**Expertise in L2 Vocabulary**

**Marijana Macis**

Manchester Metropolitan University

**Mélodie Garnier**

University of Cambridge

**Laura Vilkaite**

Vilnius University

**Norbert Schmitt**

University of Nottingham

**Introduction**

In order to communicate and express meaning in both their first (L1) and second language (L2), people need a large vocabulary. Thus, knowledge of vocabulary is fundamental to all language use and must be learned in some manner in order for learners to become communicative in a new language. However, the lexicons of most languages are extensive with English having the largest vocabulary of any known language (Schmitt & Marsden, 2006) and a vast number of loanwords from other languages. Therefore, learning vocabulary is an essential part of mastering a second language as it is likely to be the biggest hurdle for successful communication.

There is plenty of evidence pointing to the importance of vocabulary in language use. It has been acknowledged that vocabulary knowledge is a good predictor of general proficiency in a second language with strong relationships between vocabulary measures and the ability to read, write, listen, and speak (e.g. Laufer & Goldstein, 2004; Zimmerman, 2004; Stæhr, 2008). Generally speaking, the more words learners know, the more proficient they become, and the better they are likely to perform whatever the skill. Considering the multitude of factors that could affect learners’ performance (e.g. learner motivation, background knowledge, familiarity with test task), it is striking that the single factor of vocabulary accounts for such a large percentage of the variation in the performance of the four skills (e.g. 40%-62%, Alderson, 2005). In short, vocabulary knowledge is clearly a major contributor to success in language performance.

Since knowledge of vocabulary is essential to function in English, it is important to know how much vocabulary is needed to accomplish specific tasks (e.g. reading a book in English, listening to a lecture etc.). This is because mastery of the complete lexicon of English is beyond not only second language learners, but also native speakers, as English is estimated to have 54,000 word families (Goulden, Nation, & Read, 1990). A word family includes a base word, its inflected and derived forms (Hirsh & Nation, 1992); for example, the word family ‘nation’ has 26 members and *national, nationalism, nationalist, nationalize*, and *nationhood* are just few of them. This roughly corresponds to a dictionary entry, and few native speakers would be expected to know every word in a dictionary. It is estimated that well-educated native speakers know around 20,000 word families, excluding proper names and derived forms (Goulden et al., 1990; Zechmeister, Chronis, Cull, D’Anna, & Healy, 1995). However, reaching this vocabulary size would be a daunting, and probably unmanageable task for most second language learners. Fortunately, second language learners do not need to achieve native-like vocabulary sizes in order to use English well. Instead, they require much smaller amounts of vocabulary in order to be functional in specific communicative contexts.

The primary focus of this chapter is to describe the L2 vocabulary expertise necessary to be functional in a variety of situations, such as reading different types of books, watching films, and taking part in a conversation or a seminar. We believe there are three key requirements necessary to achieve this expertise. The first requirement is learning enough words to communicate successfully in English. Without a large enough vocabulary, neither language production nor language comprehension is possible. The second requirement is learning various aspects of word knowledge, which involves knowing enough information about each word. The third requirement is learning how words combine to form longer phrases. These word combinations are as important as individual words, and thus are crucial for reaching high levels of proficiency.

In the following sections, we present a more detailed account of lexical coverage and vocabulary size, vocabulary depth, and formulaic language, and highlight their importance for gaining expertise in L2 vocabulary.

**Lexical Coverage**

**How Much Coverage is Needed for Successful Comprehension?**

Before one can calculate the vocabulary size necessary to function in English, it is first necessary to determine the percentage of words in a text that a person needs to know in order to comprehend that text (*lexical coverage*). This is likely to vary between reading and listening for several reasons. Readers can refer back to lexical items in a text to decode the message, whereas this is not possible for listeners (Reves & Levine, 1998). Listeners often struggle to recognise individual spoken lexical forms correctly, while this is not the case in reading. Yet another difference is that spoken discourse tends to be lexically less dense (i.e. containing a lesser proportion of content words such as nouns, verbs, adjectives and some adverbs) than written text (Carter, 2012) and contains many fillers and repetitions (Rost, 2002). Thus, the lexical coverage required is different for written versus oral modes.

**Reading**

Most research has been conducted in relation to reading, with fewer studies focusing on coverage in listening. Hu and Nation (2000) was one of the first studies to suggest lexical coverage requirements in reading. The authors examined the relationship between coverage and reading comprehension in a fiction text. Four different coverage degrees were created (80%, 90%, 95% and 100%) by replacing some real words in the target text with non-words (e.g. *rane, sorant, yeard*), and then testing the resultant L2 comprehension. Results showed that no learner achieved adequate comprehension (conceptualized as the score that most learners in the 100% coverage group received) at 80%, while only a few did at 90% and 95%. Adequate comprehension was achieved somewhere between 95% and 100%, which led the authors to suggest that it takes 98%-99% coverage to allow unassisted reading for pleasure.

Likewise, Schmitt, Jiang, and Grabe’ study (2011) explored whether more vocabulary would lead to more comprehension. The participants completed a vocabulary checklist test based on words drawn from two academic texts, then read the texts, and completed a reading comprehension test for each. Results demonstrate a linear relationship between the percentage of vocabulary known and text comprehension, which indicates that the coverage level required depends on the degree of comprehension aimed for. If 60% comprehension is considered adequate, then 95% coverage is necessary. If the goal is 70% comprehension, then 98%-99% coverage is required. If the aim is 75% comprehension, then the data suggests 100% coverage is needed. This indicates that factors other than coverage contribute to full comprehension (e.g. background knowledge, text features, ability to infer from context etc.).

In a similar study, Laufer and Ravenhorst-Kalovski (2010) looked at the relationship between second language learners’ vocabulary level, the lexical coverage that their vocabulary provided, and their reading comprehension. Vocabulary size was measured through the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001), lexical coverage by the then newest version of Vocabulary Profile (available at http://lextutor.ca), and reading comprehension by a standardized national test based on academic texts. Two coverage threshold points were established: 98% as optimal (predicting independent reading) and 95% as minimal (which characterizes assisted reading).

Based on the results of the above studies, the consensus is that 98% is the coverage figure which is most appropriate for most reading purposes. However, this 98% figure is derived mainly from a small number of academic texts, of limited length, which were read without access to resources. Therefore, it is yet to be determined whether non-academic texts would require more or less coverage in order to be adequately understood, and whether longer passages or greater access to resources would alter the coverage figure required.

**Listening**

There has been less research on lexical coverage in the oral mode. Bonk (2000) was the first to examine lexical coverage in relation to L2 listening comprehension. In his study, the subjects listened to four short passages with increasing percentages of low-frequency words. Dictation and written free-recall tests were used to measure the participants’ lexical knowledge and comprehension. Results revealed that higher dictation scores led to better listening comprehension. Learners with lexical coverage scores below 90% showed poor comprehension, whereas learners with lexical coverage scores over 95% showed good comprehension (i.e. achieved scores of 3 and 4 on a four-point scale that was used to rate the answers on the free-recall test).

A more recent study by van Zeeland and Schmitt (2013) directly examined the lexical coverage/listening comprehension relationship by having their participants listen to four spoken informal narrative passages. The coverage of the listening material was manipulated by replacing words from the passages by non-words resulting in different percentages of known vocabulary, namely 100%, 98%, 95%, and 90%. The best comprehension for non-natives came from the 100% passage (9.62 out of 10 maximum), then the 98% passage (8.22/10), but there was no statistical difference between 95% and 90% (7.65/10 and 7.35/10 respectively). However, there was much more variability at 90% coverage than at 95% coverage, with some learners being able to cope, but some not. This led the authors to conclude that 95% lexical coverage was necessary for adequate listening comprehension, because the participants achieved reasonable comprehension (7.65/10), but with much more consistency than at 90% coverage.

However, one of the limitations of this study is that it used informal narrative stories, and did not indicate whether the same level of coverage was required for other kinds of listening contexts (e.g. academic lectures). Therefore, while there is initial evidence that 95% lexical coverage may be adequate for listening, this figure must be established across a wide range of listening contexts, and may well be higher for more complex texts, e.g. 98%.

Overall, the above research suggests that the higher the level of expected comprehension, the higher the vocabulary demands. It also shows that even high levels of lexical coverage did not lead to 100% comprehension,which suggests that vocabulary is only one aspect of comprehension. Other factors (e.g. reader purpose, task goals, and background knowledge) are equally crucial and need to be accounted for as well. However, lexical coverage was shown to be a key factor. Research points to a lexical coverage of around 98% as necessary for adequate reading comprehension (Nation, 2006, Schmitt et al., 2015). In terms of listening, research suggests that 95% coverage may be sufficient (van Zeeland & Schmitt, 2013). However, more difficult listening contexts may require higher lexical coverage (Rodgers & Webb, 2011; Schmitt et al., 2015).

**Vocabulary Size: How Much Vocabulary do Learners Need to Know?**

Based on the above lexical coverage targets, it is possible to calculate how much vocabulary is necessary to reach those targets. Varying estimates have been put forward by different researchers, but they all seem to suggest that for everyday conversation purposes (e.g. describing everyday activities, telling the time etc.), learners need around 2,000-3,000 word families, based on a 95% coverage requirement (e.g. van Zeeland & Schmitt, 2013). To engage in a wider range of more sophisticated conversation topics (e.g. politics, studying at university), knowledge of around 6,000-7,000 word families is thought necessary, based on a 98% coverage requirement (Rodgers & Webb, 2011; Schmitt et al., 2015). In order to read a wide range of authentic texts, such as novels or newspapers, one should aim for around 8,000-9,000 word families, based on a 98% coverage requirement (Laufer & Ravenhorst-Kalovski, 2010, Nation, 2006, Schmitt et al., 2015).

There has also been research on the amount of vocabulary required to watch television programs and movies. Rodgers and Webb (2011) analysed 288 television episodes and concluded that it took knowledge of the most frequent 3,000 word families to reach 95% lexical coverage, and about 6,000-7,000 word families for 98% coverage. In a similar study, the same authors (Webb & Rodgers, 2009) examined the scripts of 318 movies and found similar coverage figures. Nevertheless, there was a considerable variation among the individual programs and movies regarding the amount of vocabulary required.

It is also important to understand how vocabulary behaves in terms of frequency. It is well known that a small number of words occur very frequently and cover a large proportion of the running words (the total number of words) in a written or spoken text. Conversely, a very large number of words occur very rarely and make up the remainder of running words in discourse. This is illustrated in Figure 33.1, which shows Nation’s (2006) analyses of nine written and spoken corpora, such as novels and newspapers, everyday conversation with friends and family, and people calling in to radio talk-show programs.

A close up of a map

Description automatically generated

Figure 33.1 shows that the first 1,000 word families clearly provide a substantial amount of coverage in English, largely due to the extremely high frequency and coverage of function words (i.e. words that express grammatical relationships with other words within a sentence, such as *and, of, or*). The second 1,000 contributes a much smaller, but still useful, percentage of coverage, as does the third 1,000 to a lesser extent. But by the fourth 1,000 families, however, the coverage decreases substantially, with a maximum of only 3% for 2,000 families (4th and 5th 1,000). Beyond this, the coverage return becomes increasingly small.

Thus, the most frequent words provide the greatest amount of coverage, and this is why learners need to know these words in particular, rather than just any words. A useful way of thinking about frequency is by distinguishing between high-, mid-, and low-frequency words (Nation, 2006, Schmitt & Schmitt, 2014). The first 2,000-3,000 word families have generally been accepted as high-frequency vocabulary and with this vocabulary size, learners can communicate to some extent. These words are necessary to use English at the basic, but still useful, end of the proficiency continuum.

Despite their usefulness, however, knowledge of these words is not sufficient for successful comprehension of a text. To illustrate this point, the following example shows what a text looks like to a second language user who only knows the most frequent 2,000 words. Words that are not on the 2,000 list have been replaced by gaps:

If \_\_\_\_\_ planting rates are \_\_\_\_\_ with planting \_\_\_\_\_ satisfied in each \_\_\_\_\_ and the forests milled at the earliest opportunity, the \_\_\_\_\_ wood supplies could further increase to about 36 million \_\_\_\_\_ meters \_\_\_\_\_ in the period 2001-2015. (Nation, 1990, p. 242.)

It is evident from the above example that someone reading this text would know what the topic is, but not exactly what is being stated about it.

The mid-frequency vocabulary includes word families between the 2,000-3,000th and the 8,000-9,000th level, whereas low-frequency lexical covers words beyond the 9,000th level (Schmitt & Schmitt, 2014). Following Nation’s (2001) advice, high frequency words should be learnt and taught explicitly, whereas the mid-frequency vocabulary could be acquired through the use of strategies (e.g. guessing from context), extensive reading and watching television programs from the same genres (e.g. Webb, 2011). Extensive reading, defined as “learning of vocabulary as the by-product of any activity not explicitly geared to vocabulary learning” (Hulstijn, 2001, p. 271), is particularly important as it allows incidental vocabulary learning to occur. Because of the large number of words in English, they cannot all be explicitly taught and the majority of them have to be learned incidentally. Even though incidental learning is a slow process, it can be effective provided that learners encounter the new vocabulary enough times within a short time span (Waring & Takaki, 2003; Webb, 2007; Pellicer-Sánchez & Schmitt, 2010).

In summary, in order to become proficient in their L2 and achieve a high level of comprehension, learners need to know a large number of high-frequency and mid-frequency lexical items to operate in authentic environments without being handicapped by unknown vocabulary.

**Depth of Knowledge**

**Not Just Quantity**

In vocabulary knowledge, as in many other areas of expertise, it is not only quantity that matters, but also quality. Having a large vocabulary size is essential to function in English, but how well each individual word is known is equally crucial. In order to gain expertise in L2 vocabulary, second language learners must know both a large number of words and enough information about each word. These two dimensions of vocabulary knowledge are commonly referred to as *vocabulary breadth* and *vocabulary depth* (Schmitt, 2010). Anderson and Freebody (1981, p. 93) were the first to develop a conceptualization of vocabulary knowledge involving both breadth and depth. They defined the former as “the number of words for which the person knows at least some of the significant aspects of meaning” and the latter as “the quality or depth of understanding”. A third dimension of vocabulary knowledge is fluency (or automaticity), which refers to the ability to access words in the mental lexicon in a fast manner. This is a somewhat neglected dimension of vocabulary knowledge, even though most people would agree that knowing a lot about a word is of limited value unless that word can be used fluently in communication. The following figure illustrates the metaphor of *lexical space* by Milton and Fitzpatrick (2014, p. 9), representing the three dimensions of word knowledge: breadth, depth, and fluency. Since the fluency dimension is rarely included in discussions of vocabulary knowledge (as some researchers see fluency as part of depth of knowledge), it will not be explored further in this chapter.

A picture containing game

Description automatically generated

Although they are often described and analyzed in isolation, Schmitt (2014) argues that the two dimensions of breadth and depth might be interrelated. In other words, when one increases, the other increases as well, and we would expect someone with a very large vocabulary size to show elaborate knowledge about the words they know. This is because a large vocabulary size suggests a high level of language exposure and learning experience, which is conducive to more detailed and refined word knowledge, and vice versa (e.g. Meara, 1996; Li & Kirby, 2014). On the other hand, at the onset of language learning, only a limited number of words are known in a rather superficial manner. At the most basic level, the form-meaning link is established and a particular string of letters (or sequence of sounds) becomes associated with a referent. For example, the string of letters *cat* becomes associated with ‘a four-legged creature that purrs’. As vocabulary knowledge develops, two processes tend to occur simultaneously: more and more form-meaning links are created (i.e. more words become known), and knowledge of the previously acquired form-meaning links is developed further and becomes more detailed and precise.

In some cases, however, vocabulary size and depth might not develop in parallel (Schmitt, 2014). It is indeed possible to learn a lot about a small number of words, and conversely, a little about a large number of words. Depending on each learner’s learning goals in L2, they might approach vocabulary learning in different ways. If the goal is to learn many words in a limited amount of time (e.g. in preparation for a high-stakes proficiency test), then acquiring only the form-meaning link of those words might be enough. Memorizing wordlists or using flashcards are two popular activities for such a purpose (Nation, 2013). On the other hand, if the goal is to focus on a specialized set of words that will be particularly relevant and useful to a learner (such as in preparation for an academic course in a specific subject), then extensive study of the various aspects of those words is required. Some relevant activities might be consulting a dictionary of synonyms/antonyms or studying the word in context. If the goal is to become an expert in L2 vocabulary, then substantial exposure to the language will be necessary to enable both the acquisition of many words, and of all the different aspects involved in knowing each individual word.

**What does Depth Mean?**

Vocabulary knowledge is a complex and multifaceted construct (Schmitt, 2010). Although the notion of vocabulary size is straightforward and easy to understand, the concept of vocabulary depth is much vaguer. It is not always clear what it entails, as evidenced by the various definitions and conceptualizations of vocabulary depth in the literature. At the most basic level, depth is synonymous with quality or richness of word knowledge and is characterized as the counterpart to breadth within a two-dimensional view of vocabulary knowledge (Anderson & Freebody, 1981). It has also been conceptualized as *developmental*, and typically measured along a continuum from no knowledge at all to full mastery (Paribakht & Wesche, 1997; Read, 2000; Schmitt & Zimmerman, 2002). For example, the following table presents a list of several proposed stages of word acquisition suggested by Paribakht and Wesche (1993, p.180).

**Table 33.1. Stages of word acquisition (Paribakht & Wesche, 1993, p. 180)**

|  |  |
| --- | --- |
| Stage 1 | The word is not familiar at all |
| Stage 2 | The word is familiar but the meaning is not known |
| Stage 3 | A correct synonym or translation is given |
| Stage 4 | The word is used with semantic appropriateness in a sentence |
| Stage 5 | The word is used with semantic appropriateness and grammatical accuracy |

Another way of looking at depth of knowledge is by breaking it down into separate elements, which could be described as a *components* or *dimensions* approach (Richards, 1976; Nation, 1990, 2001, 2013). Following Richards’ original idea, Nation (2013) has proposed the most comprehensive and complete conceptualization of word knowledge to date, according to which knowing a word involves the mastery of nine aspects: spoken form, written form, word parts, meaning, concept and referents, associations, grammatical functions, collocations, and constraints of use (Table 33.2). Each of these aspects has a productive and receptive variant, and is grouped under one of three main aspects of word knowledge: form, meaning, and use.

**Table 33.2. Nation’s (2013, p. 49) framework of the dimensions involved in knowing a word**

|  |  |  |  |
| --- | --- | --- | --- |
| FORM | Spoken | [R] | What does the word sound like? |
| [P] | How is the word pronounced? |
| Written | [R] | What does the word look like? |
| [P] | How is the word written and spelled? |
| Word parts | [R] | What parts are recognisable in this word? |
| [P] | What word parts are needed to express the meaning? |
| MEANING | Form and meaning | [R] | What meaning does this word form signal? |
| [P] | What word form can be used to express this meaning? |
| Concept and referents | [R] | What is included in the concept? |
| [P] | What items can the concept refer to? |
| Associations | [R] | What other words does this make us think of? |
| [P] | What other words could we use instead of this one? |
| USE | Grammatical functions | [R] | In what patterns does the word occur? |
| [P] | In what patterns must we use this word? |
| Collocations | [R] | What words or types of words occur with this one? |
| [P] | What words or types of words must we use with this one? |
| Constraints on use (register, frequency…) | [R] | Where, when and how often would we expect to meet this word? |
| [P] | Where, when, and how often can we use this word? |

Note: R = receptive knowledge, P = productive knowledge

The above approaches highlight the ways vocabulary depth has been conceptualized in relation to individual words. However, vocabulary mastery can also be defined in terms of the overall mental lexicon, or network of interconnected words (e.g. Meara, 1996; Milton & Fitzpatrick, 2014). According to this view, having a great depth of knowledge means having strong interconnections between words (i.e. a highly organized mental lexicon). The more developed the lexical network, the quicker it becomes to recognize and retrieve words. The figure below, taken from Meara and Wolter (2004, p. 89), illustrates the network view of vocabulary knowledge.

A close up of text on a white background

Description automatically generated

From a network perspective, we would assume that the ability to use words is strongly related to how well the various words (and their various word knowledge components) are known and interrelated. To further explore what being proficient in L2 vocabulary really means, the next section will examine the various components of word knowledge (Table 33.2) in more detail.

**Knowledge of form.**

The first main aspect of word knowledge is FORM. In the minds of all speakers, concepts and referents are attached to labels and without knowing the labels, the concepts cannot be expressed. Due to their previous experience of the world in their L1, L2 learners will already be familiar with various objects and notions and will know how to express them in their native language (e.g. in French, *chat* is the form attached to ‘a four-legged animal that purrs’). Learning a new word in the L2 is first and foremost a matter of relabelling (i.e. of associating a new form to a referent). It is worth mentioning, however, that not all semantic concepts are transferable between languages and this is especially true for those that do not directly refer to physical objects (e.g. the concept of *Schadenfreude* - pleasure derived from the misfortune of others - in German).

The new L2 form of a previously known L1 concept or referent needs to be acquired in both its written and spoken forms. The spelling needs to be known so the word can be recognized at a receptive level (when reading) and used at a productive level (when writing). The pronunciation of the word must be known as well in order for learners to recognize it when heard and to produce it in speech. Research suggests that learning the spoken form of a word is more difficult for second language learners than learning the written form (Milton et al., 2014). Also, because words may be pronounced differently when used in isolation in comparison to continuous speech, some L2 learners may fail to recognize them (van Zeeland, 2014). This is where the setting of L2 acquisition becomes crucial. While naturalistic settings (i.e. direct immersion in the L2 environment) feature words in natural speech, classroom settings are characterized by words as individual written units (Nation, 2013), thereby hindering the recognition and production of spoken word forms.

Because of the spontaneous and ephemeral nature of real-time interaction, the spoken form must be recognized and recalled very quickly. The written form must be accessed just as quickly while engaging in a reading or writing task, otherwise the plot might be lost. Therefore, fluency (or automaticity) plays a crucial role in spoken and written form recognition and production.

Finally, Nation’s framework also includes knowledge of word parts, which refers to the fact that words may comprise affixes that are added to a root form (e.g. *unbelievable*, where the prefix *un*- and suffix -*able* are added to the root *believe*). Because affixes are likely to change not only the meaning of the root (e.g. *un*- means ‘not’) but also its word class (e.g. *believe* is a verb, but -*able* turns it into an adjective), proficient users need to be able to either break down or assemble the various parts of words so they can understand and use them. Knowledge of word parts has been identified as one of the most difficult word aspects, for both native and non-native speakers of English (Schmitt, 2010; Nation, 2013).

**Knowledge of meaning.**

The second main aspect of word knowledge is MEANING. As previously mentioned, L2 learners have usually gained substantial experience of the world prior to learning their L2, so their first task will be to learn the new forms attached to the meanings they already know. Receptive knowledge of the form-meaning link involves retrieving a meaning for a written or spoken L2 form, while productive knowledge entails producing the appropriate written or spoken L2 form for an intended meaning. In some cases, one form may encode multiple meanings (as is the case with polysemous words, such as *bright* and *table*) or various forms may have the same meaning, which L2 speakers need to be aware of.

Knowing a word’s concept and referents involves knowing the boundaries of the word meaning (in other words, what it includes and excludes). For example, the concept of the word *ruminant* is ‘any animal that brings back plant-based food from its stomach and chews it again’ and some referents are cows, goats and sheep, but not cats and dogs.

Word association knowledge may be defined as the ability to connect a word to other words, some of which could be used as substitutes. Those might be *synonyms* (words with the same meaning), *homophones* (words that sound the same), *syntagms* (words that occur together), *paradigms* (words that could be used as typical examples), etc. Research suggests that word association knowledge is indicative of the organization of the mental lexicon and the depth of vocabulary knowledge in general (e.g. Schmitt, 1998; Wolter, 2002; Fitzpatrick & Munby, 2014). Meaning-based associations such as synonyms or paradigms characterize a well-developed mental lexicon typical of an advanced learner, whereas form-based associations such as homophones or syntagms characterize a sparser lexicon typical of a lower level of language proficiency (Cremer, Dingshoff, Beer, & Schoonen, 2010).

**Knowledge of use.**

The third main aspect of word knowledge is USE. It includes some of the most interesting components of word knowledge and is currently attracting the most attention in vocabulary research. In order to acquire vocabulary expertise, L2 learners must not only know the formal and semantic properties of words, but also, crucially, how those words behave in context when surrounded by other words. The analysis of large samples of real-world texts (*corpora*) has revealed that much of language is composed of patterns (e.g. Erman & Warren, 2000; Oppenheim, 2000), which might be of a grammatical, semantic, or formulaic nature. All word classes (verbs, adverbs, nouns, adjectives) follow grammatical patterns; for example, verbs might be transitive or intransitive. Knowledge of the grammatical pattern (or function) of a word is necessary to understand and use the word within a sentence. Words with similar meanings, for example, commonly co-occur together (e.g. *Ice hockey is the most popular sport in Canada*). In addition, many words also follow formulaic patterns and tend to occur with the same words in discourse, like collocations (e.g. *spare time*). The widespread occurrence of collocations in language is one of the most fascinating findings in the field of applied linguistics and will be explored in greater detail in the following section.

In order to gain full knowledge of a word, L2 learners thus need to be able to recognize and use the words with which that word typically occurs in a sentence. Learners must also be aware of the constraints of use that might be associated with some words. This is known as knowledge of *appropriacy of use*, and refers to *frequency* and *register* issues. Words have different frequency distributions and the more frequent a word, the more likely it is to be understood and used in everyday speech. Having good intuitions of a word’s frequency will help a learner decide when to use it depending on the language proficiency level of the person they are addressing (e.g. a young child or older person). Gaining knowledge of a word’s register will involve, among other things, knowing whether it is more appropriate for formal or informal settings (e.g. an academic lecture or casual conversation), and for spoken or written discourse. Not all words in English have specific constraints of use. Many words (especially of the high-frequency-variety) are neutral and can be used in a wide range of situations and discourse modes. Nevertheless, ‘constraints of use’ is an essential component of word knowledge for those other words which may cause a great deal of confusion or embarrassment if used incorrectly, e.g. calling a man ‘skinny’ instead of ‘slim’.

Taken as a whole, vocabulary depth is a complex construct that is comprised of many components of knowledge. It is worth pointing out that those components are not mastered sequentially or in an ‘all or nothing’ manner. Instead, they are more likely to be interrelated and develop in parallel, although it is not yet clear in what way and to what extent. Schmitt (2010, p. 38) provides a hypothetical illustration of how well the different knowledge components for a given word might be known at a given point in time, highlighting that the components are likely to develop at different rates (Figure 33.4)

A screenshot of a cell phone

Description automatically generated

Some components are thought to be relatively easy to acquire and suitable for explicit study (e.g. form and meaning), whereas others are more difficult and require repeated exposure to the words in context in order to be fully learned (e.g. collocations and constraints on use) (Schmitt, 2014). It is obviously unrealistic to expect highly proficient L2 speakers, or even native speakers, to know all words in English in such depth. Rather, it is more likely that only a relatively small proportion of the words they use are known in all the ways described above. However, acquiring as much knowledge of those aspects as possible is a desirable goal in order to gain vocabulary expertise.

**Formulaic language**

**What is formulaic language?**

Traditionally, language has been seen as a system made of words that are arbitrary signs, and grammar rules that determine how these words can be combined to convey messages (Pinker, 1998). Based on this assumption, by combining syntactic rules with lexis, an infinite number of utterances can be created. While the creative nature of language remains unquestionable, more recent research suggests that native speakers store in memory a large number of ready-made formulaic sequences which they can rely on in order to produce and process language more efficiently (e.g. Nattinger & DeCarrico, 1992; Wray, 2002). This means that even if we could produce a large number of novel utterances, we tend not to do so, and rely on the frequent use of formulaic sequences to a considerable extent.

The term *formulaic sequence* has been defined as “a sequence <…> of words or other elements which is, or appears to be, prefabricated: that is, stored and retrieved as a whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray, 2002, p. 9). In other words, language users retain memory traces of frequently co-occurring words and are able to access these language chunks quickly when they need them. It seems that we rely on the formulaic sequences stored in the long-term memory to compensate for a limited working memory (i.e. the part of memory that we use to store and manipulate information for a short time), which makes formulaic language crucial for fluent language use (Conklin & Schmitt, 2012).

There are different types of formulaic sequences that vary in length, structure, fixedness, and frequency of use. Some examples include proverbs (e.g. *When in Rome, do as the Romans*), idioms (e.g. *kick the bucket*), phrasal verbs (e.g. *get up*), collocations (e.g. *dark night, meet expectations*), and slogans (e.g. *Just do it!*) or songs. To capture all of these sequences, different identification criteria have been suggested. The most commonly used include: degree of non-compositionality (i.e. whether the meaning of the whole phrase can be derived from the sum of the meanings of its parts); storage in memory; single intonation curve when produced in speech, and conventionalization in society (usually shown by frequent occurrence in a corpus) (e.g. Erman & Warren, 2000; Wray, 2002; Schmitt, 2010).

**Why is formulaic language important?**

Pawley and Syder (1983) suggest that adult native speakers of English know at least several hundreds of thousands of formulaic sequences, which would mean that they know more word sequences than individual words. Both written and spoken discourse contain large numbers of formulaic sequences and some of these sequences are used very frequently. It is difficult to establish exact percentages because different researchers use different definitions of formulaic language, but most research suggests that formulaic language makes up between one-third and one-half of discourse (Conklin & Schmitt, 2012). Overall, the suggested percentages vary from about 20% at the lower end (Sorhus, 1977) to up to 80% at the higher end (Oppenheim, 2000) and they vary between different types of formulaic language (Vilkaitė, 2016). However, regardless of the percentages, there is pervasive evidence that formulaic sequences are very frequent in language.

There is a general consensus that formulaic language is important for effective, appropriate, and fluent language usage (e.g. Nattinger & DeCarrico, 1992; Wray, 2002; Barfield & Gyllstad, 2009). Schmitt (2010) summarizes some of these diverse communicative purposes:

• Expressing a message or an idea: *Pull the plug on X* = stop something, such as a project or a plan;

• Functional use: *I (really) appreciate your thoughtfulness/kindness/doing X* to express gratitude (Nattinger & DeCarrico, 1992);

• Social interaction (phatic communion): *Did you really*? and *How interesting*! to provide backchannels and positive feedback to another speaker;

• Discourse organization: *as a result of X* to show cause and effect;

• Precise information transfer: *Mayday, Mayday, Mayday* as a technical phrase to indicate an emergency situation in aviation and maritime language.

Because formulaic sequences are frequently and typically used for the above communicative purposes, they have become the default means of expression for these purposes within a speech community. This means that a learner’s interlocutors will not only accept conventionalized formulaic sequences in their discourse, but will also expect them as the norm and the most appropriate means of expression. It has also been widely attested that formulaic language aids and speeds up language processing both for speaker and for hearer (Siyanova-Chanturia & Martinez, 2014). There is substantial evidence that formulaic sequences are processed faster in the brain and show clear predictability effects (as indicated by various reaction time experiments, eye-tracking and brain imaging studies). These effects seem to be driven by the frequency of formulaic sequences in language, and their degree of familiarity for speakers (Siyanova-Chanturia, 2015).

In terms of receptive skills (reading and listening), formulaic sequences are standard ways of realizing functions and expressing ideas. Therefore, knowledge of these sequences allows learners to process a piece of discourse more easily. This has been demonstrated particularly in reading with eye-tracking studies, which show that formulaic sequences are read faster than novel phrases (e.g. Siyanova-Chanturia, Conklin & Schmitt, 2011; Sonbul, 2015). In relation to language production (especially speaking), formulaic language facilitates the production of fluent speech despite the cognitive constraints of operating online. For example, Kuiper (2000) and Kuiper and Haggo (1984) found that the discourse of people who need to produce fluent speech under severe time pressure, such as auctioneers, was largely formulaic in nature. Wray (2002) also claimed that in certain situations where language has to be produced very quickly, (such as in sports commentaries or simultaneous translation), speakers tend to rely heavily on formulaic sequences in order to overcome the huge load on their working memory.

**Use of formulaic language by second language learners**

When analyzing formulaic language produced by non-native speakers, two different aspects need to be considered: amount of use and appropriacy of use. As far as the amount of use is concerned, it is not the case that L2 learners do not use formulaic sequences at all. Researchers looking at mostly written production of second language learners have shown that learners (depending on the sequences) underuse, overuse or misuse them when compared to native speakers (Paquot & Granger, 2012). They also tend to use large numbers of collocations that are easy for them (Paquot & Granger, 2012), and avoid certain formulaic sequences that they find difficult to use. For example, Dagut and Laufer’s (1985) study of the use of phrasal verbs demonstrated that L1 Hebrew learners of L2 English avoided using phrasal verbs in their writing when using English, as these did not exist in their first language. These studies suggest that failure to use formulaic sequences in a native-like way makes learners’ language sound unnatural.

However, it is not only the overall number of formulaic sequences that is important. Appropriate use of expected sequences is equally crucial, but is not an easy task to achieve. Analysis of learner language indicates that about one-third of the collocations produced by language learners are not judged acceptable by native speakers. This finding holds regardless of the learners’ level of proficiency, with even advanced learners making mistakes in their use of collocations (Paquot & Granger, 2012). This is problematic, as it was shown that malformed collocations (such as *cheap price* instead of *low price*) slows down the processing for the native speakers (Millar, 2011), and thus can hinder communication.

A very common reason for using unacceptable word sequences is the influence of the learner’s L1. If formulaic sequences are congruent in the L1 and L2, they are easier to produce and process, but if they are different in the two languages, the L1 tends to interfere and slow down processing (see Bahns & Eldaw, 1993 for producing and Wolter & Gyllstad, 2011 for processing).

**How can L2 learners improve their use of formulaic sequences?**

Because native speakers use a large number of formulaic sequences, L2 learners are faced with the enormous task of determining which word combinations sound native-like. It is not surprising that knowledge of formulaic sequences lags behind the knowledge of single words (Bahns & Eldaw, 1993). Why is acquiring formulaic language in the L2 so hard, when it is such an effortless process in the L1? Wray (2002) notes that, when starting to learn an L2, mature learners have different cognitive abilities and are usually literate. This pushes them to focus on word-level units and process language more analytically. Also, their learning situation is typically very different in that the L2 is usually learned in a classroom with much less exposure to naturally occurring language and much more social pressure. Regardless of that, L2 learners can and do acquire formulaic sequences.

Several ways of learning and teaching formulaic sequences have been suggested by previous research, with differing degrees of success. Research shows that prolonged stays in a country where the L2 is spoken, as well as the motivation to be integrated into the L2 speech community helps to increase the use of appropriate formulaic sequences (Adolphs & Durrow, 2004; Dörnyei, Durrow, & Zahran, 2004, Macis & Schmitt, 2016). Nevertheless, the vast majority of language learners across the world do not have the opportunity to live in a country where the L2 is spoken. Fortunately, exposure can be achieved easily from home via the use of technology. For example, the language used on television seems to correspond to everyday spoken English in terms of the formulaic sequences it contains (Lin, 2014).

Other more traditional methods of teaching and learning formulaic sequences also seem to be effective. Some studies suggest that the same teaching methods can be applied to both single words and formulaic sequences (Webb & Kagimoto, 2009; Alali & Schmitt, 2012). However, even if vocabulary exercises are effective for learning formulaic sequences, the problem of choosing which sequences to teach remains. In order to solve this problem, researchers have created lists of the most frequent, and therefore most important, items to learn. Examples include the *PHaVE List* for phrasal verbs (Garnier & Schmitt, 2015) and the *PHRASE List* for non-transparent formulaic sequences (Martinez & Schmitt, 2012). These lists are a valuable starting point to give learners and teachers some guidance. Nevertheless, learning phrases beyond these lists remains a challenge, especially as many other types of formulaic sequence have no available lists.

It seems that a large number of formulaic sequences have to be learned from exposure when reading or listening to language. Research on incidental acquisition (i.e. learning without intention) of formulaic language shows that, although the rate of acquisition is slow, learners do acquire formulaic sequences from reading even without consciously focusing on learning these items (Webb, Newton, & Chang, 2013; Pellicer-Sánchez, 2015). So, maximizing exposure seems to be a crucial learning principle. Overall, it appears that formulaic language takes a long time to acquire, but it is also a hallmark of the highest stages of language mastery. It is indeed the appropriate use of formulaic language that distinguishes an expert language user from a novice or intermediate one.

**Practical suggestions for gaining L2 expertise**

The above discussion has emphasized that learning enough vocabulary is a big challenge for learners. Explicitly study of individual words and formulaic sequences is the quickest and most effective way of learning this vocabulary. But this will only ever cover a limited number of the lexical items necessary to use a second language. It will also be necessary to maximize the learner’s exposure to English in a number of ways: a) reading extensively; b) watching TV and movies; c) using social media; d) interacting with native speakers of English; and e) spending time in an English-speaking country. Because spending time in an English-speaking country is impractical for most students, and access to native speakers of English may be severely limited, reading, watching TV and movies and using social media may be seen as the most convenient ways to increase a learner’s exposure to English.

For beginners, the use of graded readers (simplified books that are specially created for second language learners) is a sensible starting point, as these have the advantage of providing substantial language input at an early stage of a learner’s development. They can also help to establish a desirable long-term reading habit in the L2. As learners’ proficiency increases, they will naturally want to move on to authentic texts. However, the jump from graded readers to authentic texts can be a daunting one and narrow reading might help ease this transition. Narrow reading refers to reading numerous texts on the same topic, with the advantage that much vocabulary is recycled, making the topic increasingly easier to process. Narrow reading can be achieved by following a continuing story in a newspaper or by reading magazines focusing on a particular theme. After this stage, the most effective way to increase language input is through extensive reading, which simply refers to reading large quantities of text from diverse sources and of various topics. By reading extensively, the chances of learning new words and consolidating knowledge of partially known words increase substantially.

Watching TV and movies provides not only entertainment, but also the opportunity for learners to develop their language skills and increase their vocabulary sizes (King, 2002). Through films, learners can see how people communicate in real life in different conversational contexts, and get a feel for the kind of vocabulary they use in each context. Using subtitles is a useful source of vocabulary learning (d’Ydewalle & Pavakanun, 1995). In this way, learners can match the written subtitles to the words spoken on screen, which enables them to further develop their vocabulary. Keeping a notebook, writing down newly encountered words or phrases and looking up their meaning later is also a very useful activity for learners of all levels (e.g. Walters & Bozkurt, 2009).

Research also suggests that using social media (Facebook, Twitter, and MySpace) in the L2 has been found to facilitate knowledge of collocations and phrasal verbs (González-Fernández & Schmitt, 2015; Garnier & Schmitt, 2016). In addition, applications like *Vocabulary.com*, *Phrasal Verbs Machine*, and *English Vocabulary in Use* are an excellent resource for improving learners’ vocabulary, and are immediately accessible. They reflect the type of exposure which makes language more engaging and personal, and thus perhaps even more conducive to learning.

All of these input-maximizing approaches will not only promote the development of a larger vocabulary, but will also provide the contexts which facilitate depth of knowledge of the various lexical items.

**Conclusion**

This chapter has highlighted what we believe are the three key requirements in order to reach expert knowledge of L2 vocabulary. A large vocabulary size is necessary to be able to function well in English (Nation, 2006), but learners also need to know words well, and how they combine together in discourse, in order to use them productively, appropriately, and fluently.

The discussion presented here also calls for more research in the area. Nation’s 2006 paper is still held as the standard reference for lexical requirements in English. However, 10 years on, many of his assumptions are now being questioned, especially whether his reported figures are applicable to all kinds of reading and listening contexts and materials. Furthermore, the studies on lexical coverage reported above have not taken into account the variability among individual learners and their language skills. Consequently, we propose that future research should revisit Nation’s figures with the aim of updating estimates of how much vocabulary is needed to read and listen to a variety of discourse types, and providing more finely-tuned percentages.

The field also lacks reliable and comprehensive measures of vocabulary depth. Only few studies have examined the acquisition of multiple word knowledge components concurrently (e.g. Schmitt, 1998; Webb, 2007), so it is still unclear how the different components of vocabulary knowledge are acquired and relate to each other. González-Fernández (in preparation) explored the order of acquisition of four word knowledge components (the form-meaning link, derivatives, polysemy, and collocations) in Spanish learners of English. Results revealed an unexpected order of difficulty for some aspects, with derivative knowledge (i.e. ‘word parts’ in Table 33.2) proving to be particularly difficult. They also showed that the vocabulary knowledge components are all interrelated, but at different levels of strength. However, these are only initial results, applicable to one specific group of learners, thus indicating the need for more research in order to arrive at a more detailed and accurate description of vocabulary acquisition.

Finally, it is still unclear what the most effective way to teach formulaic language is. Although some researchers have already addressed this issue to a certain extent (Webb & Kagimoto, 2009; Alali & Schmitt, 2012), these and similar studies suffer from a number of limitations. Some of them have used a small number of participants, whereas others have examined only one particular types of formulaic language (e.g. collocations). Only further research with larger populations, a wider range of teaching methods, and a more diverse selection of formulaic sequences can help us identify the principles of truly effective teaching techniques.

**References**

Adolphs, S., & Durrow, V. (2004). Social-cultural integration and the development of

formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition,*

*processing, and use* (pp. 107-126). John Benjamins Publishing.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage and spoken discourse. *Applied*

*Linguistics,* *24*(4), 425-38.

Alali, F. A., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different

from teaching single words? *TESOL Journal*, *3*(2), 153-180.

Alderson, J.C. (2005). *Diagnosing foreign language proficiency*. London: Continuum.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.),

*Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE:

International Reading Association.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*(1),

101-114.

Bannard, C., & Lieven, E. (2012). Formulaic language in L1 Acquisition. *Annual Review of*

*Applied Linguistics*, *32*, 3-16.

Barfield, P. A., & Gyllstad, D. H. (2009). *Researching collocations in another language:*

*Multiple interpretations*. New York: Palgrave Macmillan.

Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension.

*International Journal of Listening,* *14*(1), 14-31.

Carter, J. (2012). *Vocabulary: Applied linguistic perspectives.* Routledge Classics Edition,

Abingdon: Routledge.

Cobb. T. *Lextutor*. Accessed 25 October 2016 at http:// [www.lextutor.ca](http://www.lextutor.ca)

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of*

*Applied Linguistics*, *32*, 45-61.

Cremer, M., Dingshoff, D., Beer, M., & Schoonen, R. (2010). Do word associations assess

words knowledge? A comparison of L1 and L2, child and adult word associations.

*International Journal of Bilingualism*, *15*(2), 187-204.

d’Ydewalle, G., & U. Pavakanun. (1995). Acquisition of a second/foreign language by

viewing a television program. In P. Winterhoff-Spurk (Ed.), *Psychology of media in*

*Europe: The state of the art-perspectives for the future*. (pp. 51-64). Westdeutscher

Verlag GmbH.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs-A case for contrastive

analysis. *Studies in Second Language Acquisition*, *7*(1), 73-79.

Dörnyei, Z., Durrow, V., & Zahran, K. (2004). Individual differences and their effects on

formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic sequences:*

*Acquisition, processing, and use* (pp. 87-106). John Benjamins Publishing.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text -*

*Interdisciplinary Journal for the Study of Discourse*, *20*(1), 29-62.

Fitzpatrick, T., & Munby, I. (2014). Knowledge of word associations. In J. Milton,

& T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp. 92-105).

Basingstoke: Palgrave Macmillan.

Garnier, M., & Schmitt, M. (2016). Picking up polysemous phrasal verbs: How many do

learners know and what facilitates this knowledge? *SYSTEM*, 59, 29-44.

Garnier, M., & Schmitt, N. (2015). The PHaVE List: A pedagogical list of phrasal

verbs and their most frequent meaning senses. *Language Teaching Research, 19*(6),

645-666.

González-Fernández, B. (in preparation). Towards a theory of vocabulary acquisition:

Relationships and order of acquisition of different types of word knowledge by L2

learners. Unpublished PhD thesis.

González-Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2

learners have?: The effects of frequency and amount of exposure. *ITL International*

*Journal of Applied Linguistics, 166*(1), 94-126.

Goulden, R., Nation P., & Read, J. (1990). How large can a receptive vocabulary be?

*Applied Linguistics*, *11*(4), 341-363.

Hirsh, D., & P. Nation. (1992). What vocabulary size is needed to read unsimplified texts

for pleasure? *Reading in a Foreign Language*, *8*(2), 689-696.

Hu, M., & Nation, I.S.P. (2000). Vocabulary density and reading comprehension.

*Reading in a Foreign Language, 13*(1), 403-430.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language.

*Studies in Second Language Acquisition*, *21*(2), 181-193.

Hulstijn, J.H. (2001). Intentional and incidental second language vocabulary learning: A

reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition*

*and second language instruction* (pp. 258-286). Cambridge: Cambridge University

Press.

King, J. (2002). Using DVD feature films in the EFL classroom. *Computer Assisted*

*Language Learning*, *15(*5), 509-23.

Kuiper, K. (2000). On the linguistic properties of formulaic speech. *Oral Tradition*, *15*(2),

279-305.

Kuiper, K., & Haggo, D. (1984). Livestock auctions, oral poetry, and ordinary language.

*Language in Society*, *13*(2), 205-234.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and

computer adaptiveness. *Language Learning*, *54*(3), 399-436.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical

coverage, learners’ vocabulary size and reading comprehension. *Reading in a Foreign*

*Language*, *22*(1), 15-30.

Li, M., & Kirby, J. R. (2014). The effects of vocabulary breadth and depth on English

reading. *Applied Linguistics*, 1-25.

Lin, P. M. S. (2014). Investigating the validity of internet television as a resource for

acquiring L2 formulaic sequences. *System*, *42*, 164-176.

Macis, M., & Schmitt, N. (2016). Not just ‘small potatoes’: Learning the figurative meanings

of collocations. *Language Teaching Research*. DOI: 10.1177/1362168816645957

Martinez, R., & Schmitt, N. (2012). A Phrasal expressions list. *Applied Linguistics*, *33*(3),

299-320.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjær, & J.

Williams (Eds.), *Performance and competence in second language acquisition* (pp.

35-53). Cambridge: Cambridge University Press.

Meara, P., & Wolter, B. (2004). V\_LINKS: Beyond vocabulary depth. In D.

Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), *Angles on the English*

*speaking world 4* (pp. 85-96). Copenhagen, Denmark: Museum Tusculanum.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*,

*32*(2), 129-148.

Milton, J., & Fitzpatrick, T. (Eds.) (2014). *Dimensions of vocabulary knowledge*.

Basingstoke: Palgrave Macmillan.

Milton, J., Alexiou, T., & Mattheoudakis, M. (2014). Knowledge of spoken form. In J.

Milton, & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp.13-29).

Basingstoke: Palgrave Macmillan.

Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge

University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The*

*Canadian Modern Language Review*, *63*(1), 59-81.

Nation, I.S.P. (2013). *Learning Vocabulary in Another Language.* Cambridge: Cambridge

University Press.

Nation, I.S.P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N.

Schmitt, N. and M. McCarthy (Eds.), *Vocabulary: Description, acquisition and*

*Pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford,

England; New York: OUP Oxford.

Oppenheim, N. (2000). The importance of recurrent sequences for non-native speaker fluency

and cognition. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 220-240). Ann

Arbor: University of Michigan Press.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of*

*Applied Linguistics*, *32*, 130-149.

Paribakht, T. S., & Wesche, M. B. (1997). Vocabulary enhancement activities and reading for

meaning in second language vocabulary acquisition. In J. Coady, & T. Huckin (Eds.),

*Second language vocabulary acquisition* (pp. 174-200). Cambridge: Cambridge

University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and

nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language*

*communication* (pp. 191-226). New York: Longman.

Pellicer-Sánchez, A. (2015). Learning L2 collocations incidentally from reading. *Language*

*Teaching Research*. DOI: 10.1177/1362168815618428

Pellicer-Sánchez, A, & Schmitt, N. (2010). Incidental vocabulary acquisition from an

authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*(1), 31-55.

Peters, A. M. (1983). The units of language acquisition (Vol. 198). Cambridge University

Press.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case

study. *Reading in a Foreign Language*, *18*(1), 1-28.

Pinker, S. (1998). Words and rules. *Lingua*, *106*(1-4), 219-242.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Reves, T., & A. Levine. (1988). The FL receptive skills: Same or different? *System*, *16*(3),

327-36.

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

Rodgers, M.P.H. & Webb, S. (2011). Narrow viewing: The vocabulary in related television

programs. *TESOL Quarterly, 45*(4), 689-717.

Rost, M. (2002). *Teaching and researching listening*. Longman.

Schmitt, N. (1998). Quantifying word association responses: What is native-like? *System*, *26*,

389-401.

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

Press.

Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Houndmills,

Basingstoke, Hampshire; New York: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows

*Language Learning*, *64*(4), 913-951.

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2015). How much vocabulary is needed to

use English? Replication of Van Zeeland & Schmitt (2012), Nation, (2006), and Cobb

(2007). *Language Teaching*. DOI: 10.1017/S0261444815000075.

Schmitt, N., X. Jiang, & W. Grabe. (2011). The percentage of words known in a text and

reading comprehension. *The Modern Language Journal,* *95*(1), 2643.

Schmitt, N., & Marsden, R. (2006). *Why is English like that? Historical answers to hard*

*ELT questions*. University of Michigan Press.

Schmitt, N. & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2

vocabulary teaching*. Language Teaching*, *47*(4), 484-503.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of

two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners

know? *TESOL Quarterly*, *3*6(2), 145-171.

Siyanova-Chanturia, A. (2015). On the ‘holistic’ nature of formulaic language. *Corpus*

*Linguistics and Linguistic Theory*, *11*(2), 1-17.

Siyanova-Chanturia, A., & Martinez, R. (2014). The Idiom principle revisited. *Applied*

*Linguistics*, 1-22.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An

eye-tracking study of idiom processing by native and non-native speakers. *Second*

*Language Research*, *27*(2), 251-272.

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency

effects on off-line/on-line collocational processing. *Bilingualism: Language*

*and Cognition*, *18*(3), 419-437.

Sorhus, H.B. (1977). To hear ourselves - implications for teaching English as a second

language. *English Language Teaching Journal,* *31*(3), 211-221.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing.

*Language Learning Journal*, *36*(2), 139-52.

van Zeeland, H. (2014). Second language vocabulary knowledge in and from listening.

Unpublished PhD thesis, University of Nottingham.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening

comprehension: The same or different from reading comprehension? *Applied*

*Linguistic,* *34*(4), 457-479.

Vilkaitė, L. (2016). Formulaic language is not all the same: comparing the frequency of

idiomatic phrases, collocations, lexical bundles, and phrasal verbs. *Taikomoji*

*kalbotyra*, 8, [www.taikomojikalbotyra.lt](http://www.taikomojikalbotyra.lt).

Walters, J., & Bozkurt, N. (2009). The effect of keeping vocabulary notebooks on

vocabulary acquisition. *Language Teaching Research, 13*(4), 403-423.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary

from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 1-27.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*,

*28*(1), 46-65.

Webb, S. (2011). Selecting television programs for language learning: Investigating

television programs from the same genre. *International Journal of English Studies*,

*11*(1), 117-136.

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and

meaning. *TESOL Quarterly*, *43*(1), 55-77.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language*

*Learning*, *63*(1), 91-120.

Webb, S., & Rodgers, M.P.H. (2009). The lexical coverage of movies. *Applied Linguistics*,

*30*(3), 407-427.

Wolter, B. (2002) Assessing proficiency through word associations: Is there still hope?

*System*, *30*(3), 315-329.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the

influence of L1 intralexical knowledge. *Applied Linguistics*, *32*(4), 430-449.

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

Press.

Zechmeister, E.B., Chronis, A.M., Cull, W.L., D’Anna, C.A., & Healy, N.A. (1995). Growth

of a functionally important lexicon. *Journal of Reading Behavior*, *27*(2), 201-212.

Zimmerman, K. J. (2004). The role of Vocabulary Size in Assessing Second Language

Proficiency. MA dissertation, Brigham Young University.