-363.

Waring, R. (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata [occasional papers of Notre Dame Seishin University, Okayama], 1*, 53–68.

Wells, R. (1960). Nominal and verbal style. In E. Sebeok (Ed.), *Style in Language* (pp. 213-220). Cambridge, MA: MIT Press.

Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. London: Collins COBUILD.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2018). Concluding Question: Why don't second language learners more proactively target formulaic sequences? In A. Siyanova-Chanturia & Pellicer-Sanchez, A. (Eds.),

Understanding formulaic language: A second language acquisition perspective (pp. 248-269). Routledge.

Wu, Z., & Juffs, A. (2021). Effects of L1 morphological type on L2 morphological awareness. *Second Language Research.*

Zareva, A., & Wolter, B. (2012). The "promise" of three methods of word association analysis to L2 lexical research. *Second Language Research*, *28*(1), 41–67.

Zhang, L. J. (2002). Metamorphological awareness and EFL students' memory, retention, and retrieval of English adjectival lexicons. *Perceptual and Motor Skills, 95*, 934-944.

Appendix 1: Key word families

N.B. Mid-frequency lemmas are in bold

Family	Lemmas in family
ACCEPT	accept (v), acceptability (n), **acceptable (adj)**, acceptably (adv), **acceptance (n), accepted (adj)**, acceptor (n), unacceptability (n), **unacceptable (adj)**, unacceptably (adv)
ADVISE	advisability (n), advisable (adj), **advise (v)**, advisedly (adv), advisement (n), **adviser (n)**, **advisor (n)**, advisory (adj), inadvisable (adj)
BACK	back (adj), back (n), back (adv), back (v), backed (adj), backer (n), **backing (n)**, **backward (adj)**, **backward (adv)**, backwardness (n), **backwards (adv)**
COLLABORATE	**collaborate (v)**, **collaboration (n)**, collaborationist (n), **collaborative (adj)**, collaboratively (adv), **collaborator (n)**
COMPETE	anticompetitive (adj), **compete (v)**, **competing (adj)**, competitive (adj), competitively (adv), **competitiveness (n)**, **competitor (n)**, uncompetitive (adj)
CONFUSE	**confuse (v)**, **confused (adj)**, confusedly (adv), **confusing (adj)**, confusingly (adv), **confusion (n)**
CONSTRUCT	**construct (n)**, **construct (v)**, constructed (adj), construction (n), **constructive (adj)**, constructively (adv), constructivist (adj), constructivist (n), constructor (n), **reconstruct (v)**, reconstructed (adj), **reconstruction (n)**, unreconstructed (adj)
CONTINUE	continual (adj), **continually (adv)**, continuance (n), continuation (n), continue (n), continue (v), **continued (adj)**, **continuing (adj)**, continuity (n), **continuous (adj)**, **continuously (adv)**
CORRECT	**correct (adj)**, **correct (v)**, corrected (adj), **correction (n)**, correctional (adj), corrective (adj), corrective (n), **correctly (adv)**, correctness (n), **incorrect (adj)**, incorrectly (adv), uncorrected (adj)
EMBARRASS	**embarrass (v)**, **embarrassed (adj)**, **embarrassing (adj)**, embarrassingly (adv), **embarrassment (n)**
EQUAL	equal (jj), **equal (vv)**, equaled (vv), equaled (jj), equaling (vv), **equality (nn)**, equalization (nn), equalize (vv), equalized (vv), equalizer (nn), equalizers (nn), equalizes (vv), equalizing (vv), equalled (vv), **equally (rr)**, equals (vv), equals (nn), inequalities (nn), **inequality (nn)**, unequal (jj), unequalled (jj), unequally (rr)
EXCITE	excitable (adj), excitation (n), **excite (v)**, **excited (adj)**, excitedly (adv), **excitement (n)**, **exciting (adj)**, unexciting (adj)
EXPECT	expect (v), **expectancy (n)**, expectant (adj), expectantly (adv), expectation (n), **expected (adj)**, expectedly (adv), **unexpected (adj)**, **unexpectedly (adv)**, unexpectedness (n)

FRUSTRATE	**frustrate (v)**, **frustrated (adj)**, **frustrating (adj)**, frustratingly (adv), **frustration (n)**
HEAT	heat (n), **heat (v)**, **heated (adj)**, heatedly (adv), **heater (n)**, heating (n), **preheat (v)**, preheated (adj), reheat (v), reheated (adj), unheated (adj)
INFECT	**infect (v)**, **infected (adj)**, **infection (n)**, **infectious (adj)**, infectiously (adv), infectiousness (n), infective (adj), noninfectious (adj), reinfection (n), uninfected (adj)
INTENSE	**intense (adj)**, **intensely (adv)**, intensification (n), intensified (adj), **intensify (v)**, **intensity (n)**, **intensive (adj)**, intensively (adv)
NATION	nation (n), national (adj), national (n), **nationalism (n)**, **nationalist (adj)**, **nationalist (n)**, nationalistic (adj), nationalization (n), nationalize (v), nationalized (adj), **nationally (adv)**, nationhood (n), **nationwide (adj)**, **nationwide (adv)**
OPEN	open (adj), **open (adv)**, open (v), opened (adj), **opener (n)**, opening (n), **openly (adv)**, **openness (n)**, **reopen (v)**, reopened (adj), reopening (n), unopened (adj)
PRECEDE	**precede (v)**, precedence (n), **precedent (n)**, preceding (adj), **unprecedented (adj)**, unprecedentedly (adv)
PREDICT	predict (v), predictability (n), **predictable (adj)**, predictably (adv), predicted (adj), **prediction (n)**, predictive (adj), **predictor (n)**, unpredictability (n), **unpredictable (adj)**, unpredictably (adv)
REASON	reason (nn), **reason (vv)**, **reasonable (jj)**, reasonable (rr), reasonableness (nn), **reasonably (rr)**, reasoned (vv), reasoned (jj), **reasoning (nn)**, reasons (nn), reasons (vv), unreasonable (jj), unreasonableness (nn), unreasonably (rr), unreasoning (jj)
SELECT	**select (adj)**, select (v), selectable (adj), **selected (adj)**, **selection (n)**, **selective (adj)**, selectively (adv), selectivity (n), selector (n), unselected (adj)
SPECIAL	special (adj), special (n), specialisation (n), specialised (adj), specialism (n), **specialist (adj)**, specialist (n), speciality (n), specialization (n), **specialize (v)**, **specialized (adj)**, **specially (adv)**, **specialty (n)**
STRUCTURE	poststructuralism (n), poststructuralist (adj), restructure (v), restructured (adj), **restructuring (n)**, **structural (adj)**, structuralism (n), structuralist (adj), structuralist (n), structurally (adv), structure (n), **structure (v)**, **structured (adj)**, structuring (n), unstructured (adj)
VARY	invariable (adj), **invariably (adv)**, invariance (n), invariant (adj), unvaried (adj), unvarying (adj), **variability (n)**, **variable (adj)**, variable (n), variably (adv), **variance (n)**, **variant (n)**, variate (n), **variation (n)**, **varied (adj)**, **vary (v)**, **varying (adj)**
WIDE	wide (adj), **wide (adv)**, **widely (adv)**, **widen (v)**, widened (adj), wideness (n), widening (adj), **width (n)**

Appendix 2: Concordance sample (20 lines)

["	n't have to be very	**excited**	themselves . You know that	"]
['	is not significant requirement to	**accept**	at universities . They do	']
['	like restaurant which look more	**openly**	and lightly so I like	']
['	or just boring to go	**back**	home by walking , and	']
['	Many teachers (in different	**specialties**	) have tests inside their	']
['	. 3 - The librarian	**expected**	the children to be quiet	']
['	. Sometimes , people experience	**embarrassed**	situations with their friends and	']
['	the organization with the flat	**structure**	because there is no supervisory	']
['	time I wanted to go	**back**	my country because I could	']
['	are increasing in the developing	**nations**	. They can use their	']
['	, I am angry and	**frustrated**	. As an analyst specializing	']
['	situation or they have some	**special**	problems . So , scientist	']
['	take it on the same	**structure**	and the same level although	']
['	sense of smell . It	**intensifies**	the odors , and the	']
['	many areas . Some people	**predict**	that Obama can not gain	']
['	I think much about story	**construction**	, word choice . I	']
['	have technical knowledge , and	**structural**	capacity to produce medications .	']
['	was supposed to give something	**back**	but I just felt so	']
['	words , they use it	**incorrectly**	; for example , they	']
['	. But my mother feels	**frustrated**	because she believes that she	']

Appendix 3: Comparison of part-of-speech distributions for individual forms

	Lemma	POS	COCA			PELIC			Corpus Difference
			Lemma freq	Family freq	Lemma / Family	Lemma freq	Family freq	Lemma / Family	
1	collaboration	N	8047	18741	42.9%	5	345	1.5%	**+41.5%**
2	selection	N	17831	55168	32.3%	37	419	8.8%	**+23.5%**
3	collaborative	ADJ	4396	18741	23.5%	2	345	0.6%	**+22.9%**
4	confusion	N	8421	22769	37.0%	52	354	14.7%	**+22.3%**
5	adviser	N	9841	31140	31.6%	26	248	10.5%	**+21.1%**
6	intensity	N	9871	36618	27.0%	12	129	9.3%	**+17.7%**
7	frustration	N	8465	18430	45.9%	67	236	28.4%	**+17.5%**
8	excitement	N	8055	30078	26.8%	55	589	9.3%	**+17.4%**
9	embarrassment	N	3507	14444	24.3%	11	156	7.1%	**+17.2%**
10	collaborate	V	3702	18741	19.8%	11	345	3.2%	**+16.6%**
… ITEMS 11 – 56 …									
57	nationwide	ADV	4829	255800	1.9%	4	1141	0.4%	**+1.5%**
58	acceptance	N	9059	70616	12.8%	113	989	11.4%	**+1.4%**
59	structured	ADJ	2489	66851	3.7%	12	516	2.3%	**+1.4%**
60	infection	N	13794	23259	59.3%	84	145	57.9%	**+1.4%**
61	predictable	ADJ	3877	47594	8.2%	16	236	6.8%	**+1.4%**
62	unacceptable	ADJ	2551	70616	3.6%	23	989	2.3%	**+1.3%**
63	infect	V	4140	23259	17.8%	24	145	16.6%	**+1.3%**
64	back	V	16905	446279	3.8%	72	2772	2.6%	**+1.2%**
65	nationalist	ADJ	2829	255800	1.1%	0	1141	0.0%	**+1.1%**
66	reopen	V	2535	218034	1.2%	1	1128	0.1%	**+1.1%**
… ITEMS 67 – 113 …									
114	intensive	ADJ	4937	36618	13.5%	34	129	26.4%	**-12.9%**
115	excite	V	3875	30078	12.9%	153	589	26.0%	**-13.1%**
116	competitive	ADJ	14496	47349	30.6%	152	345	44.1%	**-13.4%**
117	advisor	N	3479	31140	11.2%	64	248	25.8%	**-14.6%**
118	varied	ADJ	3629	82128	4.4%	37	186	19.9%	**-15.5%**
119	frustrate	V	4292	18430	23.3%	97	236	41.1%	**-17.8%**
120	advise	V	12216	31140	39.2%	144	248	58.1%	**-18.8%**
121	confuse	V	7086	22769	31.1%	201	354	56.8%	**-25.7%**
122	embarrass	V	2026	14444	14.0%	66	156	42.3%	**-28.3%**
123	vary	V	15532	82128	18.9%	99	186	53.2%	**-34.3%**