Teaching Collocations in a Second Language: Why, What and How?

by Zorana Vasiljevic, Bunkyo University, Japan

Abstract

Although there is little doubt that collocation knowledge is crucial for fluent language use, collocations often do not receive sufficient attention, either inside or outside the language classroom. While more recent EFL textbooks include some collocation work, this work is rarely systematic and the associated activities are often too few. Due to curriculum constraints and time pressure, teachers are frequently not in a position to create supplementary collocation study materials, and as collocations usually do not pose comprehension problems, they may opt not to attend to them at all. Furthermore, collocations may present a challenge for instructors. While they may have a good command of grammatical patterns and a large receptive vocabulary, they may lack confidence when it comes to selection of syntagmatic phrases or advising students about acceptable word combinations in the target language.

Learners, for their part, are often not even aware of the significance of collocation knowledge. They tend to identify vocabulary learning with an expansion of vocabulary size, and they perceive knowledge of difficult or specialized words as an indicator of language progress. As a result, they often focus on memorising long, decontextualized word lists, and they seldom pay attention to how these words are actually used.

This paper will review some major findings that have emerged from collocation studies in corpus and applied linguistics, and it will offer some suggestions for possible applications of these findings in the language classroom. Special attention will be given to the criteria for the selection of target collocations, and the activities that may help bridge the gap between the learners' receptive and productive vocabulary knowledge. Finally, some ideas about how student autonomy in collocation learning could be fostered will be shared.

Key words: collocation teaching activities and tests, online corpora, learner autonomy

Apstrakt

Premda kljucne za tecnu upotrebu jezika, kolokacijama se ne poklanja dovolno paznje ni u ucionici ni izvan nje. Iako noviji udzbenici za ucenje engleskog ukljucuju neke aktivnosti za vezbu kolokacija, one su nesistematicne i nedovoljne. Zbog programskih obaveza i nedostatka vremena, profesori cesto nisu u mogucnosti da pripreme dodatne materijale, i kako poznavanje kolokacija obicno nije kljucno za razumevanje teksta, one su ponekad i potpuno eliminisane iz programa. Dodatni razlog za njihovu eliminaciju je i to sto kolokacije cesto predstavljaju teskoce i za same profesore.Uprkos dobrom vladanja gramatikom i poznavanja velikog broja reci, nije neobicno da i profesori imaju nedoumica kad je u pitanju izbor sintagmatskih izraza ili upucivanje studenata ka pravom izboru reci na stranom jeziku.

Studenti, sa druge strane, cesto nisu ni svesni koliko je poznavanje kolokacija vazno. Oni poistovecuju ucenje vokabulara sa povecanjem broja reci i veruju da je poznavanje teskih reci ili specijalizovanih izraza dokaz njihovog napretka.Rezultat je njihovo koncentrisanje na memorizaciju dugih, nekontekstualizovanih lista reci i nedovoljno paznje posvecene njihovoj upotrebi.

Ovaj clanak sumira neka od vaznijih saznanja iz studija kolokacija u okviru korpusne I primenjene lingvistike, i upucuje na nacine na koje bi se ona mogla primeniti u jezickim institucijama. Posebna paznja je posvecena kriterijumima za selekciju kolokacija za direktnu nastavu i aktivnostima kojima bi se mogao smanjiti procep izmedju pasivnog i aktivnog vokabulara. Clanak takodje izlaze ideje za podsticanje studentske autonomije u ucenju kolokacija.

***Ključne reči**: poducavanje i testiranje kolokacija, elektronski korpus, autonomija studenata*

1. Introduction: The notion of collocations

One feature common to natural languages is that words tend to occur together with a restricted set of other words. These frequently co-occurring word strings are known as *collocations.* A collocation is a phrase which consists of the focus item known as the *node* or the *base*, a co-occurring word or the *collocate,* and the *span*, which is an environment in which the node and the collocate co-occur. The relationships between the phrase constituents within a span are by no means uniform and clear-cut, and a wide range of classification schemes have been developed. Handl (2009) identifies three dimensions along which collocations can be classified:1) semantic, 2) lexical and 3) statistical. On a semantic level, collocations differ in their level of transparency. As can be seen from the examples below, while the meanings of some phrases are transparent and matches that of the component words, other phrases are highly idiomatic.

Table 1. Examples of collocations of different levels of transparency (based on Reveir, 2009)

	Transparent	Semi-transparent	Non-transparent
Example 1:	*make tea*	*make a complaint*	*make the grade*
Example 2	*take the money*	*take a course*	*take sides*

Some researchers (e.g. Reveir, 2009) argue that semantic transparency affects how the collocations are processed. According to Reveir (2009), transparent collocations tend to be processed compositionally (i.e. as separate items), and their use is linked to general lexical and grammatical knowledge. However, the use of semi-transparent and non-transparent collocation requires the phrases to be stored and accessed as holistic units.

On the lexical level, collocations differ with regard to the number of words that particular phrase components tend to collocate with. While some words have a large number of syntagmatic associations, others have a highly restricted use (e.g. *a slice of bacon* vs. *a rasher of bacon*).

On the statistical level, collocations can be analyzed with regard to their combined frequency as well as the frequency of the

phrase constituents. The ratio between the two frequencies reveals the strength of the collocation and the direction of the attraction between the phrase constituents.

Collocations also differ in terms of the level of substitutability of their combinatory elements (Howarth, 1998a). While some phrases are characterized by a relatively high level of commutability of the constituent words, others tend to be more restricted, as can be seen from the following examples: *blow a trumpet* (free combination), *blow a fuse* (restricted collocation), *blow your own trumpet* (figurative idiom)*, blow the gaff* (pure idiom) (Howarth, 1998b)

As can be seen from the examples above, the concept of collocation covers a large number of multiword chunks of different compositional structure, associative strength and semantic transparency.

2. WHY should collocations be taught?

One of the most important tasks that language learners face is acquiring an extensive vocabulary. Collocation knowledge is one of the most important properties of lexical competence. Language is phrasal in its nature, and collocations represent building blocks of spoken and written discourse (Wray, 2002). They facilitate communication by

reducing the processing load of both speakers and hearers alike (Halliday, 1966).Native speakers notice, process and store formulaic word sequences as a whole, and that allows them to take advantage of language idiomaticity during language comprehension and gives them fluency during language production (Pawley & Syder, 1983; Wray 2002). Collocations are also important because they define the meaning of individual words and phrases or, as Firth (1957:179) famously put it, "You shall know a word by the company it keeps." In other words, the sense of a lexical item and its functional value are defined through its use and through the relationships it forms with other words in a specific situational context (e.g. *run a marathon* vs. *run a shop* vs. *run a risk).*

However, while collocation knowledge is an essential component of communicative competence, collocations also pose one of the biggest challenges for second language learners even at very advanced stages of language proficiency (Arnaud & Savignon, 1997; Barfield, 2009; Howarth, 1998a, 1998b).

There are several reasons why collocation development is a slow and difficult process. One problem is that collocational restrictions are not imposed by the semantic compatibility of the phrase

constituents but by the norms of language use. Knowledge of collocations is not the same as the knowledge of the phrase component parts. Even advanced learners may experience difficulty in using common delexical verbs such as 'make' and 'take', although they are familiar with their meaning (Nesselhauf, 2003).

Another problem is that learners often do not recognize the importance of collocation knowledge (Jiang, 2009; Laufer 2005; Peters 2007).One reason for this may be the fact that collocations tend to pose little difficulty in language comprehension. Learners often underestimate the difference between receptive and productive vocabulary knowledge, and they assume that if they understand the words, they will be able to use them as well. As a result, many students tend to identify vocabulary learning with the acquisition of new words, and they fail to pay sufficient attention to collocational relationships in the input.

Insufficient attention given to word combinations during class instruction and limited coverage of syntagmatic phrases in the reference books are also some reasons why L2 collocations are difficult to acquire. Due to curricular constraints and time pressure, many teachers find it difficult to integrate collocation instruction in regular classes. For example, Jiang

(2009) found that about two-thirds of 75 Chinese students she surveyed had never done any collocation-focused vocabulary practice. Similarly, due to limitations of space, lexicographers cannot include extended collocational information in learner dictionaries, despite the great progress that has been made in corpus studies in recent years.

Lack of awareness combined with a lack of attention to collocational phrases may significantly impede learners' progress. The inability to recognize collocations means that learners cannot take advantage of the lexical priming in the natural discourse. Words that they encounter in the text do not generate any expectations for them, which significantly increases the pressure on their working memory and slows down their processing of the input (Siyanova & Schmitt, 2008). Lack of collocation competence also causes problems during language production. Languages differ in the range of acceptable combinatory choices of phrase constituents. For example, according to the *Oxford Collocations Dictionary for Students of English* (2002), some common verb collocates of the noun *decision* are *make, take, arrive at, come to* and *reach.* In Japanese, however, '*decision*' collocates with *'do'* (*kettsui /kesshin/kettei* + **suru**). Spanish allows *take* but not *do* or *make* (**tomar decisión**),

while in Serbian *decision* collocates with *bring* (**doneti** *odluku*). Due to the lack of awareness and L1 interference, learners' attempts to "re-assemble" individual items often result in erroneous, unnatural or oversimplified output (Lewis, 2000; Schmitt, 2004; Singleton, 1999; Wray, 2002; Yang & Hendricks, 2004).

The findings above suggest that if learners are to benefit from the reduced processing load, fluency and idiomaticity that collocation knowledge offers, they must learn to approach L2 vocabulary as a network. This means that instruction should aim at helping students learn to process, store and retrieve word strings holistically, in chunks, rather than assembling them at the production stage. For this purpose, it is necessary to have a systematic and comprehensive analysis of common word combinations in native speakers' corpora, and develop effective teaching strategies that would make these combinations more accessible to language learners. The next section will examine the problem of the selection of collocations for targeted teaching.

3. WHAT should be taught?

The sheer number of collocations and their pervasiveness in natural language pose challenges in terms of setting the learning goals and in terms of choosing the learning priorities. The *Oxford*

*Collocations Dictionary for Students of Englis*h (2002) includes about 150,000 collocations, and that list is far from being exhaustive. In the limited class time available, teachers have to be highly selective with regard to the phrases that will become the focus of class instruction. One common criterion for collocation selection is their *frequency of occurrence* in the corpus. Developments in computer technology have enabled the analysis of very large language corpora making it possible for researchers to obtain comprehensive data on the frequency of a particular word combination in the natural language. There are a number of online corpora available to language teachers and researchers. Some of the biggest ones are the 100 million-word *British National Corpus (BNC)* and the 450 million-word *Corpus of Contemporary American English (COCA)*. Both corpora include a large collection of samples of written and spoken language from a wide range of sources. The corpora websites offer a number of useful features such as search by word or phrase, by lemma, or by a part of speech, and they enable the users to perform an analysis of word frequencies, collocates, and distribution of synonyms in different types of texts, as well as indifferent time periods. These corpora have served as a basis for a large number of research publications throughout the world, and have permitted

the development of authentic teaching materials and resources.

The COCA corpus (http://corpus.byu.edu/coca), from which the examples in this paper are taken, is balanced between five different text types: spoken, fiction, magazine, newspaper and academic, allowing a comparison of word frequencies and use in different contexts. There are four basic search functions: 1) LIST which shows a list of words or word combinations ranked according to their frequency; 2) CHART which allows a comparison of frequencies in different genres or time periods; 3) KWIC (*Key Word in Context*) which displays search words in context with colour-coding for different parts of speech, and 4) the COMPARE option which allows a comparison of two words according to their general frequency or with their specific collocates. A variety of search options available to researchers goes beyond the scope of this paper, but for readers who are interested in learning more, the three COCA tutorials shared online by Professor David Brown (http://www.thegrammarlab.com/?page_id=15) present an excellent general introduction into how corpora could be used for research purposes. As the research interfaces are similar, the tutorials will also come in useful in the search of other corpora.

This paper will now examine more closely how corpora like COCA could be used to facilitate the development of the collocation knowledge of second language learners. COCA contains the largest and one of the most accurate lists of collocations currently available, with about 4.3 million node/ collocate pairs extracted from a 450 million-word corpus. Collocation data can be of use in various strands of linguistic research such as the analysis of syntactic patterns, pragmatics, semantic prosody, and sociolinguistic and discourse analysis. One way to search collocations in COCA is to type in the target word in the search field followed by a tag for the specific part of speech. For example, in order to search for which prepositions can follow the adjective *similar* they should type in the following search string: similar [i*], where [i*] stands for "all prepositions".
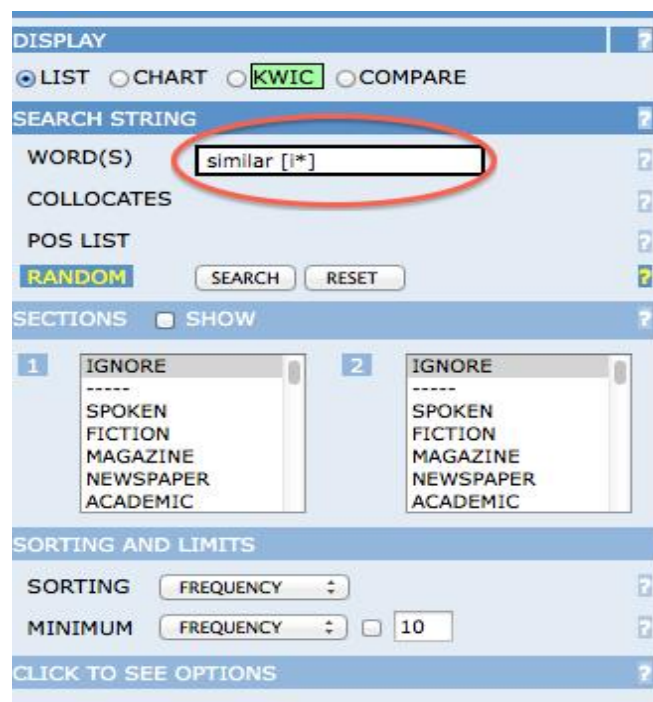
Figure 1. Searching collocations in COCA using the basic search field

The results show the most frequent collocates of '*similar*' and their frequency of occurrence.

Figure 2. Most common prepositions that collocate with *similar* and their corpus frequencies

A click on the particular collocation provides concordance data, that is, the keywords displayed in context. Concordance data includes information about the year in which when the phrase was used, the text-type and the sub-genre from which the phrase was extracted.

Figure 3. A sample of concordances *similar to*

This analysis can be extended by using the KWIC option, which displays search words in context using a colour code for different parts of speech.

Another way to search collocations in COCA is by entering a word string in the search field (e.g. *similar to*) and then use the COLLOCATES function to take advantage of additional search options. The COLLOCATES function, for instance, makes it possible to set the collocation range. For research purposes, the span is conventionally set at $\pm$ 4, that is four orthographic words to the left or to the right of the node, as this is a distance at which 95% of collocational influences have been found to occur (Jones & Sinclair, 1974). This span, however, can be modified depending on the research question that we are trying to answer.
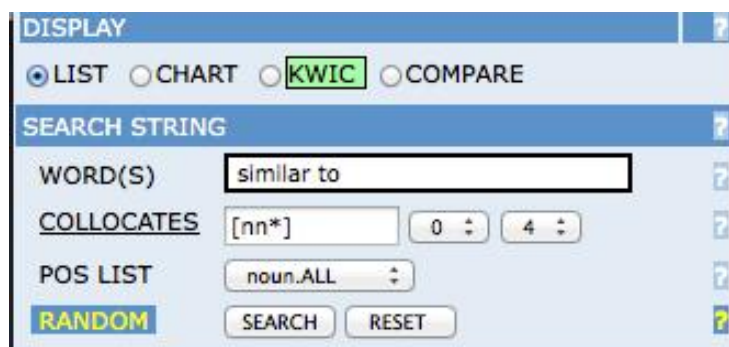
Figure 4. Collocation search with COLLOCATES option

COLLOCATES search can also be useful when users are not sure which word might be suitable in a particular phrase, but they are aware of the part of speech that they are looking for. For example, a learner may not know a suitable preposition for a sentence "I am going to Cuba _____ Christmas break". COCA has a so-called 'wildcard' search option that allows the user to search for a specific phrase collocates by part of speech only. Typing *Christmas break* in the WORD box and selecting "prepALL" tag in the POS LIST shows that *during* and *over* may be good choices in this context.

COCA can also help learners adjust their level of formality. For example, if we enter the phrase '*You're kidding* 'and select the CHART option, we will get the following results:

SECTION	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
FREQ	690	233	385	39	30	3
PER MIL	1.49	2.44	4.26	0.41	0.33	0.03

Figure 5. Collocation distribution analysis by text type

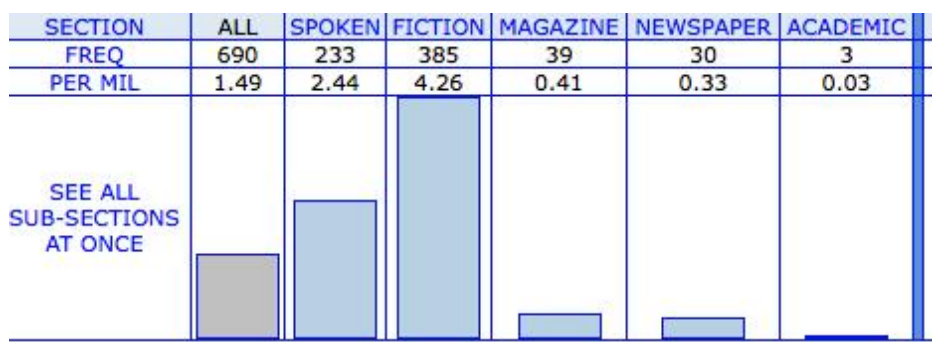The bar chart clearly shows that *'You're kidding '*is most likely to occur in fiction and spoken language and is rarely found in journalistic publications or academic work.

The use of the COMPARE button allows a comparison of collocates of two different words at the same time. This function can help learners to select a more common collocate of a particular word. For example, if a learner wants to know whether it is better to say *reduce stress*or *decrease stress*, he/she should select COMPARE, enter *reduce* and *decrease* in the search fields, type *stress* in the COLLOCATES box, set the collocate range and click on SEARCH. In the example given below, the range was set at 0:4, which means that collocates should appear within 4 word slots <u>after</u> the search words.

Figure 6. Collocation search with COMPARE option

A COCA search produced the following results.

WORD 1 (W1): REDUCE (4.45)						WORD 2 (W2): DECREASE (0.22)				
	WORD	W1	W2	W1/W2	SCORE		WORD	W2	W1	W2/W1	SCORE
1	STRESS	410	32	12.8	2.9	1	STRESS	32	410	0.1	0.3

Figure 7. Results of collocate search in COCA using COMPARE option

The word *reduce* (W1) was used 410 times, while *decrease* (W2) was used only 32 times. Therefore, a learner can conclude that *reduce* would be a better choice.

Some limitations of corpora-based collocation research are that the results obtained are based only on the statistical probability of particular words occurring within a certain span. The data reveals little about the degree of opacity and the substitutability of the combinatory elements of particular formulaic sequences. Computation scores do not account for homonymy and polysemy and the figures tell little about the phraseological significance of the extracted text. Nevertheless, there is no doubt that online corpora can be a very useful tool in foreign language instruction. Computer-based corpora such as COCA, allow teachers to select the phrases that should be given priority, and present learners with a large number of contextualized examples as well as quantitative data that can help them to make better syntactic and stylistic choices.

4. **HOW should collocations be taught?**
4.1 Dictionary training

Like online corpora, dictionaries can also be a very useful tool in collocation learning. In addition to meaning explanations and grammatical notes, learners' dictionaries today include common and useful collocations that can help learners improve their fluency, precision and naturalness of expression. Useful collocations are typically highlighted in bold type tomake them more salient. In some learners' dictionaries such as the *Cambridge Advanced Learner's Dictionary* (CALD), collocations are grouped in semantic sets making it easier for the learner to select the appropriate word combinations. The CD-ROM version of the CALD dictionary also has a special *collocation* search option which allows learners to quickly access detailed information about the collocates of the headword quickly. The collocates are grouped by part of speech and followed by examples of usage. The entries also include information about the formality level of individual expressions aimed at helping learners select the appropriate language register.

The problem, however, is that due to the limited amount of space; collocation information in general dictionaries is seldom presented in detailed and systematic way. Sometimes, there is no clear division between the literal and extended meanings (Revier, 2009), and learners are left to decide for themselves which of the word combinations that they

encounter in a lexical entry are significant and worth remembering (Handl, 2009).

In an effort to correct these shortcomings, a number of specialized collocation dictionaries have been developed in recent years.

The *Oxford Collocations Dictionary for Students of English* (OCDSE) released in 2002 defines different meanings of polysemous words and provides a list of common collocates for each word sense with some usage examples. The collocates are grouped according to their grammatical category and listed in alphabetical order. Phrasal verbs are treated separately at the end of the verb entries. The dictionary also includes some information about the register, usage restrictions and figurative meaning of the phrases.

The *Macmillan Collocations Dictionary*, first published in 2010, was the first fully corpus-based dictionary specially designed to help upper-intermediate to advanced students produce natural and more accurate English. The headwords are printed in red font, and when a word has more than one meaning, each meaning is numbered and followed by a definition. Codes show the grammatical relationship between headwords and collocates (e.g. adj+N; v+N). Collocates

are semantically grouped and followed by examples of usage in which common prepositions are highlighted in bold type. The dictionary also includes notes about the typical usage of the expressions.

As in the examples above, collocation dictionaries are comprehensive and include information that can potentially be highly useful to the learner. However, as Laufer and Kimmel (1997) point out, the usefulness of a dictionary, that is, the extent to which a dictionary provides the necessary information to the user, is not the same as its usability, which is defined by the user's willingness to use a particular dictionary and their satisfaction with it. Klotz (2003) points out that while the overall layout of the entries in OCDSE is clear, having the collocates listed in alphabetical order can make it difficult for the learner to distinguish between near-synonyms listed in the same entry. A similar observation was made by Komuro (2009) who monitored Japanese learners' OCDSE use during a translation task and found that learners often felt overwhelmed by a large number of collocates presented together. The questionable semantic grouping of the collocates combined with structural differences between L1 and L2 made it difficult for them to retrieve the correct sytnagmatic phrases.

These findings suggest that learners are likely to benefit from some dictionary use training. This can be done by presenting learners with the questions that would prompt them to examine the information about collocations in their dictionaries. For example, learners may be asked to consult their dictionary and list the collocates of the target word that they can find, or to observe how these collocates are marked and whether or not their dictionary indicates the level of formality of the collocations. A comparison of the entries for the same word in several dictionaries can also raise learners' awareness about the differences that exist between them and the ways that they can be used.

McCarthy and O'Dell (2005:11) suggest an activity in which the learners are asked to group the collocates based on their meaning. The learners are asked to classify expressions such as *to suffer pain, to alleviate pain, to be racked with pain* etc. in one of the following three groups: 1) expressions that mean 'making others experience pain', 2) phrases that refer to the experience of 'being in pain' and 3) expressions that convey the idea of 'making pain go away'. This exercise requires from learners to closely examine the usage examples in their dictionaries and can help them learn the collocates in semantic groups.

In short, both general and specialized collocation dictionaries can help students improve their vocabulary usage, but learners need to develop dictionary skills in order to take advantage of them.

4.2 Teaching Activities

The limited exposure to the target language and learners' tendency to overlook chunks when they do not cause comprehension problems are strong arguments for the explicit teaching of collocations. However, considering the limited time that most learners have at their disposal, instruction must be efficient and effective. Therefore, collocations must be a part of the planned language input. In their excellent book *Teaching Chunks of Language*, Lindstromberg and Boers (2008) proposed the following three-stage programme for teaching multiword chunks:
1) helping learners notice chunks and raising their awareness of chunk importance;
2) helping learners commit chunks to memory;
3) helping learners consolidate knowledge that they acquired through review.

Given the large number of lexical chunks that exist, teachers must be highly selective in their choice of target phrases. The frequency of use is one of the most commonly applied criteria, and as

discussed earlier, dictionaries and online corpora can help instructors make more informed choices. Teachers must also be selective with regard to the number of collocation patterns they may want to introduce. In some resource books such as *BBI Combinatory Dictionary of English* (Benson, Benson & Ilson, 1986), a distinction is made between lexical and grammatical collocations. In a grammatical collocation, a node (noun, adjective or verb) is combined with a preposition or grammatical structure. Lexical collocations are made of nouns, verbs, adjectives and adverbs only. Benson and his colleagues identified 7 types of lexical and 8 types of grammatical collocations, which, with a number of subcategories, resulted in over 30 collocation patterns. While these patterns may be interesting for research purposes, they may be overwhelming for an ordinary language learner. In my experience, for teaching purposes, it is more effective to restrict the term *collocation* to the following seven patterns:

1) Adjective + noun

2) Noun + verb

3) Noun + noun

4) Verb + noun

5) Verb + preposition

6) Verb + adverb

7) Adverb + adjective.

To help students notice the target phrases, they could be asked to read a text and then complete a chart with one or more of the patterns above. Their choices could then be examined through class discussion, and the teacher could draw their attention to some important phrases that they might have missed. Lindstromberg and Boers (2008) also suggest activities such as reading out loud with pauses and memorizing short dialogues. Another interesting activity that they propose is text reconstruction, a modified form of 'dictogloss' with the focus on lexical chunks. The teacher selects the important phrases from a text, writes them on the board and makes sure that the learners understand them. After that he/she reads the text, which contains all the chunks aloud. The learners are then asked to work in pairs or groups, and use the chunks that they can see on the board to reconstruct the text. Another activity that is both fun and conducive to learning is putting chunks into chronological order. Chunks related to romantic relationships work particularly well. Lindstromberg and Boers (2008) suggest the following phrases: *make up, pop the question, tie the knot, hang out together, chat…up, have a row, break off their engagement, catch…eye, get on, be engaged to be married, fall in love, hit it off, set a date for* and *so far so good.* The learners are presented with a list of jumbled sentences that contain the phrases above, asked to guess their meaning, and to make a

typical story. Later they can be asked to tell their stories without looking at the paper.

In addition to helping learners notice collocations, teachers should also try to help them remember common word combinations. This can be done by making the learners aware of the linguistic motivation of multiword chunks. Lindstromberg and Boers (2008) identify three types of linguistic motivation:
1) the influence of the past, of culture, and of economics;
2) the influence of register and genre;
3) the influence of the repetition of sounds. For example, if the teacher explains that words which have their origin in Norman French or Latin tend to be more formal than words of a Germanic origin, it may be easier for learners to understand and remember why it is more common to say *remain in custody* as opposed to *stay in custody.* Recognition of sound repetition patterns may also have a positive mnemonic effect. As Lindstromberg and Boers observe (2008), while words from different kinds of combinations, the ones that provide sound repetition sound more natural. For instance, sound repetition may explain why it is more common to say *boy bands* as opposed to *guy bands, head of the house* and not *boss of the house,* or *to go from hero to zero* rather than *to go from hero to nothing.*

Finally, just like in the case of individual words, review is crucial for the acquisition of multiword chunks. Learners must be given opportunities to encounter the target phrases, in different contexts and in a relatively short period of time, so that memory traces can be formed. For this reason, it is important that the instructors keep a record of the phrases that were covered in the course. One activity that could be used for the consolidation of collocations is giving learners a text that they have already encountered with slashes indicating the phrases that they should recall. Lindstromberg and Boers suggest this be done as a pair activity where one student reads the text and pauses and the other has to guess the word or the rest of the phrase. Alternatively the students could be given a list of the target phrases and asked to place them in a new text.

Another way collocations can be reviewed is by asking students to correct the sentences containing collocation errors such as in the following example:
I succeeded my dream. □*I achieved / accomplished my dream.*
Problematic collocations can also be practiced through a translation exercise, so that the students can re-notice the differences between L1 and L2 word combinations. For example, English

students of Spanish may be asked to translate a sentence such as: *I spent the whole year studying Spanish* where the literal translation of the verb *spend* as *gastar* would be inappropriate.

Vocabulary substitution exercises can also be useful ways of consolidating learners' collocation knowledge. *Focus Paraphrase* (Vasiljevic, 2008) is a pair work activity in which learners are provided with a list of words grouped by part of speech from which they need to form appropriate or adjective-noun collocations so that they can paraphrase the target L2 sentences.

Adjectives

Nouns

Example: *strong   big   heavy*

vehicles   transportation   traffic

There were many cars on the street that day. ☐ *Traffic was heavy that day.*

To make the activity more interactive, the learners work in pairs. One student has to paraphrase a sentence and the other acts as a "coach", who has the model answer and can correct his / her partner. The activity has two parts, so that each student gets to play both roles.

Collocation Bingo (Vasiljevic, 2008) is another activity that can be used to review the words that the students encountered in the class materials and to consolidate their collocation knowledge. The teacher reads a list of associates that the learners should be familiar with and the students need to identify the target word.

Sample Bingo Card

Teacher: *"delicious   exquisite   rich   strong   mellow   subtle….."*

ingredients	restaurant	menu
specialty	order	serving
flavour	portion	cuisine

In short, collocation learning requires teacher-led guidance as learners tend to overlook the common word combinations in the text and they often lack the knowledge and language sensitivity to take advantage of linguistic motivation behind the meaning and the form of multiword phrases. For learners to remember collocations, it is necessary to have well-planned and well-structured instruction that will prompt them to notice important collocations

and then help them remember them through multiple exposures and activities that promote deep level processing and draw on linguistic motivation with regard to multi-word phrases.

5. Testing

In order to assess learners' needs and their lexical progress, it is necessary to have good and reliable tests of their phrasal knowledge. Regular testing of collocations may have a positive backwash effect leading to an increase in learners' awareness of the importance of lexical chunks. However, currently there are still no standardized tests that can objectively measure learners' syntagmatic competence in the second language. As Eyckmans (2009) points out, some reasons may be the difficulties associated with defining lexical phrases. Multiword phrases differ in their lexical composition, in the function they have, as well as the collocation patterns they exhibit. Some popular ways of testing collocation knowledge are:

1) L1-L2 translation;

2) L2 sentence cloze items;

3) sentence generation tasks;

4) discrete tests where learners are presented with a node-word prompt and asked to select or supply one or more of its collocates.

However, each of the test formats above has some limitations. For example, it would not be practical to use a translation test in a context in which learners come from different language backgrounds. In sentence cloze format, leaving out the whole collocation has been found to result in multiple responses with different types of collocations. A more restricted format where the first one or two letters of the lexical constituents were provided made some learners focus on finding the words that matched the letters so much that they overlooked the propositional meaning of the sentence prompt (Revier, 2009). Sentence generation format can be time-consuming, and therefore the number of items that can be tested at one time is limited. Some studies (e.g. McNeill, 1996) also suggest that advanced learners can sometimes produce acceptable sentences even if they do not have a good understanding of the target words, which raises concerns about the validity of this test format. The scoring may be arbitrary too. Discrete tests where learners are asked to select or produce associates of a node-word do not provide sufficient information of the learner's knowledge of the whole collocation (Revier, 2009).Just like single words, collocations have their own formal, semantic and usage properties. For example, productive knowledge of a verb - object noun collocation requires knowledge of not only the phrase meaning but also its grammatical properties such as noun determination and number.

The limitations above made Revier (2009) argue that collocations should be viewed as an independent construct and tested as a whole. He developed a new test format entitled CONTRIX, which presents a modified form of a cloze test where learners are asked to select the combination of a verb, article and noun that best complete a sentence.

The quickest way to win a friend's trust is to show that you are able to	tell	a / an	joke
	take	the	secret
	keep	☐	truth

(Revier, 2009:129)

Revier argues that although the test involves selection, which is a characteristic of receptive tests, the CONTRIX format requires test takers to combine phrase constituents and grammatically encode the noun constituent for determination, providing some insight into the learners' productive vocabulary knowledge. Revier pilot-tested the CONTRIX format with a group of Danish EFL learners and found that the test met the criteria of validity and reliability and proved to be very effective in distinguishing among learners of different L2 proficiency. Some downsides of this test format are that sentence prompt writing, distractor selection, and native-speaker norming make the test somewhat time- consuming to prepare. Furthermore, as Revier himself observes, while the test validity could be improved by increasing the number of items per section, there is a risk that a longer test may have lower reliability due to the possible fatigue of test takers, in particular those at lower levels of proficiency.

Gyllsatad (2009: 157~158) offers two new test formats for testing collocation knowledge. In the first one entitled COLLEX (**co**llocating **le**xis), the learners are presented with three word sequences and asked to select the one that is most likely to be used by native English speakers.

a. drive a business   b. run a business c. lead a business   a   b   c

The second format known as COLLMATCH is basically a Yes / No test, where learners are asked to judge whether or not the given word sequences are common English collocations.

	catch a cold	*draw a limitation*
	yes	yes
	no	no

For both tests, an important issue is the selection of the target phrases. Gyllstad suggests that one possibility would be to sample a large corpus such as BNC for a particular syntagmatic pattern (e.g. V + NP)and then use a stratified random sampling technique to select the target items. This approach however, would also require some manual analysis in order to eliminate idioms and free word combinations.

Some research suggests that intonation patterns may also be indicative of the learners' collocation knowledge. In fluent, coherent speech, phraseological sequences tend to be produced with no or little hesitation and with unbroken intonation contour (Lin & Adoplphs, 2009; Moon, 1997; Wray, 2004).If multiword phrases are pronounced as single intonation units, then prosodic breaks can show where language chunking takes place. Lin and Adolphs (2009) predicted that as learners' speech tends to be slower and contains more hesitations, smooth and fluent stretches of speech are likely to be salient, which in turn would make the presence of phraseological units more explicit. This hypothesis was partially confirmed. A complete match between phraseological unit boundaries and intonation unit boundaries was found 55% of the time. While these results raise some concerns about the test validity, some mismatches were caused by the use of conjunctions at the beginning of the phrases and word repetitions. Further studies are needed in order to determine to what extent phraseological units determine intonation boundaries, and whether intonation patterns can be used to identify multiword phrases in the learner language.

6. Learner Autonomy

In spite of the great advances in corpus research and lexicography, very little is still known about how learners approach collocation learning, how they achieve a particular level of performance, and how they interpret their own learning practices. One of very few studies which has examined individual learner development of collocation knowledge was conducted by Barfield (2009), who followed collocation learning practices of four language learners over one academic year. He identified five major processes of development:

- Understanding and reconfiguring past vocabulary practices;
- Interpreting different worlds of everyday use;
- Moving from quantity of lexical knowledge to quality of collocation use;
- Reconnecting what is known and projecting new identities;
- Developing authorship(Barfield, 2009:211-212).

Barfield's observations confirmed the tendency of second language learners to identify vocabulary learning with expansion of vocabulary size. However, with practice, his students became aware of the limitations of this approach, and modified their learning strategies. For example, one of the students started to question her own learning strategies as she got more experienced with collocation learning. She noticed that "more" does not necessarily mean "better", and that indiscriminate recording of collocations just resulted in long lists of phrases that she found difficult to remember. She finally opted for a simpler form of representation that entailed creating small lexical networks that she labeled 'collocation packages' that enabled her to skillfully use a highly limited number of collocation choices. This transformation can clearly be seen in the following two figures:

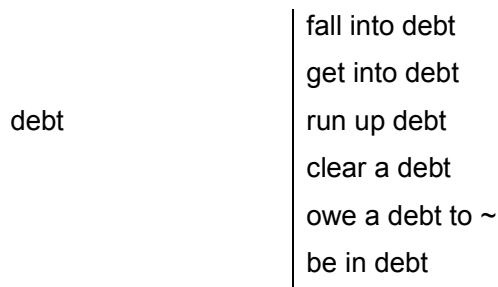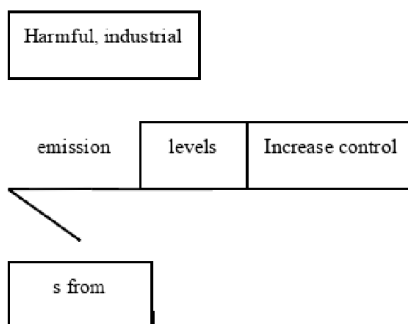|	fall into debt
debt	get into debt
	run up debt
	clear a debt
	owe a debt to ~
	be in debt

Figure 8. A sample of a learner's collocation notes at the beginning of a school year (Barfield, 2009: 214).

"Developed countries should control industrial emission levels."

Figure 9. A sample of a learner's collocation notes at the end of a school year (Barfield, 2009: 221)

Another learner reported reducing the number of collocations and moving from two-word combinations to longer phrases. This shift from quantity to quality is believed to reflect the greater control that learners establish over their learning (Benson & Lor, 1998).

This process of transformation was fostered by the teacher inviting learners' to reflect on their practices and through collaborative learning with other students. Barfield (2009) asked his students to interview each other about their collocation learning experiences. In order to prepare for the interviews they had to look into their collocation development, and reconsider the reasons behind the selection of particular phrases, their ways of recording, and the strategies they used to remember them.

Barfield also observed that learners' collocation development was to some extent affected by the process of sociocultural reorganization that they underwent in the course of language learning. The way the learners selected the vocabulary, and their willingness to use the phrases, were influenced not only by their communicative needs but also by their beliefs and attitudes, and the new social identities that they assumed during the learning process. For example, some students reported that good command of collocations helped them sound '*less Japanese'*, '*more international'* and *'cool'.*

These results are encouraging. Japanese students are known for their conformist and teacher- dependent learning style, and for being part of an English language education where vocabulary learning is often identified with memorizing long lists of decontextualised

words in preparation for the university entrance exam. The transformation of Barfield's students shows that with some guidance learners are capable of and willing to adopt more autonomous learning practices and shift their priorities from the quantity of lexical knowledge to the quality of collocational use. The results of other studies suggest that just raising learners' awareness of the importance of collocation knowledge is likely to make them become more aware of multiword phrases, even when the instruction is not collocation-oriented. For example, Peters (2009) observed that advanced EFL learners who were familiar with the concept of collocations, were likely to remember vocabulary in phrases even when collocations were not explicitly targeted as part of the learning task. This finding is significant as it suggests a long-term positive effect of consciousness-raising activities; if learners recognize the importance of collocations they may be able to overcome some of the deficiencies they may encounter in the teaching materials and in classroom instruction.

One way of helping learners become more selective and more independent in their collocation learning is by giving them some practice with the use of online corpora. Corpus analysis can help learners focus their attention on the phrases with the highest frequency of use. Corpus data can also be used to promote learners' autonomy in error correction. Self-correction is an important aspect of language learning that raises both the learner's language awareness and his / her confidence. Therefore, helping learners become more self-sufficient should be an important teaching objective and students must be given opportunities to make adjustments to their language production. However, error correction, if left unguided, can turn into a very frustrating experience. For the reasons discussed earlier, collocation errors may be particularly difficult to detect. Therefore, the teachers can help learners by letting them know that they have a problem in terms of word choice. For example, a student may say *I succeeded my dream.* The teacher should underline the incorrect word and ask the student to look for a more appropriate expression. This can be done by using the earlier explained 'wildcard search' or doing a synonym inquiry as below.

Figure 10. Synonym search in COCA

The results of the search show that *achieved* may be the best synonym in this context.

In short, considering the amount of language that needs to be covered in the classroom, it is essential to equip learners with the skills that will enable them to learn collocations and confirm their vocabulary choices outside the classroom. By encouraging learners to reflect on their learning practices and offering them some training in the use of online corpora, teachers can help students take control over their vocabulary learning.

7. Conclusion

This paper has highlighted the importance of incorporating collocation instruction in the L2 classroom, discussed the resources that could help teachers and learners select their vocabulary learning targets, introduced some activities that could help learners to notice and remember the L2 collocations, and

examined some test formats through which learners' needs and progress could be assessed. Attention was also paid to the challenges that learners face in reconfiguration of their learning practices during their transition to more autonomous productive word learning and resources that could help them in that process.

As discussed above, many of the features relevant to vocabulary learning in general, such as noticing, depth of processing, and multiple encounters with the target phrases are applicable to collocation learning as well. First, in order to revise their learning priorities, students must recognize the value of a chunk-based approach to vocabulary learning. Unselective, mechanical recording of teacher-selected phrases is not likely to produce the desired effect. What is needed is learners' active engagement in the learning process. They need to redefine what word knowledge entails and recognize that vocabulary learning is more

than just adding single new words to the lexicon. It is only when learners can see the significance of collocations that they will persist in their efforts to master them. Teachers play an important role in guiding students to appreciate the importance of collocation knowledge.

Second, learners need class instruction where selected multiword phrases are targeted explicitly. Learners sometimes lack the cognitive resources to notice new language elements in the input (Laufer & Hulstijn, 2001), and therefore simple exposure to authentic language may not be sufficient for phrase uptake. Planned and structured in-class learning tasks are necessary to help learners commit language chunks to memory and to ensure that they have multiple encounters with the target phrases in order to ensure consolidation of their collocation knowledge.

Finally, learners need to acquire the cognitive and metacognitive strategies that will help them to identify and remember multiword combinations effectively outside the classroom, and enable them to monitor their progress and take control of their vocabulary learning. Given the enormous number of collocations in the English language, even the most dedicated teachers will only be able to teach a small fraction of them.

Furthermore, while collocation learning may begin with the acquisition of the most frequent word combinations, in the end, it is the learners' communicative needs and their perception of usefulness that will determine which word combinations will be attended to. To this end, in addition to explicit teaching of the selected multiword phrases, classroom instruction should incorporate dictionary use training and practice in using online corpora such as COCA. In addition, students should be introduced to different ways of making collocation notes and prompted to discuss their strengths and weaknesses. They should also be encouraged to experiment with various collocation recording formats until they develop the style that best suits their needs. Class practice should be complemented with homework activities designed to promote more autonomous learning. For example, students could be asked to keep journals in which they reflect on their learning. These reflections should prompt learners to reexamine their learning strategies and make adjustments where necessary.

8. Future directions

One of the remaining challenges for teachers and for textbook writers is establishing better connections between EFL materials and the results of experimental research. In spite of the great progress in corpus studies and

lexicography, a large number of the existing course books target too few language chunks and do not do enough to help learners remember them (Lindstromberg & Boers, 2008). New insights about the various kinds of collocational relationships should be reflected in the teaching materials as well as in classroom methodologies.

More research is needed to find out how learners use dictionaries and other reference materials, and which strategies they employ to master L2 collocations.

There is also a need for new, more sensitive ways of tracking the development of learners' collocation knowledge, and assessing the effectiveness of different kinds of input on learners' uptake of lexical phrases. It is hoped that this paper will encourage teachers to adopt the chunk-based approach to vocabulary instruction and prompt them to further explore creative and communicative ways through which such an approach can be implemented in the language classroom.

References:

Arnaud, P. J. L. & Savignon, S. J. (1997). Rare words, complex lexical units and the advanced learner.
In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 153-173). Cambridge University Press.

Barfield, A. (2009). Exploring productive L2 collocation knowledge. In T. Fitzpatrick and A. Barfield (Eds.), *Lexical processing in language learners: Papers and perspectives in honour of Paul Meara* (pp. 95-110). Clevendon: Multilingual Matters.

Cambridge Advanced Learner's Dictionary (2014). Cambridge: Cambridge University Press.

Benson, M. Benson, E., & Ilson, R. (1986). *The BBI combinatory dictionary of English.* Amsterdam: John Benjamins.

Benson, P. & Lor, W. (1998). *Making sense of autonomous language learning: conceptions of learning and readiness for autonomy.* English Centre Monograph, No.2, Hong Kong: University of Hong Kong.

Eyckmans, J. (2009). Toward an assessment of learners' receptive and productive syntagmatic knowledge. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 139-152)*.* London: Palgrave Macmillan.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. In F.R. Palmer (Ed.), *Selected papers of J.R. Firth 1952-1959* (pp.168-205). London: Longman.

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

Halliday, M.A.K. (1966). Lexis as a linguistic level. In C.E. Bazell, C. Catford, M.AK. Halliday, & R.H.

Robbins (Eds.), *In memory of J.R. Firth* (pp. 148-162). London: Longman.

Handl, S. (2009). Towards collocational webs for presenting collocations in learners dictionaries. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 69-85),

London: Palgrave Macmillan.

Howarth, P. (1998a). Phraseology and second language proficiency. *Applied Linguistics 19* (1): 24-44.

Howarth, P. (1998b). The phraseology of learners' academic writing. In A.P. Cowie (Ed.), *Phraseology:*

theory, analysis and applications (pp. 161-186). Oxford: Oxford University Press

Jiang, J. (2009). Pedagogic materials for L2 collocation use. In A. Barfield & H. Gyllstad (Eds.),

Researching collocations in another language (pp. 99-113), London: Palgrave Macmillan.

Jones, S. & Sinclair, J. (1974). English lexical collocations: A study in computational linguistics. *Cahiers*

de Lexicologie 24 (1): 15-61.

Klotz, M. (2003). Review of "Oxford Collocations Dictionary for Students of English"*. International*

Journal of Lexicography 6 (4): 300-304.

Komuro, Y. (2009). Collocation dictionary accessibility. In A. Barfield & H. Gyllstad (Eds.), *Researching*

collocations in another language (pp. 86-98). London: Palgrave Macmillan.

Laufer, B. (2005). Focus on form in second language vocabulary learning. In S.H. Foster-Cohen, M.P. Garcia-Mayo & J. Cenoz (Eds.), *EUROSLA Yearbook 5*: 223-250. Amsterdam: John

Benjamins.

Laufer, B. & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics 22* (1): 1-26.

Laufer, B. & Kimmel, M. (1997). Bilingualised dictionaries. How learners really use them. *System 25* (3): 361-369.

Lewis, M. (2000). Learning in the lexical approach. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp.155-185). Croatia: Thomson.

Lin, P.M.S. & Adolphs, S. (2009). Sound evidence: Phraseological units in spoken corpora. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 34-48).

London: Palgrave Macmillan.

Lindstromberg, S. & Boers, F. (2008). *Teaching chunks of language.* Cambridge: Cambridge University

Press.

Macmillan Collocations Dictionary (2010). UK: Macmillan.

McCarthy, M. & O'Dell, F. (2005). *English Collocations in Use.* Cambridge: Cambridge University

Press.

McNeill, A. (1996). Vocabulary knowledge profiles: evidence from Chinese-speaking ESL teachers. *Hong Kong Journal of Applied Linguistics 1*: 39-63.

Moon, R. (1997). Vocabulary connections: Multiword-Items in English. In N. Schmitt & M. McCarthy

(Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 40-63). Cambridge: Cambridge University Press.

Oxford Collocations Dictionary for Students of English (2002). Oxford: Oxford University Press.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications

for teaching. *Applied Linguistics 24* (2): 223-242.

Pawley, A. & Syder, F. H. (1983). Two puzzle for linguistic theory: Nativelike selection and native like fluency. In J. C. Richards & R.W. Schmidt (Eds.) *Language and communication* (pp.

191-226). London: Longman.

Peters, E. (2007). Manipulating L2 learners' dictionary use and its effect on L2 word retention. *Language, Learning & Technology 11* (2): 36-58.

Peters, E. (2009). Collocation and attention-drawing techniques. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 194-207). London: Palgrave Macmillan.

Revier, R.L. (2009). Evaluating a new test of *whole* English collocations. In A. Barfield & H. Gyllstad
        (Eds.), *Researching collocations in another language* (pp. 49-59). London: Palgrave
Macmillan.

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

Singleton, D. (1999). *Exploring the second language mental lexicon.* Cambridge: Cambridge University
        Press.

Siyanova, A. & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review 64* (3): 429-458.

Vasiljevic, Z. (2008). Developing collocational competence of second language learners. *The East Asian Learner 4* (1): 46-50.

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

Wray, A. (2004). "Here's one I prepared earlier": Formulaic language learning on television. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 249-268). Amsterdam: John Benjamins.

Yang, Y. & Hendricks, A. (2004). Collocation awareness in the writing process. *Journal of Reflections on English Language Teaching* 3: 51-78.

*****

Zorana Vasiljević is associate professor at the Faculty of Language and Literature at Bunkyo University, Japan. She holds a PhD in Applied Linguistics (University of Queensland). Her research interests include vocabulary acquisition, EFL methodologies and learner autonomy.