

Multiword Sequences and Language Learning Pedagogy: Bridging the Research-Practice Gap

Abdullah A. Alasmary¹

¹ Department of English Language & Translation, Collegel of Languages and Translation, King Saud University, Riyadh, Saudi Arabia

Correspondence: Abdullah A. Alasmary, King Saud Universty, Riyadh, Saudi Arabia.

Received: October 8, 2018 Accepted: December 24, 2018 Online Published: December 26, 2018

doi: 10.5539/elt.v12n1p204 URL: <https://doi.org/10.5539/elt.v12n1p204>

Abstract

Strong claims are made regarding the significant role that multiword sequences play in the comprehension and production of speech and writing. Although the literature is replete with research-informed, evidence-based guidelines on how to maximize the learning of such sequences, such guidelines need to be synthesized, analyzed and evaluated for learning purposes. This paper attempts to fill this gap, addressing the numerous instructional options that ESL/EFL practitioners, curriculum designers and materials authors have at their disposal while dealing with lexically bundled sequences in several learning contexts. Another purpose of this paper, is to explain the problems that may arise as a result of teaching these sequences and the ways to solve them. The bulk of discussion will be centered on the tasks, activities and techniques suggested by researchers for deepening learners' understanding of these patterns across a wide range of contexts.

Keywords: multiword sequences, lexical bundles, recurrent patterns, language pedagogy

1. Introduction

The past four decades have seen a growing interest in the study of language vocabulary beyond the word level (e.g., Nattinger & DeCarrico, 1992; Wood, 2015; Wray, 2000). While known by different terms, including formulaic sequences (Ellis, Simpson-Vlach, & Maynard, 2008), lexical bundles (Biber & Barbieri, 2007), and multi-word constructions (Liu, 2012), the phenomenon where a recurrent string of lexical items behaves like a single lexical unit has been the focus of intense research activity spanning fields such as Applied Linguistics (e.g., Wood, 2015), English for specific /Academic Purposes (e.g., Biber, Conrad, & Cortes, 2004), corpus linguistics (Flowerdew, 2015), and Pragmatics (Bardovi-Harlig, 2012). The research in this area appears to benefit greatly from advances in computer technology and software programming, both of which allow researchers from distinct epistemological and conceptual backgrounds to explore different aspects of a recurrent unit, giving a precise measurement of its frequency, dispersion across text types, and the degree to which its subparts are held together.

At the turn of the last decade, (Wood, 2002) warned that not much research was available on the teaching of repetitive sequences despite the mounting evidence that these patterns pose a serious challenge, particularly for second/foreign language learners. A decade later, Wray (2012) revisited this claim, offering a somewhat different opinion by asserting that much progress has been made despite the complex nature of research in this area.

Given that recurrent word combinations are problematic even for advanced learners, it is important that language instructors, textbook authors and materials designers become aware of the wide-ranging, empirically grounded approaches on how to deepen the learners' knowledge of these sequences in second/foreign language contexts. This line of research has been pursued by key researchers (Cortes, 2006; Meunier, 2012; Wood, 2002; Wray, 2000) and this article is a contribution to an ongoing discussion on the various interventions conducive to better learning and retention of academic multiword sequences. Several terms are used to describe this phenomenon, some of which are used as "umbrella terms" aimed to account for as many instances of occurrence as possible (e.g., multiword units, recurrent word combinations and formulaic sequences), but some others are used to refer to a special set of sequences sharing certain characteristics (e.g., academic formulas, lexical bundles). Here, I will use *multiword expressions* as a cover term under which a whole range of notions such as lexical bundles, collocations, formulaic sequences and academic constructions are included and discussed. This does not imply that we blind our eyes to the real differences existing between these terms. Rather, the impetus is to use a term that is neutral and can account for the many attributes featuring sequences.

In this paper, I will address the teaching and learning of multiword expressions. First, the role of the learner's first language in the language classroom will be addressed in the light of numerous studies which highlight L1 as an impediment to the learning of L2 formulaic sequences. Next, the discussion will revolve around the lists of sequences compiled by several researchers, which can be used by language teachers to foster the learning of the most pedagogically useful sequences. Two interrelated teaching approaches are then discussed: form-focused instruction and function-focused instruction. The discussion will be supplemented by categorization schemes discussed in the literature.

2. Rethinking the Role of L1

The impact of L1 on the learners' recognition and production of multiword units in a second/foreign language is undeniably great, as is pointed out by several researchers who realized that some errors in the use of L2 patterns can be traced to the influence of the students' first language (Brown, 2014; Laufer, 2011; Nesselhauf, 2003; Paquot, 2008; Sadeghi, 2009). In his study of Japanese EFL learners, Brown (2014) gave learners of English a list of words and instructed them to write three collocates of each word, provided that they do so within a time frame of 30 seconds. Negative transfer of Japanese combinatory patterns was identified, resulting in phrases such as *keep health*, which is unusual in English, but "a literal translation of the Japanese phrase *kenkou o mamoru*" (p. 136). In much the same way, Laufer (2011) found that L2 learners combine verbs and nouns that normally do not go together. Examples include *notice attention** and *try an attempt**. However, when learners are allowed to access collocational patterns using a bilingual dictionary and two monolingual dictionaries, their correct use of collocations has significantly increased. However, the author warns of what she calls "deceptive compatibility", that is, the misguided perception that a collocation in one language has its equivalent in another language. In a corpus-based study, Paquot (2008) examined lexical patterns fulfilling an exemplification function by non-native speakers representing different L1 backgrounds. The Analysis of results has shown that non-native writers tend to rely heavily on strings such as *for example* and *for instance*, apparently unaware that other synonymous patterns exist. Over-reliance on these two structures is probably a result of their existence in the learners' native language. To help learners become more creative and to make their writing less repetitive, the author suggests that instructors and materials authors draw their learners' attention to verbs such as *illustrate* and *exemplify* as they occur in authentic contexts. Strings like *let us take as an example*, *let's take the example*, *to illustrate the truth of*, and *what I wanted to illustrate* are rhetorically possible alternatives as well.

Nesselhauf (2003, p. 140) seems to be an advocate of this line of thinking, suggesting the use of a rather L1-based approach to alert learners to the L1-L2 collocational differences. Yamashita and Jiang (2010) investigated whether L2 collocations with L1 parallels are recognized faster and more accurately than combinations without L1 equivalents. Results show that multiword units without corresponding strings in the learners' L1 pose a considerable challenge, even for learners demonstrating higher proficiency. As a pedagogical implication to their study, the researchers have recommended that learners are presented with strings in the target language and are then prompted to come up with an L1 equivalent. In modern times where classrooms constitute learners with diverse first-language backgrounds, this procedure can also be implemented, as the numerous cross-linguistic comparisons will enrich the classroom discussion. In a rather comprehensive study, Paquot, (2013) found that 54% of mistakes in the use of bundles could be attributed to the influence of the students' first language. Four elements contributing to the transfer effect are identified: the transfer of collocational and colligational preferences, transfer of syntactic constructions, transfer of functions, and transfer to frequency. To address the negative effects of L1 transferability, the author emphasizes the need for learners to attend to the existing differences and similarities between their native language and the language they opt to learn. To foster such approach, a series of classroom-based activities that help raise learners' consciousness of these differences would be of great value. One way to do that is to give a "brief explanation of L1-L2 differences ... which proves more effective than other teaching methods that ignore the cross-linguistic differences" (Paquot, 2013, p. 666). Such an explanation is helpful in showing that some words in the native language can easily create a lexical bond, while the combination of the same words in the target language is odd and nonnative. The use of L1 in an L2 context is not problematic as far as learners share the same L1. The problem arises when the classroom has learners with diverse L1 backgrounds. Yamashita and Jiang (2010) argue that such diversity is an asset, not a disadvantage, as instructors have an abundant number of real examples that they can draw on while explaining why a specific string of words is comprehensible in one language, but impossible in another language.

To conclude this part, it is important to note that the influence of the learners' first language in learning L2 recurrent bundles should never be underestimated, nor is it to be completely seen as negative and inappropriate. It is essential in this case that the instructional intervention is driven by the need to explicitly draw attention to the similarities and differences shaping the L1-L2 collocational patterns. Instructors need not be deterred by the

fact that their classrooms are becoming more linguistically diverse. L1 diversity, as pointed out by Yamashita and Jiang (2010), who claimed that cross-linguistic comparisons initiated and monitored by the language instructor will help diversify the discussion and will also contribute to the realization that languages, though have some similarities, exhibit a certain degree of variation. However, this procedure should be pursued with caution, as there is a possibility that the learning session may turn into a translation session in which differences, rather than

3. Lists of Multiword Units

One way to maximize the learning of various types of multiword units is to make use of the growing number of lists that are easily accessible and readily available. Some of these lists are compiled to provide learners and practitioners with a more comprehensive resource of pedagogically useful wordstrings (Ackermann & Chen, 2013; Durrant, 2009; Hsu, 2014; Martinez & Schmitt, 2012; Shin & Nation, 2008; Simpson-Vlach & Ellis, 2010). While some others are generated for sampling purposes; that is, the aim is to address the research questions using a manageable size of patterns. Regardless of the intent of the list compilers, these lists offer language instructors with ready-made sequences that can be incorporated into a language learning program. Although the latter type can also be helpful for educators, I will only discuss these lists which are created for pedagogical purposes.

The list created by Martinez and Schmitt (2012) comprises 505 different recurrent expressions which, according to the authors, may pose a challenge to learners. It is thus “compiled with pedagogic purposes in mind, and it is hoped that it will be used to incorporate multiword items into existing wordlists” (p. 315). What makes the Phrasal Expressions List a good resource for use by instructors is the fact that each lexical expression is supplemented by an example sentence that captures its meaning. A sentence such as *we shared a room in order to reduce costs* is used to exemplify the meaning of the sequence *in order to*. The classification of items in the list takes into consideration the type of discourse that such sequences are commonly found. A sequence such as *going to* is normally encountered in general spoken and written English but is rarely used in the written academic discourse. Given the huge number of items on the list, the language teaching professionals should give priority to sequences that best address the learners’ immediate needs as well as their learning purposes.

One of the most comprehensive attempts to generate a list of recurrent sequences is carried out by Simpson-Vlach and Ellis (2010). Informed by results analyzed from a 2-million corpus of written and spoken academic data, the list is created based on a complex set of criteria: frequency of occurrence, the degree of association between lexical items comprising a sequence and the opinions gleaned from a panel of experts regarding the usefulness of each sequence. Items on the final list are functionally grouped into referentials, text-organizers and stance expressions, each of which is divided further into smaller subcategories.

Following the same path, Ackermann and Chen (2013) have utilized a set of statistical measures combined with expert opinions to create a cross-disciplinary list of useful collocations pertaining to the academic domain. This listing comprises 2,468 entries representing four categories: noun, verb+noun/adjective, verb+adv and adv+adj combinations. This list, which is also known as the Academic Collocation List (ACL), is hoped to offer “a further tool for EAP teachers to construct appropriate teaching materials and help students focus on frequent lexical items beyond individual words.” (p. 246).

Hsu (2014) built a corpus of written textbooks from a range of disciplines, including humanities, engineering and health care and retrieved multiword sequences which are more likely to pose a challenge for the students, given their odd and unfamiliar meaning. To elicit sequences as such, criteria of frequency, distribution across the corpus subparts, meaningfulness, grammatical correctness and semantic non-compositionality are applied. The list includes sequences of various lengths, ranging from two-word to five-word patterns. The final list incorporates patterns that are of greater value to the students. Examples include *according to*, *by virtue of*, *on the other hand* and *so far as~ be concerned*.

As the focus in the previous studies has been on both the written and spoken English, the study conducted by Shin and Nation (2008) draws attention to the most recurrent collocations commonly found in speech. The list contains a total of 4,698 different collocations which meet six criteria, the most important of which are frequency of occurrence, grammatical well-formedness, the meaning that the sequence carries in the text. The list includes collocations such as *you know*, *I think that*, *as well*, *in fact* and *at the moment*. The list compilers conclude by warning that using the frequency of occurrence as the only criterion in teaching vocabulary is not methodologically appropriate, suggesting that other factors such as the learners’ needs, level of study and their academic experience and expertise.

While each list type is a powerful means for enriching learners’ vocabulary knowledge, care should be taken as to what to include in an EAP course. Drawing on the course goals, instructors may prioritize sequences based on

the objectives of their individual courses, the needs of learners and the different characteristic inherently associated with multi-word expressions such as frequency, semantic opaqueness, and distribution among registers. Realizing these challenges, Martinez (2013) proposed a framework based on the frequency and semantic opaqueness of the target sequence. A sequence merits the utmost attention if it is both frequent and opaque. Infrequent/transparent sequences, on the extreme end of the spectrum, deserve the least attention from instructors. Frequent/transparent and opaque/frequent clusters occupy the second and third ranks of sequence, respectively.

The number of sequences in a given discourse is impossible to pin down, as is noted by Wray (2002) who asserts that “there is not simply a single stock of formulaic sequences which all speakers first learn, and then draw upon, but rather that the store is constantly changing, to meet the changing needs of the speaker” (p. 101). This implies that stockpiling sequences is not a good technique, given the ever-changing nature of recurrent bundles in a given discourse on the one hand and the shifting needs of learners on the other. In this case, it is the responsibility of the course instructors or pedagogical material authors to prioritize items that have the utmost pedagogical value for the students.

4. Emerging Technologies and Multiword Units Instruction

The term emerging technology is commonly used to refer to the set of computer-based applications that enable ordinary users to create, revise and disseminate content using online platforms. Computer technology has been used for a wide range of learning purposes and several researchers seem to concur that such technology can be successfully used to improve different language-related skills and subskills (for thorough review see Golonka, Bowles, Frank, Richardson, & Freynik, 2014). As the Internet becomes easily accessible in many places, language educators are more likely to harness its power in creating interactive content, asking for greater participation from the part of the students and making the learning process both rewarding and interesting. The Internet is still evolving, and several platforms are constantly emerging (social media is a recent phenomenon), a fact that makes it hard to draw a full picture of its contribution to language learning. Apart from the fact that countless number of studies on multiword expressions involve the use of electronic corpora, several other studies have examined the potential of the computer technology as a tool for the learning and teaching of multiword expressions (Chan & Liou, 2005; Grami & Alkazemi, 2016; Lin, 2014; Walker, 2011; Wu, Witten, & Franken, 2010). This latter feature is what concerns us in this section.

Wible (2008) argues that there are two reasons why incorporating the Internet in language teaching is pedagogically useful. First, the Internet serves as a massive repository of authentic materials. This could be implemented by individual teachers designing lessons and activities, or by an institution running a learning management system. So if a particular learner is not sure whether a set of adjacent words makes a recurrent cluster, he or she can search the web and obtain countless number of instances showing not only that the recurrent cluster is possible but also the way this string of words is used in distinct contexts. Second, the Internet is replete with content that is not created purely for learning purposes, but can be accommodated to fulfill a learning objective. Wible seems to suggest that this latter perspective is more congruent with the nature of formulaicity, as it is hard to pin down the number of recurrent clusters of potential use for learners. He then introduces *Collocator*, a software installed in the user’s browser which, first, detects word strings in a webpage and, second, offers learners further input of the selected cluster by hyperlinking other instances where it occurs. Care should be taken with such programs, as they may provide learners with erroneous examples or ill-formed patterns since content there is sometimes created by individuals with less education.

Wu, Witten and Franken (2010) have built and subsequently tested a web-derived frequency-based corpus comprising 380-million five-word sequences. Language learners can search for different collocational patterns, including verb-nouns, noun-nouns, and adjective-nouns, to name but a few. Unlike many learner corpora, this one is credited for allowing learners to further enhance their exposure to collocations by offering them extra examples derived either from the Web or the British National Corpus (BNC).

Bilingual concordancers which can be accessed online may also be to enhance the learners’ awareness of the combinatory features of some linguistic elements. Chan and Liou (2005) have created a Chinese-English software that helps Chinese learners improve their knowledge of verb-noun combinations in English. The group of the students who received training on this program showed significant gains at the end of the training session, as is reflected in their posttest results. Although these gains seem to decrease with time, as is reflected in the delayed posttest results, it is important to note that providing the students with a web-based content on how words are joined to form comprehensible strings would certainly raise their awareness of the combinatory features characterizing sequences in a language.

The study by Lin (2014) has elicited the most recurrent sequences in the iTV corpus to determine their validity as useful patterns to learn by EFL learners. The same target sequences in the iTV corpus are found to recur frequently in the spoken part of the British National Corpus (BNC), a finding that led the researcher to maintain the pedagogical value of video materials streaming online in enhancing learners' knowledge of formulaic language. Videos which are more conducive to better learning come from three genres: religion, news and comedy.

Another potential use of electronic corpora involves checking whether certain lexical items tend to co-occur with specific words and expressions. The study by Walker (2011) is an attempt to raise business professionals' awareness of the way corpora can be used to show the co-occurrences of certain words. The expressions *run* and *in charge of*, for example, are associated with power, as is exemplified in statements such as *run the company* and *in charge of the country*. Expressions such as *manage* and *responsible for*, in contrast, do not carry the sense of power as is the case in *run* and *in charge of* but are more likely to denote responsibility towards people.

The use of an online checker to help non-native writers use multiword sequences is the subject of a study conducted by Grami and Alkazemi (2016). Here, two groups of undergraduate students participated in an experiment where members of one group were trained on how to use an online checker for using words that normally co-occur, whereas members of the other group were not allowed to access that checker. The results maintain that the students using the software well outperformed those who did not, a result attesting to the usefulness of such program in enhancing the students' use of correct word combinations while writing. The online program can also provide instructors with information on the nature of errors that are committed by individual students. Such information can be used to determine the type of intervention that is more appropriate to address these errors and help make their writings error-free.

5. Form-Focused Instruction

The retrieval of a wide range of bundles requires a structural analysis to uncover their underlying grammatical patterns. Biber, Johansson, Leech, Conrad, and Finegan (1999) developed a classification scheme that accounts for the structural correlates of multiword patterns in academic prose. Bundles are thus grouped into noun-based, verb-based and preposition-based. In most cases, Noun-based bundles are initiated by a noun and end with *-of-phrase* (e.g., *the number of the*) or a post-modifier fragment (e.g., *the ways in which, the fact that the*). multiword patterns headed by a preposition also vary noticeably, but the greatest number of these patterns begin with a preposition and end up with an embedded *of-phrase*. Verb-based lexical bundles include several subgroups such as the anticipatory *It*-clause (e.g., *it is possible to*), passive constructions (e.g., *is based on the*), adverbial clauses (e.g., *as we have seen*). Hyland's (2008b: 12) description of the structural patterns of multiword constructions is based on that of Biber and colleagues, thus classifying recurrent patterns into nominal, prepositional and verbal patterns. The majority of noun-based patterns incorporate *of-construction* (e.g., *the presence of the, a function of the*), while most verbal patterns involve the passive bundles followed by a prepositional phrase (e.g. *are shown in figure, is based on the*). Patterns headed by a preposition are also common, including sequences such as *at the end of* and *in this case the*. The structural framework developed by Biber, Conrad, and Cortes (2004) is slightly different, dividing lexical patterns into three major categories, each of which is further divided into smaller subgroups. In the first category, verb phrase fragments, recurrent multiword constructions comprise of 1st/2nd person pronoun followed by verb phrase fragment (e.g., *you don't have to*), connector+ 3rd person pronoun+ VP fragment (e.g. *it's going to be*), discourse marker +VP fragment (e.g., *I mean you know*), verb phrase with passive and non-passive verbs (e.g., *is going to be, is based on the*) and Wh-, yes/no questions (e.g., *are you going to, what do you think*). The second category consists of patterns incorporating depended clause patterns such as *If-clause* fragments (e, g., *if you want to, verb/adjective*) + to-clause fragment (e.g. *to be able to*), that-clause fragments (e.g., *that there is a*). As for the noun/prepositional phrase fragments, the third category, recurrent patterns take several structural forms, including noun phrases followed by an *-of* phrase construction (e.g., *the end of the*), noun phrase with post-modifiers (e.g. *a little bit about*) and preposition-based word strings (e.g., *at the end of*).

6. Function-Focused Instruction

A core challenge facing second language educators and materials writers is the identification of the underlying functions served by a set of recurrent multiword sequences. Although the literature is replete with several function-based classification schemes, there is still a need to situate functions in the contexts in which bundles occur. Corpus tools offer a powerful means by which recurrent bundles can be identified and examined. Cortes (2004) sheds doubts on the claim that functions associated with certain lexical bundles would be accidentally noticed by learners and calls for an explicit, function-based approach that makes the students aware of the

functional patterns associated with multiword units. This is why some learners tend to employ an avoidance strategy; that is, lexical bundles whose functions are not known are deliberately skipped in favor of those with familiar meanings. This gives rise to the student's writings being replete with odd (Schmitt, 2004) and redundant (Cortes, 2004) sequences.

An earlier attempt to suggest a functional classification of bundles was carried out by Biber et al. (2004) who classified patterns into three major categories: referentials, discourse markers and stance expressions. While referentials are used to "make direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important", discourse markers "reflect the relationships between prior and coming discourse". Stance markers, the third category, are utilized to "express attitudes or assessments of certainty that frame some other propositions" (Biber et al., 2004, p. 384). Each of these functional categories is further divided into several subgroups. Referentials, for example, can allude to imprecision (e.g., and things like that), identification (e.g., that's one of the) and the specification of attributes (e.g., there's a lot of). Stance expressions are divided into personal epistemic stance (e.g., I don't know if), impersonal epistemic stance (e.g., the fact that the) or attitudinal stance (e.g., I don't want to). Multiword sequences serving a discourse-organizing function are functionally used to introduce a topic (e.g., If you look at) or to elaborate on a topic (e.g., on the other hand). In much the same vein, Hyland (2008) proposed a similar framework, dividing bundles into research-oriented, text-oriented and participant-oriented, each of which further branches out into smaller sub-categories. Recurrent sequences fulfilling a research-oriented function are functionally used to "help writers to structure their activities and experiences in the real world" whereas those sequences serving a text-oriented function are "concerned with the organization of the text and its meaning as a message of argument" (Hyland, 2008b, p. 13). Participant-oriented expressions are "focused on the writer or reader of the text" (Hyland, 2008b, p. 14). Within research-oriented group, sequences can help indicate time (e.g., at the same time) or place (e.g., in the present study). Some other sequences are used to allude to a specific procedure (e.g., the role of the) or description (e.g., the structure of the).

Since the development of these two functional schemes, a wealth of different studies have investigated the functions of multiword patterns in different contexts.

Ädel and Erman (2012) have examined four-word sequences in the Stockholm University Student English Corpus, totaling one million words of essays written by multi-level Swedish students and British students. They then classified the sequences functionally in order to compare the use of sequences among the two groups of students. The overall comparison does not reveal much difference as referential bundles are evidently more commonly used in 47% of natives' writings and 45 of nonnatives' writings. Stance bundles are the second most commonly used bundles

Durrant (2017) studied the distribution of lexical bundles across different disciplines and found that they can be grouped into research-oriented, text-oriented and stance. Research-oriented bundles fall into sub-groups: location (e.g., on the surface of), procedure (e.g., be used in the), quantification (e.g., a large number of), description (e.g., the presence of a) and intangible framing attributes (e.g., the idea of a). Text-oriented bundles show the relationships between different parts of the text. These bundles are also divided into four sub-categories: transition signals (e.g., an example of the), resultative markers (e.g., this means that the), structuring devices (e.g., is shown in figure) and framing signals (e.g., with respect to the). The third major functional category involves stance markers which can be further grouped into centrality (e.g., at the heart of), epistemic (e.g., it could be argued) and modality (e.g., it is impossible to).

These functions are by no means exhaustive, as research into recurrent word combinations is increasingly unveiling new patterns performing distinct functions. While some patterns are intended to serve specific functions irrespective of the topic or discipline (e.g., *on the other hand*, *the basis of the*), some other patterns are typical of the disciplines under examination (e.g. *turn of the century*, *the court of appeal*).

7. Attention-Drawing Tasks

Several researchers have pointed out to the importance of deliberately drawing learners' attention to recurrent multiword sequences (AlHassan & Wood, 2015; Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Cortes, 2004; Fordyce, 2014; Lewis, 2000; Nekrasova, 2009; Sonbul & Schmitt, 2013) (AlHassan & Wood, 2015; Cortes, 2004; Nekrasova, 2009). Creating cross-linguistic comparisons to foster learning can also take place, provided that learners come from different linguistic backgrounds. There are several attention-drawing techniques that can be weighted by instructors. Sonbul and Schmitt (2013) investigated the effect of three different teaching approaches on the learning of two-word collocations by native English speakers and nonnatives representing diverse linguistic backgrounds. In one task, participants were presented with the target

collocations embedded into a text, but were given a distinct color so as ease noticing by learners. Two test measuring the ability to recognize and recall such collocations from exposure were administered and the results pointed out to the positive effects of this technique on both two levels of collocational knowledge. Helping learners take notice of the embedded stance bundles was the subject of a study conducted by Fordyce (2014). Here, two groups of Japanese learners of English took part in the study that aims at comparing the results gleaned from two instructional models: explicit and implicit. In the explicit model, learners' attention was deliberately drawn to stance bundles by means of highlighting, underlining and bolding. This attention-drawing technique was followed by a number of carefully designed activities: guided listening, bundle-focused assignments, explanation of the general uses of the target bundles and teacher-initiated classroom-based discussion of the form-function characteristics of stance bundles. While this is the case for the explicit instruction group, the attention of the implicit group was thus driven to the overall meaning of the texts and approaches needed to analyze it. All of the follow-up activities are intended to help learners comprehend the texts, discuss their content and respond to general questions on accuracy and organizational patterning. Results from post- and delayed-tests indicate that drawing learners attention into the target stance bundles by means of highlighting, bolding and underlining is more conducive to immediate and long-term learning than merely seeing them in the context.

Another possible approach to draw the learners' attention to a set of strings embedded in a text is to guide them through a process where they first recognize the node word and then identify any neighboring items that, together, would form a collocational pair. Lewis (2000) recommends that, in the case of verb-noun collocations, learners are first instructed to underline all given nouns in a text and then move to determine co-occurring verbs. Similarly, Coxhead (2008) proposes a three-stage process that, ultimately, enhances the learning of lexical bundles. In the first stage, which she calls "noticing", learners are asked to read authentic texts where key bundles are made clear by means of highlighting. Then, in the retrieval stage, the same learners are instructed to recall the text they have just read, making use of cards to jot down keywords and phrases. These cards, together with any written materials, are passed from one student to the other so that struggling students can catch up with high-proficient ones. While "noticing" and "retrieval" stages are key to help learners deepen their receptive knowledge of the target lexical strings, "generation" is aptly necessary for turning receptive knowledge into productive one, by means of providing learners with opportunities for trying out newly learned strings in novel ways.

AlHassan and Wood (2015) trained twelve L2 students on how to use formulaic sequences to produce a coherent text. The training sessions have two phases: a presentation phase where sequences are given as discrete items, and an activity phase, where participants engage in tasks such as filling in, matching, and completion. Ultimately, participants are prompted to produce a novice text that utilizes the set of sequences. By comparing the results across the pretest, posttest and delayed posttest, results show that L2 learners have made some important gains and that these gains tend to become part of their mental lexicon, as is clearly evident from the delayed test. By comparing the two paragraphs, which are produced by the same student, it is obvious that the student has utilized a number of key lexical devices in two places of the text: introduction and conclusion. (AlHassan & Wood, 2015: 59).

- It show how are some materials throw around one year In fact, when the price became higher that's give pad effect. For example. The poor people will be increased, and many people will do criminal thing like theft.
- The line paragraph presents the rise and fall in the general price of commodity between May 2010 and April 2011.... In conclusion, it can be seen that the commodity prices have generally increased which leads to a decline in the purchasing power of the [sic] consumers.

Boers et al. (2006) focusing on fluency, found that drawing learners' attention to useful formulaic sequences is a pedagogically useful technique for helping L2 learners become more fluent and their speech more hapitual. Similarly focusing on fluency, Wood (2010) made the case for considering sequences as target vocabulary, suggesting some strategies that can be incorporated in the teaching of recurrent sequences (p. 196):

- Watch for the use of formulaic sequences in daily life encounters.
- Make a list or notebook of formulaic sequences heard.
- Listen to the media and note how recurring formulaic sequences are used.
- Interpret functions and meaning of sequences from the context or from analyzing their component parts.
- Explore the cultural metaphors underlying sequences.
- Use mnemonics to aid in the storage of sequences.

A final way of heightening the learners' awareness of sequences is through extensive exposure. Peters (2012) reported substantial gains after a series of non-communicative, vocabulary-focused activities in which lexical items and collocations were recycled five times and that these gains tend to last for a considerably long time, as is evidently clear from the delayed posttest. Durrant and Schmitt (2010) examined the production and processing of adjective-noun collocations and found considerable gains when the target sequence is recycled twice.

8. Conclusion

The overall picture emerging from the previous discussion is that although multiword sequences pose a challenge for learners, particularly for those whom English is a second or foreign language, still researchers have suggested a wide range of empirically grounded and pedagogically compelling approaches conducive to better learning of these sequences. Language instructors will find these approaches useful for helping their learners become aware of the range of recurrent multiword patterns available to them while writing a term paper, taking an entry exam or producing a thesis. Materials designers and textbooks authors can draw on the readily available lists of multiword constructions that have been compiled by various researchers. Moreover, classroom instructors may find the attention-drawing techniques discussed in this article of greater value for fostering the learning of such patterns in a pedagogically useful manner.

References

- Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) - A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235-247. <https://doi.org/10.1016/j.jeap.2013.08.002>
- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81-92. <https://doi.org/10.1016/j.esp.2011.08.004>
- AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes, 17*, 51-62. <https://doi.org/10.1016/j.jeap.2015.02.001>
- Bardovi-Harlig, K. (2012). Formulas, Routines, and Conventional Expressions in Pragmatics Research. *Annual Review of Applied Linguistics, 32*, 206-227. <https://doi.org/10.1017/S0267190512000086>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*, 371-405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, Johansson, Leech, Conrad, & Finegan. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Pearson.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (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>
- Brown, D. (2014). Knowledge of collocations. In J. Milton, & F. Tess (Eds.), *Dimensions of vocabulary knowledge* (pp. 123-139). Basingstoke, Hampshire: Palgrave Macmillan. https://doi.org/10.1007/978-1-137-36831-7_9
- Chan, T., & Liou, H.-C. (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>
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education, 17*(4), 391-406. <https://doi.org/10.1016/j.linged.2007.02.001>
- Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and opportunities. In *Phraseology in Foreign Language Learning and Teaching* (pp. 149-161). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/z.138.12cox>
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*(3), 157-169. <https://doi.org/10.1016/j.esp.2009.02.002>

- Durrant, P. (2017). Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. *Applied Linguistics*, 38(2), 165-193. <https://doi.org/10.1093/applin/amv011>
- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163-188. <https://doi.org/10.1177/0267658309349431>
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375-396. <https://doi.org/10.1002/j.1545-7249.2008.tb00137.x>
- Flowerdew, L. (2015). Using corpus-based research and online academic corpora to inform writing of the discussion section of a thesis. *Journal of English for Academic Purposes*, 20, 58-68. <https://doi.org/10.1016/j.jeap.2015.06.001>
- Fordyce, K. (2014). The differential effects of explicit and implicit instruction on EFL learners' Use of Epistemic Stance. *Applied Linguistics*, 35(1), 6-28. <https://doi.org/10.1093/applin/ams076>
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70-105. <https://doi.org/10.1080/09588221.2012.700315>
- Grami, G. M. A., & Alkazemi, B. Y. (2016). Improving ESL writing using an online formulaic sequence word-combination checker. *Journal of Computer Assisted Learning*, 32(2), 95-104.
- Hsu, W. (2014). The most frequent opaque formulaic sequences in English-medium college textbooks. *System*, 47, 146-161. <https://doi.org/10.1016/j.system.2014.10.001>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Laufer, B. (2011). The Contribution of Dictionary Use to the Production and Retention of Collocations in a Second Language. *International Journal of Lexicography*, 24(1), 29-49. <https://doi.org/10.1093/ijl/ecq039>
- Lewis, M. (2000). Materials and resources for teaching collocation. In M. Lewis (Ed.), *Teaching collocations: Further developments in the lexical approach* (pp. 186-204). Bostin, MI: Heinle.
- Lin, P. M. S. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System*, 42, 164-176. <https://doi.org/10.1016/j.system.2013.11.010>
- Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25-35. <https://doi.org/10.1016/j.esp.2011.07.002>
- Martinez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT Journal*, 67(2), 184-198. <https://doi.org/10.1093/elt/ccs100>
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299-320. <https://doi.org/10.1093/applin/ams010>
- Meunier, F. (2012). Formulaic Language and Language Teaching. *Annual Review of Applied Linguistics*, 32, 111-129. <https://doi.org/10.1017/S0267190512000128>
- Milbury, K., Spelman, A., Wood, C., Matin, S. F., Tannir, N., Jonasch, E., ... Cohen, L. (2014). Randomized controlled trial of expressive writing for patients with renal cell carcinoma. *Journal of Clinical Oncology*, 32(7), 663-670. <https://doi.org/10.1200/JCO.2013.50.3532>
- Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nekrasova, T. M. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Language Learning*, 59(3), 647-686. <https://doi.org/10.1111/j.1467-9922.2009.00520.x>
- Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, 24(2), 223-242. <https://doi.org/10.1093/applin/24.2.223>
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier, & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 101-119). Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.14>
- Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics*, 18(3), 391-417. <https://doi.org/10.1075/ijcl.18.3.06paq>

- Peters, E. (2012). Learning German formulaic sequences: the effect of two attention-drawing techniques. *The Language Learning Journal*, 40(1), 65-79. <https://doi.org/10.1080/09571736.2012.658224>
- Sadeghi, K. (2009). Collocational Differences Between L1 and L2: Implications for EFL Learners and Teachers. *TESL Canada Journal*, 26(2), 100-124. <https://doi.org/10.18806/tesl.v26i2.417>
- Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal*, 62(4), 339-348. <https://doi.org/10.1093/elt/ccm091>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512. <https://doi.org/10.1093/applin/amp058>
- Sonbul, S., & Schmitt, N. (2013). Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions. *Language Learning*, 63(1), 121-159. <https://doi.org/10.1111/j.1467-9922.2012.00730.x>
- Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English for Specific Purposes*, 30(2), 101-112. <https://doi.org/10.1016/j.esp.2010.12.003>
- Wible, D. (2008). Multiword expressions and the digital turn. In F. Meunier, & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 163-180). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/z.138.13wib>
- Wood, D. (2002). Formulaic Language in Acquisition and Production : Implications for Teaching. *TESL Canada Journal*, 20(1), 1-15. <https://doi.org/10.18806/tesl.v20i1.935>
- Wood, D. (2010). *Formulaic language and second language fluency*. London: Continuum.
- Wood, D. (2015). *Fundamentals of Formulaic Language: An introduction*. London: Bloomsbury.
- Wray, A. (2000). Formulaic Sequences in Second Language Teaching: Principle and Practice. *Applied Linguistics*, 21, 463-489. <https://doi.org/10.1093/applin/21.4.463>
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511519772>
- Wray, A. (2012). What Do We (Think We) Know About Formulaic Language? An Evaluation of the Current State of Play. *Annual Review of Applied Linguistics*, 32, 231-254. <https://doi.org/10.1017/S026719051200013X>
- Wu, S., Witten, I. H., & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL*, 22(1), 83-102. <https://doi.org/10.1017/S0958344009990218>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).