Mind the gap: Towards determining which collocations to teach

Déogratias Nizonkiza

English Language Learning and Acquisition (ELLA), Douglas College, Canada | School of Languages, North-West University, Potchefstroom Campus, South Africa
E-mail: nizonkizad@douglascollege.ca; deo.nizonkiza@nwu.ac.za

Kris Van de Poel

University of Antwerp, Belgium | School of Languages, North-West University, Potchefstroom Campus, South Africa
E-mail: kris.vandepoel@uantwerpen.be

Abstract

Collocations form part of formulaic language use that is considered by many scholars as central to communication (Henriksen 2013; Wray 2002). Today, most scholars agree that teaching collocations to second and/or foreign language users (henceforth "L2 students") is a must. This study offers a reflection on the directions L2 researchers and teachers may explore, and that could contribute to modelling the teaching of collocations or at least spark the debate on this issue. The fundamental point raised here is the extent to which pedagogy may be informed by knowing the most common lexical collocations (combinations of content words) and using frequency of collocates as a key factor in selecting which collocations to bring to learners' attention. The results from this study indicate that out of the eight different lexical collocations, adjective+noun and verb+noun collocations are the most common, and should therefore be introduced first. Furthermore, most collocates ("co-occurring words" in Sinclair's (1991) terms) come from the 1,000 and 2,000 most frequent words. Therefore, this study suggests that the same way that "[u]sing the computational approach as a starting point makes it possible to distinguish between collocations of varying frequency of use" (Henriksen 2013: 32), frequency may be used to select the target words and their collocates once collocations have been identified. This could potentially contribute to addressing the issue of selection criteria of which collocations to teach.

Keywords: formulaic language, collocations, teaching collocations, word frequency.

1. Introduction

Formulaic language is considered by many scholars as central to language in use, and communication in particular (Henriksen 2013; Wray 2000, 2002), which is the reason why it has

attracted growing research interest over the past few decades (Barfield 2009; Barfield and Gyllstad 2009; Boers, Demecheleer, Coxhead and Webb 2014; Henriksen 2013). Wray (2000: 465), among others, notes that the concept has not been easy to define, and points to a plethora of terms used to refer to this phenomenon, including "formulaic language", "collocations", "chunks", "multiword units", "fixed expressions", "idioms", "preassembled speech", etc.

The lack of consensus on the definition of "formulaic language" (Gyllstad 2009; Nesselhauf 2005; Wray 2000, 2002, 2013), referred to by some scholars as "definitional heterogeneity" (Gyllstad 2007: 6; see also Eyckmans 2009; Revier 2009), can be attributed to the different angles from which the topic has been investigated (Eyckmans 2009; Nesselhauf 2005; Wray 2000, 2002, 2013). Henriksen (2013: 29) points to six main types of formulaic sequences: "idioms (*if life deals you with lemons make lemonade*), figurative expressions (*to freeze to the spot*), pragmatic formulas (*have a nice day*), discourse markers (*let me see now*), lexicalized sentence stems (*this means that…*), and collocations (*rough crossing, remotely clear*) […]".

Following Henriksen's examples above, collocations are the focus of this paper. Since it is beyond the scope of this paper to provide an overview of the different definitions and the ways in which research on collocations has evolved through time, it is sufficient to note here that this study retains the definition provided in the *Oxford Collocations Dictionary for Students of English* (McIntosh, Francis and Poole 2009), namely:

> the way words combine in a language to produce natural-sounding speech and writing. For example, in English you say *strong wind* but *heavy rain*. It would not be normal to say *heavy wind* and *strong rain*.
>
> (McIntosh et al. 2009: v)

The results from research on collocations among L2 students indicate that collocations are important to gain competence in a second language (Barfield and Gyllstad 2009; Granger and Meunier 2008; Lewis 1993, 1997, 2000; Meunier and Granger 2008; Nattinger and DeCarrico 1992; Nesselhauf 2005; Pawley and Syder 1983; Schmitt 1998, 2004; Wray 2000, 2002). Indeed, collocations characterise nativelike fluency (Boers, Eyckmans, Kappel, Stengers and Demecheleer 2006; Ellis 2001; Nation 2001; Pawley and Syder 1983; Schmitt 2004; Wray 2000, 2002) and overall proficiency for both comprehension (Eyckmans 2009; Gyllstad 2007, 2009; Keshavarz and Salimi 2007; Nizonkiza 2011a, 2015) and production (Boers et al. 2006; Bonk 2001; Eyckmans, Boers and Demecheleer 2004; Gitsaki 1999; Gledhill 2000; Granger 1998; Howarth 1998; Nizonkiza 2011b, 2012b, 2014; Paquot 2008). Collocations also contribute to signalling identity and reducing processing effort (Wray 2002).

For Ellis (2001), among others, formulaic language – collocations in particular – should be central to language learning. This observation echoes earlier calls to teach collocations based on the growing importance and significance attributed to them. As early as the 1990s, scholars such as Lewis (1993), Nattinger and DeCarrico (1992), and Willis (1990) made a strong argument that the very least that should be done when teaching in L2 contexts is to introduce the formulaic dimension of language to learners. This observation fits Palmer's (1933) definition that collocations are "successions of words [that] must or should be learnt as an integral whole or independent entity, rather than by the process of piecing together their component parts" (Palmer 1933: 4).

These calls were not responded to immediately, however, because collocations are semantically transparent and do not seem to cause any problem for comprehension. They therefore remain unnoticed as problematic by both teachers and learners (Biskup 1992; Gouverneur 2008; Henriksen and Stoehr 2009; Laufer and Waldman 2011; Paquot 2008). Another, more fundamental, reason why collocations have not been a subject of focus in the classroom is that some scholars believe that formulaic language is implicit in nature and thus should be taught implicitly (e.g. Boers and Lindstromberg 2008, 2009; Ellis 2001; Wray 2002). However, given the lack of sufficient exposure to the target language in L2 and foreign-language contexts, this approach may be called into question. Even for single words, scholars such as Nation (2001) recommend recycling the words through multiple exposures. This view is supported by Schmitt (2008) according to whom learners need more explicit reference to the target words even for receptive purposes. The same then can be expected of more encompassing formulaic language of which collocations form the main representatives (Alali and Schmitt 2012; Henriksen 2013; Henriksen and Stoehr 2009). Henriksen and Stoehr (2009: 227), for instance, contend that learners do not have enough exposure regarding both amount and range of input needed for developing nativelike collocational competence. They further argue that even in cases of sufficient input, learners would still be confronted with the challenge of selecting collocations on which to focus. In this line of thinking, Henriksen (2013) observed that mere exposure to collocations is not enough, and suggests adopting more explicit approaches to teaching them. Furthermore, some studies have demonstrated that collocations remain problematic even at an advanced level (Laufer and Waldman 2011; Nesselhauf 2005). Therefore, expecting exposure alone to help learners master collocations could be a utopian idea.

Today, most scholars agree that teaching collocations to L2 students is a must. The main questions that have yet to be addressed are how to teach collocations and exactly which ones to teach (e.g. Durrant and Schmitt 2010; Granger and Meunier 2008; Jones and Durrant 2010; Nizonkiza 2012a, 2017; Nizonkiza and Van de Poel 2014). Although there is still disagreement on whether to adopt explicit or implicit approaches to teaching collocations, a number of empirical studies indicate that teaching collocations explicitly through awareness-raising might go some way to making a difference in students' proficiency gain (e.g. Barfield and Gyllstad 2009; Webb and Kagimoto 2009, 2011; Ying and O'Neill 2009). Adopting an explicit approach to teaching collocations has recently gained the support of studies using corpora as a source for teaching collocations (e.g. Chan and Liou 2005; Sun and Wang 2003). The approach basically adopts awareness-raising of the phenomenon of collocations by involving students in using corpora to identify these collocations. Thanks to corpora, collocations can be identified from their real contexts of use, which may facilitate learning (Biber, Conrad and Reppen 1998).

For many scholars (e.g. Biber et al. 1998; Davies 2010; Hunston 2002), corpora represent an excellent source for teaching collocations. Collocations indeed constitute the aspect of linguistics that has most benefited from advances in corpus linguistics. Even with these advantages from corpora and the tendency to agree to teach collocations explicitly, research still has yet to determine the number of collocations that should be taught to students based on their needs and proficiency levels (e.g. Durrant and Schmitt 2010; Granger and Meunier 2008; Jones and Durrant 2010; Nizonkiza 2014, 2017; Nizonkiza and Van de Poel 2014, Nizonkiza, Van Dyk and Louw 2013). This is what Durrant and Schmitt (2010), paraphrasing Wray (2002: 207), refer to as the difficulty

> [...] to see how we could specify a level of collocational knowledge that would be properly 'on a par with' any particular level of grammatical/vocabulary knowledge. There is simply no baseline defining what level of collocational knowledge learners should achieve after a certain level of experience with a language.
>
> (Durrant and Schmitt 2010: 169)

Furthermore, Benson, Benson and Ilson's (2010) *BBI Combinatory Dictionary of English* indicates that collocations are of different types. These scholars have distinguished between seven types of lexical collocations and eight major types of grammatical collocations. For them, "[a] grammatical collocation is a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or clause" (Benson et al. 2010: xix). They contend that any native speaker of English can tell that *account for*, *accuse (somebody) of*, *adapt to*, and *agonize over* are correct combinations whereas **account over a loss*, **accuse somebody on a crime* and **adapt towards new conditions* are wrong. Lexical collocations, in contrast to grammatical collocations, normally do not contain prepositions, infinitives or clauses. Typical lexical collocations consist of nouns, adjectives, verbs and adverbs. An example of an adjective+noun collocation is *warmest regards*, as in *I send warmest regards*. Typical violations of lexical collocability are **I send hot regards* and **I send hearty regards* (Benson et al. 2010: xxxi).

Other collocations dictionaries, such as *Oxford Collocations Dictionary for Students of English* (McIntosh et al. 2009), basically use the same main entries although they do not make any distinction between grammatical and lexical collocations – noun, verb and adjective. McIntosh et al. (2009) suggest the following classification. Under noun entries, the possible collocations are:

(i)   adjective+noun (e.g. *bright/harsh/intensive/strong **light***),

(ii)   quantifier+noun (e.g. *a beam/ray of **light***),

(iii) verb+noun (e.g. *cast/emit/give/provide/shed **light***),

(iv)   noun+verb (e.g. ***light** gleams/grows/shines*),

(v)   noun+noun (e.g. *a **light** source*),

(vi)   preposition+noun (e.g. *by the **light** of the moon*), and

(vii) noun+preposition (e.g. *the **light** from the window*).

Verb entries have the following collocations:

(i)   adverb+verb (e.g. ***choose** carefully*),

(ii)   verb+verb (e.g. *be free to **choose***), and

(iii) verb+preposition (e.g. ***choose** between two things*).

As far as adjective entries are concerned, these include:

(i)   verb+adjective (e.g. *make/keep/declare sth **safe***),

(ii)   adverb+adjective (e.g. *perfectly/not entirely/environmentally **safe***), and

(iii) adjective+preposition (e.g. ***safe** from attack*).

As observed by scholars, the different types of collocations have not been investigated to the same extent, with only verb+noun collocations being the most widely investigated (Boers et al. 2014; Durrant and Schmitt 2010; Nizonkiza and Van de Poel 2014). This suggests that, up until now, research cannot tell which type of collocations is acquired first nor the number of collocations – not even for the type that appears to have been most extensively investigated, (verb+noun) – that should be taught at the different learning stages (Durrant and Schmitt 2010; Nizonkiza 2017; Nizonkiza and Van de Poel 2014; Nizonkiza, Van Dyk and Louw 2013). We seem to be at a turning point, with the role of collocations in L2 teaching and learning contexts being widely acknowledged, and the calls to explicitly teach collocations increasing in number (Boers et al. 2014; Howarth 1998; Lewis 1993, 1997, 2000; Nattinger and DeCarrico 1992; Willis 1990). Therefore, not knowing which collocations to teach at this point is a gap that needs to be bridged.

This study offers a reflection on the directions that researchers and L2 practitioners could take, and could help model the teaching of collocations or at least spark the debate on its modelling. The reflection is based on lexical collocations as defined and classified in the *Oxford Collocations Dictionary for Students of English* (McIntosh et al. 2009). The fundamental question raised here is the extent to which knowing the most common lexical words, their most common collocations, and the collocates' (co-occurring words) frequency could be used as key factors in selecting which collocations to bring to learners' attention, thereby possibly informing pedagogy. Frequency is indeed an important factor in the learnability of words (e.g. Milton 2009) as well as for the teaching of collocations (e.g. Granger 2011; Nizonkiza 2012a, 2017; Nizonkiza and Van de Poel 2014). Frequency may also be an important factor in the learning of collocations. Siyanova-Chanturia (2015) has revealed that even with no explicit teaching of collocations, Chinese learners of Italian improved their use of frequent and strong collocations over a period of just five months. This is a clear indication that learners notice and pay attention to frequent combinations, and actually go beyond single words (Siyanova-Chanturia 2015). Frequency is by no means the sole factor, but it is an important one, and could be useful in showing more common lexical words as well as the most frequent of their collocations.

The sample words guiding this reflection constitute an attempt to design a lexical syllabus based on words selected from the Academic Vocabulary List (AVL) developed by Gardner and Davies (2014). This selection was inspired by a course which one of the authors of this paper taught to graduate students from across disciplines at McGill University, Canada. One of the aims of the course was to introduce collocations to these students, and highlight the importance of collocations in building students' productive academic vocabulary. Students were presented with the major types of academic vocabulary of which the AVL is one. They were also introduced to sources of collocations, of which the main one is the Corpus of Contemporary American English (COCA)[1], along with its related corpus for academic vocabulary[2]. Another major source of collocations is an online collocations dictionary, the *Ozdic*[3]. However, based on observations and experiences shared by some of the students, it is the researchers' belief that students may be overwhelmed by the hundreds of collocations that are identified through these sources. Students are faced with a tremendous threshold as they themselves have to define the selection criteria. It is believed that "[u]sing the computational approach as a starting point makes it possible to

---

[1] Available at http://corpus.byu.edu/coca/.
[2] Available at http://www.wordandphrase.info/.
[3] Available at http://www.ozdic.com/.

distinguish between collocations of varying frequency of use" (Henriksen 2013: 32), and thus frequency may be used to select the target words and the collocates once collocations have been identified. This could potentially contribute to addressing the issue of defining selection criteria.

2.   Lexical syllabus sample based on the Academic Vocabulary List

2.1   Common lexical words from the Academic Vocabulary List

The AVL consists of a core academic vocabulary amounting to approximately 3,000 of the most frequent words in the 120 million words of the COCA academic texts. The AVL consists of lexical words, i.e. nouns, verbs, adjectives, and adverbs. For the purpose of this study, and in order to have a sense of the proportion of lexical words – classified according to their number – the top 100 most frequent words of the AVL were selected. The words were examined in terms of their lexical category and then classified accordingly. It was found that nouns account for 66% of the words from the AVL, while verbs account for 17%, and adjectives and adverbs account for 14% and 3%, respectively. This proportion may entail that nouns be given priority should we be teaching words from the AVL, for example. Furthermore, this could be the underlying reason why verb+noun collocations constitute the most investigated type of collocations (Henriksen 2013; Paquot and Granger 2012).

2.2   Common lexical collocations

The lexical words – for example, *system* (noun), *provide*, (verb), *social* (adjective), and *particularly* (adverb) – described in the previous section were used as key words for the collocations. However, only nouns, verbs, and adjectives were retained, and the six most frequent for each category were selected. Adverbs were not considered for the purpose of this study since collocation dictionaries do not have entries for them as key words. The nouns selected were *study*, *group*, *system*, *research*, *level*, and *result* while the selected verbs included *provide*, *suggest*, *require*, *report*, *describe*, and *indicate*. The adjectives selected were *important*, *low*, *significant*, *likely*, *similar*, and *common*. Once the words were selected, their lexical collocations were identified from the *Ozdic* online collocation dictionary. This dictionary basically uses the same entries as the *Oxford Collocations Dictionary for Students of English* (McIntosh et al. 2009) as well as the same classifications. Once the collocations were identified, they were manually counted. This means that the collocates as presented in the online collocation dictionary were counted (simple counting), and then a comparison was made between the different types of collocations for each target word. For example, nouns are collocated with adjectives (e.g. *large group*), verbs [e.g. verb+noun (*form group*), and noun+verb (*group split up*)], and other nouns (e.g. *group member*). The comparison aimed to determine which type is more common than others. Put differently, the comparison aimed to rank the collocates according to their number.

For example, for the noun *group* used above, all of the collocates on the list (e.g. adjectives such as *big*, *large*, *wide*, *coherent*, *cohesive*, *tight*, *minority*, *cultural*, *ethnic*, *racial*, *family*) were checked in terms of frequency using Lextutor. Lextutor is a vocabulary profiler (VP) linked to the British National Corpus (BNC), COCA, and other corpora, and classifies words in bands according to their frequency of occurrence. VP-Compleat is one of the software

packages[4] which classifies words into bands of 1,000 words each. It is helpful to identify word-frequency bands that could be used to support language development of L2 students from Grade 9 through to university[5]. The bands range from the most frequent 1,000 words (K1) to the twenty-fifth band, those very infrequent words (K25). The question worth raising here is how one should go about comparing the different collocates for each word entry. For example, a noun is collocated with different types of words, such as adjectives (adjective+noun), verbs (verb+noun and noun+verb), nouns (noun+noun), and prepositions (noun+preposition and preposition+noun). In particular, it is worth looking into the proportions of these types, for example, are there more verb+noun than adjective+noun collocations? Knowing which type has the highest number of occurrences, and thus the order in which the different types are ranked, might be important owing to the cost-benefit law (Laufer and Nation 1999); in other words, those that are more frequent might be more important, and thus spending more time on what is more important may make more sense.

As previously explained, the collocations of the selected words identified from the collocations dictionary were counted, and the results are presented below in Table 1 for nouns, Table 2 for verbs, and Table 3 for adjectives.

Table 1: Nouns collocated

Noun collocates	*Study*[6]	*Group*	*System*	*Research*	*Level*	*Result*	Tot	Average
Adjective+Noun	30	25	106	23	33	8	225	38
Verb+Noun	12	10	57	5	16	10	110	19
Noun+Verb	28	2	13	8	8	7	66	11
Noun+Noun	1	1	0	18	0	0	20	4

As can be seen from Table 1, lexical collocates of nouns include adjectives, verbs, and nouns (see first column). The next six columns present the nouns collocates in this order. Adjectives (e.g. *present study*) top the list of nouns collocates with on average 38 adjective collocates per noun (right-hand column). They are followed by verbs in the verb+noun combination (e.g. *conduct a study*) with about 19 verbs collocates on average. This is then followed by the verbs in the noun+verb combination (e.g. *a study examines sth*) which are in turn followed by noun+noun collocations (e.g. *study group*) with about 11 verb and four noun collocates, respectively.

Table 2: Verbs collocated

Verb collocates	*Provide*	*Suggest*	*Require*	*Report*	*Describe*	*Indicate*	Tot	Ave.
Adverbs	2	4	2	1	7	2	18	3
Verbs	10	4	0	8	5	3	30	5

As can be seen from Table 2, verbs are collocated with adverbs (e.g. *kindly provide*) and verbs (e.g. *be designed to provide*). On average, the number of collocates does not seem to differ greatly, with about three adverb and five verb collocates.

---

[4] VP-Kid and VP-Classic are the other two software packages used mainly for the same purpose of language development support for Grades 0 to 3 and Grades 4 to 8, respectively.

[5] Information retrieved on 26 August 2018 from https://www.lextutor.ca/vp/.

[6] For copyright reasons, we do not list all of the collocates here. We only select the noun *study* for illustration; all of the collocates for the noun *study* are listed in Appendix A.

Table 3: Adjectives collocated

Adjective collocates	*Important*	*Low*	*Significant*	*Likely*	*Similar*	*Common*	Tot	Average
Verbs	11	3	10	3	6	4	37	7
Adverbs	24	6	8	5	13	5	61	11

Table 3 shows that adjectives are collocated with verbs (e.g. *consider sth important*) and adverbs (e.g. *critically important*) with slightly more adverbs than verbs.

2.3   Teaching collocations: Role of collocates' frequency

In addition to ranking collocations in respect of their number per type (i.e. how many), the frequency of collocates could also help in determining which collocations to teach. The collocates can be weighed against frequency bands using Lextutor for collocates' individual frequency. This was done for the collocates of the selected words. The collocates were copied and pasted in the relevant Lextutor search window. Clicking on the "submit" button generates an output showing the frequency bands of each of the collocates and the relative percentages. The output also highlights words in colours, with each band having its own. The results are summarised in Table 4 below, and a detailed example for the noun *study* is presented in Appendix B.

Table 4: Collocates classified in frequency bands

Target words		Frequency bands of collocates in %								
		K1	K2	K3	K4	K5	K6	K7	K8	K9
Nouns	*Study*	65	13	16	5	2	0	0	0	0
	Group	69	19	6	2	0	2	2	0	0
	System	41	24	17	8	2	4	3	0	0
	Research	53	24	19	3	0	0	1	0	0
	Level	54	22	17	5	0	0	2	0	0
	Result	51	12	27	6	3	0	0	0	0
	Average	56	19	17	5	1	1.06	1	0	0
Verbs										
	Provide	73	23	3	0	0	0	0	0	0
	Suggest	67	20	0	0	7	0	7	0	0
	Require	50	0	50	0	0	0	0	0	0
	Report	91	9	0	0	0	0	0	0	0
	Describe	76	10	10	5	0	0	0	0	0
	Indicate	100	0	0	0	0	0	0	0	0
	Average	76	10	10	1	1	0	1	0	0
Adjectives										
	Important	64	20	14	2	0	0	0	0	0
	Low	90	10	0	0	0	0	0	0	0
	Significant	79	14	4	4	0	0	0	0	0
	Likely	89	11	0	0	0	0	0	0	0
	Similar	58	26	0	0	11	0	5	0	0
	Common	67	33	0	0	0	0	0	0	0
	Average	74	19	3	1	2	0	1	0	0

As can be seen from Table 4, the target words are presented in the first column in the order nouns, verbs, and adjectives. The other columns consist of the word bands from K1 to K9. K1 and K2 represent the frequent words while K3 to K9 represent mid-frequency words, and for each category (nouns, verbs, and adjectives), an average is presented. As the results show, most of the collocates come from K1 and K2. For the case of verb+noun collocations, a little over 55% of the verbs come from K1, while about 19% of them come from K2. About 17% of the collocates come from K3, with each of the other bands accounting for less than 5%. Collocates of verbs come from K1 and K2, about 76% and 10% respectively. This overall tendency is confirmed for adjectives collocates wherein K1 has 74% of the collocates, and K2 has 19%. Briefly, more than 70% of the nouns collocates come from K1 and K2; these bands account for over 80% for verbs collocates, and over 90% for the adjectives collocates. The teaching implications of this are discussed in the next section.

3. Discussion and conclusion

This study gives a reflection on the selection criteria of collocations to teach in L2 contexts by examining three lexical word categories in terms of their distribution, the way they are collocated, and the frequency of their collocates. It was the researchers' belief that these three factors, once uncovered, could contribute to the teaching of collocations in L2 contexts.

Exploring the lexical words in terms of their distribution indicates that nouns are by far the biggest category. This means that nouns constitute the most common lexical word category in English or at least top the list of the 100 most frequent lexical words from the AVL. While this is based on a small sample of words – the 100 most frequent words from the AVL – we believe that it is indicative of the most frequent lexical words in English. This implies that learners are likely to encounter nouns more often than other word categories, and thus, nouns should be given priority in teaching. While verb+noun collocations seem to have been the most widely investigated type because they convey the most important information (cf. Gyllstad 2007, 2009; Henriksen 2013), the fact that there are many nouns in English could be another reason.

Regarding lexical collocations, the sample words were analysed in terms of the eight lexical collocations presented in the *Oxford Collocations Dictionary for Students of English* (McIntosh et al. 2009) − adjective+noun, verb+noun, noun+verb, noun+noun, verb+verb, adverb+verb, verb+adjective, and adverb+adjective. As the results show, nouns prove to be the most collocated words as they account for about 50% of the lexical collocations (four of the eight lexical collocations), while adjective+noun and verb+noun collocations are the most common. This could be yet another reason why nouns, and especially verb+noun collocations, constitute the most widely investigated type of collocations (e.g. Gyllstad 2007, 2009; Henriksen 2013). Based on these results, we as teachers should concern ourselves first with nouns when it comes to the teaching of collocations, and then maybe introduce collocations in the order of distribution (i.e. those higher in number come first). However, high frequency should not be the only guiding principle. For example, even though adjectives outnumber verbs, the latter part in the verb+noun combination conveys the most important information (e.g. Gyllstad 2007) and should probably be introduced before or at least at the same time as adjective+noun collocations. This study concurs with Nizonkiza and Van de Poel's (2014) observation that adjective+noun and verb+noun collocations should be taught simultaneously. Adjective+noun collocations are the largest in number, while verb+noun are the most important, which implies that introducing both types as early as possible will benefit learning. Furthermore, in some cases, the order of the way

in which lexical words are collocated may change depending on individual words. For example, even though the results from this study indicate that nouns have more verbs as their collocates in the verb+noun combination than noun+noun collocations, the noun *research* has more noun+noun collocations (e.g. *research programme*). The noun+noun collocation is, however, the least frequent combination for the other nouns (ranked fourth). The same holds true for the noun *study*, which has more noun+verb collocations (e.g. *study reveal[s] sth*) than verb+noun collocations (e.g. *carry out a study*).

With the frequency of collocates, results from this study indicate that most collocates came from K1 and K2. This suggests that most of the collocates are likely to be known as individual words by learners. According to Gyllstad (2007), it would be pointless to include collocates which are less frequent than key words in a collocation test. The argument here is that learners cannot be expected to know the combinations without knowing the individual words. It would therefore make more sense to introduce more or equally frequent collocates together with the key words when teaching. Otherwise, less frequent collocates are likely to distract learners as the learners will want to focus on the individual words, the meanings of which they may not know, instead of focusing on the collocations' constituents. This finding that most collocates come from K1 and K2 is good news for teachers. However, it represents quite a challenge because, while learners do not seem to have difficulty with the comprehension thereof, these collocates are omitted in their production (e.g. Eyckmans 2009; Paquot 2008). It is the teachers' responsibility to raise learners' awareness, and this study suggests introducing collocations as early as possible, and possibly at the same time as the new word(s) are being taught, if K1 and K2 collocates are introduced before other collocates. What matters when teaching collocations is raising students' awareness of the expected combinations – collocations – by first identifying them and then suggesting activities that help students to notice them. This study supports research suggesting the explicit teaching of collocations through awareness-raising (e.g. Barfield 2009; Nizonkiza 2017; Ying and O'Neill 2009) and complements this body of literature by addressing the collocations' selection criteria.

The lexical syllabus discussed in this study is based on the AVL, and is therefore suitable for advanced students in higher education. However, we believe that the steps suggested in this study may be applied to different levels of proficiency. The main limitation along this line is that the study does not address the question of how many collocates learners should be exposed to. Based on Hill's (1999) observation that learners may be marked down because they do not know the top five collocates of a word, Nizonkiza (in press) requires students to select five collocates of target nouns. Even though participants in his study show a high level of satisfaction regarding their learning of collocations, there is no guarantee that these collocates are well mastered. Requiring students to master five collocates per target word in one semester may be too demanding and therefore counter-productive. We believe that the learning/teaching process of collocations must be gradual, but the question of how to implement it needs further research. Our suggestion is to trial the lexical syllabus with different numbers of collocates – for example, two, three, four, or five – which may give us insight into what students can actually manage. In our discussion above, we suggest introducing collocates at the same time as the new words as long as the collocates come from K1 or K2. However, some words have many K1 and K2 collocates, for example, the noun *system*, where just the first meaning ("set of ideas/rules for organizing sth[7]") has 95 adjectives collocates, with 31 and 26 of them coming from K1 and

---

[7] This information was accessed on 19 September 2018 from http://www.ozdic.com/collocation-dictionary/system.

K2, respectively. The main limitation of this study is that it does not offer any solution when it comes to the number of collocates. The topic suggested above for further research – trialling the lexical syllabus with different numbers of collocates – may give us insight into what students can actually manage and may contribute to addressing this limitation. Another limitation that we foresee is that the idea of checking the frequency of collocates may not appeal to teachers, let alone learners, because it is time consuming. Even with limitations and much more work still to be done, the results discussed in the paragraphs above are encouraging, and lay groundwork for the modelling of teaching collocations.

The present study has examined selection criteria for lexical collocations which could inform the teaching of these collocations. The study has shown that nouns constitute the largest lexical category and the most collocated word category, while most collocates come from K1 and K2 and are therefore also frequent. These results are important and contribute to the debate surrounding the question of modelling the teaching of collocations. Researchers and L2 practitioners have the daunting task of determining which collocations may be more useful than others, and which ones should be focused on first. To this end, many scholars, for example, Nizonkiza (2012b, 2014), Nizonkiza and Van de Poel (2014), Durrant and Schmitt (2010), Jones and Durrant (2010) and Wray (2002), among others, suggest establishing a baseline for collocation competence. The latter may prove useful when it comes to determining which collocations to teach at different proficiency levels. This study concurs with this recommendation, and argues in favour of a thorough assessment of the different types of collocations. Such an assessment seems to be warranted, and addressing the question of selection criteria is an important step in this direction.

References

Alali, F.A. and N. Schmitt. 2012. Teaching formulaic sequences: The same as or different from teaching single words? *TESOL* 3(2): 153–180. https://doi.org/10.1002/tesj.13

Barfield, A. 2009. Following individual L2 collocation development over time. In A. Barfield and H. Gyllstad (eds.) *Researching collocations in another language*. New York: Palgrave Macmillan. pp. 208–223. https://doi.org/10.1057/9780230245327_16

Barfield, A. and H. Gyllstad (eds.) 2009. *Researching collocations in another language*. New York: Palgrave Macmillan. https://doi.org/10.1057/9780230245327

Benson, M., E. Benson and R.F. Ilson. 2010. *The BBI combinatory dictionary of English*. Amsterdam: John Benjamins.

Biber, D., S. Conrad and R. Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511804489

Biskup, G.D. 1992. L1 influence on learners' rendering of English collocations: A Polish/German empirical study. In P.J.L. Arnaud and H. Béjoint (eds.) *Vocabulary and applied linguistics*. London: Macmillan. pp. 85–93. https://doi.org/10.1007/978-1-349-12396-4_8

Boers, F. and S. Lindstromberg (eds.) 2008. *Cognitive approaches to teaching vocabulary and phraseology*. Berlin: Mouton de Gruyter.

Boers, F. and S. Lindstromberg. 2009. *Optimizing a lexical approach to instructed language acquisition*. New York: Palgrave Macmillan. https://doi.org/10.1057/9780230245006_8

Boers, F., M. Demecheleer, A. Coxhead and S. Webb. 2014. Gauging the effects of exercises on verb–noun collocations. *TESOL* 18(1): 54–74. https://doi.org/10.1177/1362168813505389

Boers, F., J. Eyckmans, J. Kappel, H. Stengers and M. Demecheleer. 2006. Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. *Language Teaching Research* 10(3): 245–261. https://doi.org/10.1191/1362168806lr195oa

Bonk, W.J. 2001. Testing ESL learners' knowledge of collocations. In T. Hudson and J.D. Brown (eds.) *A focus on language test development: Expanding the language proficiency construct across a variety of tests*. Honolulu: University of Hawaii Second Language Teaching and Curriculum Center. pp. 113–142.

Chan, T.P. and H.C. Liou. 2005. Effects of web-based concordancing instruction on EFL students' learning of verb–noun collocations. *Computer Assisted Language Learning* 18(3): 231–251. https://doi.org/10.1080/09588220500185769

Davies, M. 2010. The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing* 25(4): 447–464. https://doi.org/10.1093/llc/fqq018

Durrant, P. and N. Schmitt. 2010. Adult learners' retention of collocations from exposure. *Second Language Research* 26(2): 163–188. https://doi.org/10.1177/0267658309349431

Ellis, N.C. 2001. Memory for language. In P. Robinson (ed.) *Cognition and second language instruction*. Cambridge: Cambridge University Press. pp. 33–68. https://doi.org/10.1017/cbo9781139524780.004

Eyckmans, J. 2009. Toward an assessment of learners' receptive and productive syntagmatic knowledge. In A. Barfield and H. Gyllstad (eds.) *Researching collocations in another language*. New York: Palgrave Macmillan. pp. 139–152. https://doi.org/10.1057/9780230245327_11

Eyckmans, J., F. Boers and M. Demecheleer. 2004. *The Deleted-Essentials Test: An effective and affective compromise*. Available online: www.hltmag.co.uk/nov04/mart04.htm (Accessed 22 May 2013).

Gardner, D. and M. Davies. 2014. A new Academic Vocabulary List. *Applied Linguistics* 35(3): 305–327. https://doi.org/10.1093/applin/amt015

Gitsaki, C. 1999. *Second language lexical acquisition: A study of the development of collocational knowledge*. San Francisco: International Scholars Publications.

Gledhill, C.J. 2000. *Collocations in science writing*. Tubingen: Gunter Narr Verlag.

Gouverneur, C. 2008. The phraseological patterns of high-frequency verbs in advanced English for general purposes: A corpus-driven approach to EFL textbook analysis. In F. Meunier and S. Granger (eds.) *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins. pp. 223–243. https://doi.org/10.1075/z.138.17gou

Granger, S. 1998. Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A.P. Cowie (ed.) *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press. pp. 145–160.

Granger, S. 2011. From phraseology to pedagogy: Challenges and prospects. In T. Herbst, P. Uhrig and S. Schller (eds.) *Chunks in the description of language. A tribute to John Sinclair*. Berlin and New York: Mouton de Gruyter. pp. 123–146. https://doi.org/10.1515/9783110257014.123

Granger, S. and F. Meunier (eds.) 2008. *Phraseology. An interdisciplinary perspective*. Amsterdam: John Benjamins.

Gyllstad, H. 2007. Testing English Collocations. Unpublished PhD dissertation. Lund: Lund University. Available online: http://lup.lub.lu.se/search/record/599011 (Accessed 22 September 2017).

Gyllstad, H. 2009. Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In A. Barfield and H. Gyllstad (eds.) *Researching collocations in another language*. New York: Palgrave Macmillan. pp. 153–170. https://doi.org/10.1057/9780230245327_12

Henriksen, B. 2013. Research on L2 learners' collocational competence and development – a progress report. *EUROSLA MONOGRAPHS SERIES* 2: 29–56.

Henriksen, B. and L.S. Stoehr. 2009. Processes in the development of L2 collocational knowledge: A challenge for language learners, researchers and teachers. In A. Barfield and H. Gyllstad (eds.) *Researching collocations in another language*. New York: Palgrave Macmillan. pp. 224–231. https://doi.org/10.1057/9780230245327_17

Hill, J. 1999. Collocational competence. *English Teaching Professional* 11: 3–6.

Howarth, P. 1998. Phraseology and second language proficiency. *Applied Linguistics* 19(1): 24–44. https://doi.org/10.1093/applin/19.1.24

Hunston, S. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Jones, M. and P. Durrant. 2010. What can a corpus tell us about vocabulary teaching materials? In A. O'Keeffe and M. McCarthy (eds.) *The Routledge handbook of corpus linguistics*. Cambridge: Cambridge University Press. pp. 387–400. https://doi.org/10.4324/9780203856949.ch28

Keshavarz, M.H. and H. Salimi. 2007. Collocational competence and cloze test performance: A study of Iranian EFL learners. *International Journal of Applied Linguistics* 17(1): 81–92. https://doi.org/10.1111/j.1473-4192.2007.00134.x

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

Laufer, B. and T. Waldman. 2011. Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning* 61(2): 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

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

Lewis, M. 1997. *Implementing the lexical approach: Putting theory into practice*. Hove: Language Teaching Publications.

Lewis, M. 2000. *Teaching collocations: Further development in the lexical approach*. Hove: Language Teaching Publications.

McIntosh, C., B. Francis and R. Poole (eds.) 2009. *Oxford collocations dictionary for students of English*. Oxford: Oxford University Press.

Meunier, F. and S. Granger (eds.) 2008. *Phraseology in foreign language learning and teaching.* Amsterdam: John Benjamins. https://doi.org/10.1017/s0272263109090500

Milton, J. 2009. *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Nation, I.S.P. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nattinger, J.R. and J.S. DeCarrico. 1992. *Lexical phrases and language teaching*. Oxford: Oxford University Press. https://doi.org/10.1017/s0272263100013000

Nesselhauf, N. 2005. *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Nizonkiza, D. 2011a. The relationship between lexical competence, collocational competence, and second language proficiency. *English Text Construction* 4(1): 113–146. https://doi.org/10.1075/etc.4.1.06niz

Nizonkiza, D. 2011b. The relationship between controlled productive collocational competence and L2 proficiency. *TTWiA* 84/85: 29–37. https://doi.org/10.1075/ttwia.84-85.04niz

Nizonkiza, D. 2012a. The relationship between lexical competence, collocational competence and second language proficiency. Unpublished PhD thesis, University of Antwerp.

Nizonkiza, D. 2012b. Quantifying controlled productive knowledge of collocations across proficiency and word frequency levels. *Studies in Second Language Learning and Teaching* 2(1): 67–92. https://doi.org/10.14746/ssllt.2012.2.1.4

Nizonkiza, D. 2014. The relationship between productive knowledge of collocations and academic literacy among tertiary level learners. *Journal for Language Teaching* 48(1): 149–171. https://doi.org/10.4314/jlt.v48i1.8

Nizonkiza, D. 2015. Measuring receptive collocational competence across proficiency levels. *Stellenbosch Papers in Linguistics Plus* 44: 125–146. https://doi.org/10.5774/44-0-186

Nizonkiza, D. 2017. Improving academic literacy by teaching collocations. *Stellenbosch Papers in Linguistics* 47: 153–179. https://doi.org/10.5774/47-0-267

Nizonkiza, D. (in press). Students' perceptions of corpora and online tools in teaching collocations to improve academic written communication. In A. Tsedryk and C. Doe (eds.) *Words: Description, acquisition, and pedagogy*. Cambridge: Cambridge University Press.

Nizonkiza, D. and K. Van de Poel. 2014. Teachability of collocations: The role of word frequency counts. *Southern African Linguistics and Applied Language Studies* 32(1): 301–316. https://doi.org/10.2989/16073614.2014.997061

Nizonkiza, D., T. Van Dyk and H. Louw. 2013. First-year university students' productive knowledge of collocations. *Stellenbosch Papers in Linguistics* 42: 165–181. https://doi.org/10.5842/42-0-143

Palmer, H. 1933. *Second interim report on English collocations*. Tokyo: Kaitakusha.

Paquot, M. 2008. Exemplification in learner writing: A cross-linguistic perspective. In S. Granger and F. Meunier (eds.) *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins. pp. 101–119. https://doi.org/10.1075/z.138.09paq

Paquot, M. and S. Granger. 2012. Formulaic language in learner corpora. *Annual Review of Applied Linguistics* 32: 130–149. https://doi.org/10.1017/s0267190512000098

Pawley, A. and F.H. Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J.C. Richards and R.W. Schmidt (eds.) *Language and communication*. London: Longman. pp. 191–227.

Revier, R.L. 2009. Evaluating a new test of 'whole' English collocations. In A. Barfield and H. Gyllstad (eds.) *Researching collocations in another language*. New York: Palgrave Macmillan. pp. 125–152. https://doi.org/10.1057/9780230245327_10

Schmitt, N. 1998. Measuring collocational knowledge: Key issues and an experimental assessment procedure. *ITL Review of Applied Linguistics* 119-120: 27–47. https://doi.org/10.1075/itl.119-120.03sch

Schmitt, N. (ed.) 2004. *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins.

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. https://doi.org/10.1016/j.system.2015.07.003

Sun, Y.C. and L.Y. Wang. 2003. Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning* 16(1): 83–94. https://doi.org/10.1076/call.16.1.83.15528

Webb, S. and E. Kagimoto. 2009. The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly* 43(1): 55–75. https://doi.org/10.1002/j.1545-7249.2009.tb00227.x

Webb, S. and E. Kagimoto. 2011. Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics* 32(3): 259–276. https://doi.org/10.1093/applin/amq051

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

Wray, A. 2000. Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics* 21(4): 463–489. https://doi.org/10.1093/applin/21.4.463

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

Wray, A. 2013. Formulaic language. *Language Teaching* 46(3): 316–334.

Ying, Y. and M. O'Neill. 2009. Collocation learning through an 'AWARE' approach: Learner perspectives and learning process. In A. Barfield and H. Gyllstad (eds.) *Researching collocations in another language*. New York: Palgrave Macmillan. pp. 181–193. https://doi.org/10.1057/9780230245327_14

Appendix A: Example of nouns collocates

Target words		Collocates			
		Adjectives	**Verbs (V+N)**	**Verbs (N+V)**	**Nouns**
Nouns	*study*	full-time, part-time, graduate, postgraduate, independent, private, academic, language, further, business, management, religious, women's, new, present, recent, earlier, previous, careful, close, comprehensive, detailed, intensive, serious, major, definitive, initial, pilot, preliminary, research, empirical, experimental, systematic, theoretical, anthropological, historical, scientific, sociological, field, feasibility, literary, case, comparative	take up, undertake, continue, pursue, complete, finish, resume, return to, do, lecture in, teach, commission, fund, support, carry out, conduct, make, work on, publish	take place, aim at/to, be aimed at sth, attempt to, be designed to, set out to, be based on sth, cover sth, concern sth, deal with sth, examine sth, focus on sth, investigate sth, look at sth, compare sth, conclude sth, demonstrate sth, find sth, indicate sth, report sth, reveal sth, say sth, show sth, suggest sth, warn sth, provide sth, highlight sth, identify sth,	group, leave, skills

Appendix B: Collocates categorized in frequency bands

Frequency bands	Collocates of the noun study			
	Adjectives	**Verbs (V+N)**	**Verbs (N+V)**	**Nouns**
K1	business, careful, case close, earlier, field, full-time, further, historical, major, management, new, part-time, present, recent, scientific, serious, systematic, women's	carry out, complete, continue, do, finish, make, return, support take up, teach, work on	take place, aim at/to, be aimed at sth, attempt to, be designed to, set out to, be based on sth, cover sth, concern sth, deal with sth, examine sth, focus on sth, investigate sth, look at sth, compare sth, conclude sth, demonstrate sth, find sth, indicate sth, report sth, reveal sth, say sth, show sth, suggest sth, warn sth, provide sth, highlight sth,	group, leave
K2	detailed, intensive, language previous, private, research	fund	identify sth	skills
K3	academic, comprehensive, experimental, graduate, independent, initial, literary, pilot, religious, theoretical	commission, conduct, lecture, publish, pursue, resume, undertake		
K4	anthropological, comparative, empirical, preliminary, sociological			
K5	feasibility			
K6	-			
K7	-			
K8	postgraduate			
K9	-			