ReCALL (2021), 33: 1, 4–17
doi:10.1017/S0958344020000117

ARTICLE

Narrow reading, vocabulary load and collocations in context: Exploring lexical repetition in concordances from a pedagogical perspective

Oliver James Ballance
Victoria University of Wellington, New Zealand (Oliver.Ballance@vuw.ac.nz)

Abstract

Narrow reading has the potential to reduce vocabulary load and to provide rich opportunities for developing collocation knowledge, but these benefits rely on narrow reading increasing lexical repetition within a text. Hence, interest in narrow reading has been limited by the relatively small lexical effect of narrowing reading by topic (Nation, 2013). Nevertheless, research in data-driven learning and teaching and language corpora has reported positively on learners using concordances in a manner comparable to narrow reading. However, the potential for concordances to provide an increased lexical-repetition effect has not been assessed. This study bridges this gap by exploring the degree of lexical repetition available in concordances and identifies corpus composition as a key predictor of lexical repetition. The study uses standardised type-token ratio (sTTR) to analyse concordances extracted from corpora at three different levels of homogeneity/heterogeneity. The results show large, reliable variations in lexical repetition resulting from variation in corpus homogeneity/heterogeneity, and so identifies concordance-based narrow reading as a possible means of overcoming the limitations of traditional narrow reading by topic. The results are discussed with reference to pedagogical implications for language learners, teachers, and researchers.

Keywords: corpus linguistics; data-driven learning; depth of vocabulary knowledge; narrow reading; type-token ratio; vocabulary load

1. Introduction

There are strong links between learner use of concordances, vocabulary learning and development, and reading research. Concordances are often recommended for vocabulary study (Johns, 2002; Lee, Warschauer & Lee, 2019), vocabulary knowledge is a prerequisite of reading comprehension (Grabe & Stoller, 2011), and reading is seen as a core driver of vocabulary development (Nation, 2013). Unsurprisingly then, a number of researchers have reported on and recommended concordancing as an effective means of second language acquisition through reading (Bernardini, 2000, 2002, 2004; Cobb, 1997, 1999). This paper presents as an empirical exploration of concordances as a resource for narrow reading.

Traditionally, narrow reading is the practice of reading multiple texts around a single topic or theme. One major advantage is that it may help to reduce the difficulty of reading in a foreign language (Nation, 2013). On the one hand, it increases the amount of background knowledge that a learner has available, a kind of top-down processing advantage; on the other, narrow reading may reduce the number of words a learner needs to recognise by reducing lexical variation in a

Cite this article: Ballance, O.J. (2021). Narrow reading, vocabulary load and collocations in context: Exploring lexical repetition in concordances from a pedagogical perspective. *ReCALL* 33(1): 4–17. https://doi.org/10.1017/S0958344020000117

© European Association for Computer Assisted Language Learning 2020.

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

text, a type of bottom-up processing advantage (Grabe & Stoller, 2011). This is because a comparable amount of writing on a single topic will have less lexical variety than an equivalent amount of writing on a range of different topics (Kennedy, 1998). As Nation and Waring (1997: 10) point out, "Within narrowly focussed areas of interest, such as an economics text, a much smaller vocabulary is needed than if the reader wishes to read a wide range of texts on a variety of different topics".

A second reason why narrow reading is advocated is that reading in a narrow field increases the number of times that particular words and their collocations are encountered in novel contexts. Multiple encounters with words in novel contexts are theoretically important in both vocabulary studies and corpus linguistics. In vocabulary studies, meeting words in novel contexts is connected to the idea of depth of vocabulary knowledge: knowing how and when words can be used (Nation, 2013). A very similar idea is important in corpus linguistics, where meeting words in multiple contexts is how we develop lexical primings: knowledge of how, where and when a word can be used, derived inferentially from experiencing the word's use in different linguistic and social contexts (Hoey, 2005). As such, narrow reading has the potential to fast track mastery of lexical items by providing reading that is focused on a comparatively narrow range of lexical items. In relation to this study in particular, we can think of the potential for narrow reading to provide rich exposure to collocations in context.

Despite the potential value of narrow reading, it has been relatively neglected in the applied linguistics literature. Nation (2013) suggests that this may be because the effect of narrow reading on lexical variation/repetition is relatively modest because of the natural distribution of low-frequency words in text. However, research has not yet explored the potential for concordances to provide sources of narrow reading with unusually high degrees of lexical repetition.

2. Conceptual background

2.1 Concordances as reading resources

People may typically envision a concordance in the key word in context (KWIC) format in which a keyword is displayed in the centre of a screen and only one line of text is shown for each instance of the keyword. However, concordances can also return results that show a sentence or an entire paragraph of text for each instance of the keyword. Hence, concordances can be used for quite naturalistic reading experiences.

Within the sub-field of learner use of corpora/data-driven learning (DDL), there have been two main proponents of reading concordances. Cobb (1997, 1999) has argued that concordances can provide learners with *massed exposure* to a word in context, and that massed exposure to a word in context is what is needed for depth of vocabulary knowledge to develop. Although his research is couched in terms of massed exposure, the underlying logic is comparable to that of narrow reading. Bernardini (2000, 2002, 2004) has also argued for the value of exploring corpora through a concordancer in an open-ended reading procedure she terms either *discovery learning* or *serendipitous exploration*. It is an approach that emphasises attention to usage, learner autonomy and high levels of task engagement. It is compatible with narrow reading in that she reports on learners using a concordancer to pursue language or topics that they find interesting via a concordancer. Thus, although concordances may not typically be conceived of as a source of reading material, there are pedagogical motivations for viewing them as such, and these motivations are comparable to the motivations behind narrow reading.

Concordance-based narrow reading has the potential to provide even narrower reading than reading by theme or topic because a concordance returns multiple text extracts (citations, results, hits or lines) that contain a user-specified linguistic category, such as a word or collocation. Hence, concordances allow learners to read lexical items in novel contexts in much higher concentrations than would be possible in any normal text. In other words, concordance-based narrow reading

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

may be able to increase lexical repetition to the point at which it overcomes the barrier to more widespread use that Nation (2013) identifies.

2.2 Lexical repetition and type-token ratio

The extent to which a text presents sufficient lexical repetition to support narrow reading can be operationalised as a measure of its type-token ratio (TTR). A text's TTR can be determined by dividing the count of the number of word types by the count of the word tokens. For ease of discussion, this figure can then be multiplied by 100. The closer a text's TTR score is to 100, the less lexical repetition there is in the text. The nearer a text's TTR score is to 0, the more repetition there is. A text that consisted of 250 different words would have a TTR of 250/250 x 100 = 100. In contrast, a text that consisted of one word repeated 250 times would have a TTR of 1/250 x 100 = 0.4. Consequently, TTR has been described as a simple and intuitively appealing measure (Jarvis, 2013). However, it has also been criticised for two main reasons.

One criticism of TTR relates to comparing texts of greatly different lengths (Jarvis, 2013; Malvern, Richards, Chipere & Durán, 2004; McCarthy & Jarvis, 2010). However, one way to avoid this issue is to use texts of very similar length. The problem can then be further reduced by calculating a standardised TTR in which the figure is averaged for every 100 words it contains (Scott, 2016).

A second possible criticism of TTR relates to the distinction between measures of lexical diversity and measures of lexical variation/repetition. TTR has frequently been used as a measure of lexical diversity in fields such as stylistics, neuropathology, first language acquisition, second language acquisition, data mining and textual forensics, and interpreted as an index of a wide range of underlying constructs, such as writing quality, vocabulary knowledge, speaker competence, Alzheimer's onset, hearing variation and socioeconomic status (McCarthy & Jarvis, 2010). However, as Jarvis (2013) explains, TTR is not a good measure of lexical diversity because it does not take account of how repetition is perceived, such as when it helps to create coherence in a text. Consequently, Jarvis claims that TTR is not a good measure of lexical diversity, but only of lexical variability – that is, how often words are repeated. This is an important distinction for the study reported here, which is not concerned with lexical diversity (quality of repetition), only lexical variability (quantity of repetition), and, as such, TTR captures the construct of interest precisely.

2.3 Lexical repetition and vocabulary load

In terms of TTR as a measure of text difficulty, higher levels of lexical repetition can be interpreted as presenting learners with lower vocabulary burdens than texts with greater lexical variation (Nation, 2013; Schmitt, 2000). However, as noted above, TTR has not been a popular measure of vocabulary load due to the modest degree of increased repetition that reading by topic is likely to provide to learners. This is because the topic effect occurs in much larger quantities of naturally occurring text than many learners are likely to read. The hypothesis behind this study is that concordances may be able to provide sources of narrow reading that deliver significantly increased levels of repetition. However, even if concordance-based narrow reading does deliver significant changes in TTR, there are several important factors to consider when interpreting TTR as a measure of vocabulary load.

The relationship between TTR and the notion of vocabulary load may appear straightforward initially. Less vocabulary variety can be interpreted as indicative of a lower vocabulary load in the straightforward sense of requiring learners to recognise fewer word types. Furthermore, if texts are on similar topics, this may help a learner bring their pre-existing schemata to bear on text comprehension. However, it should be acknowledged that the extent to which low TTR scores represent a lower vocabulary load would appear to be contingent upon the text having already met the more

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

basic criterion of appropriate proportions of known and unknown words in a text: the concept of vocabulary coverage affecting text comprehension, as discussed by Nation (2013).

TTR makes no assumptions about which words a reader knows and which words are present in a text. Indeed, this can be seen as the main drawback of TTR as a measure of vocabulary load. Schmitt (2000) gives the example of an academic text that repeats key terms and hence scores higher than a children's text with a simpler but more diverse vocabulary. He points out that TTR scores may not always reflect our intuitions as to the relative difficulties of texts. If a text has a relatively low TTR, a reader will have to deal with a relatively narrow range of vocabulary as compared with a text of similar length but a higher TTR; this does not mean, however, that the words are likely to be known to the reader. The utility of TTR as a measure of vocabulary load depends on the assumption that the text consists primarily of word types that the reader is likely to know. Nation (2013) suggests that narrow reading may be most appropriate for advanced learners reading in a subject area that they are familiar with. The need to assume that most words are known is probably Nation's rationale for this prescription.

With this issue in mind, research by Ballance and Coxhead (2020) has shown that recognition of 98% of the words in the average citation in a concordance only requires knowledge of around 5,000 word families, and that 4,000 may be sufficient if lexically simpler texts are being concordanced. That is, if a reader is not obliged to read every citation in a concordance, a concordance provides text with a considerably lower vocabulary load than a naturally occurring whole text (typically 8,000 word families or more). Furthermore, we should also consider the implications of research that has shown that second language learners' vocabulary knowledge often deviates from that of native speakers, with learners often having mixed word knowledge profiles – that is, profiles wherein they know words at a variety of frequency levels but do not necessarily possess complete knowledge of even the highest frequency word lists (Cobb, 2010; Milton, 2009; Webb & Chang, 2012). Hence, for learners in languages for specific purposes (LSP) contexts who have professional or field-specific knowledge, we may find that measures of vocabulary load based on general word frequencies do not reflect their vocabulary knowledge well. Consequently, in LSP contexts in which vocabulary knowledge can be expected to match the vocabulary presented in a text, learners may be expected to be familiar with technical terms and other topic-related low-frequency vocabulary, and so TTR may be considered either a complementary or preferable measure of vocabulary load to those based on word frequency bands.

2.4 Lexical repetition and collocations in context

In terms of opportunities for developing knowledge of collocations in context, TTR indicates lexical repetition in a text, and hence it provides an indication of opportunities for learning how a word collocates. Concordances have attracted considerable research interest because of their potential to illustrate such lexical patterning (Hunston, 2002; Sinclair, 1991), and just as concordances have helped researchers observe the patterns of collocation that a search word occurs in, many researchers have suggested that concordances may also be able to help learners acquire such patterns (Cobb & Boulton, 2015; Lee *et al.*, 2019). Fundamentally, all such lexical patterns depend on repeated co-occurrence; collocations, n-grams, clusters, skipgrams and concgrams are all based on the analysis of recurring orthographic forms. Because concordances present citations, all of the word tokens in a concordance co-occur within the span of the search word specified in extraction of the citation and, consequently, the TTR of a concordance not only reports the extent of lexical repetition in the text (Jarvis, 2013), but also the extent of lexical repetition co-occurring with the search word. Thus, concordances with lower TTR scores contain more occurrences of the same collocates than concordances with higher TTR scores, and so afford learners more opportunities to observe patterns of co-occurrence. Indeed, identifying recurrent lexis in concordance lines is something of a mainstay of the literature on learner use of concordances, and TTR scores provide a useful measure of the extent to which a concordance affords

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

learners the opportunity to observe collocation, in the corpus linguistic sense of repeated lexical co-occurrence.

2.5 Lexical repetition and corpus composition

Finally, corpus composition is a key variable in concordance generation generally (Ballance, 2017) and a fundamental consideration in corpus-based LSP (Gavioli, 2005). However, as previously indicated, we should expect it to be of particular importance in measures of TTR because the original idea behind narrow reading was reading on a narrow theme or topic. Hence, this study explores differences between the TTR of concordances extracted from corpora that differ in terms of the level of generality in their composition: whether a corpus is composed of a more or less homogeneous or heterogeneous collection of texts. This reflects the relationship between topic homogeneity and lexical homogeneity observed in previous work on TTR and narrow reading (Nation, 2013). However, it also relates to the different conditions under which concordancing has been operationalised in studies of learner use of concordances.

Some studies report having used concordances of corpora that seek to represent a language in general, such as British English or American English (Boulton, 2010; Gordani, 2013). Other studies have presented learners with concordances from a more narrowly defined language variety of one type or another – for instance, concordances of only the written section of the British National Corpus (BNC) (Kaur & Hegelheimer, 2005), concordances of newspapers representing a national variety of English (Boulton, 2009; Chan & Liou, 2005), or concordances of various collections of academic English (Charles, 2011). And then, at a further level of specificity, we find studies in which classes of learners were able to concordance the reading materials available to them on the course they were studying (Cobb, 1997, 1999), or corpora of writing relevant to a task they were working on (Chambers & O'Sullivan 2004; O'Sullivan & Chambers, 2006; Park, 2012). Finally, at the finest level of corpus homogeneity are studies in which learners compiled their own corpora of texts related to their individual field of language use (Charles, 2015; Lee & Swales, 2006). Because such a wide range of corpora at different levels of generality have been used, and because TTR scores have implications for vocabulary load and their capacity to afford learners observation of lexical patterning, it is important to explore the level of corpus generality as a potential determinant of the TTR scores of concordances.

2.6 Research questions

This paper explores two research questions:

1. Do concordances extracted from corpora with differing levels of generality have significantly different sTTR?
2. What is the magnitude of difference between the TTR scores of concordances extracted from corpora of differing levels of generality?

The reliability of the results obtained is examined through replication and interpreted with reference to concordances as a source of narrow reading material.

3. Methodology

3.1 Corpora

The experimental conditions used in this study were three different corpora that were used to generate concordances. The three corpora each consisted of different samples of texts from the BNC, and each corpus represented a different level of generality. That is, the corpora were composed of more or less heterogeneous or homogeneous text types. At the highest level of

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

Table 1. Summary of corpora used

Corpus	Size	Composition
Corpus of general written English	~16 mil.	Texts from 54 non-academic subgenres
Corpus of written academic English	~14.5 mil.	Academic texts
ESP corpus of written academic English	~700,000	Five texts from a single journal

Note. ESP = English for specific purposes.

generality, texts were selected to represent written language in general: *a corpus of general written English*. At the middle level of generality, texts were selected to represent written academic language: *a corpus of written academic English*. At the lowest level of generality, all the texts in the corpus were extracts from a single academic journal: *an English for specific purposes (ESP) corpus of written academic English*. This last corpus is intended to represent the kind of learner-compiled, LSP corpora discussed previously. Table 1, and the information thereafter, present a brief summary of the size and the composition of the corpora used in the study.

For the corpus of general written English, it was not possible to use the full written subcorpus of the BNC due to the file size constraints of the Python script used to extract concordances. Consequently, it was necessary to sample texts from within the written section of the BNC. Texts were sampled from a variety of genre categories (as defined in Lee, 2002) from the nine domain categories of the BNC: applied science, arts, belief and thought, commerce and finance, imaginative, leisure, natural and pure science, social science, and world affairs (Burnard, 2002). This created a subcorpus of the written BNC within which each of the nine domains contributed an approximately equal quantity of data and thereby represented a very broad scope of written English. The resulting subcorpus comprised approximately 16 million words from 54 non-academic subgenres (Lee, 2002) of written English.

The corpus of academic English consisted of all texts from the written subsection of the BNC identified as academic (Lee, 2002), with the exception of five text files used to form the ESP corpus of written academic English discussed below. The corpus of academic English comprised approximately 14.5 million words.

The corpus constructed to represent the lowest level of generality was formed from five text files from the BNC: HU2, HU3, HU4, HWS and HWT. They were identified via Lee and Rayson's BNC Web Indexer (http://ucrel.lancs.ac.uk/bncindex/). All five files are extracts from the academic journal *Gut: Journal of Gastroenterology and Hepatology*, and together they consisted of approximately 700,000 words. This corpus represents the kind of self-compiled corpora discussed in Charles (2015) and Lee and Swales (2006).

3.2 Search terms

After compiling the three corpora that formed the independent variable in this study, the next step was to select the search terms that would be used to extract concordances from the corpora. Selection of search terms to use in this study addressed two main data collection concerns: availability and dispersion. On the one hand, it was important that there were sufficient occurrences of the search words used in each of the three corpora. On the other hand, it was necessary to ensure that the search terms generated citations that had sufficient dispersion within the more general corpora to meaningfully represent citations drawn from the general corpora. That is, that the citations were not extracted from just one or two texts within the larger, more general corpora. To address the availability issue, because the ESP corpus was the smallest corpus used in this study, the availability of words in the ESP corpus was used as the starting point for identifying search terms to use; words available in sufficient quantities in the smallest corpus were likely to be

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

Table 2. Search terms used in standardised type-token ratio studies

Search terms 1–50	Search terms 51–100
able, action, added, age, agreement, air, allow, almost, alone, already, also, amount, animal, apparent, appearance, area, base, become, better, blood, body, bone, breath, carried, case, caused, change, children, class, clear, collection, complete, considered, continuous, control, cost, count, course, cross, daily, day, death, degree, dependent, died, differences, difficult, double, drugs, during	earlier, end, especially, even, except, expected, experience, explanation, expressed, fact, fall, family, fast, fat, fed, final, finding, fixed, follow, food, forms, found, free, full, further, gave, general, good, greater, group, growth, health, high, history, hospital, hour, human, important, indeed, information, informed, interest, involvement, known, large, last, later, lead, least, left

available in sufficient quantities in the two much larger corpora. To provide some measure of control over the dispersion of search terms in the two larger corpora, high-frequency words were also chosen. Thus, each of the 100 search terms used were word types found in the most frequent 1,000 word families in Nation's BNC/Corpus of Contemporary American English (COCA) 25 word-family frequency lists. Each set of search terms is shown in Table 2. As hypothesised, they were found to occur sufficiently frequently in the more general corpora, and visual inspection of the citations in the concordances was used to confirm an appropriate degree of dispersion.

3.3 Generating concordances

For each of the 100 search terms used in this study, a concordance of 70 citations was extracted from each of the three corpora (example concordances are available online as supplementary material). Citations were selected randomly, and each citation was extracted with as many complete words of co-text as found within an 80-character span of the search term. This helped to ensure that the concordances were all of similar length. Duplicate citations were manually identified, removed and replaced with randomly selected non-duplicates. The data were then split into two halves to provide two sets of concordances consisting of 50 search words each.

3.4 Scoring

To generate consistent measures of lexical variation and repetition, each concordance generated was scored using standardised type-token ratio (sTTR) in WordSmith Tools (Scott, 2016) as displayed in the statistics tab when a word list is generated from an input text. The scoring proceeded on the basis of word type (not lemma or family), no stop list was used, and the program was set to standardise scores to TTR per 100 words. Using standardised TTR compensated for any minor difference between concordances in terms of their exact word length.

3.5 Data analysis

Data for item sets 1–50 and 51–100 were both examined before conducting statistical analyses. Both data sets satisfied the assumption of normal distribution but failed the assumption of equal variance. However, because group sizes were equal and $n = 50$ in each test, the effect of unequal variance is a very minor concern: a distortion of α by a few hundredths (Sheskin, 2004). More importantly, failure to satisfy the assumption of equal variances indicated that Dunnett's T3 was the appropriate post hoc test to use; this is a conservative test that can be used when all pairwise comparisons can be made but the assumption of equal variance is not met (Larson-Hall, 2010). The text statistic used was Welch's adjusted F because this test statistic can also accommodate samples with unequal variance (Field, 2009).

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

4. Results

4.1 Results for search terms 1–50

A one-way ANOVA was conducted to compare the effect of *level of corpus generality* on *sTTR*. Descriptive statistics for the groups were *written*, X = 70.74, *SD* = 1.97, *n* = 50; *academic*, X = 67.78, *SD* = 2.57, *n* = 50; *ESP*, X = 64.12, *SD* = 3.57, *n* = 50. Since the assumption of equal variances was not met, the test statistic used for the ANOVA was Welch's adjusted *F* ratio. ANOVA showed that the effect of corpus generality of sTTR was significant, Welch's $F(2, 93.267) = 70.667$, $p = < .001$. Comparisons using Dunnett's T3 with 1,000 bootstrap samples found statistically significant differences between all groups: *written* and *academic*, MD = 2.96, 95% CI [2.04, 3.88], $p = < .001$, $d = 1.5$; *academic* and *ESP*, MD = 3.66, 95% CI [2.15, 5.17], $p = < .001$, $d = 1.42$; *written* and *ESP*, MD = 6.63, 95% CI [5.22, 8.03], $p = < .001$, $d = 3.36$. This result indicates that concordances extracted from more general corpora have higher sTTR levels than concordances extracted from more specialised corpora, supporting the hypothesis that level of corpus generality has a very large effect on the sTTR of concordances.

4.2 Results for search terms 51–100

A second one-way ANOVA was conducted to compare the effect of *level of corpus generality* on sTTR in the second data set. Descriptive statistics for the groups were *written*, X = 71.07, *SD* = 2.31, *n* = 50; *academic*, X = 68.31, *SD* = 2.10, *n* = 50; *ESP*, X = 64.14, *SD* = 4.30, *n* = 50. Again, since the assumption of equal variances was not met, the test statistic used for the ANOVA was Welch's adjusted *F* ratio. ANOVA showed that the effect of corpus generality of sTTR was significant, Welch's $F(2, 92.786) = 54.254$, $p = < .001$. Comparisons using Dunnett's T3 with 1,000 bootstrap samples found statistically significant differences between all groups: *written* and *academic*, MD = 2.76, 95% CI [1.89, 3.63], $p = < .001$, $d = 1.32$; *academic* and *ESP*, MD = 4.17, 95% CI [2.89, 5.61], $p = < .001$, $d = 1.99$; *written* and *ESP*, MD = 6.93, 95% CI [5.24, 8.61], $p = < .001$, $d = 3.01$. The analysis of the second data set confirmed the results of analysis of the first: concordances extracted from more general corpora had higher sTTR levels as compared with concordances extracted from more specialised corpora, supporting the hypothesis that level of corpus generality has a very large effect on the sTTR of concordances. The results of both data sets are presented in Table 3.

5. Discussion

The results clearly indicate that the concordances of high-frequency words extracted from the three different corpora differed significantly on a standardised measure of TTR. Not only are all of the effect sizes reported very large (anything over .08 is typically considered large), but they represent significant differences in real terms. Because ratios were standardised to 100 word tokens, average score differences between conditions are equivalent to differences in the average number of word types per 100 words. Hence, taking the average of the mean difference between scores for both data sets, we can say that concordances from the corpus of general written English contained an average of 2.86 (*SD* = 0.1) more word types per hundred word tokens than concordances from a corpus of written academic English. This figure approximates to around three more word types per eight citations. Comparing concordances from a corpus of written academic English with concordances from an ESP corpus composed of extracts from a single academic journal, we observe an average difference of 3.92 (*SD* = 0.4). That is, the more general corpus contains almost four more word types per eight citations. Unsurprisingly then, when concordances from the general written corpus are compared with those from the ESP corpus, the average is 6.78 (*SD* = 0.2): nearly seven more word types per eight citations in the general written English concordances than the ESP concordances. To put this another way, learners would be confronted

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

Table 3. Summary of standardised type-token ratio comparisons for level of corpus generality

		MD	*d*	95% CI
Written cf. academic	Items 1–50	2.96	1.5	2.04, 3.88
	Items 51–100	2.76	1.32	1.89, 3.63
	AVG	2.86		
	SD	0.1		
Academic cf. ESP	Items 1–50	3.66	1.42	2.15, 5.17
	Items 51–100	4.17	1.99	2.89, 5.61
	AVG	3.92		
	SD	0.4		
Written cf. ESP	Items 1–50	6.63	3.36	5.22, 8.03
	Items 51–100	6.93	3.01	5.24, 8.61
	AVG	6.78		
	SD	0.2		

Note. ESP = English for specific purposes.

with nearly one fewer word type per citation in the ESP concordance. This finding has important implications for both the vocabulary load of the concordances and their capacity to provide opportunities for learning collocations.

5.1 Vocabulary load

With regard to vocabulary load, these findings show that concordancing may be an effective way of realising the goal of providing text with a reduced vocabulary load. While TTR score reductions may be difficult to operationalise through narrow reading of whole texts (Nation, 2013), these results show that significant reductions in lexical variability can be achieved by concordancing corpora composed of closely related texts. Unfortunately, because TTR measures are sensitive to text, it is difficult to identify a whole-text TTR baseline to compare these results with. To make such a comparison, we would need to compare the TTR of equivalent quantities of whole texts and concordances, but such a comparison might be considered somewhat dubious given the different characteristics of whole texts and concordances. However, given the fact that the average citation in a concordance has a lower vocabulary load than that of a whole text (Ballance & Coxhead, 2020), it seems clear that concordances of more homogeneous texts on familiar topics have real potential for providing learners with texts that have a lower vocabulary load. As such, it suggests that concordances represent an accessible source of narrow reading, lending support to Cobb's advocacy of concordances as a source of massed exposure (Cobb, 1997, 1999) and Bernardini's (2000, 2002, 2004) work on discovery learning and serendipitous exploration.

At the same time though, it should also be remembered that the potential for a reduction in lexical variability to represent a lower vocabulary load is dependent on the vocabulary contained being known; reduced TTR is probably irrelevant to readability when many words are unknown. Hence, readability gains reflected in reduced TTR scores for concordances extracted from corpora at lower levels of generality are likely to depend on a match between learners' internal lexicons (Cobb, 2010; Milton, 2009) and the particular words contained in a given concordance. In other words, from the perspective of readability, these results are most pertinent in relation to the use of concordances in LSP settings in which it is feasible to narrow the composition of a corpus in such a

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

way that it reflects the learners' areas of familiarity or expertise (cf. Chambers & O'Sullivan, 2004; Charles, 2015; Lee & Swales, 2006; O'Sullivan & Chambers, 2006; Park, 2012).

Another important consideration pertaining to increased readability is the type of concordance used in this study. To control for the effects of text length, this study operationalised concordance extracts generated via a fixed citation length in the KWIC format. However, research has shown citation format to be an important variable within learner use of concordances (Ballance, 2017), and future studies might consider exploring lexical variation in relation to sentence format or paragraph format citations. Although these citation formats are methodologically problematic because they vary greatly in length, they are the preferred citation format for learner concordance use that is based on reading (Ballance, 2017). There is no strong reason to assume the results obtained in this study would be vastly different to those obtained using variable length citations, but variation in citation length may have some moderating effect on TTR differences. For example, we might expect the TTR scores of concordances composed of longer citations to have slightly more varied TTR scores because they will typically include more word tokens at greater distances from the search word, reducing the collocational pull of the search term itself, and so increasing variance. Exactly how much effect this might have is hard to predict, but it is unlikely to contravene the findings reported here because the effect sizes reported are very large.

5.2 Opportunities for learning collocations in context

Lexical repetition underlies the concept of one word being a common collocate of another. If a certain word only co-occurs with the search word a single time in a concordance, the concordance user can only observe that these words *can* be used together. However, if a concordance shows the word co-occurring with the search word in multiple citations, then the concordance user has a basis on which to estimate the strength of the association between these two words as compared with other words that occur in the same syntactic relation to the search word a different number of times. Of course, the generalisability of this observation depends upon the composition of the corpus, but the point stands: if there is no repetition, one can only infer that a certain pair of words can co-occur with the search word, not that any particular co-occurrence is more frequent than another. This point is fundamental to corpus linguistic perspectives on collocation (Sinclair, 1991). Lexical repetition provides the underlying basis for one type of generalisation – how frequently different words co-occur with the search term. Noticing this is an important aspect of many DDL-type exercises (see, for instance, Thurstun & Candlin, 1997, 1998), and more homogeneous corpora can provide concordances that display stronger patterns of collocation.

This study has demonstrated that concordances generated to different specifications are not equal in terms of illustrating patterns of lexical repetition, and as such, the degree of lexical repetition in a concordance must be an important variable in studies that explore learning collocations from concordances. Hence, it would be highly desirable for researchers to report sTTR scores for the concordances used in studies of learning collocations through concordances.

Nevertheless, the relationship between learning and the degree of lexical variability in a concordance is complex. Although concordances with lower variability presumably make repeated co-occurrence easier to observe, whether this has a primarily positive or negative effect on learning outcomes is hard to predict. On the one hand, it seems reasonable to assert that the increased lexical repetition indicated by lower TTR scores would make lexical patterns more noticeable and that encountering co-occurrences more frequently would aid retention, but at the same time, the results could also indicate that concordances of more homogeneous corpora expose learners to a more limited range of co-occurrences, perhaps leading to the learner developing a skewed or limited representation of the word (cf. Hoey, 2005, and the notion of *lexical priming*). Exactly how the degree of lexical repetition interacts with the type and extent of learning outcomes is an area of research where more work is needed.

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

Furthermore, although TTR scores do underlie patterns of lexical repetition, TTR is a somewhat crude indicator of linguistic patterning in some important respects. Lower TTR scores indicate less lexical variation and more lexical repetition, and so do indicate an increased opportunity to observe a word co-occurring with the search term in a concordance, but several points need to be remembered about exactly what it is that TTR measures.

First, it is important to note that TTR as used in this study does not distinguish between word types that are repeated within a citation and word types that are repeated across citations. Some proportion of repeated word types may be within a single citation, and hence not representative of the type of collocation observed when a word type is repeated in a number of different citations. However, such words are likely to be function words (it is relatively unlikely to find lexical words repeated within a very short span of words, though phrases such as *more and more* and *better and better* are an obvious exception). There is no strong reason to think that the extent of this phenomenon should vary much between conditions. Repeated lexis within citations probably has a very marginal effect on the results obtained as compared with words repeated across citations.

Second, TTR does not take co-occurrence of lexical sets (e.g. *oil*, *water*, *liquid* and *fluid*) or grammatical categories (e.g. *adjectives*, *nouns*, *verbs*) into account, only collocation of orthographic forms – in this study, lexical repetition at the level of word types. As such, it cannot effectively assess the extent of semantic or grammatical repetition in a concordance.

Third, TTR makes no warrant as to the value or significance of co-occurrence, even when a word co-occurs with the search term in multiple citations. There may be a sentence boundary such as a full-stop between the repeated co-occurring word and the search word, the rate of collocation may be unexceptional in terms of mutual information, or the word may occur at a greater distance from the search word than is typically considered in studies of collocation (for instance, six or seven words away from the search term). From this perspective, it is interesting to note that there is increasing interest in collocates that occur at greater removes from the search term than the traditional five-word span (Hoey, 2015), and so this last issue may not be as significant a disadvantage of TTR as a measure of lexical patterning as it may at first appear. It would, however, still be interesting to explore these issues further, perhaps by measuring at the citation level the frequency of concgrams (Cheng, Greaves & Warren, 2006) occurring in each concordance. An analysis based on concgrams would provide not only a much finer-grained quantitative account of the types of lexical repetition contained in the concordances, but also a starting point for a more qualitative analysis of the types of co-occurrences prevalent in concordances.

6. Limitations

It should be remembered that this study examined concordances of high-frequency words. This was done to ensure that the search words used to generate concordances occurred in sufficient quantities in the ESP corpus examined but also had a reasonable range of distribution within the general corpora. Consequently, it is not clear how generalisable these results are to less frequent search terms. For instance, if we search the whole BNC for an extremely technical term such as endoprosthesis, we find that all 43 occurrences are in two texts: HU4 and HWS – that is, two of the five texts that formed the ESP corpus in this study. This suggests that concordances of very low-frequency words in large general corpora are likely to form their own de facto specialised concordance in virtue of their limited distribution. It seems likely that the effect of level of corpus generality will interact with the dispersion of a search word within a corpus. As less frequent words are generally more narrowly dispersed, we may expect the magnitude of the differences in TTR scores of concordances extracted from corpora at different levels of generality to diminish when less frequent, less dispersed words are concordanced. A better understanding of this phenomenon would help to clarify the circumstances in which the level of corpus generality can be expected to affect a concordance's TTR.

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

Finally, although the experimental results obtained were replicated in two sets of data derived from two different sets of search terms, the corpora examined remained constant. Therefore, we cannot be sure that similar results would be obtained if level of generalisability were operationalised differently. From this perspective, a cross-validation study using different corpora might be useful. But perhaps more importantly, it is not completely obvious which factors in compiling a corpus at a lower level of generality have an effect on the TTR scores of concordances. For instance, the BNC was originally divided into nine domain categories (Burnard, 2002) without a distinction being drawn between academic and non-academic texts (Lee, 2002). Lee and Rayson's BNC Web Indexer provides a plethora of options for compiling narrower subcorpora of texts from the BNC: medium, domain, genre, audience age, audience sex, audience level, author age, author sex, author type, text sampling type, circulation status, interaction type, time period and mode, as well as customisable criteria for identifying texts on the basis of library catalogue keyword fields (COPAC keywords), an indicator of a text's topic. It not clear which ways of narrowing a corpus can be expected to affect the TTR of concordances extracted from them, or the relative magnitude of any effect that each may have. Consequently, exploring the effects of different ways of narrowing the focus of a corpus is an interesting avenue for future research.

7. Conclusion

This study raises many questions about the use of concordances with language learners: To what extent do lower TTR scores affect what can be learned from a concordance? To what extent do TTR differences actually represent more or less readable text? What types of lexical co-occurrence and repetition are low TTR scores indicative of? To what extent is there an interaction between level of corpus generality, the distribution of a search term within a corpus, and TTR scores? Which factors within level of corpus generality are operative upon the TTR scores of concordances extracted from said corpora? However, this study has also answered some important questions. Corpora representing different levels of generality consistently produce concordances with markedly different TTR scores, and this has important implications for using concordances for narrow reading, developing vocabulary knowledge, and learning how a word collocates, as well as many other uses of concordances for language learning. If teachers or learners are seeking a source of narrow reading, concordances can provide narrow reading conditions that are not provided by natural text, and concordancing more homogeneous corpora can maximise the narrow reading effect.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0958344020000117

Ethical statement. The data used was available free of charge from the Oxford Text Archive (https://ota.bodleian.ox.ac.uk/repository/xmlui/). The research did not involve any conflict of interest, and it was conducted in accordance with ethical practices in New Zealand.

References

Ballance, O. J. (2017) Pedagogical models of concordance use: Correlations between concordance user preferences. *Computer Assisted Language Learning*, 30(3–4): 259–283. https://doi.org/10.1080/09588221.2017.1307228

Ballance, O. J. & Coxhead, A. (2020) How much vocabulary is needed to use a concordance? *International Journal of Corpus Linguistics*, 25(1): 36–61. https://doi.org/10.1075/ijcl.17116.bal

Bernardini, S. (2000) Systematising serendipity: Proposals for concordancing large corpora with language learners. In Burnard, L. & McEnery, T. (eds.), *Rethinking language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt: Peter Lang, 225–235.

Bernardini, S. (2002) Exploring new directions for discovery learning. In Kettemann, B. & Marko, G. (eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora*. Amsterdam: Rodopi, 165–182. https://doi.org/10.1163/9789004334236_015

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

Bernardini, S. (2004) Corpora in the classroom: An overview and some reflections on future developments. In Sinclair, J. McH. (ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins, 15–36. https://doi.org/10.1075/scl.12.05ber
Boulton, A. (2009) Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1): 37–54. https://doi.org/10.1017/S0958344009000068
Boulton, A. (2010) Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3): 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x
Burnard, L. (2002) The BNC: Where did we go wrong? In Kettemann, B. & Marko, G. (eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora*. Amsterdam: Rodopi, 51–70. https://doi.org/10.1163/9789004334236_007
Chan, T.-P. & Liou, H.-C. (2005) Effects of web-based concordancing instruction on EFL students' learning of verb–noun collocations. *Computer Assisted Language Learning*, 18(3): 231–250. https://doi.org/10.1080/09588220500185769
Chambers, A. & O'Sullivan, Í. (2004) Corpus consultation and advanced learners' writing skills in French. *ReCALL*, 16(1): 158–172. https://doi.org/10.1017/S0958344004001211
Charles, M. (2011) Using hands-on concordancing to teach rhetorical functions: Evaluation and implications for EAP writing classes. In Frankenberg-Garcia, A., Flowerdew, L. & Aston, G. (eds.), *New trends in corpora and language learning*. London: Continuum, 26–43.
Charles, M. (2015) Same task, different corpus: The role of personal corpora in EAP classes. In Leńko-Szymańska, A. & Boulton, A. (eds.), *Multiple affordances of language corpora for data-driven learning*. Amsterdam: John Benjamins, 131–154. https://doi.org/10.1075/scl.69.07cha
Cheng, W., Greaves, C. & Warren, M. (2006) From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics*, 11(4): 411–433. https://doi.org/10.1075/ijcl.11.4.04che
Cobb, T. (1997) Is there any measurable learning from hands-on concordancing? *System*, 25(3): 301–315. https://doi.org/10.1016/S0346-251X(97)00024-9
Cobb, T. (1999) Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12(4): 345–360. https://doi.org/10.1076/call.12.4.345.5699
Cobb, T. (2010) Learning about language and learners from computer programs. *Reading in a Foreign Language*, 22(1): 181–200.
Cobb, T. & Boulton, A. (2015) Classroom applications of corpus analysis. In Biber, D. & Reppen, R. (eds.), *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press, 478–497. https://doi.org/10.1017/CBO9781139764377.027
Field, A. P. (2009) *Discovering statistics using SPSS* (3rd ed.). London: SAGE.
Gavioli, L. (2005) *Exploring corpora for ESP learning*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.21
Gordani, Y. (2013) The effect of the integration of corpora in reading comprehension classrooms on English as a foreign language learners' vocabulary development. *Computer Assisted Language Learning*, 26(5): 430–445. https://doi.org/10.1080/09588221.2012.685078
Grabe, W. & Stoller, F. L. (2011) *Teaching and researching reading* (2nd ed.). London: Routledge.
Hoey, M. (2005) *Lexical priming: A new theory of words and language*. Abingdon: Routledge.
Hoey, M. (2015) Words and their neighbours. In Taylor, J. R. (ed.), *The Oxford handbook of the word*. Oxford: Oxford University Press, 141–153. https://doi.org/10.1093/oxfordhb/9780199641604.013.39
Hunston, S. (2002) *Corpora in applied linguistics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524773
Jarvis, S. (2013) Defining and measuring lexical diversity. In Jarvis, S. & Daller, M. (eds.), *Vocabulary knowledge: Human ratings and automated measures*. Amsterdam: John Benjamins, 13–44. https://doi.org/10.1075/sibil.47.03ch1
Johns, T. (2002) Data-driven learning: The perpetual challenge. In Kettemann, B. & Marko, G. (eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora*. Amsterdam: Rodopi, 107–117. https://doi.org/10.1163/9789004334236_010
Kaur, J. & Hegelheimer, V. (2005) ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, 18(4): 287–310. https://doi.org/10.1080/09588220500280412
Kennedy, G. D. (1998) *An introduction to corpus linguistics*. London: Routledge.
Larson-Hall, J. (2010) *A guide to doing statistics in second language research using SPSS*. London: Routledge. https://doi.org/10.4324/9780203875964
Lee, D. (2002) Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle. In Kettemann, B. & Marko, G. (eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora*. Amsterdam: Rodopi, 247–292. https://doi.org/10.1163/9789004334236_021
Lee, D. & Swales, J. (2006) A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1): 56–75. https://doi.org/10.1016/j.esp.2005.02.010
Lee, H., Warschauer, M. & Lee, J. H. (2019) The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5): 721–753. https://doi.org/10.1093/applin/amy012

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press

Malvern, D., Richards, B., Chipere, N. & Durán, P. (2004) *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230511804

McCarthy, P. M. & Jarvis, S. (2010) MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2): 381–392. https://doi.org/10.3758/BRM.42.2.381

Milton, J. (2009) *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781847692092

Nation, I. S. P. (2013) *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Nation, I. S. P. & Waring, R. (1997) Vocabulary size, text coverage and word lists. In Schmitt, N. & McCarthy, M. (eds.), *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press, 6–19.

O'Sullivan, Í. & Chambers, A. (2006) Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1): 49–68. https://doi.org/10.1016/j.jslw.2006.01.002

Park, K. (2012) Learner–corpus interaction: A locus of microgenesis in corpus-assisted L2 writing. *Applied Linguistics*, 33(4): 361–385. https://doi.org/10.1093/applin/ams012

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

Scott, M. (2016) *WordSmith Tools (Version 7)*. Stroud: Lexical Analysis Software.

Sheskin, D. J. (2004) *Handbook of parametric and nonparametric statistical procedures* (3rd ed.). London: Chapman & Hall/CRC. https://doi.org/10.4324/9780203489536

Sinclair, J. (1991) *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Thurstun, J. & Candlin, C. N. (1997) *Exploring academic English: A workbook for student essay writing*. Sydney: National Centre for English Language Teaching and Research.

Thurstun, J. & Candlin, C. N. (1998) Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, 17(3): 267–280. https://doi.org/10.1016/S0889-4906(97)00013-6

Webb, S. A. & Chang, A. C.-S. (2012) Second language vocabulary growth. *RELC Journal*, 43(1): 113–126. https://doi.org/10.1177/0033688212439367

About the author

Oliver Ballance teaches at Victoria University of Wellington where he earnt his PhD. He works on the English Language Training for Officials Programme, an ESP-focused, capacity-building programme for government officials from developing countries in the Asia-Pacific region. Before this, he worked in a variety of EFL, ESOL and EAP settings. He occasionally lectures in various fields of applied linguistics.

Author ORCiD. Oliver James Ballance, https://orcid.org/0000-0003-4695-6406

https://doi.org/10.1017/S0958344020000117 Published online by Cambridge University Press