The Routledge Handbook of Vocabulary Studies

Stuart Webb

The Relationship Between Vocabulary Knowledge and Language Proficiency

Publication details
David D. Qian, Linda H. F. Lin
Published online on: 13 Aug 2019

Accessed on: 05 Dec 2023

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
The Relationship Between Vocabulary Knowledge and Language Proficiency

David D. Qian and Linda H. F. Lin

Introduction

The relationship between vocabulary knowledge and language proficiency is multifaceted. Vocabulary knowledge is “fundamental to all language use” (Schmitt, Cobb, Horst, & Schmitt, 2015) and thus is a crucial part of language learning. Achieving certain levels and qualities of lexical knowledge is one of the important prerequisites for successful language learning. This makes lexical knowledge a powerful predictor of learners’ language proficiency and even their academic achievement (Lin & Morrison, 2010). Meanwhile, learners’ lexical competence tends to improve as their language proficiency develops (Zareva, Schwanenflugel, & Nikolova, 2005) since the process of applying the four macro language skills (i.e., reading, listening, speaking, and writing) in communication is most conducive to imprinting newly learned words into memory (Laufer, 2013).

Grounds for the Prominent Role of Vocabulary Knowledge in Language Learning

One of the reasons for the widely recognized importance of lexical knowledge in language learning is the inextricably intertwined relationship between lexis and grammar. The strong association between the two language forms has led to the appearance of lexicogrammar (Halliday & Matthiessen, 2013), a linguistic view that focuses on the integration rather than the disparity between grammar and vocabulary. Nevertheless, when the two language forms are compared, in particular in the process of second language (L2) acquisition, lexical competence is often considered more important than syntactic knowledge in achieving effective communication. The critical role of lexical knowledge in L2 learning is further confirmed by studies investigating factors that affect learners’ communication. Lexical errors have been found to be the most numerous type of error in L2 learners’ language production, and were identified as a major factor hindering L2 learners’ communication (Llach, 2011). A notable example is Santos (1998), who elicited responses from tertiary teachers in reaction to mistakes made by their L2 students in academic writing. The study reveals that lexical errors were the most severe obstruction to the understanding of the students’ texts. This finding
was in line with Djokic’s (1999) investigation, which reports that vocabulary errors “cause momentary confusion” and “bring about misunderstanding” (p. 128). This misunderstanding not only interrupts the flow of communication but may also cause a communication breakdown (Llach, 2011). Difficulties in lexical use have also been deeply felt by many L2 students who believe that, of all the error types in their language production, the ones pertaining to vocabulary use are most damaging. Some types of lexical errors such as those in word collocation and word choice are persistent in the language production of some learners. They not only frequently appear in the language production of L2 beginners but also in that of high-proficiency learners (Gass & Selinker, 2008).

**Vocabulary Knowledge of L1 and L2 Learners**

Although an adequate knowledge of words is important for both first language (L1) and L2 learners, researchers (e.g., Jiang, 2000) believe that inappropriate lexical use causes more communication problems in the language production of L2 learners than of L1 learners, largely due to different lexical development processes of the two types of learners.

The lexical development of an L1 child is relatively fast due to his/her extensive and highly contextual exposure to the language, which makes it possible for the child to develop the semantic, syntactic, and morphological knowledge of a word. These three types of information about the word form an integrated part of the lexical entry in the child’s lexicon, which is automatically activated when the word is used. This automatic and simultaneous activation of all three types of information is crucial for the learners’ appropriate and efficient contextual use of the word.

The lexical development of L2 learners, on the other hand, is much more onerous. When L2 learners, in particular those who learn the target language in classroom settings, learn a word, they face two practical constraints. The first constraint is a lack of input opportunities in terms of both quantity and quality. In other words, there is a shortage of opportunities for learners to have sufficient and highly contextual exposure to the target language. This restriction causes significant difficulties in the learners’ extraction of the semantic, syntactic, and morphological knowledge about a word. The second constraint, which is likely to have a more serious impact on L2 learning, is the existence of an established conceptual/semantic system in the learners’ lexicon. The presence of the established L1 lexical system can greatly facilitate the acquisition of L2 words by providing a source for learners to draw on, but, meanwhile, it could also cause L2 learners, in particular adult learners, to over-depend on the L1 semantic and syntactic knowledge about a word while learning new words (Ellis & Shintani, 2013).

**Quality and Quantity of Vocabulary Knowledge**

The quantitative aspect of lexical knowledge refers to the number of words existing in a learner lexicon. This comprises the learner’s knowledge of the form and meaning of a word which can be simply recalled. The qualitative aspect of vocabulary knowledge, on the other hand, refers to a learner’s deep and extensive lexical knowledge and ability to use the word appropriately and efficiently. Having learned a word does not only mean the learner can recall the form and retrieve the meaning of the word in a vocabulary test, but it also implies the effective use of the word for authentic communication purposes, such as reading a newspaper article or writing an email message.

Researchers also divide word knowledge into two facets: receptive vocabulary and productive vocabulary. The receptive vocabulary forms the *size*, or *breadth*, of a learner’s
vocabulary in the mental lexicon. This part of the vocabulary knowledge needs to be further developed into productive knowledge in order to be used in a communication context. The development from receptive to productive vocabulary can be seen on a continuum, starting from superficial familiarity with the word and ending with an ability to use the word correctly in free production. The process of progressing on this continuum is the development of qualities or depth of one’s lexical knowledge, also known as vocabulary depth (Henriksen, 1999; Qian, 1999, 2002, 2005; Read, 2007).

Despite the importance of mastering depth of vocabulary knowledge in language learning, investigations into L2 vocabulary acquisition have mostly focused on the size rather than depth of vocabulary knowledge, probably due to the difficulty in developing quality measures for assessing the depth dimension. On the other hand, the straightforward construct of vocabulary size tests, which normally measure only one superficial dimension of a learner’s lexical knowledge, results in a variety of measures for assessing breadth of vocabulary knowledge. In research, vocabulary size tests reportedly play an important role in predicting success in learners’ proficiency. Zareva et al. (2005), for example, found that vocabulary size was one of the strongest discriminating factors of L2 learners’ proficiency level. Vocabulary size measures were also found to correlate well with scores on proficiency tests such as the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL). Stæhr (2008), for example, compared the vocabulary size of 88 lower secondary EFL students in Denmark with their examination grades on listening, reading, and writing skills. Results of the study show a strong positive association of their vocabulary size with their reading and writing abilities and a moderate positive association with their listening comprehension.

The construct of the quality dimension of a learner’s lexical knowledge, on the other hand, is complex and multifaceted. It is not only unquantifiable but also difficult to conceptualize. Schmitt (2014) regards it as “the wooliest, least definable, and least operationalisable construct in the entirety of cognitive science” (p. 920). Consequently, Daller, Milton, and Treffers-Daller (2007) subdivided the quality aspect of lexical knowledge into depth and fluency. This division has created a three-dimensional space to view a learner’s vocabulary knowledge, namely, breadth, depth, and fluency. Here, fluency is defined as the ease, or speed and accuracy, with which words are used in language production. This division is important because it “moves the conceptualisation of lexical proficiency onward from simple knowledge to the ability to use that knowledge” (Schmitt, 2014, p. 920). To improve learners’ vocabulary fluency is to increase the automaticity (Qian, 2002) of their lexical use and should be the ultimate goal of most language learners.

Notwithstanding the complexity in conceptualizing the construct of depth of lexical knowledge, vocabulary depth has also been found to be closely associated to learners’ language skills, in particular productive skills (i.e., speaking and writing). For example, Zareva et al. (2005) found L2 learners’ knowledge in word associations was one of the strongest indicators of their proficiency level. Crossley, Salsbury, and McNamara (2014) analyzed spoken and written texts produced by L2 learners and found that a very important component in the depth construct, collocation accuracy, accounted for 84% of the variance in the writing scores and explained 89% of the holistic speaking scores. This finding underscores the importance of knowledge of lexical collocation in speaking and writing proficiency. Achieving a high collocation accuracy demands both syntagmatic and paradigmatic information of the involved words. Acquiring such knowledge requires extensive exposure to a context which can provide sufficient and meaningful target language input and enable activities leading to quality output in the target language. For this purpose, Qian and Schedl (2004) even explored...
the possibility of developing such a measure in the reading comprehension part of a new TOEFL test.

Reading, listening, speaking, and writing are the four main language modalities. The former two, reading and listening, mainly require learners’ receptive language skills in comprehending written and aural texts, respectively, while the latter two require learners’ productive skills in producing written and oral texts, respectively. The productive skills are thus more demanding. The four language skills will be examined separately in their association with vocabulary knowledge in the next two sections.

Critical Issues and Topics

Vocabulary Knowledge and Receptive Language Skills

Vocabulary knowledge is at the heart of the two receptive skills. Vocabulary size is found to be closely associated with the two receptive skills (Laufer & Ravenhorst-Kalovski, 2010; Qian, 1999, 2002) and vocabulary depth has been identified to be a better predictor of these two skills than vocabulary size (Han, 2017; Qian, 2002). The nature of these two modalities has made it possible for scholarly work to investigate the percentage of words a learner needs to know in order to comprehend a stretch of written and spoken discourse, known as lexical coverage (Nation, 2006). The establishment of lexical coverage allows researchers to determine the number of words learners need to know to achieve a required comprehension level, or vocabulary size.

Reading

Reading has attracted much research attention. Qian (1999, 2002) and Qian and Schedl (2004) conducted a series of studies with samples ranging from 74 to 217 young adult English-as-a-second-language (ESL) learners in Canada to examine the role of lexical knowledge in reading comprehension. Findings from these investigations indicated that both breadth and depth of vocabulary knowledge play a major role in ESL learners’ reading processes since the scores on tests measuring depth and breadth of vocabulary knowledge correlated significantly and positively with readings scores. More importantly, the studies also revealed that the depth dimension is not only an effective predictor of reading comprehension but can also provide additional prediction of reading comprehension over and above the prediction already made by the size dimension. Additional studies have also been conducted to explore the role of depth of vocabulary knowledge in determining ESL learners’ success in lexical inferencing (Nassaji, 2006; Qian, 2005). Results from these studies indicated that depth of vocabulary knowledge also plays a critical role in lexical inferencing, as it accounts for a significant portion of variance in learners’ ability to successfully infer the meanings of unknown words. These findings have since been supported empirically by a number of recent studies (e.g., Laufer & Ravenhorst-Kalovski, 2010; Prior, Goldina, Shany, Geva, & Katzir, 2014).

Given the centrality of lexical knowledge in reading comprehension, researchers looked into a possible lexical coverage threshold that learners should reach in order to achieve adequate comprehension of written texts. The earliest research pertaining to lexical coverage suggested a need for 95% lexical coverage (Laufer, 1989). Accepting this figure, however, means that there are over 15 unknown words on every page, which could hinder the reading comprehension of some learners. A follow-up study by Hu and Nation (2000) recommended...
David D. Qian and Linda H. F. Lin

a 98% coverage as the standard for unassisted L2 reading. To make the concept more practical for L2 teaching and learning, Laufer and Ravenhorst-Kalovski (2010) suggested two lexical coverage figures based on their findings: a minimal coverage of 95% for basic comprehension and an optimal coverage of 98% for a better comprehension. A linear relationship between lexical coverage and reading comprehension suggests that the establishment of a lexical threshold should depend on the level of comprehension that is aimed for (Schmitt, Jiang, & Grabe, 2011). The higher the requirement for the comprehension level, the larger the coverage necessary, and in turn the vocabulary size that is required. The 95% coverage figure requires a vocabulary size of the 4,000 to 5,000 word families and the 98% coverage 8,000 word families for written text. This finding is in line with Nation’s (2006) study, which posits that a vocabulary size of around 8,000 to 9,000 families plus proper nouns is needed to reach the 98% coverage necessary for adequate reading comprehension.

Listening

The process of listening comprehension differs from that of reading comprehension, and vocabulary knowledge plays different roles in these two processes (van Zeeland & Schmitt, 2013). Compared with the robust research initiatives on the relationship between vocabulary knowledge and reading proficiency, the number of studies on the association between lexical knowledge and listening proficiency is rather limited. Unlike reading processes which can be either bottom-up, top-down, or integrative, depending on the reader’s proficiency level, listening is more of a top-down process due to the nature of the input. While similar factors are at work in the process of reading and listening, such as lexical knowledge, genre knowledge, topic familiarity, text organization, and world/subject knowledge, the fleeting nature of spoken language, which demands the activation of a short-term working memory, requires a different role of vocabulary knowledge in the listening process. When reading, learners can refer back to lexical items in the text to facilitate their comprehension, but this is not possible during the fast-moving process of listening, which can be onerous to many learners. This could explain the considerable variation identified in the listening comprehension scores awarded to learners with a similar vocabulary size (e.g., Stæhr, 2009; van Zeeland & Schmitt, 2013) since their general language proficiency levels can be different. However, a counterargument holds that spoken language is usually not as lexically dense as written language. This feature, coupled with the assistance of nonverbal communicative devices, such as tone, gesture, and facial expression, can alleviate the difficulty in lexical processing, and such facilitation is nonexistent in a typical reading process.

Stæhr (2008) explored the relationship between vocabulary size and three language modalities – reading, writing, and listening – with a sample of 88 lower secondary EFL students in Denmark. Results of the study indicated that the size dimension of the participants’ vocabulary was significantly, but variously, correlated with the three language skills (reading at \( r = 0.83, p < .01 \); writing at \( r = 0.73, p < .01 \); listening at \( r = 0.69, p < .01 \)). A later study (Stæhr, 2009) with 115 advanced EFL learners in Denmark probed the relationship between both vocabulary breadth and depth and listening comprehension, and detected a correlation similar to that of his previous study (Stæhr, 2008) between vocabulary size and listening comprehension. As regards vocabulary depth, Stæhr (2009) reported a positive correlation of 0.65 (\( p < .01 \)) with listening comprehension. A follow-up stepwise multiple regression analysis showed that the variable of vocabulary size was a better predictor of listening comprehension than vocabulary depth, which led Stæhr to claim that vocabulary size is the fundamental dimension of vocabulary knowledge in listening comprehension.
Somewhat conflicting results have also been reported in a number of very recent studies (e.g., Dabbagh, 2016; Han, 2017; Wang & Treffers-Daller, 2017). Examining the listening comprehension performance of 151 EFL university students in China, Wang and Treffers-Daller (2017) found that vocabulary size was a strong predictor of listening comprehension. Dabbagh (2016), when investigating the listening comprehension performance of 73 university EFL students in Iran, found that only depth of vocabulary knowledge could provide significant prediction of learners’ listening comprehension; vocabulary size, on the other hand, was not a strong predictor of the learner’s listening success. Han (2017) furthered this line of research by developing audio forms of the Vocabulary Levels Test and the Word Associates Test, to represent depth and size of vocabulary knowledge respectively. Han administered the newly developed aural tests together with the traditional written forms of the same measures to 718 EFL learners from four universities in Mainland China, and confirmed that scores on written and aural vocabulary knowledge measures were both positively and significantly correlated with listening scores to different degrees: (1) the vocabulary depth score from the written-form Word Associates Test and the vocabulary size score from the audio-form Vocabulary Levels Test were highly predictive of listening performance, (2) scores on written vocabulary depth were found to be more predictive than scores on written vocabulary size, and (3) aural vocabulary size scores were more predictive than written vocabulary size scores. Furthermore, (4) both written and aural vocabulary depth scores provided a better prediction of listening performance of EFL advanced learners than written and aural vocabulary size scores, which interestingly provided a better prediction of listening performance of low proficiency learners than vocabulary depth scores. The comprehensive results of the study suggest that the relationship between vocabulary knowledge and listening proficiency can change along with the development of a learner’s proficiency level. As these results somewhat differ from previous research findings, replication studies are desirable so that a more profound understanding of these issues can be achieved.

Another noteworthy finding from Stæhr’s (2009) study was that a vocabulary size of 5,000 word families could enable adequate listening comprehension of academic text. This contrasts with the suggestions of 6,000 to 7,000 word families made by Nation (2006) and that of 3,000 to 4,000 word families from Nation and Webb (2011) for an adequate comprehension of spoken texts. The figure, nonetheless, is rather distant from the size of 2,000 to 3,000 word families suggested by van Zeeland and Schmitt (2013). Further investigation is highly desirable to either confirm or disconfirm these claims.

Vocabulary Knowledge and Productive Language Skills

Corpora and Lexical Analysis

Compared with the scholarly work pertaining to the relationship between vocabulary knowledge and the two receptive language skills, research on the relationship between vocabulary knowledge and the two productive language skills, namely, speaking and writing, has been rather insufficient (Saito, Webb, Trofimovich, & Isaacs, 2016). This paucity is possibly owing to the complication of measuring the quality of learner output; it is more challenging to assess the relationship between vocabulary knowledge and writing and speaking proficiency than between vocabulary knowledge and reading and listening proficiency. However, with the availability of various corpora and computer programs specifically designed for analyzing corpus data, there has been an increasing amount of research examining the relationship between vocabulary knowledge and productive language skills. The availability
of such electronic data and tools have made possible large-scale investigations into lexical features and patterns, such as “word collocations, formulaic language in different registers, and lexical bundles in spoken and written English” (Paribakht & Webb, 2016, p. 121).

Corpora that have been used for lexical analysis include the British National Corpus (BNC), COCA (Corpus of Contemporary American English), and American National Corpus (ANC). Learner corpora, such as International Corpus of Learner English (ICLE), appear to be more useful in this area of research as they enable comparative interlanguage analysis between learner language and native speakers’ output. Results from such analysis can reveal patterns and problems that are typical in the learner language, such as overuse, underuse, or misuse of a lexical item. Further, with the large amount of data available, corpus-driven comparative analysis may possibly help teachers and researchers trace the cause of problems in learners’ language use.

Computer programs, such as RANGE (www.victoria.ac.nz/lals/about/staff/paul-nation/#vocab-programs), AntWordProfiler (www.laurenceanthony.net/software/antwordprofiler/), VocabProfile (www.lextutor.ca/vp/), and P_Lex (Meara & Bell, 2001), enable researchers to explore lexical richness in learners’ language production, both in written and spoken forms, although the number of studies on written texts is much higher than that of spoken texts, probably due to the relative ease in collecting and processing written data. Some of these programs are freely available on Tom Cobb’s website, Compleat Lexical Tutor. These programs can measure lexical features that are considered dimensions of lexical richness, including lexical frequency, or “words from different frequency levels” (Laufer & Nation, 1995, p. 311); lexical variation, or the variety of activated words in language production; and lexical sophistication, or the proportion of infrequent words in a text. Information on lexical richness has been found to relate closely to learners’ language proficiency (Daller & Phelan, 2007; Laufer & Nation, 1995; Lin & Morrison, 2010).

Vocabulary knowledge is a crucial component of the two productive language skills. For L2 learners, vocabulary use is one of the most important factors that determine their writing quality (Leki & Carson, 1994). Effective speaking also depends heavily on word knowledge, particularly so with L2 learners (de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012). Compared with receptive language skills, productive language skills exert a high demand on the quality dimension of learners’ lexical knowledge. This is evident in Crossley et al.’s (2014) study which demonstrated that collocation accuracy, a crucial component in the depth dimension of vocabulary knowledge, explained 84% of the variance in evaluations of writing and 89% of the variance in evaluations of speaking. These findings underscore the importance of collocation accuracy in learner production and suggest that superficial knowledge of words primarily learned in isolation, for example from decontextualized word lists, may not lead to successful language production, be it in written or spoken form. On the other hand, lexical fluency, or the ease (speed and accuracy) with which words are used in language production, plays a more determining role (Schmitt, 2014). Improving learners’ lexical fluency is to increase the automaticity of their lexical use, which requires “intensive exposure to meaningful L2 input and interaction” (Saito et al., 2016, p. 697). The importance of exposure to contextualized vocabulary learning therefore rests in the complex organization of the multidimensional aspects of vocabulary knowledge in a learner’s mental lexicon (Milton, 2009; Schmitt, 2010, 2014). Factors such as the number of nodes, or links, a word has with other words, the way these nodes are organized and the level of strength between them, could all have an impact on the learner’s language production.

Many of the above aspects of vocabulary knowledge (e.g., collocation, connotation, and register constraints) are implicit in character. Their acquisition thus requires extensive and
highly contextual exposure to the target language. Some aspects of a learner’s word knowledge, such as polysemous senses, may have been well developed because these parts of vocabulary knowledge are more amenable to intentional learning. Other aspects, such as word association and word collocation, which require much exposure to diverse contexts, however, may have not been acquired by the learner in the same stage and thus links between these aspects may have not been fully developed.

Meara (1982) argues that words are connected very differently in the mental lexicon of an L2 learner and a native speaker of the language. A native speaker’s lexis is primarily organized in paradigmatic associations while that of an L2 learner is largely in syntagmatic associations. This disparity in organizing word knowledge components could affect the appropriateness and effectiveness of L2 learners in the use of the target language. Another factor is that some learners partially develop their L2 vocabulary links in the lexicon by transferring their existing L1 lexical knowledge to newly learned L2 words (Webb, 2007). This may mean that many L2 learners develop their vocabulary networks in the lexicon according to their existing L1 lexical links (Milton, 2009). These L1-related networks could lead to further differences in L2 learners’ language performance.

Some types of vocabulary depth knowledge such as word association and connotation are intuitive and difficult to explain and thus demand extensive exposure to diverse contexts. Insufficiency in either frequency or duration in the exposure could affect the strength of knowledge in some parts. This lack of strength could weaken the learner’s ability to effectively employ many words in the lexicon.

Theories of working memory also lend explanation for the higher demand on lexical competency in L2 speaking and writing. Language production is a complex process for L2 learners. This complexity can be manifested in the writing process. One of the most influential models in this regard (Kellogg, 1996) divides writing into three, albeit recursive, processes: formulation, execution, and monitoring. The first process involves writers’ formulating and organizing ideas, which demands their cognitive and metacognitive resources. The second process involves the execution of the formulated plans, i.e., to translate ideas into linguistic forms. This stage requires writers’ cognitive and linguistic knowledge, including retrieving related lexical items, encoding clauses and sentences syntactically, and establishing cohesive relationships in the written text. The last process is to monitor the quality of the created text, to ascertain whether the composed text effectively expresses the writer’s intention. If not, revision needs to be undertaken. This stage requires the use of writers’ linguistic, cognitive, and metacognitive resources. Kellogg’s model indicates a writing task requires different facets of the writer’s knowledge, some at lower-order (linguistic) and some at higher-order (cognitive and metacognitive) levels. Studies have provided empirical evidence showing that cognitive and metacognitive knowledge is a major discriminating factor in the performance of an L1 writer (Victori, 1999) whereas linguistic knowledge is more instrumental in predicting the writing proficiency of L2 learners (Sasaki & Hirose, 1996). This difference can be explained by the theory of working memory capacity.

Working memory refers to a person’s mental capacity to store and orchestrate resources connected to a task (Baddeley, 2003). The role of working memory is described as “a bottleneck for learning” (Gathercole & Alloway, 2008, p. 12). During the process of composing a spoken or written text, for example, a learner’s linguistic, cognitive, and metacognitive knowledge stored in long-term memory, has to be processed by working memory at the same time. If a learner has obtained automatized linguistic knowledge, as is often the case in L1 writing, the learner can attend more closely to the cognitive aspects of the task such as organizing ideas and enhancing the persuasiveness of the text. However, many L2 learners are
found to be “tied up with word- or sentence-level processes” (Sasaki & Hirose, 1996, p. 158). When this lower-order processing load increases, the working memory capacity available for the higher-order (cognitive) aspects of writing/speaking, which are required in all the three processes in Kellogg’s model, is reduced. The consequence of this insufficient activation of cognitive knowledge is very likely reduced quality of the written/spoken product.

The preceding working memory theories point to the importance of automatization of L2 learners’ lexical and syntactic knowledge. Of these two language domains, lexical knowledge deserves more attention from advanced L2 learners. At this proficiency level, most of these learners have gained a sufficient understanding of syntactic rules (You, 2010). Their level of automaticity in lexical use has thus become more instrumental in composing a spoken or written text.

Speaking

The demanding nature of lexical automaticity in the productive skills has resulted in challenges to some learners who aim at sophisticated communication, for example, speaking in an academic context in the target language. However, if learners only need to engage in “a simple conversation”, knowing the 2,000 most frequent words could be “adequate” (Laufer & Nation, 2012, p. 169). This figure is possibly derived from a large-scale investigation (Schonell, 1956) which examined the spoken interaction of Australian workers and found that 99% of the words used in their conversations were from the most frequent 2,000 word families. This figure, however, was challenged by Adolphs and Schmitt (2003). By analyzing the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) and spoken component of the British National Corpus (BNC), Adolphs and Schmitt contend that conducting everyday conversation competently would require more than what is available at the 2,000-word level, which can only allow learners to engage in simple conversations. They suggest that sophisticated oral communication requires a much larger vocabulary.

The demanding nature of lexical automaticity is evident in Saito et al. (2016), which examined the lexical factors that determined the comprehensibility (ease of understanding) of oral narratives delivered by 40 French learners of English. The results indicate that lexical appropriateness, fluency and diversity are the main discriminating factors of the learners’ comprehensibility. The study also found that comprehensibility for beginner-to-intermediate learners was closely associated with their fluent and correct use of L2 words, whereas in the case of intermediate-to-advanced learners, comprehensibility was more related to their level of sophistication in using L2 lexis, meaning “morphologically accurate use of complex, less familiar, and polysemous words” (Saito et al., 2016, p. 678). These results were in line with findings from Crossley et al. (2014), which found collocation accuracy (a fundamental element in lexical appropriateness) critically important in their learners’ speaking performance.

Lu (2012) corroborates Saito et al. (2016) and in addition argues that lexical variation can be a strong predictor of learners’ speaking ability. By analyzing the Spoken English Corpus of Chinese Learners (Wen, Wang, & Liang, 2005), Lu explored the relationship between lexical richness (lexical diversity, density and sophistication) and the speaking performance of Chinese learners of English in mainland China. Results showed that lexical variation (diversity) correlated strongly with learners’ scores on evaluations of speaking performance. Lexical density was found to have no impact and lexical sophistication (words beyond the 2,000 frequency level in BNC) “a very small” impact on learners speaking performance (p. 14). Lexical sophistication has long been regarded as an effective indicator of learners’ language proficiency (Daller & Phelan, 2007; Laufer & Nation, 1995). This very small impact made
Vocabulary Knowledge and Language Proficiency

by lexical sophistication on the learners’ speaking ability was possibly a consequence of learners’ lack of exposure to contextual language learning activities in mainland China (Lin, 2015).

Writing

The demand on lexical competence in writing is as high as that in speaking if not more. Writing, defined as the production of an original meaningful written text (Laufer, 2013), is a highly complex task due to the conventions required by a large variety of genres. Tasks such as filling in blanks and correcting lexical or grammatical errors in a passage, connecting two simple sentences to make a complex sentence, or copying a written text, are not considered “writing” in this chapter, although some of these task formats have appeared in the writing component of some high-stakes English proficiency tests such as the National Matriculation English Test in China (Lin, 2015; Qi, 2007; You, 2010).

Vocabulary knowledge has been found to correlate significantly with learners’ writing performance (Lin, 2015; Shi & Qian, 2012; Stæhr, 2008). Lin analyzed the performance of 67 Hong Kong university students on three language measures: Vocabulary Level’s Test, Word Associates Test, and a writing test. Results showed that scores on the two vocabulary tests representing learners’ lexical knowledge could predict a quarter (25%) of their writing scores. The study also found that scores on vocabulary depth added 11% variance over and above the variance explained by vocabulary size, thus contributing an additional unique prediction of the participants’ writing scores. Results of Shi and Qian (2012) and Stæhr (2008) also confirm that vocabulary knowledge makes an important contribution to L2 learners’ writing proficiency.

Research on lexical richness in L2 learners’ writing draws from the notion that high-frequency words are learned at the beginning of their L2 acquisition. As the learners become more proficient, new properties of words and polysemous words start to be gathered in their mental lexicon. This allows for sophistication in their lexical use, for example, employing lower frequency words, having more diversity in word choice and achieving more appropriate word collocation. This notion is empirically supported by Laufer and Nation (1995), who found that the higher the proficiency level, the more likely the learner was able to deploy low-frequency words in writing. Two later studies, Daller and Phelan (2007) and Lerenzo-Dus (2007), provide further empirical evidence for this notion, confirming that lexical sophistication serves as an effective indicator of a learner’s writing. A recent study (Bestgen, 2017) also found that both lexical diversity and lexical sophistication are significant predictors of ESL learners’ writing quality, and competence in using formulaic language is the best predictor of their writing ability. However, it is worth noting that measures of lexical richness may produce different results when applied to different genres. For example, Olinghouse and Wilson (2013) report that vocabulary diversity has a unique predating power for the quality of learners’ story texts, whereas sophisticated use of content words is a strong predictor of the writing quality of persuasive and informative texts.

Future Directions

Many topics related to the examination of the relationship between vocabulary knowledge and language proficiency are under-researched, in particular those pertaining to the two productive language skills. The area with perhaps the greatest value for further research is examining the extent to which learners are able to use memorized words appropriately in
speaking and writing. Learners may occasionally have to rely on dictionaries and vocabulary lists in association with preparation for high-stakes examinations (Lin, 2015; Qi, 2007; You, 2010). Words learned through such channels often cannot be used fluently in their language production. Many of these learners finally have a relatively large vocabulary size but do not know much about the words that they have studied and are unable to use them effectively (Schmitt, 2014). Milton (2009) suggests that they have a large number of words stored in the mental lexicon but these words are poorly organized for retrieval. This issue emerged in Lin’s (2015) study which explored the relationship between learners’ lexical knowledge and their writing proficiency. The participants in this study included two groups of first year university students. One group was from mainland China where there is a shortage of L2 input and output opportunities. The other group was from Hong Kong where there are abundant L2 input and output opportunities and an increasing demand for advanced English language proficiency learners to fill prestigious jobs in the territory. The mainland participants reported that their vocabulary knowledge was derived mainly from vocabulary lists. Some of the participants reportedly memorized as many as 200 new words per day. Their Hong Kong counterparts, however, relied much less on vocabulary lists. A large proportion of their vocabulary knowledge was acquired from reading and listening. This supports the notion that vocabulary size is the basic dimension of learners’ vocabulary knowledge and vocabulary depth serves to further refine their word knowledge and facilitates automaticity in retrieving words for language production (Qian, 2002; Qian & Schedl, 2004; Schmitt, 2008; Stehr, 2009).

Lu (2012) also found that lexical sophistication only had “a very small” impact on the speaking proficiency of the English learners in mainland China (p. 14). Results of these two studies, however, deviate from other studies on the relationship between learners’ lexical knowledge and their ability in language production (e.g., Laufer, 1994; Laufer & Nation, 1995). This deviation is very likely a consequence of the unconducive English learning context in mainland China (Lin, 2015). More research is desirable to gain a better understanding of the impact of language learning context on learners’ ability to use the target language.

Furthermore, there is still a paucity of research on automaticity of vocabulary knowledge and its role in language performance. While investigation into automaticity has been challenging due to its difficulty in operationalization, the advancement of modern technology, in particular eye tracking and brain imaging techniques (e.g., EEG and fMRI) makes it possible to research this dimension of vocabulary knowledge in relation to language processing and proficiency levels.

The second area that requires further investigation is the contribution depth of vocabulary knowledge can make to explaining the variance in learners’ speaking production. The contributions of depth of vocabulary knowledge that go beyond vocabulary size have ranged from 5% for listening comprehension (Stæhr, 2009) to 13% for reading comprehension (Qian, 2002) and 11% for writing performance (Lin, 2015). It would be interesting to see if the construct of depth of vocabulary knowledge can also make a unique and significant contribution to the prediction of learners’ speaking ability. In addition, the figures derived from Han (2017), Qian (1999, 2002), Stæhr (2008, 2009), and Lin (2015) varied concerning the predictive power of vocabulary depth relative to that of vocabulary size. Even though explanations may have been provided for the discrepancies between these figures in the context of individual studies, replicated studies, and meta-analysis would be useful to clarify the findings.

Another area that needs scholarly attention is the role of the first 3,000 word families in predicting the language proficiency of advanced L2 learners. Laufer and Ravenhorst-Kalovski (2010) studied 745 Israeli students in an academic college, and both teachers and
learners informally reported that students with a vocabulary size of 3,000 families still could not manage required reading assigned to them even after having completed three semesters of mandatory English support classes. Lin (2015) analyzed the writing performance of 150 university students in Hong Kong. The learners’ lexical knowledge at the 3,000-word level was weakly associated with their writing scores. When a regression analysis was performed to identify the contribution that the learners’ word knowledge at the 3,000 and 5,000 frequency levels could make to their writing performance, the 3,000 level was forced out from the equation, indicating its insignificant role in predicting the learners’ writing scores. A recent paper by Dabbagh (2016), which examined the listening comprehension of 73 EFL students in an Iranian university, found only a weak correlation between the students’ vocabulary knowledge of the 3,000 word families and their listening comprehension. Results from these and other similar studies indicate that word knowledge at the 3,000-word-family level has a rather limited role in predicting learners’ language proficiency. Therefore, more investigations should be conducted to determine which word frequency levels can effectively predict L2 learners’ language proficiency for various purposes. Findings from such investigations will possibly provide significant pedagogical implications. A further reason for conducting such investigations is that there is an increasing agreement in the literature that the 5,000-word level constitutes the minimum threshold required for undertaking academic studies in an English-medium university (Roche & Harrington, 2013). Based on this threshold, the 3,000-word level is far below the required level. Given these findings, it is not surprising to see a suggested new boundary of high-frequency English words (see Schmitt & Schmitt, 2014). Following this suggestion, the level of 3,000 word families should be included in the high-frequency vocabulary. This may mean the first 3,000 word families play a very limited role in predicting the language proficiency of advanced L2 learners.

Further Readings


Based on a synthesis of related conceptual and empirical studies, this paper was among the first to recognize the importance of depth of vocabulary knowledge in second language reading comprehension. It provides an in-depth empirical analysis of the complex relationships between the size and depth of vocabulary knowledge in a learner’s lexicon.


This article critically reviews recent empirical studies on the relationship between size and depth of vocabulary knowledge. The article evaluates the merits of such conceptualization but also points out issues arising from this line of research.

Related Topics

The different aspects of vocabulary knowledge, L1 and L2 vocabulary size and growth, measuring depth of vocabulary knowledge, measuring lexical richness, key issues in measuring vocabulary knowledge

References


