

This item is the archived peer-reviewed author-version of:

Teachability of collocations : the role of word frequency counts

Reference:

Nizonkiza Déogratias, Van de Poel Kris.- Teachability of collocations : the role of word frequency counts
Southern African Linguistics & Applied Language Studies - ISSN 1607-3614 - 32:2(2015), p. 301-316
Full text (Publishers DOI): <http://dx.doi.org/doi:10.2989/16073614.2014.997061>

Teachability of collocations: The role of word frequency counts

Abstract

This study aims to gain insights in which collocations to teach at which proficiency levels by answering the following questions: (i) Of the Verb+Noun and Adjective+Noun types of collocations which one is more difficult (and should therefore come at a later learning stage) and (ii) What is the size of collocations of the types Verb+Noun and Adjective+Noun mastered across proficiency levels? Besides, the paper suggests an approach that should be trialled in English as a Foreign Language teaching context. Two collocation tests (modelled after Laufer & Nation, 1999) with words selected from Nation's (2006) word list and Coxhead's (2000) Academic Word List were administered to two groups of English majors at different learning stages from the University of Antwerp and the University of Burundi. Results show that Adjective+Noun combinations are more difficult than Verb+Noun combinations. However, in both collocations types, only upper-intermediate and low-advanced students master the 2000-word band. We therefore suggest teaching collocations from the 2000-word band to beginners/low-intermediates and first exceed the 2000-word band from upper-intermediate onwards, a suggestion in line with Nation's (2006) on how to teach vocabulary¹.

Key words: collocation size, controlled productive knowledge, teachability of collocations, word frequency counts, English as a Foreign Language (EFL)

1. Introduction

Over the past few years, collocations in a second or foreign language teaching and learning context have been the centre of attention of foreign language practitioners (Barfield & Gyllstad, 2009). The main topics addressed can be grouped roughly into four categories, i.e. the importance of collocations in a foreign or second language, the relevance of teaching collocations, the way in which to teach collocations, and the selection of collocations to teach. While the importance of collocations (see among others Pawley & Syder, 1983; Cowie, 1998; Howarth, 1998; Wray, 2002) and the relevance of teaching them (see among others Nattinger & DeCaricco, 1992; Lewis, 1993, 1997, 2000; Nesselhauf, 2005; Boers, Eyckmans, Kappel, Stengers & Demecheleer, 2006; Ozaki, 2011) have been addressed to a great extent, it is still not clear how collocations should be taught and exactly which ones should be taught (Granger & Meunier, 2008; Ozaki, 2011).

The importance of collocations was addressed as early as the 1920s (Gyllstad, 2007), but growing interest into the topic dates back to the 1980s with researchers such as Pawley and Syder (1983) convincingly demonstrating that collocations play an important role in characterising overall proficiency (see also, among others, Howarth, 1998; Cowie, 1998; Granger, 1998; Gitsaki, 1999; Bonk, 2001; Wray, 2002; Nesselhauf, 2005; Gyllstad, 2007, 2009; Keshavarz & Salimi, 2007; Nizonkiza, 2011). Following this observation, many foreign language practitioners unanimously agreed that collocations should be taught explicitly and systematically (see among others Nattinger & DeCaricco, 1992; Lewis, 1993, 1997, 2000; Wei, 2000; Martynska, 2004; Nesselhauf, 2005; Boers et al., 2006; Ozaki, 2011). This call has been responded to with empirical studies on the ways in which to teach collocations and the criteria for selecting which collocations to teach.

The pedagogical experiments on teaching collocations conducted so far suggest teaching methods that are promising for the future of teaching foreign languages; much of what is available being on English as a foreign language (EFL). Notable examples include awareness-raising and attention-drawing techniques, which are basically similar in nature. Both approaches consist of strategies aimed to bring collocations to the learners' attention with the expectation that learners who are aware of collocations will better use them. These approaches have been trialled in many contexts using different techniques and involving EFL learners from completely different backgrounds (cf. Wray & Fitzpatrick, 2008; Barfield, 2009; Jiang, 2009; Ying & O'Neill, 2009). The conclusions arrived at indeed indicate that the more learners are sensitised to collocations, the better they use them.

However, the question of selecting collocations to teach seems to us equally important and it has not yet been properly addressed. One of the major attempts to find out which collocations to teach is a web-based project commonly referred to in the literature as 'collocator' (Wible, Kuo, Chen, Tsao & Hung, 2006). The collocator highlights the collocations from a page being viewed and is intended to help learners and teachers know which collocations to attend to. However, it highlights many collocations which may require a further selection for teaching and learning purposes, but the question which remains unanswered is on which basis. Collocation dictionaries that list words and their collocates with examples provided may be relied on when teaching collocations. However, like the collocator, they do not tell us what collocations to teach at which learning stages, leaving both teachers and learners confronted

with the selection criteria. Even the approaches to teaching collocations mentioned above do not seem to have laid down strict criteria for selecting collocations to present to learners. Furthermore, very little, if anything at all, is known about the size of collocations needed for both language comprehension and production.

In addition to the selection criteria that have remained undefined so far, collocations are of different types. According to Benson, Benson and Ilson (2010), there are eight types of grammatical collocations and seven types of lexical collocations. Although it was not the main focus of her study, Nesselhauf (2005), who conducted a study on German EFL students at tertiary level, has shown that verb-noun (V+N) combinations are more difficult than adjective-noun (Adj+N) combinations. This would mean that if we teach a foreign language with an explicit focus on collocations, Adj+N combinations should be introduced earlier than V+N combinations. However, not only have her findings not been validated by other studies, but the corpus-driven approach she adopted also represents only one type of collocational competence. She investigated free productive knowledge by analysing EFL students' essays compiled in a corpus of about 154,191 words, from which she manually extracted combinations to be examined.

The present study will try to complement Nesselhauf and track controlled productive knowledge of collocations of the types V+N and Adj+N combinations along proficiency levels among Belgian and Burundian EFL students, taking into account word frequency, which is a 'determining factor in words' learnability' (Milton, 2007, 2009; Granger in press), with the intention to determine which collocations to focus on at which learning stage. Furthermore, we aim to suggest a teaching approach to adopt and discuss its practicalities. To this end, the following research questions will be investigated: (i) Of the two types of collocations, V+N and Adj+N combinations, which one is more difficult than the other and (ii) What is the size of collocations of the types V+N and Adj+N mastered across proficiency levels?

2. Definition of collocations

Even though defining collocations has not been an easy task (Wray, 2002), researchers tend to agree on the traditions under which collocations have been investigated. According to Nesselhauf (2005), Gyllstad (2007), and Granger and Paquot (2008) among others,

collocations have been researched under three main traditions. First, we have what is known in the literature as the frequency based tradition. Its characteristic feature is the frequency of co-occurrence of collocations' constituents. Secondly, collocations have been investigated under the phraseological tradition, a tradition characterised by the varying degrees of fixedness and substitutability of collocations' constituents. The third tradition is conciliatory in nature and retains elements from the other two traditions, which is the reason why it is referred to by some scholars as 'the best of the two worlds' (Gyllstad, 2007; Granger & Paquot, 2008). From the perspective of syntactic categories of collocations constituents, collocations can be further grouped in two categories, i.e. grammatical and lexical collocations (Benson et al., 2010), which can in their turn be divided in subcategories.

Grammatical collocations consist of a content word such as a verb, an adjective, or a noun plus a preposition or a grammatical structure such as an infinitive or a clause. Examples of grammatical collocations include: *account for*, *to be afraid that*, etc. Lexical collocations consist of nouns, adjectives, verbs, and adverbs with the following seven possible combinations:

1. Verb + noun: *take/sit an exam*, *mark an exam*, etc;
2. Noun + verb: *the exam begins ...*, *as soon as the exams are over*, ... etc;
3. Noun + noun: *exam format*, *exam time*, etc;
4. Adverb + adjective: *perfectly aware*, *fully aware*, etc;
5. Adjective + noun: *difficult exam*, *final exam*, etc;
6. Verb + adverb: *place gently*, *strategically placed*, etc;
7. Verb + expression with preposition: *run out of money*, *burst into tears*, etc.

For comparative reasons explained in section one, the present study will investigate two types of lexical collocations, i.e. verb + noun and adjective + noun that are referred to in this paper as V+N and Adj+N, respectively.

3. Modelling the teaching of collocations: Evidence from size estimates

3.1 Instruments: The collocation tests

In order to test the research questions this study purports to answer, two collocation tests were developed and presented to participants. The first test consists of V+N combinations and was developed and used in an earlier study (Nizonkiza, 2011). As explained in Nizonkiza (2011),

the stimulus words are nouns and were selected according to their frequency at the 2000-word, 3000-word, and 5000-word levels from Nation's (2006) word frequency countⁱⁱ and the Academic Word List (Coxhead, 2000). Ten words were selected from each of the levels (cf. Nation & Beglar, 2007), making a total of 40 items. The word selection followed a systematic random sampling (Babbie, 1990), i.e. from a random starting point, each 100th word was selected and in case the latter was not a noun, the next one was selected. Once the target words were selected, their collocates (verbs) were selected from the *Oxford Collocations Dictionary for Students of English* (2002). All the verbs collocating with the noun in the V+N combination were listed. Then, their frequency level was checked in Nation's (2006) frequency list, retaining the verbs of higher frequency level than the target word (noun). If no verbs of higher frequency were found, verbs belonging to the same frequency band were selected (cf. Gyllstad, 2007).

Finally, in order to retain stronger collocations, an online 'collocation sampler'ⁱⁱⁱ was used. The latter lists the 100 most co-occurring words (collocates) of the node and their respective distribution (i.e. how many times they appear) in the Bank of English. It especially indicates the strength of their co-occurrence, i.e. how often they co-occur with the node and how significantly they do so. The most significant collocate (with the highest z-score) was retained. The words were embedded in a sentential context with sentences selected from the *Oxford Collocations Dictionary for Students of English* (2002). In each sentence, the collocate was deleted, but in order to avoid other possible words (cf. Laufer & Nation, 1999), the two first letters were provided. They were underlined and test-takers were instructed to fill in the missing letters (see Appendix A). They were awarded 1 point per correct answer and 0 point per wrong or no answer.

The second test consists of Adj+N combinations and was basically developed along the same lines with the same nouns retained as target words for easy comparison. Both tests measure controlled productive knowledge that refers to,

the ability to use a word when compelled to do so by a teacher or researcher, whether in an unconstrained context such as a sentence writing task, or in a constrained context such as a fill in task where a sentence context is provided and the missing target word has to be supplied (Laufer & Nation, 1999: 37).

3.2 Participants

The present study involved English majors from Burundi and Belgium. The Burundian participants use Kirundi as their native language (L1), and French as official language. They also have some basic knowledge of Kiswahili, the lingua franca of the East African region. They were recruited from the Department of English Language and Literature of the University of Burundi (N = 92, from Year 1 and Year 4) and their age range was 20-26. They sat the tests on two different days. They had sat TOEFL and these scores^{iv} were used to allocate them to proficiency levels that were mapped onto the Common European Framework of Reference (CEFR) proficiency scale using the equivalence table developed by the Vancouver Language Centre^v. They were found to belong to three proficiency groups, i.e. upper-beginners (score: 347 - 393), low-intermediate (score: 397 - 473), and upper-intermediate (score: 477 - 547), respectively referred to as A2, B1, and B2 in the CEFR terms.

The Belgian participants (N = 117) were English majors from the University of Antwerp. They speak Dutch as their L1 and were aged between 18 and 20. They were first year students and participated in this study as part of a writing class. Their proficiency level was determined through their DIALANG^{vi} scores with four proficiency levels distinguished, i.e. low-intermediate (B1), upper-intermediate (B2), low-advanced (C1), and upper-advanced (C2). When taking both groups together, the following five proficiency levels were distinguished; which are summarised in Table 1.

[Table 1 should be here]

As Table 1 shows, the five proficiency levels are: A2 (N = 69), B1 (N = 33), B2 (N = 47), C1 (N = 54), and C2 (N = 6). In both groups of participants, the tests were presented in the order V+N and Adj+N combinations in one session. Each test lasted for 15 to 30 minutes among Burundians and 5 to 15 minutes among Belgians, which can be explained by the difference in proficiency between the two groups. Even though DIALANG is a low stake and diagnostic test (Alderson, 2000), the levels it identifies can be mapped onto those set by the TOEFL not for the purpose of establishing cut scores, but for interpretation of the TOEFL scores in relation to the CEFR, the levels of which are measured by DIALANG^{vii}.

4. Results

4.1 Adj+N collocations more difficult than V+N collocations

One of the research questions investigated in the study is which of the V+N and Adj+N types of collocations poses more problems to learners. This was tested by running a paired sample t-test involving scores on the two types of collocations^{viii}. Both total scores as well as scores at each frequency band were compared; the results of which are summarised in Table 2.

[Table 2. should be here.]

As can be seen in Table 2, the results indicate that the Adj+N type of collocations was more difficult than the V+N collocations. The means are 22.61 and 29.21 for Adj+N and V+N, respectively; and the difference is statistically significant, with a Sig. (2-tailed) of .000. Comparisons at the level of frequency bands also indicate that the scores are much higher on V+N than on Adj+N collocations with 8.34 vs. 7.83 at the 2000-word; 8.37 vs. 4.86 at the 3000-word; 7.06 vs. 6.22 at the AWL; and 5.39 vs. 3.68 at the 5000-word. The differences are statistically significant everywhere with a Sig. (2-tailed) of .000. These findings answer the first research question and indicate that Adj+N collocations are more difficult than V+N collocations, which contradicts earlier findings (Nesselhauf, 2005).

4.2 Collocation size estimates

The second question addressed in this study is how we can make size estimates of collocations alongside proficiency level. It was achieved at each proficiency level using Laufer and Ravenhorst-Kalovski's (2010) formula^{ix} - $Tot\ score \times 5000 \div Maximum\ score$ - and results are presented in Table 3 for V+N collocations and Table 4 for Adj+N collocations. In both tables, the first column consists of word frequency bands with the 1000-word and 4000-word bands added and the AWL excluded as Laufer's (1998) formula suggests^x. The maximum score in this study will therefore be 50, i.e. 10 at each of the five frequency bands. The next columns consist of scores per proficiency level, levels presented in the first row, where we have upper-beginners (A2), low-intermediate (B1), upper-intermediate (B2), and low-advanced (C1); N represents the number of students allocated to each proficiency level^{xi}.

[Table 3 should be here.]

As can be seen from the last row of Table 3, the means were converted into size estimates at the 5000-word families. The size of V+N collocations likely to be known is 3173 for upper-beginners; 3298 for low-intermediate; 4068 for upper-intermediate; and 4336 for low-advanced. Upper-beginner and low-intermediate learners seem to have reached the same size of 3000-word families while the size of upper-intermediate and low-advanced learners seems to have grown to the 4000-word level. These figures were obtained by applying Laufer and Ravenhorst-Kalovski's (2010) formula above ($Tot\ score \times 5000 \div Maximum\ score$), which, at the upper-beginner level for instance, can be numerically expressed as: $31.73 (7 + 7 + 7.81 + 5.91 + 4.01) \times 5000 \div 50 = 3173$.

However, weighing the means against Schmitt's cut-off point shows different results. According to Schmitt (2003) quoted in Xing and Fulcher (2007), the expected score at an acquired word frequency level should be 80%^{xii}, which means 8 out of 10 in the present case at each frequency band. The scores in Table 3 show that upper-beginners (A2) and low-intermediate (B1) fall short of the 2000-word band. Only upper-intermediate learners (B2) master the 2000-word and the 3000-word bands; no change for low-advanced (C1) students who seem to have achieved mastery of collocations at the different word frequency bands except the 5000-word level.

As done for the V+N collocations, the means of Adj+N collocations were converted into size estimates.

[Table 4 should be here.]

As can be seen from Table 4 (last row), students have achieved a size of 2290 for upper-beginners; 2677 for low-intermediate; 3042 for upper-intermediate; and 3415 for low-advanced. Upper-beginners and low-intermediate students fall short of the 3000-word families, a size reached by upper-intermediate as well as low-advanced learners. The Adj+N scores were also weighed against Schmitt's cut-off point, which shows that only students at upper-intermediate (B2), and low-advanced (C1) levels master the 2000-word band.

5. Discussion

The present study attempted to explore the extent to which ranking collocations in terms of difficulty and making size estimates of collocations known at different proficiency levels can

contribute towards the teaching of collocations. To this end, two collocation tests were developed and presented to Belgian and Burundian English majors (EFL students).

The first question sets out to determine which of the two types of collocations, V+N or Adj+N collocations, considered in this study poses more problems. Scores from both types of collocations were compared and Adj+N were found to be more difficult than V+N collocations, contradicting earlier findings (Nesselhauf, 2005). A possible explanation of the conflicting results is the nature of the collocations investigated. While Nesselhauf (2005) analysed students' essays and therefore studied free productive knowledge of collocations, the present study tested controlled productive knowledge of collocations, which represents a different level of knowing words. Furthermore, Nesselhauf studied adjectives used in the V+N collocations and observed that V+N collocations do not always allow adjectives that otherwise modify the noun and constitute acceptable Adj+N collocations. She gives the example of the collocation *take advantage* that accepts *full* and *unfair*, but not *considerable*, even though *considerable advantage* is a very common collocation. Even though Adj+N collocations were found to be more difficult than V+N collocations, the 2000-word band is mastered in both types of collocations by upper-intermediate (B2) and low-advanced (C1) EFL learners. We therefore do not make any distinction between types of collocations to be introduced earlier than others in teaching but suggest focusing on the 2000-word band as described below.

Making size estimates of collocations known at different proficiency levels is the second question addressed in this study. Results indicate that only upper-intermediate (B2) and low-advanced (C1) learners master the 2000-word band (following Schmitt's cut-off point). Considering the size estimates of the present study, we could be tempted to say that all the participants have reached the 2000-word band. However, free productive knowledge always lags behind controlled productive knowledge (Laufer, 1998) and even though our findings indicate that they have achieved the 2000-word, it certainly is the case that learners will not use all collocations from this word band without any prompt. This implies that the size of what they can actually use freely and productively is much lower than what they can use in a controlled (and experimental) context. Furthermore, taking up Nation's (1990) suggestion that in order for L2 learners to lexically function well without any help, they must have a control of the 3000-word band, we believe that the size reached by our participants is not enough for 'good' use of the foreign language. We therefore join in the call to explicitly teach

collocations and support the view that collocations should be introduced as early as possible. On the basis of these findings, we propose to focus on the 2000-word band at the beginner (A1 and A2) and low-intermediate (B1) levels and only go beyond the 2000-word from upper-intermediate onwards. This is in line with Nation's (2006) suggestion as regards teaching vocabulary that the 2000-word level should be explicitly taught while the other vocabulary levels can simply be taught through reading. Furthermore, it makes sense to select the collocations to teach from frequent words up to the 2000-word level in terms of 'cost-benefit' as this level is the cut-off point for frequency (Schmitt et al., 2001) and covers more than 84% of a running text *no matter what kind of text is being focused on* (Nation & Chung, 2009: 545). The crucial role played by frequency in teaching words is emphasized in Granger (in press: 9) according to whom, *far too much time is wasted on words and phrases that are not even worth bringing to learners' attention for receptive purposes, let alone for productive purposes.*

6. Teaching implications

In light of the discussion above, we suggest developing a collocation teaching and learning syllabus, using two types of sources, a collocation dictionary, for instance the *Oxford Collocations Dictionary for Students of English* (2002), from which to select the collocations to teach; and a frequency word count such as Nation's (2006) in which the frequency of both target words and their collocates should be checked. We advise syllabus designers in particular and teachers alike to depart from existing material (texts) and going about teaching words in a different manner by observing the seven steps below. We believe that the proposed steps will match Nation's (2001) psychological conditions which should be met when teaching vocabulary in general, i.e. noticing, retrieving, and generation; which constitute the theoretical foundation of this approach. The collocation web format, proposed in step six (below), allows learners to notice the collocations constituents/directions/syntactic categories, etc.; recreating and creating the word webs in their minds as Handl (2009) suggests, helps retrieving the collocations, while the exercises proposed in step seven contributes to generation.

Below are the proposed seven steps:

1. Selecting a target noun from a text;

2. Checking its frequency in a word frequency count;
3. Consulting a collocation dictionary and listing all the collocates of the target word following:
 - a. their meaning;
 - b. their lexical class;
 - c. the direction of the collocation;
4. Checking the frequency level of the collocates against a word frequency count;
5. Deciding on which collocates to retain depending on the learning level; we suggest retaining collocates of higher or similar frequency at beginner and low-intermediate levels;
6. Choosing the appropriate format for presenting the words to the learners (see point 6 below; we suggest adopting a collocation web format); and
7. Selecting exercises in order to reinforce the learning of collocations.

The above steps can be illustrated by means of an example taken from a Burundian coursebook used at upper-beginner level (Bureau d'Etudes et Programmes de l'Enseignement Secondaire^{xiii}, 1998). In order to teach the word 'competition' used in Unit One 'Sport Day in Burundi', teachers are instructed in the teacher's guide to explain the word to the students and to provide a definition. All they can do further is to invite learners to use the word in a sentence, but they never make reference to words with which it collocates; an approach that we consider ineffective.

1. The word 'competition' is selected from the text;
2. Its frequency is checked in Nation's (2006) word frequency count. Competition belongs to the 1000-word level;
3. The *Oxford Collocations Dictionary for Students of English* (2002) is consulted and shows that the word:
 - a. Is a noun that means:
 - i. Event in which people try to win something;
 - ii. Trying to achieve the same thing/gain advantage;
 - b. Competition, meaning (i) collocates with the following words:
 - i. In Adj+N combination: *international, national; knock-out; dancing, piano, sporting etc.*

- ii. In V+N combination: *win; lose; have, hold, stage; enter, take part in; withdraw from;*
 - iii. In N+V combination: *take place; be open to sb;*
 - iv. In Prep+N combination: *in a/the; between; for;*
- c. Competition, meaning (ii) collocates with the following words:
- i. In Adj+N combination: *cut-throat, fierce, intense, keen, serious, severe, stiff, strong, tough; fair, free, healthy, open; domestic; foreign, global, international, overseas; economic;*
 - ii. In V+N combination: *be up against, face; go into; beat off; fight off;*
 - iii. In Prep+N combination: *against; in the face of;*
 - iv. In Prep+N+Prep combination: *in...with;*
 - v. In N+Prep combination: *among, between; for, from;*
- d. In all the cases, the syntactic category of the collocates shows the direction of the collocation;
4. The frequency band of each collocate is checked in Nation's (2006) word frequency count. The following words: *national, dancing, piano, win, lose, have, hold, stage, enter, knock (knock-out), take (take part in and take place), be (be open to sb)*, belong to the 1000-word level; *international* belongs to the 2000-word level; while *sporting* and *withdraw (withdraw from)* belong to the 3000-word level;
5. The collocates *dancing, piano, win, lose, have, hold, stage, enter, knock-out, take part in, take place, be open to somebody*, that belong to the 1000-word level; and *international* that belongs to the 2000-word level, will be retained;
6. The words are put in a word web format that clusters the collocates of the same lexical class in the same box and following their meaning as exemplified in Figure 1 (see Appendix C).

This format is an adapted version of McCarthy and O'Dell's (2005) collocation web model. The original model suggests putting the target word at the centre of a word web with lines that connect it to its collocates that (each) have to be put in a box that contains a definition of the word (collocate) learners have to supply. We believe that finding the appropriate word may be challenging for EFL learners, who may understand very well the definitions provided, but still fail to find the relevant collocates. Furthermore, showing the direction of the collocation (collocation to the right or to the left of the target word), for instance 'win a competition' vs.

‘a competition takes place’... is what Handl (2009) refers to as collocational webs which allow learners to create and recreate such webs in their minds.

7. The collocations are presented to students in the form of exercises with the collocates deleted; but the first two letters may be provided.

This type of exercise is in line with McCarthy and O’Dell’s (2005) collocation web model, with, however, a different approach. For instance McCarthy and O’Dell (2005: 35) give an exercise on collocations of ‘countryside’ where three collocations have to be supplied. An example is given, i.e. ‘surrounding’ which is put in a box in which the following definition is provided: ‘adjective meaning “lying around it”’. The collocates that learners have to supply are ‘unspoilt’, ‘destroy’, and ‘tranquil’, the definitions of which are provided, which are: ‘adjective meaning “beautiful because it has not been changed or damaged by people”’, ‘verb meaning “damage something so badly that it no longer exists or cannot be used”’, adjective meaning “peaceful, quiet”, respectively. While for instance learners can go for ‘ravage’ or ‘spoil’ instead of ‘destroy’ for the verb, and which are correct collocations of ‘countryside’, they can also easily go for ‘extinguish’, ‘break down’, ‘dismantle’, etc, all of which are its synonyms (synonyms found at <http://dictionary.reference.com/browse/destroy?s=t>).

We advise embedding the target word in a sentential context, deleting the collocate, but with/or without the first two letters provided, the same way Laufer and Nation (1999) suggests testing controlled productive knowledge of vocabulary. This is teaching through cued recall, somewhat productive in nature, which is likely to result in more positive outcomes. In this vein, Nation and Chung (2009: 546) state that teaching multi-word units through production has four major effects, stressing that they have their receptive equivalents:

1. *Learners will be able to produce grammatically correct utterances;*
2. *Learners will be able to produce utterances that are native-like;*
3. *Learners will be able to produce utterances fluently;*
4. *Learners will be able to communicate very early in their language learning.*

Teaching collocations productively has an added value as it has been established that collocational knowledge is more important at the productive level (Schmitt, 1998; Howarth, 1998; Bonk, 2001; Boers et al., 2006; Nation, 2006) and most collocation errors occur in

production (Eyckmans, 2009). Activities intended to lead to more and better engagement of the learners, and therefore to improving learning as Schmitt (2010) recommends to syllabus developers, teachers, and learners, should be well thought-out and presented at this stage.

7. Conclusion

As discussed above, results of the present study suggest that Adj+N collocations are more difficult than V+N collocations and that only upper-intermediate (B2) and low-advanced (C1) learners master the 2000-word band, which answers the questions addressed in the study. At the same time the present study raises questions that are worth exploring in follow-up studies.

First of all, the present study considered only two types of collocations, i.e. V+N and Adj+N collocations. A follow-up study involving other types of collocations like V+N and N+V may contribute towards gaining more insights in the collocations that cause more problems and how much is known at the different word-bands.

Secondly, even though the size of collocations known at different proficiency levels was estimated, the participating students did not spread across all the proficiency levels from beginners up to the most advanced; we therefore propose to replicate the study for all proficiency levels and we hope that this will bring more detailed results which will advance our understanding of which collocations are needed at every learning stage.

Finally, we propose to explore the different teaching and learning phases discussed above in a more experimental study, for which course materials will be developed based on the 2000-word level. They will be presented to beginners and low-intermediate learners in a pre- and post- experimental design in order to gain insights in the effect of the exposure, which may help determine the exact role of explicit teaching of collocations and the effect on the teachability of collocations taking into account both proficiency and word frequency counts.

All in all, we believe that the present study provides basic groundwork for modelling the teaching of collocations, especially using what word frequency counts can offer today. In this perspective, our study complements previous studies that have investigated ‘how’ to teach collocations by bringing in approaches to what collocations to teach at which learning stages, which is our modest contribution to the field.

References

- Alderson JC. 2000.** Technology in testing: The present and the future. *System* 2000 (28): 593-603.
- Babbie ER. 1990.** *Survey Research Methods*. California and Belmont: Wordsworth Company.
- Barfield A. 2009.** Following individual L2 collocation development over time. In Barfield A & Gyllstad H (eds) *Researching Collocations in Another Language*. New York: Palgrave Macmillan, pp 208-223.
- Barfield A & Gyllstad H. (eds) 2009.** *Researching Collocations in Another Language*. New York: Palgrave Macmillan.
- Benson M, Benson E & Ilson R. 2010.** *The BBI Combinatory Dictionary of English: Your Guide to Collocations and Grammar*. Amsterdam: John Benjamins Company.
- Boers F, Eyckmans J, Kappel J, Stengers S & Demecheleer M. 2006.** Formulaic sequences and perceived oral proficiency: Putting the Lexical Approach to the test. *Language Teaching Research* 10(3): 245-261.
- Bonk WJ. 2001.** Testing ESL learners' knowledge of collocations. In Hudson T & Brown JD (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.
- Cowie AP. (ed) 1998.** *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press.
- Coxhead A. 2000.** A new academic word list. *TESOL Quarterly* 34(2): 213-239.
- Coxhead A. 2008.** Phraseology and English for academic purposes: Challenges and opportunities. In Granger S & Meunier F (eds) *Phraseology. An Interdisciplinary Perspective*. Amsterdam: John Benjamins Publishing Company, pp 149-161.
- Ebel RL. 1979.** *Essentials of Education Measurement*. New Jersey: Prentice Hall.
- Eyckmans J. 2009.** Toward an assessment of learners' receptive and productive syntagmatic knowledge. In Barfield A & Gyllstad H (eds) *Researching Collocations in Another Language*. New York: Palgrave Macmillan, pp 139-152.
- Gitsaki C. 1999.** *Second Language Lexical Acquisition: A Study of the Development of Collocational Knowledge*. San Francisco: International Scholars Publications.
- Granger S. 1998.** Prefabricated patterns in advanced EFL writing: Collocations and formulae. In Cowie AP (ed) *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press, pp 145-160.

- Granger S. (in press).** From phraseology to pedagogy: Challenges and prospects. In Uhrig P & Schüller S (eds) *The Phraseological View of Language. A Tribute to John Sinclair*. Berlin and New York: Mouton de Gruyter.
- Granger S & Meunier F. (eds) 2008.** *Phraseology. An Interdisciplinary Perspective*. Amsterdam: John Benjamins Publishing Company.
- Granger S & Paquot M. 2008.** Disentangling the phraseological web. In Granger S & Meunier F (eds) *Phraseology. An Interdisciplinary Perspective*. Amsterdam: John Benjamins Publishing Company, pp 27-49.
- Gyllstad H. 2007.** Testing English collocations. Unpublished PhD Dissertation, Lund University.
- Gyllstad H. 2009.** Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In Barfield A & Gyllstad H (eds) *Researching Collocations in Another Language*. New York: Palgrave Macmillan, pp 153-170.
- Handl S. 2009.** Towards collocational webs for presenting collocations in learners' Dictionaries. In Barfield A & Gyllstad H (eds) *Researching Collocations in Another Language*. New York: Palgrave Macmillan, pp 69-85.
- Howarth P. 1998.** Phraseology of second language proficiency. *Applied Linguistics* 19(1): 24-44.
- Jiang J. 2009.** Designing pedagogic materials to improve awareness and productive use of L2 collocations. In Barfield A & Gyllstad H (eds) *Researching Collocations in Another Language*. New York: Palgrave Macmillan, pp 99-113.
- Keshavarz MH & Salimi H. 2007.** Collocational competence and cloze test performance: A study of Iranian EFL learners. *International Journal of Applied Linguistics* 17(1): 81-92.
- Laufer B. 1998.** The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics* 19(2): 255-271.
- Laufer B & Nation P. 1999.** A vocabulary size test of controlled productive ability. *Language Testing* 16(1): 33-51.
- Laufer B & Paribakht TS. 1998.** The relationship between passive and active vocabularies: Effects of language learning contexts. *Language Learning* 48(3): 365-391.
- Laufer B & Ravenhorst-Kalovski GC. 2010.** Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22(1): 15-30.
- 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.
- Martynska M. 2004.** Do English language learners know collocations? *Investigationes Linguisticae* 11: 2-12.
- McCarthy M & O'Dell F. 2005.** *English Collocations in Use*. Cambridge: Cambridge University Press.
- Melka TF. 1997.** Receptive versus productive vocabulary. In Schmitt N & McCarthy M (eds) *Vocabulary: Description, Acquisition, and Pedagogy*. New York: Cambridge University Press, pp 84-102.
- Milton J. 2007.** Lexical profiles, learning styles and the construct validity of lexical size tests. In Daller H, Milton J & Treffers-Daller J (eds) *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press, pp 47-58.
- Milton J. 2009.** *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Nation P. 1990.** *Teaching and Learning Vocabulary*. New York: Heinle and Heinle.
- Nation P. 2001.** *Learning Vocabulary in Another Language*. New York: Cambridge University Press.
- Nation P. 2006.** How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63(1): 59-82.
- Nation P & Beglar D. 2007.** A vocabulary size test. *The Language Teacher* 31(7): 9-13.
- Nation P & Chung T. 2009.** Teaching and testing vocabulary. In Long MH & Doughty C (eds) *The Handbook of Language Teaching*. Malden: Wiley-Blackwell, pp 543-559.
- Nattinger JR & DeCarrico JS. 1992.** *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nesselhauf N. 2005.** *Collocations in a Learner Corpus*. Amsterdam: John Benjamins Publishing Company.
- Nizonkiza D. 2011.** The relationship between controlled productive collocational competence and L2 proficiency. *TTWiA* 84/85: 29-37.
- Oxford Collocations Dictionary for Students of English*, 2002.
- Ozaki S. 2011.** Teaching collocations effectively with the aid of L1. *The Language Teacher*

- 35(3). Available at <jalt-publications.org/tlt> (accessed 5 June 2012).
- Pallant J. 2007.** *SPSS Survival Manual*. Buckingham and Philadelphia: Open University Press.
- Pawley A & Syder FH. 1983.** Two puzzles for linguistic theory: Native like selection and native like fluency. In Richards JC & Schmidt RW (eds) *Language and Communication*. London: Longman, pp 191-226.
- Peters E. 2009.** Learning collocations through attention-drawing techniques: A qualitative and quantitative analysis. In Barfield A & Gyllstad H (eds) *Researching Collocations in Another Language*. New York: Palgrave Macmillan, pp 194-207.
- Schmitt N. 1998.** Measuring collocational knowledge: Key issues and an experimental assessment procedure. *ITL* 119-120: 27-47.
- Schmitt N. 2010.** *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke and New York: Palgrave Macmillan.
- Schmitt N, Schmitt D & Clapham C. 2001.** Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing* 18(1): 55-88.
- Wei Y. 1999.** Teaching collocations for productive vocabulary development. Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (New York, March 9-13, 1999).
- Wible D. 2008.** Multiword expressions and the digital turn. In Granger S & Meunier F (eds) *Phraseology. An Interdisciplinary Perspective*. Amsterdam: John Benjamins Publishing Company, pp 163-181.
- Wible D, Kuo CH, Chen MC, Tsao NL & Hung TF. 2006.** A computational approach to the discovery and representation of lexical chunks. In Mertens P, Fairon C, Dister A & Watrin P (eds) *Verbum ex machina. Actes de la 13e Conférence sur le Traitement Automatique des Langues Naturelles*. Louvain-la-Neuve: Presses universitaires de Louvain, pp 868-875.
- Wray A. 2002.** *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray A & Fitzpatrick T. 2008.** Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In Granger S & Meunier F (eds) *Phraseology. An Interdisciplinary Perspective*. Amsterdam: John Benjamins Publishing Company, pp 123-147.
- Xing P & Fulcher G. 2007.** Reliability assessment for two versions of Vocabulary Levels Tests. *System* 35(2007): 182-191.

Ying Y & O'Neill M. 2009. Collocation learning through an 'AWARE' approach: Learner perspectives and learning process. In Barfield A & Gyllstad H (eds) *Researching Collocations in Another Language*. New York: Palgrave Macmillan, pp 181-193.

Accompanying Tables

Proficiency levels	N
A2	69
B1	33
B2	47
C1	54
C2	6

Table 1. Participants' Proficiency Levels

Type of Collocations	2000-word	3000-word	AWL	5000-word	Total
V+N	8.34	8.37	7.06	5.39	29.21
Adj+N	7.83	4.86	6.22	3.68	22.61
Sig. (2-tailed)	.000	.000	.000	.000	.000

Table 2. Scores at Different Word Levels Paired

Word bands	A2, N=69	B1, N=33	B2, N=47	C1, N=54
1000	7	7.33	9.30	9.65
2000	7	7.33	9.30	9.65
3000	7.81	7.36	8.85	9.17
4000	5.91	6.11	7.36	8.02
5000	4.01	4.85	5.87	6.87
Total	31.73	32.98	40.68	43.36
Size	3173	3298	4068	4336

Table 3. V+N Scores across Word and Proficiency Levels

Word bands	A2, N=69	B1, N=33	B2, N=47	C1, N=54
1000	6.53	7.82	8.19	9.03
2000	6.53	7.82	8.19	9.03
3000	3.85	4.36	5.32	5.90
4000	3.28	3.71	4.68	5.36
5000	2.71	3.06	4.04	4.83
Total	22.9	26.77	30.42	34.15
Size	2290	2677	3042	3415

Table 4. Adj+N Scores across Word and Proficiency Levels

Appendix A: V+N Collocation Test



Productive Vocabulary Test

Name:

Level of study (year):

University:

Date:

Start hour:

End hour:

Instruction: Complete the underlined words in the sentences below.

Example: She is conducting campaigns to at..... new clients.

She is conducting campaigns to attract new clients.

1. I ha..... no intention of changing jobs because I am happy where I am.
2. Enemy planes were seen dr..... bombs along the railway line.
3. They always pa..... a 10% commission on every sold encyclopaedia.
4. I wonder, this unusual building seems to barely fi..... the definition of a house.
5. Better sa..... your energy not trying to persuade people who are not interested.
6. She asked him if he could ke..... a secret before telling him the horrible story.
7. Great care is being taken to en..... the accuracy of research data with good planning, several revisions and rewrites as part of the procedure.
8. She felt she would ma..... a terrible mess of her life if she were to throw everything overboard now.
9. They did not ge..... the permit for a street demonstration against university fees they had applied for a couple of months ago.
10. Her appointment will fi..... the gap created when the marketing manager left.
11. They held celebrations to ma..... the anniversary of Mozart's death.
12. It is common practice that when a song ends, the performer has to ta..... a bow.
13. They plan to se..... congratulations to Tony on his new job and bought a nice card.
14. We could he..... a faint echo, before it slowly died away.
15. Victory will br..... glory, fame, and riches to the football team.
16. She inherited all the family precious stones, but she does not like to we..... jewellery.

17. In May and June, females leave the males to bu..... a nest and incubate their eggs.
18. She joined the navy where she expects to re..... the rank of captain before retiring.
19. He is a person who can se..... his soul to the devil provided he gets money.
20. Why didn't the referee bl..... the whistle just before he shot the goal; it would have prevented the clash between rival supporters.
21. When she got pregnant at the age of 16, she decided to ha..... an abortion.
22. The estate expects to ho..... an auction to raise money.
23. Our party should en..... diversity, not division, in order to attract new members.
24. How do you ex..... the discrepancies between the money and the receipts?
25. Jumbo jets somehow la..... the glamour of the transatlantic liner which has an impact on the number of passengers.
26. She had a short time to dress and ap..... lipstick before rushing out to the party.
27. The burglars had to br..... a pane of the front window to enter the house.
28. He vowed to ta..... revenge on the man who had killed his brother.
29. They have decided to ch..... the catwalk stereotype of the skinny model.
30. They called on the government to help pro..... native wildlife as a response to the major environmental concerns of the century.
31. She was hoping she would not have to gi..... evidence in court.
32. I can't re..... any conclusions from their vague observations.
33. She had to pa..... some compensation for the damages she had caused.
34. With the new computer, you can ha..... access to all the files.
35. The mechanic can ma..... the necessary adjustments to the broken engine.
36. Many universities in the UK ch..... special fees to overseas students.
37. His sound argument will la..... the foundations for future cooperation between the two countries.
38. We have to fo..... the safety guidelines laid down by the government.
39. It is the duty of the local community to pr..... accommodation for the homeless.
40. He was found to su..... from clinical depression after several months of hospitalisation.

Appendix B: Adj+N Collocation Test



Productive Vocabulary Test

Name:
Level of study (year):
University:

Date:
Start hour:
End hour:

Instruction: Complete the underlined words in the sentences below.

Example: The company needs to focus on ne..... clients.
The company needs to focus on new clients.

1. My or..... intention was to study all morning, but I had to change my plans.
2. India started to build a nu..... bomb.
3. That bank charges a hi..... commission for cashing travellers' cheques.
4. The author provides a cl..... definition of the term culture.
5. There was a lack of cr..... energy in the British film industry.
6. The film stars were married in gr..... secret to avoid publicity.
7. The needle has to be positioned with ab..... accuracy.
8. They left the kitchen in a to..... mess after dinner.
9. A wo..... permit is required for foreign students to get a job.
10. His death left a bi..... gap in my life.
11. He bought her a ring on their fi..... wedding anniversary.
12. After the performance, he gave a li..... bow and left the room.
13. May I offer my wa..... congratulations on your promotion.
14. There are st..... echoes of Elvis Presley in his vocal style.
15. They are driven by a craving for pe..... glory.
16. She inherited some go..... jewellery from her mother.
17. The bi..... nest is mainly made by the females.
18. He held of..... rank in the air force for many years.
19. Missionaries saw it as their task to save in..... souls.
20. They scored their only goal just before the fi..... whistle.
21. Women's groups are calling for free contraception and le..... abortion.

22. He regularly attended ch..... auctions.
23. Our party believes in preserving cu..... diversity in the city.
24. Ma..... discrepancies were found between the written and electronic records.
25. Her long dark hair lent her unr..... glamour.
26. She put on br..... lipstick and rushed out for the party.
27. The burglars got in the house by smashing a wi..... pane.
28. Killing the man who had killed his brother was judged by many as at..... revenge.
29. He is a writer who rejects se..... stereotypes.
30. They called on to the government to help preserve the end..... wildlife.
31. The witness was expected to provide cl..... evidence in court.
32. The ob..... conclusion from their observations is that nobody knew what was going on.
33. She got fi..... compensation for the damages caused.
34. I lived deep in the country without ea..... access to shops.
35. The garage owner required sm..... adjustments to the broken engine.
36. The club's me..... fee went up by 3% last year.
37. With this agreement, we now have a so..... foundation to build on.
38. The same ge..... guidelines on when to dress formally apply to both men and women.
39. Many old people choose to live in pri..... accommodation.
40. When his wife left him, he fell into de..... depression.

Appendix C: Collocation Web Model

Format: Collocation Web Model

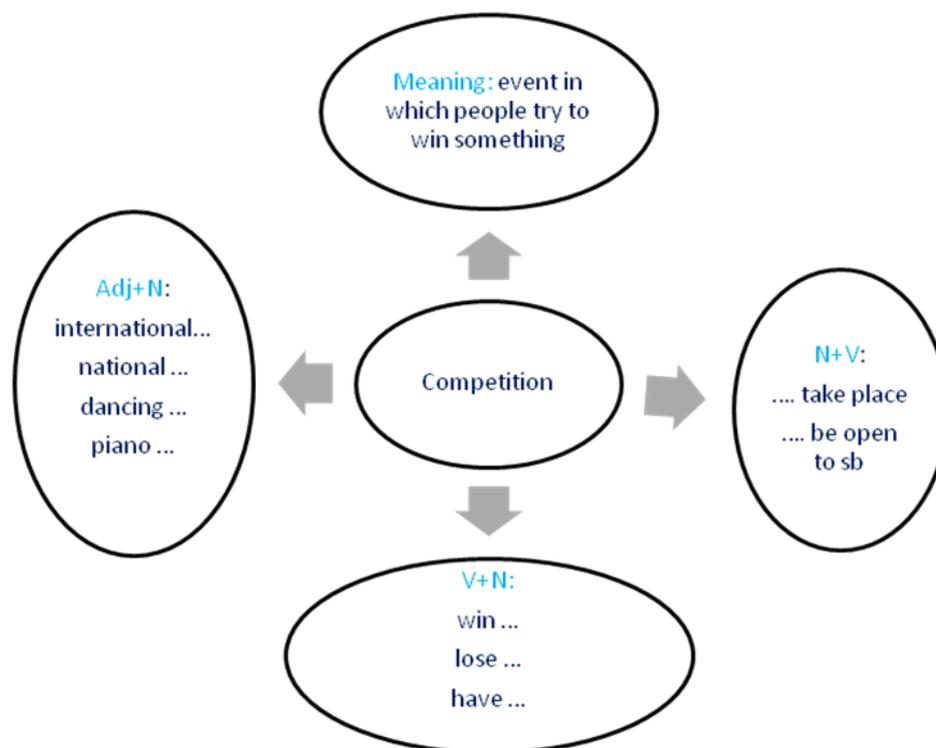


Figure 1. Collocation Web Model

ⁱ Acknowledgement

We sincerely thank Prof. Johann Van der Walt from the North West University for his comments and advice on the manuscript of the paper. We also thank Ms Hua Zhong from the University of Sydney for her advice on data processing. Our sincere thanks are also due to two anonymous reviewers for their constructive comments that helped improve the paper.

ⁱⁱ Nation (2006) is a database of word families based on the British National Corpus (BNC) and organised in word frequency bands of 1000 words each.

ⁱⁱⁱ The collocation sampler is available online at: <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>. (Accessed 21 September 2009).

^{iv} Paper based TOEFL total scores range between 310 and 677.

^v The conversion table is available at: <http://secure.vec.bc.ca/toefl-equivalency-table.cfm>. (Accessed on 15 November 2011).

^{vi} DIALANG is a web-based project that uses CEFR scale and thus tests proficiency at six proficiency levels, i.e. C2, C1 (proficient users), B2, B1 (independent users), and A2, A1 (basic users). More information can be obtained from the following website:

<http://www.lancs.ac.uk/researchenterprise/dialang/about>. (Accessed 12 October 2010.)

^{vii} Information retrieved from [http://www.britishcouncil-ieltsforusa.com/sg-en/compare IELTS and TOEFL.html](http://www.britishcouncil-ieltsforusa.com/sg-en/compare_IELTS_and_TOEFL.html). (Accessed on 25 November 2011). The same information was found in an interim report mapping TOEFL scores onto the CEFR scales. It was retrieved from http://www.ets.org/Media/Research/pdf/CEF_Mapping_Study_Interim_Report.pdf. (Accessed on 25 November 2011).

^{viii} Although not reported here, Cronbach's Alpha was calculated for both tests and is as high as .856 and .828 for V+N and Adj+N respectively, which is acceptable (Pallant, 2007). The items were also weighed against Ebel's (1979) scale and found to function well with 89.37% for V+N and 76.91% for Adj+N.

^{ix} Laufer and Ravenhorst-Kalovski (2010: 21) state that, "...we filled in the missing 4,000 level by averaging the scores received on the 3rd 1,000 and 5th 1,000. The score at each frequency level represents an approximate

knowledge of 1,000 words, except the first 2,000; where the score represents knowledge of 2,000 words. If, for example a learner received 28 on the second 1,000, 22 on the third, and 8 on the fifth, his score would be $28+28+22+15+8=101$. (The figure 28 appears twice as it represents 2,000 words, (i.e. two frequency levels), while the other scores represent 1,000 words each. The figure 15 is the average of 22 and 8. Since each frequency level has 30 items, the maximum score, which represents knowledge of 5,000 words, would be $30 \times 5 = 150$. The score in our example would represent $101 \times 5,000 / 150 = 3,366$ word families”.

^x Laufer (1998: 269) uphold that: “The UWL is not a separate frequency level, but consists mainly of words at 4th and 5th level. Therefore it is added to the 5,000 sample”. The UWL stands for the University Word List, today’s AWL.

^{xi} C2 was excluded from the analysis because it consisted of 6 students only.

^{xii} This information is Schmitt’s (personal communication), found in Xing and Fulcher (2007: 184) according to whom Schmitt’s cut-off point is 80% and actually the best we have even though it is not from a published source.

^{xiii} Elaborating programmes in secondary schools in Burundi is centralised, which is coordinated by the ‘Bureau of Studies and Programmes for Secondary School Studies (our translation).