Effect of Frequency and Idiomaticity on Second Language Reading Comprehension

RON MARTINEZ
University of Nottingham
Nottingham, England

VICTORIA A. MURPHY
University of Oxford
Oxford, England

A number of studies claim that knowledge of 5,000–8,000 of the most frequent words should provide at least 95% coverage of most unsimplified texts in English, arguably enough to guess or ignore most unknown words while reading (Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer, 1991; Nation, 2006). However, perhaps hidden in that 95% figure are other kinds of words—multiword expressions—not accounted for by current estimates based on frequency lists. Such expressions are often composed of highly frequent words, and therefore it is possible that such items may go unnoticed by learners reading in the second language. To test this assertion, a two-part test was taken by 101 adult Brazilian learners of English: One part contained short texts composed of the top 2,000 words in English; the second part contained the exact same words, however the arrangement of these same words constituted multiword expressions (e.g., large, and, by → by and large). Tests of reading comprehension indicated that learners' comprehension not only decreased significantly when multiword expressions were present in text but students also tended to overestimate how much they understood as a function of expressions that either went unnoticed or were misunderstood.
doi: 10.5054/tq.2011.247708

Much research into second language (L2) learning points to reading as an important source for linguistic development in the target language (e.g., Elley & Mangubhai, 1983; Krashen, 1993; Nation, 1997). However, unlike native speakers who can usually understand close to 100% of nonspecialized texts (Carver, 1994), readers processing text in a foreign language are often faced with a comparatively laborious and cumbersome job that at times might seem more like an unpleasant guessing game. In much of the literature to date, researchers have suggested that, by passing certain vocabulary "thresholds" (Nation, 2001, p. 144), L2 readers' comprehension of texts in the target language will naturally increase. Although the preceding claim is undoubtedly true to a degree, there is some question as to what exactly constitutes *vocabulary*.

When most language students and teachers hear vocabulary, they think *words* (Hill, 2000). However, studies of large bodies of naturally occurring textual data, or *corpora*, have shown that some words commonly occur with other words, and these combinations actually form unitary and distinct meanings (Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Sinclair, 1991). Such multiword expressions are increasingly being viewed by researchers as a central part of the mental lexicon and even language acquisition itself (Ellis, 2008; Wray, 2002). Therein lies the current chasm between research into the relationship between vocabulary and reading comprehension and research into vocabulary: Clearly, vocabulary is more than individual words, but individual words are all that is mentioned in current research on vocabulary thresholds.

A possible reason for the aforementioned dichotomy is the fact that word coverage estimates (how many words are known in a text) are based on lists of the most common orthographic words alone. This monolexical tendency in word lists is merely a reflection of current technological limitations: There is simply no easy way to automatically extract frequency lists that are inclusive of meaningful multiword lexical items. Nonetheless, this convenient exclusion has also meant that little research to date has occurred into what effects such expressions might have on reading comprehension, as it is often simply assumed that idiomatic expressions are fairly rare in language, and still fewer are the ones which cannot be decoded through context or other semantic clues (Grant & Bauer, 2004). However, it is argued throughout the present article that not only are multiword expressions much more common than popularly assumed, but they are also difficult for readers to both accurately identify and decode—even when they only contain very common words.

As supporting evidence for the above assertions, we describe and present data from a study that put to test multiword expressions and the possible effects they have on reading comprehension.

RELATIONSHIP BETWEEN THE NUMBER OF WORDS KNOWN AND READING COMPREHENSION

According to informed estimates (e.g., Goulden, Nation, & Read, 1990), the average educated native speaker of English possesses a

[1] According to Nation (2001), a *word family* consists of "a headword, its inflected forms, and its closely derived forms" (p. 8).

receptive knowledge of around 20,000-word families[1]—a number which may seem daunting to a learner of the language. Perhaps for that reason, a number of researchers (e.g., Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer, 1989b) have endeavored to answer the following question: How many words are really necessary in order to comprehend most texts? The answer to that question is of interest to a wide range of parties, from developers of English as a foreign language textbooks, to writers of graded readers, to practicing classroom teachers and their students (Nation & Waring, 1997). After all, to be able to put a concrete number on the words one needs to know to function in the target language is to be able to set teaching goals, divide proficiency levels, and see a proverbial light at the end of the L2 learning tunnel.

However, the answer to that question has also proved somewhat complex, requiring identifying not so much how many words one needs to know in absolute terms, but rather how many words a learner needs to know in order to understand a text in spite of unknown vocabulary. Basing their assertions mostly on the assumption that pleasurable reading occurs only when a reader knows almost all the words in a text, Hirsh and Nation (1992) stipulated the ideal percentage of words known in an unsimplified text at around 98%, which the authors claimed could be reached with a knowledge of 5,000-word families. However, one limitation of the Hirsh and Nation study was that the texts used were novels written for teenagers and adolescents. To determine whether the same word family figure would apply to authentic texts designed for general (i.e., adult native speaker) consumption, Nation (2006) conducted a new analysis of fiction and nonfiction text (e.g., novels and newspapers). The trialing showed that, if 98% coverage of a text is needed for unassisted comprehension, then an 8,000- to 9,000-word family vocabulary is needed. Therefore, assuming the 98% figure is valid (as supported by Hu & Nation, 2000, and, most recently, Schmitt, Jiang, & Grabe, 2011), a learner requires a knowledge of at least 8,000-word families in order to adequately comprehend most unsimplified fiction and nonfiction text.

Word Counts and Their Limitations

The estimates of how many words a reader needs to know in order to read unsimplified text may actually be somewhat misleading, without critical examination of the underlying constructs. The main problem lies in the compilation of word frequency lists themselves, including what constitutes a word (cf. Gardner, 2007).

As mentioned earlier, current research suggests that 8,000–9,000 words can provide around 98% coverage of most texts (Nation, 2006). However, Nation's recommendations are for an "8,000–9,000 word-family vocabulary" (p. 79), which does not necessarily mean knowing 8,000 words:

> From the point of view of reading, a word family consists of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately. [. . .] The important principle behind the idea of a word family is that once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort. (Bauer & Nation, 1993, p. 253)

In the lists that Nation and other researchers have used to calculate word knowledge (e.g., Nation, 2006; West, 1953), a word can include a base form and over 80 derivational affixes (Nation, 2006, p. 66), resulting in "some large word families, especially among the high-frequency words" (Nation, 2006), but there may be an issue of overconflation of forms. Consider, for example, the semantic distance between the following pairs of words: *name → namely; price → priceless, fish → fishy; puzzle → puzzling*. Each of the preceding pairs would be grouped into the same respective word family, but it is unlikely that a learner of English would require "little or no extra effort" (Bauer & Nation, 1993, p. 253) to derive the meaning of a word like *fishy* from *fish*.[2] It is therefore conceivable that a number of those 8,000- to 9,000-word families do not have the psycholinguistic validity that is sometimes assumed, and some of the 30,000 (or so) separate words subsumed in those families would in fact need to be learned separately.

Similar to the semantic distance between *fish* and *fishy*, there is often an equal or greater disparity of meaning when a word is juxtaposed with another or more words and a new expression forms (Moon, 1997; Wray, 2002). For example, the words *fine*, *good*, and *perfect* each have meaning; however, those meanings do not remain in the expressions *finely tuned*, *for good*, and *perfect stranger*.

Nation (2006) recognized this limitation of current word lists; however, he did not consider it a problem. Nation based this assertion on the assumption that most learners will be able to guess the meaning of multiword expressions that have some element of transparency, and since the number of "truly opaque" phrases in English is relatively small, for the purposes of reading they are "not a major issue" (p. 66). However, it is debatable just how "small" in number those opaque expressions are, and, much like the previously discussed derived word forms that are actually semantically dissimilar, just how easy it is for a learner of English to accurately guess the meaning of more "transparent" expressions.

[2] According to the *Cambridge Advanced Learner's Dictionary* (Walter, Woodford, & Good, 2008), which is informed by the one-billion-word Cambridge International Corpus, the first sense of *fishy* is "dishonest or false" (p. 537), and not "smelling of fish."

Martinez and Schmitt (2010), for example, sought to compile a list of the most common expressions derived from the British National Corpus[3] (BNC). Their main criteria for selection was frequency and relative noncompositionality; in other words, the items chosen for selection needed to possess semantic and grammatical properties that could pose decoding problems for learners when reading. Their exhaustive search rendered over 500 multiword expressions that were frequent enough to be included in a list of the top 5,000 words in English—or over 10% of the entire frequency list. A sample of the list is provided in Table 1.

As an example of how taking word frequency into account alone potentially leads to very misleading estimates of text comprehensibility, consider the following text taken from *The Economist*:

> But over the past few months competing 3G smartphones with touch screens and a host of features have been coming thick and fast to the American market. And waiting in the wings are any number of open-source smartphones based on the nifty Linux operating system. Apple will need to pull out all the stops if the iPhone is not to be swept aside by the flood of do-it-all smartphones heading for America's shores. ("The iPhone's second coming," 2008)

The above paragraph contains a number of expressions which are partly or totally opaque, including the following seven:

- *over the past*
- *a host of*
- *thick and fast*
- *waiting in the wings*
- *any number of*
- *pull out all the stops*
- *swept aside*

TABLE 1
Highly Frequent Idioms With Words From Beginner EFL Textbooks for Comparison

Frequency (BNC)	Idioms	Frequency (BNC)	Words (for comparison)
18,041	*as well as*	466	*steak*
14,650	*at all*	455	*niece*
12,762	*in order to*	400	*receptionist*
10,556	*take place*	387	*lettuce*
7,138	*for instance*	385	*gym*
4,584	*and so on*	377	*carrot*
4,578	*be about to*	341	*snack*
3,684	*at once*	337	*earrings*
2,676	*in spite of*	302	*dessert*
1,995	*in effect*	291	*refrigerator*

Note. EFL = English as a foreign language; BNC = British National Corpus.

[3] A 100-million word corpus, predominantly composed of written English (The British National Corpus, 2007).

Current text word coverage calculations ignore such expressions. According to software commonly used to analyze the word family frequency distribution of text (VocabProfile, [Cobb, n.d.]), the same *Economist* paragraph is broken down as shown in Table 2.

So, if comprehension of a text were based on word coverage alone, current methods of text analysis (Table 2) suggest that a learner with a vocabulary of at least the top 2,000 words in English should be able to understand 95.52 % of the lexis in the *Economist* text (64 of the 67 words [tokens] counted), by some estimates (e.g., Laufer, 1989b) enough for adequate comprehension. If that same learner also knew just two words in the text on the Academic Word List (Coxhead, 2000)—*features* and *source*—s/he would understand an additional 2.99%, affording that learner a knowledge of 98.51%—theoretically approximating nativelike levels of comprehension (Carver, 1994, p. 432). However, a closer look at the breakdown in Table 2 shows that words like *pull*, *out*, *all*, *the* and *stops* were all considered as separate and very common words, when in reality they form one noncompositional expression: *pull out all the stops*. In fact, all of the seven expressions have mistakenly been fragmented and categorized as pertaining to the top 2,000 words. Therefore, assuming that a learner who knows only the 2,000 most common words in English would not understand those expressions without the help of a dictionary, and if we reconduct the analysis taking those seven expressions into account (constituting 23 words), the total number of words fitting into the top 2,000 goes down to 41 (64 minus 23), and that 95.52% figure actually drops from adequate comprehension down to 61.19% (41 ÷ 67 = 0.6119)—well below acceptable levels of reading comprehension (Hu & Nation, 2000).

TABLE 2
Word Frequency Breakdown of *The Economist* text (Not Including Proper Nouns)

Frequency	Words (67 tokens, 54 types)	Text coverage
0–1,000	*a all and any are based be been but by coming do fast few for have heading if in is it market months need not number of on open operating out over past pull shores stops system the to touch waiting will with*	83.58%
1,001–2,000	*aside competing flood host screens swept thick wings*	11.94%
Academic Word List (AWL)	*features source*	2.99%
Off list	*Nifty*	1.49%
Words in top 2,000		95.52%
+ AWL words		2.99%
Total text coverage		98.51%

RELATIONSHIP BETWEEN FORMULAIC LANGUAGE AND READING COMPREHENSION

Considering the relative wealth of research and literature on L2 reading comprehension and, separately, multiword expressions in English, there is a surprising dearth of information regarding the role, if any, formulaic language plays in the comprehension of texts in a foreign language. The relatively few studies that do exist (e.g., Cooper, 1999; Liontas, 2002) seem to confirm that it is especially the more semantically opaque idioms that pose interpretability problems for L2 readers, and, as these more core idioms are relatively rare (Grant & Nation, 2006), Nation (2006) could be right in attenuating their significance in reading comprehension.

Nevertheless, as discussed earlier, there is evidence that a significant number of relatively opaque expressions occur frequently in texts in English. One commonly cited estimate (Erman & Warren, 2000) is that somewhat more than one-half (55%) of any text will consist of formulaic language (p. 50). Naturally, the opacity of those expressions will lie on what Lewis (1993, p. 98) called a “spectrum of idiomaticity” (Figure 1), a kind of continuum of compositionality.

Furthermore, even when an expression does not meet the criteria of core or nonmatching idiom, the relative ease or difficulty with which a learner will unpack its meaning is less inherent to the item itself and more a learner-dependent variable. Just as knowing *fish* may or may not translate into understanding *fishy*, knowing *perfect* does not necessarily mean understanding *perfect stranger*.

What is more, although previous studies (Cooper, 1999; Liontas, 2002) have found that in textual context more transparent idioms were more easily understood than their opaque counterparts, it should also be noted that the participants in those studies were aware that they were being tested specifically on their ability to correctly interpret idioms. In other words, what cannot be known from most existing studies on idiom interpretation is how well the participants would have been able to identify and understand the idioms in the first place had they not been aware of their presence in the text.

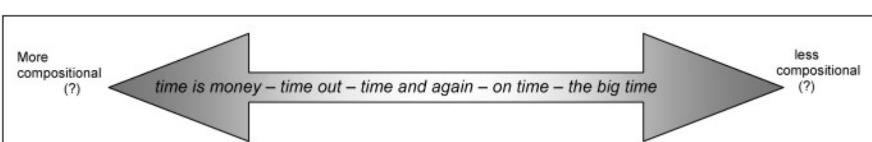


FIGURE 1. A spectrum of idiomaticity (compositionality).

A notable exception is Bishop (2004), who investigated the differential look-up behavior of participants who read texts that contained unknown words and unknown multiword expressions synonymous with those words. Bishop confirmed that, even though both words and multiword expressions were unknown, readers looked up the meaning of words significantly more often. He concludes that learners "do not notice unknown formulaic sequences as readily as unknown words" (p. 18).

Therefore, idiomatic language in text, irrespective of compositionality, might most usefully be classified as what Laufer (1989a) called "deceptively transparent" (p. 11). Laufer found that many English learners misanalyze words like *infallible* as *in+fall+ible* (i.e., "cannot fall") and *nevertheless* as never+less (i.e., "always more"; p. 12). Likewise—although they were not part of her study—she found that idioms like *hit and miss* were being read and interpreted word for word. These lexical items that "learners think they know but they do not" (Laufer, 1989a, p. 11) can impede reading comprehension in ways not accounted for in lists of common word families. Nation (2006) seemed to assume that multiword expressions that have some element of transparency, however small, will be reasonably interpretable through guessing. However, the Laufer (1989a) study may provide evidence to the contrary:

> But an attempt to guess (regardless of whether it is successful or not) presupposes awareness, on the part of the learner, that he is facing an unknown word. If such an awareness is not there, no attempt is made to infer the missing meaning. This is precisely the case with deceptively transparent words. The learner thinks he knows and then assigns the wrong meaning to them [. . .]. (p. 16)

Substitute "idiom" for "word" above—not an unreasonable conceptual stretch—and it becomes clear that multiword expressions just may present a larger problem for reading comprehension than accounted for in the current literature.

In fact, such "deception" seems even more likely to occur with multiword expressions, because such a large number of them are composed of very common words a learner would assume he or she knows (Spöttl & McCarthy, 2003, p. 145; Stubbs & Barth, 2003, p. 71). Moreover, there is evidence that learners are reluctant to revise hypotheses formed regarding lexical items when reading, even when the context does not support those hypotheses (Haynes, 1993; Pigada & Schmitt, 2006).

Summary and Research Questions

In summary, the following has been argued thus far:

- Current estimates of how many words one needs to know in order to comprehend most texts may be inaccurate due to overinclusion of derived word forms and a total exclusion of multiword units of vocabulary.
- Contrary to some research, there is evidence in corpus data that the number of frequently occurring noncompositional multiword expressions in English is higher than previously believed.
- Even when an expression is partly or even fully compositional, there is no way of knowing how accurately an L2 reader will interpret (or even identify) that item.
- The claim that special attention to the 2,000–3,000 most common words in English is pedagogically sound since they provide around 80% of text coverage (e.g., O'Keeffe, McCarthy, & Carter, 2007; Read, 2004, p. 148; Stæhr, 2008) deserves closer scrutiny, because those words are often merely tips of phraseological icebergs.

It is therefore clear that there is a need for further investigation into how common words and the multiword units they form can affect reading comprehension when reading in English as a foreign language. To that end, we conducted a study to answer the following research questions:

1. Are two texts, written with the exact same high-frequency words, understood equally well by L2 learners, when one of the texts is more idiomatic than the other?
2. Can the presence of multiword expressions in a text lead L2 learners to believe they have understood that text better than they actually have?

METHOD

Participants

Brazilian adult learners of English (n = 101), all native speakers of Brazilian Portuguese, were selected to participate in the study. The sample ranged in age from 18 to 64 years (M = 25.76, SD = 9.31) and consisted of 43 men and 58 women, representing seven different regions of Brazil.

All participants in the study had had a minimum of 80 hr of tuition in English prior to the start of the research and had been tested as possessing intermediate or higher levels of proficiency. However, because all participants attended private language schools, the actual instruments by which their proficiency was assessed varied widely.

Regardless, uniformity in proficiency was not of prime importance in the study, because the research questions are concerned with a significant change in the paired samples within the same group.

The Test

To write the eight texts (four texts in each test), a corpus of words was carefully chosen from the list of the 2,000 most frequent words in the BNC. The reading comprehension of each text was tested by seven true or false items, totaling 28 per test part, or 56 overall. The test, when administered, appears as one but is actually in two parts (Test 1 and Test 2), each part containing the exact same words, with some words in Test 2 forming multiword expressions. Great care was taken to ensure that the texts are otherwise equal. There is no visual difference between the two parts, and the texts are of almost uniform length in both parts (Table 3, and Figure 2). In addition, care was taken not to include any extra cultural references in any of the texts, and the comprehension task itself did not change across the test parts. The texts are stated to have come from people's description of themselves in "Friendsbook" (intentionally similar to Facebook). On the whole, therefore, it could be said that the style of the text is personal and informal.

The vast majority of the words are in the top 1,000, with approximately 98.5% of the words occurring in the BNC top 2,000 (Table 3). These results were in turn compared with the General Service List (West, 1953) using software developed by Heatley and Nation (*Range*, 1994) and another package using the same corpus created by Cobb (n.d.). The results were practically identical in all cases (*Range*: 98.5%; *VocabProfile*: 98.49%[4]).

Another key feature of the test is the rating scale which requires the test taker to circle what s/he believes is his or her comprehension of each text, from 5% to 100%.[5] This self-reported comprehension is designed to help answer the second research question of whether the presence of multiword expressions in a text can lead L2 learners to believe they have

TABLE 3
Summary of Test 1 and Test 2 Text Data

	Total word count	Total clause count	T-unit count	Top 1,000 words coverage	Top 1, 001–2,000 words coverage
Test 1	416	56	45	95.7373%	2.7650%
Test 2	412	54	42	95.7373%	2.7650%

Note: Data measured using vocabulary profiler (Cobb, n.d.). A T-unit, or a minimal terminable unit, is "one main clause plus any subordinate clause" (Hunt, 1968, p. 4).

Text From Test 1	Text From Test 2
The good part about living here is nature. I need to be free in the world. As I like to make and grow things, I don't live by any large cities. The grounds here on my block are fairly good (and they are pretty). I've been across the country and around the whole world a few times, but I think that here is better than any place I know on earth. And I don't ever need to go to the city to buy things – I don't like it there, actually. That's the good thing about living here, I feel: it's neither the country nor the city. I like to show it off. Call me to come down and see it if you like. I like company! (And I'm nice – really!)	About me? I'm living large in the city, better than ever. Actually, I've been around the block a few times and I think I make a pretty good living on the whole, but that's neither here nor there. If you don't know any better, I come across as a show off. The thing is, my nature is to be fairly down to earth, by and large, on the grounds that I call it like I see it. I know my place in the world. If things are down and I need to part company with the world, I go to the country. I don't really need to buy nice things to feel better. But I live off cities – they grow on you. Feel free to call me.

FIGURE 2. Side-by-side comparison of matched texts from the two test parts (full version of the test available on request).

understood that text better than they actually have. Testing what comprehenders believe they understood—compared to what they actually understood—by comparing test scores to self-assessment of comprehension via a rating scale is often referred to as "comprehension calibration," and is a method that has been used extensively in psychological research (e.g., Bransford & Johnson, 1972; Glenberg, Wilkinson, & Epstein, 1982; Maki & McGuire, 2002; Moore, Lin-Agler, & Zabrucky, 2005), but somewhat less so in applied linguistics (cf. Brantmeier, 2006; Jung, 2003; Morrison, 2004; Oh, 2001; Sarac & Tarhan, 2009).

Finally, participants were asked to record their start and finish times for each part of the test.

Procedure

Following an initial field test, an item analysis was conducted to establish the facility value[6] of the test items, and a discrimination index for both test parts, to discern whether the test was discriminating between stronger and weaker participants. Items that were found to have exceptionally high or low scores were carefully analyzed and the wording of both test items and reading texts adjusted accordingly.

Finally, as advocated in Schmitt, Schmitt, and Clapham (2001), one important requirement of an L2 test of reading comprehension is that it be answerable by native or nativelike speakers of the language (p. 65). To that end, a smaller group of native speakers (n = 8) was also tested as an additional check of the test's validity. That group produced a mean

[4] These are vocabulary profilers that essentially use word frequency lists to break all the words in a text down into those that occur, for example, in the top 1,000, second 1,000, third 1,000, and so on.

[5] "0%" would be virtually impossible, because learners would be familiar with most words in the texts, if not all.

[6] A test item's facility value is calculated by dividing the number of participants who answered that item correctly by the total number of participants (Hughes, 2003).

score of 28 in Test 1 (the maximum score) and 27.75 in Test 2, showing that both test parts posed very little difficulty for people for whom English is a first language.

The students received the two parts of the test in alternating order (i.e., counterbalanced) to control for any effect of taking Test 1 before Test 2 (and vice versa), because by counterbalancing one is able to include a variable, order, into an analysis and identify the extent to which order affects performance on the dependent measures.

RESULTS

All the scores and self-reported comprehension measures were recorded in a statistical analysis software program (SPSS). Each participant's test results were transcribed into the software item by item (i.e., correct and incorrect), text by text (e.g., Text A, Text B, etc., number correct out of 7 for each), and test by test (i.e., Test 1 and Test 2, number correct out of 28 possible for each). Also recorded were each participant's self-reported comprehension assessments (a rating scale from 5 to 100%) as they pertained to each text, as well as which version (Test 1 or Test 2 first) of the instrument each candidate had received. To assess the effect of counterbalancing, a repeated measures analysis of variance (ANOVA) was conducted with one within-subjects factor (TEST) with two levels (Test 1 score, Test 2 score), and one between-subjects factor: (VERSION) with two levels (Test 1 first, Test 2 first). This analysis revealed a robust main effect of test ($F\,(1.99) = 593.38$, $p < 0.001$, $\eta^2 = 0.86$) and a discrete effect of version ($F\,(1.99) = 4.05$, $p = 0.047$, $\eta^2 = 0.04$), but importantly there was no significant Test × Version interaction ($F\,(1.99) = 2.81$, $p = 0.097$, $\eta^2 = 0.02$), illustrating that participants' scores did not vary as a function of which version of the test they were completing.

Comprehension of Test 1 versus Test 2

The central tendencies from both tests are presented in Table 4.

As predicted, participants' scores were significantly lower on Test 2 relative to Test 1 ($t(100) = 24.10$, $p < 0.001$) with a strong effect size ($\eta^2 = 0.828$),[7] confirming that even when two sets of texts contain the exact same words, and even if those words are very common, comprehension of those texts will not be the same when one contains idiomaticity.

[7] Confidence intervals at 95% for all *t*-tests.

TABLE 4
Descriptive Statistics (Tests 1 and 2)

	Mean	Median	Mode	Minimum	Maximum	*SD*
Test 1[a]	24.09	25.00	25.00	18.00	28.00	2.44
Test 2[b]	14.76	15.00	15.00	6.00	25.00	3.93

Note. *SD* = standard deviation. [a]Texts composed of common words, however with little or no opaque phraseology. [b]Texts composed of common words which comprise collocations and idiomatic language.

Self-Reported Comprehension of Test 1 versus Test 2

A *t*-test was also performed on the difference between what participants believed they understood (i.e., their reported comprehension) and what they actually understood in both Tests 1 and 2 (Table 5). Mean reported comprehension (MRC) was arrived at by calculating the means of all the self-reported comprehension percentages (5–100%), and the mean actual comprehension (MAC) is the mean of the score on each text divided by seven (the maximum score).

MAC was lower than MRC in both Test 1(MRC: $M = 0.8738$, SD 0.13; MAC: $M = 0.8603$, $SD = 0.09$) and Test 2 (MRC: $M = 0.6029$, SD 0.19; MAC: $M = 0.5258$, $SD = 0.14$); however, the difference between MRC and MAC was statistically reliable in Test 2 only where the test was composed of multiword expressions ($t(100) = 3.95$, $p < 0.001$, $\eta^2 = 0.07$).[8] These data, therefore, appear to support a positive answer to the second research question—that learners may believe they understand more than they actually do by virtue of their simply understanding the individual words in a text.

Variation in the Data

MRC versus MAC

Naturally, there were some individual differences in the test results. Table 6 shows that roughly one-half of all participants ($n = 50$) overestimated their comprehension in both tests (indicated by MRC

TABLE 5
Reported Versus Actual Comprehension

	Test 1	Test 2
Mean reported comprehension (MRC)	87.38%	60.29%
Mean actual comprehension (MAC)	86.03%	52.58%

[8] By comparison, the results of the paired sample *t*-test for the Test 1 MRC-MAC means were $t(100) = 1.04$, $p = 0.302$, $\eta^2 = 0.005$.

greater than MAC). On the surface, this might suggest that those 50 participants did not overestimate their comprehension in Test 2 as a result of its relative noncompositionality but rather as a general tendency, regardless of idiomaticity. However, when MAC is subtracted from MRC for that group and the difference compared across the tests, the difference is twice as large in Test 2 ($t(49) = 4.92$, $p < 0.001$, $\eta^2 = 0.108$). This difference is even more evident in the same analysis of Test 2 alone. Isolating the results of that test, there was a total of 70 participants (over 69% of the entire sample) whose MRC was greater than their MAC. The difference between MRC minus MAC for the subgroup of 70 participants reveals that their overestimation of their comprehension in Test 2 was four times greater than in Test 1 ($t(69) = 7.70$, $p < 0.001$, $\eta^2 = 0.228$). Further, among that group of 70 whose Test 2 MRC was greater than their MAC, 20 students either underestimated or actually accurately guessed their comprehension in Test 1, but significantly overestimated their comprehension in Test 2 ($t(19) = 8.49$ $p < 0.001$, $\eta^2 = 0.264$). Only 30 participants in the present study actually underestimated their comprehension on Test 2.

Variation in Time Spent on the Tests

All candidates were also asked to record their start and finishing times of each test (Table 7). It is not surprising that, on average, participants spent more time on the texts which contained idiomatic expressions. However, this trend was not uniform among the participants, and the time data may help provide some insight into just how aware some participants were of the presence of idiomaticity in the texts. This possibility is explored further in the Discussion.

TABLE 6
Trend of Increase in MRC-MAC Discrepancy Across Test Parts

	N	MRC minus MAC (Test 1)	MRC minus MAC (Test 2)
Students with MRC > MAC (both test parts)	50	0.09	0.19
Students with MRC > MAC (Test 2)	70	0.04	0.17
Students with MRC > MAC (Test 2 only)	20	−0.10	0.14
Students with MRC ≤ MAC (Test 2)	31[a]	−0.05	−0.15

Note. MRC = mean reported comprehension; MAC = mean actual comprehension. [a]Only one participant in this group had MRC = MAC.

TABLE 7
Time It Took Participants to Take Tests

	N	Mean time (minutes)	*SD*
Time it took to take Test 1	91[a]	10.85	4.36
Time it took to take Test 2	91	14.65	3.77

Note. *SD* = standard deviation. [a]Some participants (n = 10) forgot to write in the time, or left it incomplete.

DISCUSSION

To reiterate, the present study sought to address the following research questions:

1. Are two texts, written with the exact same high-frequency words, understood equally well by L2 learners, when one of the texts is more idiomatic than the other?
2. Can the presence of multiword expressions in a text lead L2 learners to believe they have understood that text better than they actually have?

A group of 101 adult Brazilian learners of English was tested using two separate measures of reading comprehension (one with idiomatic language and one without), and the results of the two tests were then compared. Although verbal reports would provide even greater insight, the resultant data from this study do seem to support a positive answer to the first research question: Participants achieved significantly lower scores on the comprehension measure which assessed their understanding of texts that contain multiword expressions. It should be reiterated that this result was obtained in spite of both tests being written with the exact same high-frequency words.

Likewise, through participants' self-reported comprehension, there seems to be evidence that the presence of multiword expressions in a text can lead L2 readers to overestimate their assessment of how much they have actually understood, suggesting a positive answer to the second research question.

Although it seems apparent that participants' reading comprehension in the present study was adversely affected by idiomaticity, this section explores and possibly explains variation in comprehension of the tests and further discusses evidence that multiword expressions negatively affected participants' comprehension.

On the Stronger and Weaker Performance in the Tests

There are at least three important revelations emerging from the MRC-MAC data analysis presented in Table 6. First, nearly two thirds of all participants believed they understood more than they actually did in Test 2. Second, the data show that over 28% of those students ($n = 20$) did not overestimate their comprehension in Test 1, providing evidence that, whereas they were apparently able to gauge their comprehension fairly accurately when the texts were relatively devoid of idiomaticity, they were significantly less able to do so when relative noncompositionality was present. Finally, the MRC-MAC data presented so far indicate that, although not all participants tended to overestimate their comprehension in Test 2, those who did (again, the majority of cases) generally did so by a much larger margin. Put simply, it would seem that, not only are many learners misunderstanding idiomaticity in text but they may be doing so more than they (or researchers) realize.

Further Evidence of Idiomaticity Negatively Affecting Comprehension

The mean difference between MRC and MAC for the 70 students who overestimated their comprehension in Test 2 (Table 6) was 0.17. For the sake of argument, one could establish an MRC-MAC difference above 0.20 as a cutoff for apparently marked lack of awareness of the extent of idiomaticity in Test 2,[9] comparatively speaking. This in turn would mean that there were 30 participants who ostensibly demonstrated a particularly high overestimation of how much of Test 2 they had actually understood. A sample of just the top 10 is provided in Table 8.

Table 8 shows not only how different those participants' MRC was from their Test 2, it also lists the same figures for Test 1 for comparison.

TABLE 8

Partial List of Participants With Exceptionally High Overestimation of Test 2 Comprehension (Difference > 0.20)

Participant	MRC Test 1	MAC Test 1	Difference	MRC Test 2	MAC Test 2	Difference
118	1.00	0.89	0.11	0.94	0.32	0.62
6	0.75	0.68	0.07	0.81	0.28	0.53
92	0.81	0.89	−0.08	0.75	0.25	0.50
90	1.00	0.89	0.11	0.81	0.39	0.42
115	1.00	0.89	0.11	0.87	0.50	0.37
22	1.00	0.96	0.03	0.94	0.57	0.37
80	0.94	0.68	0.26	0.69	0.32	0.37
59	0.94	0.89	0.05	0.69	0.36	0.33
21	1.00	0.93	0.07	0.87	0.57	0.30
50	0.87	0.78	0.09	0.69	0.39	0.29

[9] A t-test reveals that the mean is in fact statistically significant even at the 0.03 difference level ($t = 6.53$, $p = 003$).

Whereas for Test 1 those 30 participants' assessment of reading comprehension was usually only slightly overestimated[10] (and in one instance even underestimated), the same participants in Test 2 had a markedly overinflated conception of how much they had understood.

Although not intended as a principal source of data for our analysis, the start and finish times recorded by participants on the tests in a few instances helped confirm the above assertions.[11] Table 7 shows that there was a significant difference between how long participants spent taking each test ($t = 8.52$, $p < 0.001$). However, there were a number of participants ($n = 26$) who reported actually spending either the same or less time on Test 2. On the surface at least, this fact would indicate that those 26 participants did not notice much difference between the tests. Of course, in reality that datum alone does not provide such evidence, because that time may simply reflect a number of other confounding variables, including affective (e.g., simply giving up) or even social (e.g., the need to leave early to meet a friend), and for that reason also was not considered to be a reliable source of data from those participants. However, as a complementary source, when triangulated with the other variables, some of the data analyzed earlier in the current study are occasionally fortified further still with the time data (Table 9).

The data in Table 9 in many ways tie together the elements discussed in the analysis thus far in the present study: A much lower score in Test 2 than in Test 1 (columns B and C), an increase in the mean MRC-MAC difference across the tests (columns H and K), a relatively moderate decrease in MRC in Test 2 vis-à-vis Test 1 (columns F and I), a marked difference between MRC and MAC in Test 2 (column K), and the occasional lack of increase (in Table 9, participants 118, 80, 59, and 50) in time it took to take Test 2 (columns D and E). These findings, on their own, in combination, and especially in concert, provide evidence that (1) the students in Table 9 did not seem to appreciate the full extent of the idiomaticity in Test 2, and (2) they appeared to think that they understood more than they actually did in Test 2.

Perhaps most important, 98% of all participants ($n = 99$) scored significantly lower on Test 2 (mean difference of 9.53 points out of 28, $SD = 3.63$), which shows that learners are generally unable to guess the meaning of idiomatic language, even in the (largely inconsistent) instances in which they become aware of its presence.

[10] As can be seen in Table 8, in fact only one participant—number 80—actually had an MRC-MAC difference in Test 1 that was also >0.20. Those of the other 29 participants were all lower.

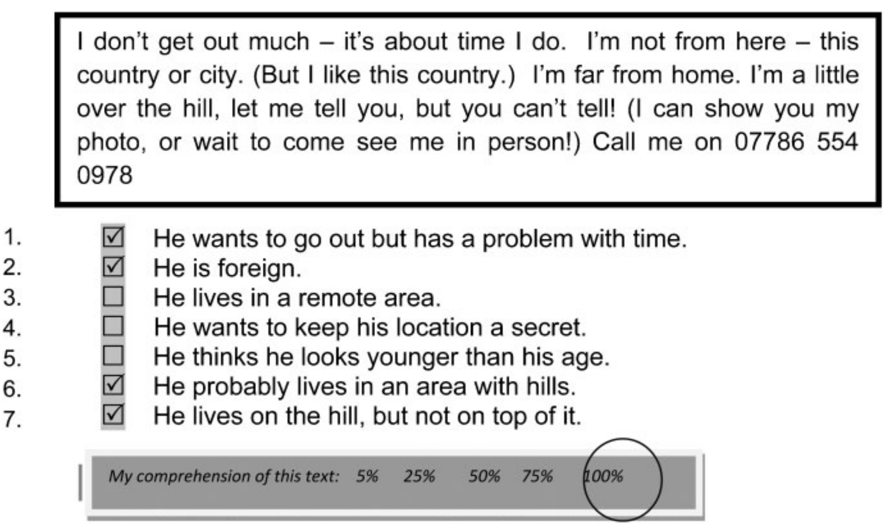
[11] It should be noted that because 10 of the participants' time data were left incomplete, a full group analysis could not be carried out; thus the time data are only included as auxiliary information.

TABLE 9
Triangulated Data Showing Negative Effect of Idiomaticity on Comprehension

A	B	C	D	E	F	G	H	I	J	K
	Score		Time		Test 1			Test 2		
Participant	Test 1	Test 2	Test 1	Test 2	MRC	MAC	Difference	MRC	MAC	Difference
118	25	9	20.00	18.00	1.00	0.89	11	0.94	0.32	0.62
6	19	8	Missing	Missing	0.75	0.68	0.07	0.81	0.28	0.53
92	25	8	9.00	11.00	0.81	0.89	−0.08	0.75	0.25	0.50
90	25	11	9.00	15.00	1.00	0.89	0.11	0.81	0.39	0.42
115	25	14	13.00	15.00	1.00	0.89	0.11	0.87	0.50	0.37
22	27	16	Missing	Missing	1.00	0.96	0.03	0.94	0.57	0.37
80	19	9	10.00	10.00	0.94	0.68	0.26	0.69	0.32	0.37
59	25	10	15.00	14.00	0.94	0.89	0.05	0.69	0.36	0.33
21	26	16	Missing	Missing	1.00	0.93	0.07	0.87	0.57	0.30
50	22	11	20.00	12.00	0.88	0.79	0.09	0.69	0.39	0.30

For example, the first participant in Tables 8 and 9 (number 118), who consistently overrated his own comprehension relative to his actual comprehension, answered for Text A of Test 2 as shown in Figure 3.

On the basis of the responses, it is likely that participant 118 associated *it's about time* with "has a problem with time," *I'm a little over the hill* with "he probably lives in an area with hills" and "he lives on the hill,


I don't get out much – it's about time I do. I'm not from here – this country or city. (But I like this country.) I'm far from home. I'm a little over the hill, let me tell you, but you can't tell! (I can show you my photo, or wait to come see me in person!) Call me on 07786 554 0978

1. ☑ He wants to go out but has a problem with time.
2. ☑ He is foreign.
3. ☐ He lives in a remote area.
4. ☐ He wants to keep his location a secret.
5. ☐ He thinks he looks younger than his age.
6. ☑ He probably lives in an area with hills.
7. ☑ He lives on the hill, but not on top of it.

My comprehension of this text: 5% 25% 50% 75% 100%

FIGURE 3. Sample answer to Text A of Test 2.

[12] Of the 24 multiword expressions targeted for assessment of comprehension in Test 2, only four were core idioms.

but not on top of it"—a literal reading of those expressions. His 100% self-assessment is an indication that his answers were not guesses, but instead based on confidence that he had understood all the individual words in the text. The same answering dynamic occurred time and time again in Test 2.

Returning, therefore, to the issues raised in the introduction, the present research appears to provide evidence that idioms—opaque or otherwise[12]—can in fact cause major problems for L2 learners of English when reading. In general, the participants in our study were less successful at guessing the meaning of idioms than those in previous idiom-centered research, such as Cooper (1999) and Liontas (2002). Because the learners in our study were not alerted to the presence of idiomaticity in the text prior to reading, this study may more accurately reflect what takes place in more naturally occurring L2 reading. Finally, the difficulties that the participants encountered occurred in texts composed purely of very common words, showing that word frequency alone is not necessarily a good measure of text coverage. Laufer's (1989a) hypothesis that much lexis in text can be deceptively transparent is confirmed in the present research, extended to multiword expressions.

Limitations of the Research

Although the research presented here does provide some significant evidence that the presence of idiomaticity in a text can negatively affect the comprehension of that text by an L2 reader, little is known about the strategies the test takers employed. Even though metacognition was outside the scope of examination of the present study, such information is highly relevant, and future research should include protocols, such as those included in previous research on idioms (e.g., Liontas, 2003, 2007), which may provide crucial qualitative data regarding the individual differences and behavior of the participants themselves. Overall, future work in this area should consider making use of different measures of self-assessment in order to explore the limitations of the methodology used in the present study.

Also, a question can be asked concerning the relative density of multiword expressions in each text. In other words, what is not known is, on average, how many such items a reader can expect to encounter. The ratio of idioms to orthographic words in the texts tested in the present study is very likely higher than average—but how much higher is as yet unknown. Likewise, the passages themselves were relatively short, so it is unknown whether or not texts of greater length (and therefore contextualization) would render the same or similar results.

CONCLUSION

The present study provides some empirical evidence that a focus on word in L2 pedagogy may be detrimental to learners' development of reading comprehension in the target language. In most cases, participants overestimated how much they had understood the texts tested, when those texts contained multiword expressions of varying degrees of compositionality, comprised of very common individual words. Even when participants showed evidence of awareness of a presence of idiomaticity in the texts, they generally were not very good at guessing the meaning of expressions. Although Grant and Nation (2006) could be right that the teaching of individual idioms may not be "deserving of class time" (p. 5), it is clear that, at the very least, raising learners' consciousness as to their prevalence and frequently deceptive transparency in text certainly is.

Concomitantly, the present research serves to demonstrate the incompleteness of current lists of common words which now inform important pedagogical instruments such as vocabulary tests, graded readers, and course syllabuses. Under current methods, as illustrated in the Table 2 analysis of the *Economist* paragraph, a text measured against the word lists now commonly in use can be mistakenly analyzed as being easily understood, when in fact, once multiword expressions are taken into account, the actual text coverage of easily understood words is far less. Existing estimates of how many words one needs to know in order to adequately process naturally occurring texts (Hu & Nation, 2000; Nation, 2006) may need to be revisited to determine how many words and multiword expressions one needs to know.

In practice, research shows that learners will continue to believe they understand items that they actually do not, even after repeated exposures to that same item (Bensoussan & Laufer, 1984; Haynes, 1993; Pigada & Schmitt, 2006). Language teachers and related professionals, such as textbook writers, need to consider ways to continuously help learners notice the gap between what they think they know and understand, and what they actually do (Laufer, 1997). One way this can be achieved has already been thoroughly discussed in the present paper: devise texts (and tests) designed to lead readers down an idiomatic garden path. Norbert Schmitt (as cited in Weir, 2005) made a similar recommendation:

> Perhaps the best and most valid type of vocabulary test is a reading passage with comprehension questions, but with the items requiring a full understanding of particular words or phrases in the text. This would mimic the real world task of reading for comprehension and also the loss of comprehension when key vocabulary is not known. (p. 123)

Therefore, a further implication this research has for L2 pedagogy is the more careful design of the reading component of proficiency tests, such as IELTS (International English Language Testing System) and Test of English as a Foreign Language, to be more strategic regarding the amount and kind of phraseology contained in the texts. Proficiency was not the central focus of our study, and future research could adopt a similar methodology, but with the inclusion of proficiency measures of participants in order to explore how various types of expressions (e.g., noncompositional, figurative, imageable, etc.) may differentially affect preidentified proficiency levels.

Thus, in order to be more methodical about such things as the inclusion of formulaic language in L2 tests and reading input in general, an important preliminary step will be to establish (a) what type of formulaic sequences are good discriminators of proficiency threshold levels; and (b) which expressions of that identified type are most common. In other words, what will be useful—and is indeed now beginning to emerge (Martinez & Schmitt, 2010; Shin & Nation, 2008; Simpson-Vlach & Ellis, 2010)—are lists of formulaic sequences that serve a function similar to that of the General Service List, not to revolutionize current language pedagogy, but to more finely tune it.

THE AUTHORS

Ron Martinez is currently pursuing a doctorate at the University of Nottingham, Nottingham, England, on the topic of inclusion of phraseology in language assessment. He specializes in second language materials development.

Victoria Murphy is a lecturer in Applied Linguistics and Second Language Acquisition at the University of Oxford, Oxford, England. Her research interests include the lexical development of first language or second language learners and the cognitive processes that underlie language acquisition.

REFERENCES

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*, 253–279, doi:10.1093/ijl/6.4.253.

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, *7*, 15–32. doi:10.1111/j.1467-9817.1984.tb00252.x.

Bishop, H. (2004). Noticing formulaic sequences: A problem of measuring the subjective. *LSO Working Papers in Linguistics*, *4*, 15–19.

Bransford, D. B., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726. doi:10.1016/S0022-5371(72)80006-9.

Brantmeier, C. (2006). Advanced L2 learners and reading placement: Self-assessment, CBT, and subsequent performance. *System*, *34*, 15–35. doi:10.1016/j.system.2005.08.004.

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior, 26*, 413–437.
Cobb, T. (n.d.). *Web VocabProfile* [Online program]. Retrieved from http://www.lextutor.ca/vp/, [an adaption of A. Heatley & P. Nation, *Range* (1994)].
Cooper, T. C. (1999). Processing of idioms by L2 learners of English. *TESOL Quarterly, 33*, 233–262. doi:10.2307/3587719.
Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213–238. doi:10.2307/3587951.
Elley, W., & Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly, 19*, 53–67. doi:10.2307/747337.
Ellis, N. C. (2008). Phraseology: The periphery and the heart of language. In F. Meunier & S. Granger (Eds.), *Phraseology in language learning and teaching* (pp. 1–13). Amsterdam, The Netherlands: John Benjamins.
Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*, 29–62.
Gardner, D. (2007). Validating the construct of *Word* in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics, 28*, 241–265. doi:10.1093/applin/amm010.
Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory and Cognition, 10*, 597–602. doi:10.3758/BF03202442.
Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11*, 341–363. doi:10.1093/applin/11.4.341.
Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics, 25*, 38–61. doi:10.1093/applin/25.1.38.
Grant, L., & Nation, P. (2006). How many idioms are there in English? *ITL - International Journal of Applied Linguistics, 151*, 1–14. doi:10.2143/ITL.151.0.2015219.
Haynes, M. (1993). Patterns and perils of guessing in second language reading. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46–64). Norwood, NJ: Ablex Publishing Corporation.
Heatley, A., & Nation, P. (1994). *Range* [Computer program]. Wellington, New Zealand: Victoria University of Wellington. Retrieved from http://www.vuw.ac.nz/lals/
Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further developments in the Lexical Approach* (pp. 47–69). Hove, England: LTP.
Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*, 689–696.
Hu, H. M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*, 403–429.
Hughes, A. (2003). *Testing for language teachers.* Cambridge, England: Cambridge University Press.
Hunt, K. (1968). *An instrument to measure syntactic maturity.* Tallahassee, FL: U.S. Department of Health, Education and Welfare.
Jung, E. H. (2003). The effects of organization markers on ESL learners' text understanding. *TESOL Quarterly, 37*, 749–759. doi:10.2307/3588223.
Krashen, S. (1993). *The power of reading: Insights from research.* Englewood, CO: Libraries Unlimited.
Laufer, B. (1989a). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review, 6*, 10–20.

Laufer, B. (1989b). What percentage of lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). London, England: Multilingual Matters.
Laufer, B. (1991). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). Oxford, England: Macmillan.
Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 20–34). Cambridge, England: Cambridge University Press.
Lewis, M. (1993). The Lexical Approach: The state of ELT and a way forward. Hove, England: LTP.
Liontas, J. I. (2002). Context and idiom understanding in second languages. *EUROSLA Yearbook, 2*, 155–185.
Liontas, J. I. (2003). Killing two birds with one stone: Understanding Spanish VP idioms in and out of context. *Hispania, 86*, 289–301. doi:10.2307/20062862.
Liontas, J. I. (2007). The eye never sees what the brain understands: Making sense of idioms in second languages. *Lingua et Linguistica, 1*), 25–44.
Maki, R. H., & McGuire, M. J. (2002). Metacognition for text: Findings and implications for education. In T. Schwarz & L. Bennett (Eds.), *Applied metacognition* (pp. 39–67). Cambridge, England: Cambridge University Press.
Martinez, R., & Schmitt, N. (2010). *A phrasal expressions list.*Manuscript submitted for publication
Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40–63). Cambridge, England: Cambridge University Press.
Moore, D., Lin-Agler, L. M., & Zabrucky, K. M. (2005). A source of metacomprehension inaccuracy. *Reading Psychology, 26*, 251–265. doi:10.1080/02702710590962578.
Morrison, L. (2004). Comprehension monitoring in first and second language reading. *The Canadian Modern Language Review, 61*, 77–106.
Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge, England: Cambridge University Press.
Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*, 59–82.
Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6–19). Cambridge, England: Cambridge University Press.
Nation, P. (1997). The language learning benefits of extensive reading. *Language Teacher, 21,* 13–16.
Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching.* Oxford, England: Oxford University Press.
Oh, S. Y. (2001). Two types of input modification and EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly, 35*, 69–96. doi:10.2307/3587860.
O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge, England: Cambridge University Press.
Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191–226). New York, NY: Longman.
Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*, 1–28.
Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146–161. doi:10.1017/S0267190504000078.

Sarac, S., & Tarhan, B. (2009). Calibration of comprehension and performance in L2 reading. *International Electronic Journal of Elementary Education, 2,* 167–179.
Schmitt, N., Jiang, N., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal, 95*(1), 26–43.
Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18,* 55–88.
Shin, D., & Nation, I. S. P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal, 62,* 339–348. doi:10.1093/elt/ccm091.
Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487–512.
Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford, England: Oxford University Press.
Spöttl, C., & McCarthy, M. (2003). Formulaic utterances in the multi-lingual context. In J. Cenoz, B. Hufeisen, &U. Jessner (Eds.), *The multilingual lexicon* (pp. 133–151). Dordrecht, The Netherlands: Kluwer.
Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36,* 139–152. doi:10.1080/09571730802389975.
Stubbs, M., & Barth, I. (2003). Using recurrent phrases as text-type descriminators: A quantitative method and some findings. *Functions of Language, 10,* 61–104. doi:10.1075/fol.10.1.04stu.
The British National Corpus. (2007). *The British National Corpus*(version 3, BNC XML ed.) [CD-ROM]. Distributed by Oxford University Computing Services on behalf of the BNC Consortium.
The iPhone's second coming. (2008 July), *The Economist* Retrieved from http://www.economist.com/science-technology/displaystory.cfm?story_id=E1_TTSDDJTG
Walter, E., Woodford, K., & Good, M. (Eds.). (2008). *Cambridge advanced learner's dictionary* (3rd ed.). Cambridge, England: Cambridge University Press.
Weir, C. J. (2005). Language testing and validation: An evidence-based approach. New York, NY: Palgrave Macmillan.
West, M. (1953). A general service list of English words. London, England: Longman
Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge, England: Cambridge University Press.