Tom Cobb
Linguistique et didactique des langues
UQAM, Montréal

Necessary or nice?
Computers in second language reading

Introduction

Does the computer have any important role to play in the development of second language (L2) reading ability? One role seems uncontroversial; networked multimedia computers can do what traditional media of literacy have always done, only more so. They can expand the quantity, variety, accessibility, transportability, modifiability, bandwidth, and context of written input, while at the same encouraging useful computing skills. Such contributions are nice if available, but hardly necessary for reading development to occur. A more controversial question is whether computing can provide any unique learning opportunities for L2 readers. This chapter argues that the role computing has played in understanding L2 reading, and the role it can play in facilitating L2 reading, are closer to necessary than to nice.

This chapter departs from two framing ideas in the introduction, one that L2 reading research continues to borrow too much and too uncritically from L1 reading research, the other that within the communicative paradigm reading is seen as a skill to be developed through extensive practice. The argument here is that L2 reading ability is unlikely to reach native-like competence simply through skill development and practice, and that the reason anyone ever thought it could is

precisely because the realities of L1 and L2 reading development have been systematically confused. However, we now have a dedicated body of L2 reading research, along with some more careful interpretations of L1 research, that together provide a detailed task analysis of learning to read in a second language and show quite clearly why reading has long been described as "a problem" (e.g., Alderson, 1985) and reading instruction a "lingering dilemma" (Bernhardt, 2005). The L2 findings, based on empirical research and supported by the analysis of texts by computer programs, detail both the lexical knowledge that is needed to underpin reading and the rates at which this knowledge can be acquired. While strong and relatively well known, however, these findings tend not to be incorporated into practice because there has seemed no obvious way to do so. However, the same computational tools that helped produce the findings can also help with exploiting them, and indeed that there is probably no other way to exploit them within a classroom context and time frame.

There are four relevant contexts to this chapter. The first is the large body of high quality L2 vocabulary and reading research, or rather vocabulary in reading research, that has come together since about 1990. The second is the spread of networked computing throughout much of the education system since about 1995. The third is the rapidly increasing number of non-Anglophone students worldwide who are attempting to gain an education through English, which of course largely means reading in English, i.e. reading English to learn. At one end, 30 per cent of PhD students enrolled in US universities in 2002/03 were international students (almost 50% in areas like engineering) according to Open Doors 2003 (Institute of International Education, 2003), and figures are similar or higher in other English speaking countries like Canada, The United Kingdom, Australia, and New Zealand. At the other end, in the K-12 population 12 per cent are currently classified as having limited English proficiency (LEP; US Census, 2000), and this figure is predicted to increase to 40% by the 2030's (Thomas & Collier, 2002). We owe it to these people to know what we are doing with their literacy preparation.

And finally a less obvious context is the longstanding debate among educational media researchers about the contribution to learning that can be expected from instructional media, particularly those that involve computing. One camp in this debate argues that while such media may improve access or motivation (i.e., they could be nice), they can in principle make no unique contribution to any form of learning that could not be provided in some other way (i.e., they are not necessary; Clark, 1983, 2001). Another camp argues that, while the no-unique-learning

argument often happens to be true, there are specific cases where media can indeed make unique contributions to learning (Cobb, 1997; 1999; in review) and that L2 reading is one of them. Discussed in generalities, there is no conclusion to this debate; discussed in specific cases, the conclusion is clear.

An advance organizer for the argument is as follows. Thanks to the extensive labours of many researchers in the field of L1 literacy, we now know about the primacy of vocabulary knowledge in reading (e.g., Anderson & Freebody, 1981). And thanks to the extensive labours of many in the field of L2 reading (as brought together and focused by Nation, e.g., 1990, 2001), we now know about the minimum amount of lexical knowledge that competent L2 reading requires, and in addition we know the rate at which this knowledge can be acquired in a naturalistic framework. As a result, we can see that while time and task tend to match up in an L1 timeframe (Nagy & Anderson, 1984), they do not match up at all in a typical L2 timeframe. First delineating and then responding to this mismatch is the theme of this chapter, with a focus on the necessary role of the computer in both parts of the process.

The chapter is intended to address practice as much as theory. All of the computational tools and several of the research studies discussed are available to teachers or researchers at the author's *Compleat Lexical Tutor* website (www.lextutor.ca, with individual pages indicated in the references; for a Site overview see Sevier, 2004). These tools can be used to test many of the claims presented here, and to perform concrete tasks in research, course design, and teaching.

Part I: The role of computing in defining the problem

How many words do you need to read?

Here are two simple but powerful findings produced by L2 reading researchers. The first is from Laufer (1989), who determined that an L2 reader can enjoy, answer questions on, and learn more new words from texts for which they know 19 words out of 20. This finding has been replicated many times and can be replicated by readers for themselves by trying to fill the gaps in the two versions of the same text below. The first text has 80% of its words known (one unknown in five), the

second has 95% of its words known (one unknown in twenty). For most readers, only a topic can be gleaned from the first text, and random blanks supplied with effort and backtracking; for the second text, a proposition can be constructed and blanks supplied with clear concepts or even specific words.

Figure 1: Reading texts with different proportions of words known

Text 1 (80% of words known - 32:40):

If _____ planting rates are _____ with planting _____ satisfied in each _____ and the forests milled at the earliest opportunity, the _____ wood supplies could further _____ to about 36 million _____ meters _____ in the period 2001-2015.

Text 2 (95% of words known – 38:40)

If current planting rates are maintained with planting targets satisfied in each _____ and the forests milled at the earliest opportunity, the available wood supplies could further _____ to about 36 million cubic meters annually in the period 2001-2015.

(From Nation, 1990, p. 242, and elaborated at Web reference [1].)

The second finding comes from a study by Milton and Meara (1995), which establishes a baseline for the amount of lexical growth that typically occurs in classroom learning. They found the average increase in basic recognition knowledge for 275 words in a six-month term, or 550 words per year. Readers can confirm their own or their learners' vocabulary sizes and rates of growth over time using the two versions of Nation's Vocabulary Levels Test (1990, provided online at Web reference [2]).

Thus, if we have a goal for L2 readers (to know 95% of the words in the texts they are reading), a way of determining how many words they know now (using the Levels test), and a baseline rate of progress toward the goal (550 new words per year), then it should be possible to put this information together in some useful way, for example to answer practical questions about which learners should be able to read which texts, for which purposes (consolidation, further vocabulary growth, or content learning), and how many more words they would need in order to do so.

Do learners who know 2,500 words thereby know 95% of the words on the front page of today's *New York Times*?

In fact, we cannot answer this type of question yet because there is a hole in the middle of the picture as presented so far. On one side, we have the numbers of words learners know, and on the other we have the percentages of words needed to read texts, but we have no link between words and percentages. Which words provide which percentages in typical texts, and is it the same across a variety of texts? Producing such a link requires that we inspect and compare the lexical composition of large numbers of large texts, or text corpora - so large, in fact, that they can only be handled with the help of a computer.

Corpus and computing are not needed to see that natural texts contain words that are repeated to widely different degrees, from words that appear on every line (*the* and *a*) to words that appear rarely or in specialized domains (*non-orthogonal* in statistics). Before computers were available, researchers like Zipf (Web reference [3]) developed different aspects of this idea, showing for example that oft-repeated *the* accounts for or *covers* a reliable 5 to 7% of the running words in almost any English text, and just 100 words provide coverage for a reliable 50%. Readers can confirm this type of calculation in a hand count of *the* from the previous paragraph, with six instances in 110 words or a coverage of just over 5%. Or they can investigate other coverage phenomena using texts of their own with the help of a text frequency program (at Web reference [4]). It seems quite encouraging that just a few very frequent words provide a surprisingly high coverage across a wide variety of texts, as the recent data from the 100 million-word British National Corpus, provided in Table 1, shows. If learners know just these 15 words, then they know more than a quarter of the words in almost any text they will encounter. Thus, in principle it can be calculated how many words they will need to know in order to achieve 95% coverage in any text.

Table 1: Typical coverages in a corpus of 100 million words

Word	PoS	Frequency/ million	Coverage (%) Word	Cumulative
1. the	Det	61847	6.18	-
2. of	Prep	29391	2.93	9.11
3. and	Conj	26817	2.68	11.79

4. a	Det	21626	2.16	13.95
5. in	Prep	18214	1.82	15.77
6. to	Inf	16284	1.62	17.39
7. it	Pron	10875	1.08	18.47
8. is	Verb	9982	0.99	19.46
9. to	Prep	9343	0.93	20.39
10. was	Verb	9236	0.92	21.31
11. I	Pron	8875	0.88	22.19
12. for	Prep	8412	0.84	23.17
13. that	Conj	7308	0.73	23.95
14. you	Pron	6954	0.69	24.64
15. he	Pron	6810	0.68	25.33

Source: Leech, Rayson, & Wilson (2001), or companion site at Web reference [5].

Earlier educators like Ogden (1930, Web reference [6]) and West (1953, Web reference [7]) had attempted to exploit the idea of text coverage for pedagogical purposes, but with conceptual techniques alone, in the absence of corpus and computing, this was only a partial success. (For an interesting discussion of early work in the vocabulary control movement, see Schmitt, 2000). The pedagogical challenge was to locate, somewhere on the uncharted lexical oceans between the extremes of very high and very low frequency words, a cut-off point that could define a basic lexicon of a language, or a set of basic lexicons for particular purposes, such as reading or particular kinds of reading. This point could not be found, however, until a number of theoretical decisions had been made (whether to count *cat* and *cats* as one word or two), until usable measurement concepts had been developed (*coverage* as a measure of average repetition), and until large text corpora had been assembled and the computational means devised for extracting information from them.

It was only quite recently that corpus researchers with computers and large text samples or corpora at their disposal, like Carroll, Davies and Richman (1971), were able to determine reliable coverage figures, such as that the 2000 highest frequency word families of English reliably cover 80% of the individual words in an average text (with minor variations of about 5% in either direction). Subsequent corpus analysis has confirmed this figure, and readers can reconfirm it for themselves by entering their own texts into the computer program at Web reference [8]. This program, Vocabprofile, provides the coverage in any text of these most frequent 2000 words of English Readers will discover that for most

texts, 2000 words do indeed provide about 80% coverage. For the previous paragraph, for example, it shows that the 2000 most frequent words in the language at large account for 81.35% of the words in this particular text. Here, then, is the missing link between numbers of words known and percentages of words needed.

With reliable coverage information of words of different frequencies across large numbers and types of texts, we are clearly in possession of a useful methodology for analyzing the task of learning to read in a second language. If learners know the 2000 most frequent words of English, then they know 80% of the words in most texts, and the rest of the journey up to 95% can be calculated. But first, do learners typically know 2000 word families?

What the coverage research tells us

What the coverage research mainly tells us is that there is no mystery why L2 reading should be seen as a problem area of instruction normally ending in some degree of failure. This is because 95% coverage corresponds to a vast quantity and quality of word knowledge and L2 learners tend to have so little of either.

Just within the 2,000 word zone already mentioned, intermediate classroom ESL learners typically do not know such a number of words, even at the most basic level of passive recognition. It is often the case that upper intermediate learners know many more than 2,000 words but not the particular 2000 complete word families that would give them 80% coverage. They often know words from all over the lexicon, which is a fine thing in itself, but nonetheless not have covered the basic level that gives them four words known in five. In several studies conducted by the current writer in several ESL zones (Canada, Oman, Hong Kong), academic learners were tested with different versions of Nation and colleagues' frequency based Levels Test, and a similar result was invariably produced: through random vocabulary pick-up, intermediate learners have at least recognition knowledge of between 4000 and 8000 word families, but this knowledge is distributed across the frequency zones – say, following interests in sports, hobbies, or local affairs – but is incomplete at the 2000 frequency zone.

A study by Zahar, Cobb & Spada (2001) shows the results of frequency-based vocabulary testing with Francophone ESL learners in Montreal, Canada. The test samples word knowledge at five frequency levels, as shown in Table 2. The high group (Group 5) are effectively bilinguals, and Groups 1 and 2 are intermediate

learners. The Total figure on the right of the table refers to the total number of word families out of 10,000 that these learners know, so that learners in Groups 1 and 2 have recognition knowledge of 3,800 and 4,800 words respectively. But despite this, these learners only know about half the words at the 2000 level. These skewed profiles are the typical products of random pick-up, with a possible contribution in the case of Francophone or Spanish ESL learners from easy-to-learn (or anyway easy-to-interpret) loan words or cognates which are mainly available at level 3000 and beyond (*absent, accident, accuse, require*), the 2000 level itself consisting largely of Anglo-Saxon items (*find, need, help, strike*) that are non-cognate.

Table 2: Levels scores by proficiency: Many words, low coverage for some

Group By proficiency	Vocabulary level scores (%)					Words known
	2000	3000	5000	UWL	10,000	Total
1 (low)	50	56	39	33	17	3800 wds
2	61	72	44	39	22	4800 wds
3	72	83	56	56	39	6000 wds
4	83	89	67	62	39	6900 wds
5 (high)	94	100	83	72	56	8000 wds

Note. UWL = University Word List (referred to below)

Are the learners in Groups 1 and 2 in good shape for reading texts in English? Despite the number of L2 words they apparently know, the answer is probably No, as was confirmed empirically with these particular learners, but probably No in principle. That is because the words they know are mainly medium frequency, low coverage words that do not reappear often in new texts and hence do not increase the known-to-unknown ratio.

There is a rapid fall in text coverage after the 2000 mark on the frequency list, as can be seen in Table 2 and its graphic representation in Figure 3. While 100 words give 50% coverage, and 2000 words give 80% coverage, after that the curve flattens out rather dramatically, so that learning another 1000 word families gives only a further 4-5% coverage, another 1000 only a further 2-3%, and so on. In other words, knowing a substantial number of even slightly lower frequency words does not necessarily affect the key known-to-unknown word ratio. As they read, these learners are facing texts with *at least* one unknown word in five, in other words with more dark spots than the first of the Forestry texts in Figure 1 above.

With such a small lexical knowledge base, both comprehension and further lexical growth through reading can only be sporadic.

Table 3: Average coverage based on a corpus of 5 million words

Number of words	Coverage provided
10	23.7%
100	49%
1,000	74.1%
2,000	81.3%
3,000	85.2%
4,000	87.6%
5,000	89.4%
12,448	95%
43,831	99%
86,743	100%

Source: Carroll, Davies & Richman (1971).

Figure 2: Graphic representation of coverage figures
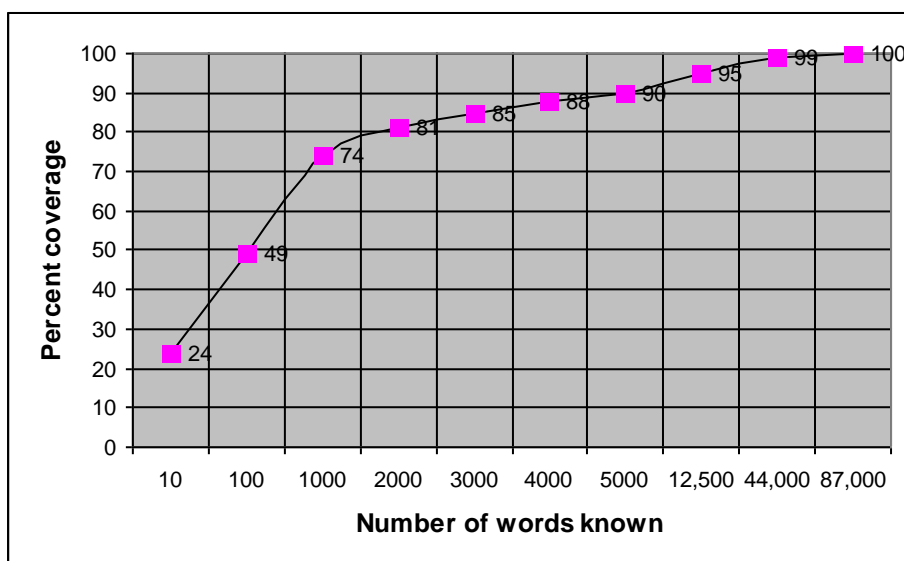
But if the coverage figures expose potential problems at the 2000 level, they expose far worse problems beyond the 2000 level. Suppose a learner were attempting to reach 95% coverage on the basis of naturalistic expansion, a prescription which of course is implicit in the skills and practice model. Figure 2

predicts a rather slow climb from the 80% to the 95% mark, which on the basis of naturalistic growth or extensive practice would require the learning of some additional thousands of words, specifically more than 12,000 word families, to reach 95% coverage. Let us now build a logical scenario for how this further growth could happen. Large numbers of post-2000 words would clearly need to be learned, but unfortunately these words present themselves less and less frequently for learning. How infrequently? Again this can be determined by corpus analysis. Let us take the Brown Corpus as representing a (rather improbable) maximum amount and variety of reading that an L2 learner could do over the course of a year. The Brown Corpus is one million words sampled from a wide variety of not very technical subjects. A description of this corpus can be found at Web reference [9], and the concordance analysis program used in this part of the analysis, called Range, at Web reference [10].

Range takes a word or word-root (roughly, a family) as input and returns the number of times and the number of sub-domains in the Brown corpus (from a total of 15) in which the input appears. Together, these counts give a maximum estimate of the number of times the input would be likely to appear in even the most diligent L2 learner's program of extensive reading. Table 4 shows the number and range of occurrences of a sample of words from below 2000 and above 2000 on the frequency list (see Web reference [11] for all lists discussed in this chapter). It seems quite clear that below 2000, words appear often and in a wide variety of domains, but that at some point quite soon after 2000 words appear much more rarely and only in some domains. Members of the *abort'* family appear only 10 times in 1 million words, and in fewer than half of the sub-domains. Readers can extend and test this information on Range by entering their own words of from different frequency levels. The conclusion seems obvious, that words will be encountered more and more sporadically after 2000, and progress toward 95% coverage will be slow to non-existent (with the possible exception of cognates, as mentioned). And the picture presented in Table 4 may be even more dire than it appears, since counts are based on families, as indicated with apostrophes (*arriv' = arrive, arrives, arrival*), yet as Schmitt and Zimmerman (2002) have shown, learners cannot be assumed to recognize the different members of a family as being related.

Table 4. Decreasing likelihood of meeting words

0-1000	1000-2000	4000-5000

Word family	Occur-rences	Domains / 15	Word family	Occur-rences	Domains / 15	Word Family	Occur-rences	Domains / 15
able	216	15	accustom'	15	10	abort'	10	7
accept'	270	14	admir'	66	14	adher'	26	9
agree'	286	15	afford'	58	12	ambigu'	40	7
answer'	277	15	amus'	38	13	analog'	29	8
appear'	426	15	annoy'	26	9	arbitrar'	27	7
arriv'	134	15	argu'	158	15	aspir'	28	9
MEAN	268.17	14.83		60.17	12.17		26.67	7.83
(SD)	(96)	(0.41)		(51.59)	(2.32)		(9.63)	0.98

The number of possible encounters with words clearly decreases, but does it decrease to the point where there is a learning problem? Are 10 occurrences many or few? Vocabulary acquisition research has by now told us quite a bit about the conditions of L2 word learning, and of the various kinds of lexical knowledge that are produced by different kinds and number of encounters. The number of occurrences question is reviewed in Zahar, Cobb & Spada (2001), and the overall determination seems to be that an a minimum 10 occurrences are needed in most cases just to establish a basic representation in memory. As Table 2 suggests, after the 2000 point this many encounters could not always be guaranteed even with 1 million words of wide reading. However, that is not the end of the problem.

But of course just having a basic representation for words in memory or a vague sense of their meaning in certain contexts is not quite all that is needed for their effective use, especially in oral or written production, but also in effective reading comprehension. A common theme in the L1 reading research of the 1980s was that certain types of word knowledge, or certain ways of learning words, somehow did not improve reading comprehension for texts containing the words (e.g., Mezynski, 1983). These ways of learning mainly involved meeting words in single contexts, looking up words in dictionaries, or being taught words in a language classroom. In fact, the only learning method that did affect reading comprehension was meeting new words not only several times but also in a rich variety of contexts and even a rich variety of distinct situations. The reason appears to be that most words have a range of facets or meanings, such that if only one of these is known, then it is unlikely to be fully applicable when the word is met in a new text.

In the L1 research, however, inert learning was only a problem of passing interest. Direct vocabulary instruction and dictionary look-ups are relatively rare in L1 language instruction, and most L1 learners who are reading at all are doing so roughly within their 95% zones and meeting words in varied contexts as a matter of course. It is rather L2 learners who are quite likely to be learning words from infrequent, impoverished contexts, in texts chosen for interest rather than level or learnability, aided (or not) by dictionaries of varying qualities, one-off classroom explanations, and so on. In other words, here is a case where L1 reading research is more relevant to L2 than to L1. Some of the writer's own research (e.g., Cobb 1999) extends this line of investigation to an L2 context and indeed finds that learning new words in rich, multiple contexts and situations, compared to learning the same words from small bilingual dictionaries, reliably produces about 30% better comprehension for texts that incorporate the words. The problem, of course, is where to find a steady supply of such contexts for relatively infrequent words in a time frame of less than a lifetime. For the purposes of the experiments, the contexts were artificially assembled; in nature, even a million words of reading per year would not necessarily provide them in the case of post-2000 words.

A further dimension of post-2000 word knowledge that could be predicted to be weak on the basis of insufficient encounters is lexical access. Rapid lexical access is vital for reading ability, and its development is mainly a function of number of encounters with words. Vast numbers of encounters are needed to produce instantaneous lexical access for words (Ellis, 2002). With testing that relies on basic recognition as its criterion of word knowledge, learners may look as if they are well on the way to 12,000 word families, but how accessible is this knowledge? Again some of the writer's own research may suggest an answer.

The importance of lexical access in reading was a landmark discovery in L1 research (e.g., Perfetti, 1985) and is now being adapted to the contours of L2 reading by Segalowitz and colleagues (e.g., 1998) and others. A good measure of how well a word is known is the length of time in milliseconds (ms) that it takes a reader to recognize the word or make some simple decision about it. Educated L1 readers produce a baseline reaction time of about 700 ms for common words, a time that rises slightly with frequency to about the 10,000-word mark, rarely surpassing 850 ms. But for even advanced L2 learners, the base rate is not only slower, at about 800 ms for the most common words (Cobb, in preparation; Segalowitz & Segalowitz, 1993), but also rises as words become even slightly less frequent, including between the 1000-2000, and 2000-3000 word levels. Results

from an experiment with 19 advanced francophone L2 learners are shown in Figure 3. As can be seen, even medium frequency words (in the 3k or 3000 word zone) are taking just under 1000 ms, almost a full second, to recognize; this is 30% over the L1 baseline. These are times associated with "problem reading" in L1 (Perfetti & Roth, 1981); that is because lexical access is stealing resources from meaning construction. There is no reason to think the situation is any different in L2. Teachers can test for frequency effects in their own learners' reaction times, for L1 or L2 words or both, using tools available at Web reference [12].

Figure 3: Reaction times in L1 and L2 at three frequency levels
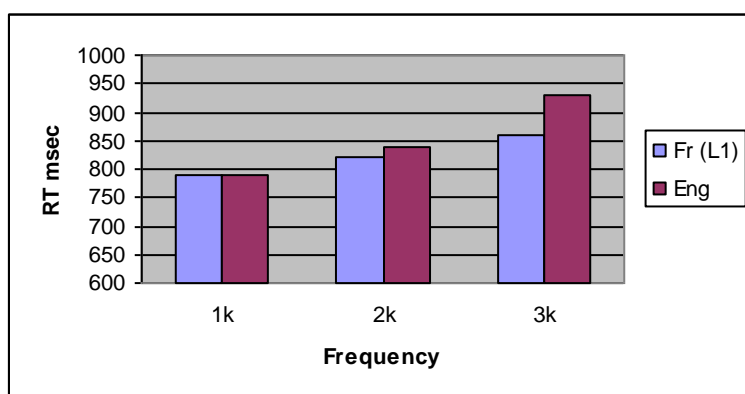
Note: L1 differences n.s.; all Eng (L2) differences significant at p<.05

In other words, the prospects for arriving at 95% coverage by slogging through the outer reaches of the lexicon are not good, as can be predicted logically, and confirmed empirically. With words distributed as they are in the print lexicon, as shown by coverage analysis, sufficient encounters for deep learning can only happen over extremely long periods – in fact, as long as it takes to grow up in a language.

Learning to read in L1 and L2: More and more different

Much of the shared knowledge of a culture is stored in its lexicon, and much of a lexicon is stored in written texts. Spoken language, especially conversation, does not normally require a great number of lexical items, because distinctions of meaning can be conveyed extralinguistically or pragmatically or can be co-constructed through negotiation. The sparse nature of the conversational lexicon

can be confirmed by running conversational texts through Vocabprofile (Web reference [8]), which typically shows 90% or greater coverage from just 1000 or even fewer words (cf. 80% for 2000 words in written texts).

Because lexicons attempt to encode all of the possible distinctions that are relevant to successful functioning in their typical environments, they can grow to considerable sizes. In the case of English, and other languages with significant written components, this has led to the evolution of truly enormous lexicons. Limiting the proliferation of lexicons, however, is the competing requirement of learnability. A lexicon cannot become so enormous that no child can learn it. Borrowing from the analysis above, we can infer that the literate adult lexicon must be minimally sufficient to provide 95% coverage within a wide variety of texts, such that almost any but the most specialist texts can be read and the literally hundreds of thousands of low frequency items at the peripheries of the lexicon interpreted with varying degrees of effort in context. As already suggested, about 12,000 word families seems to be the minimum number needed for this, with most literate adults having well more than that. How much more? Empirical size estimates are compatible with this proposal in that they typically determine the vocabulary size of literate adults to be 20,000 word families in their L1s (Nation, 2001). This seems, then, to be the zone where the demands of proliferation and learnability intersect – somewhere between 12,000 and 20,000 word families.

And how long does it take to learn either number of word families to the degree that they are useful in reading comprehension? A classic study by Nagy, Herman and Anderson (1985) detailed rates and sources of lexical growth, showing that while the average probability of learning any word from a single contextual encounter was rather low, it was nonetheless sufficient to account for the measured lexical growth of first-language learners on the basis of an average reading intake of 1 million words per year over 10 years of schooling. Between the ages of 6 and 16, an average of 3,000 to 4,000 literate words per year are learned, which adds up to far more than 20,000 word families. The L2 learner, in contrast, as we remember from the Milton and Meara (1994) study, learns on average 550 words per year, as measured on a basic yes-no vocabulary test (Do you know this word, Yes or No), i.e. with no clue as to whether this knowledge was of the rapid-access, multicontextual variety that can affect comprehension or provide the basis for further learning. In other words, the L1 rate is greater than the L2 rate by a factor of at six or seven. At best, L2 learners can end up with 5,000 or so word families in total, often not known very well, as already shown, and with gaps in high coverage

zones, again as already shown. Even with 5,000 well known words and no gaps, L2 learners are working with only 90% coverage, that is, with one word missing in ten – or one per line of printed text, as opposed to one per two lines at 95%. If we do the math (words needed, learning opportunities, learning rates) the numbers might add up for L1 learners, but they do not add up for L2 learners.

So the task of acquiring a literate lexicon is radically different for L1 and L2 learners, such that becoming a fully competent L2 reader is effectively impossible on the basis of input and practice alone. This runs against the assumption of communicative approaches to reading, namely the sufficiency of input and practice, which in turn is based on the assumption that learning to read in L1 and L2 are essentially the same. *Impossible* is of course a word the needs interpretation. Clearly some form of L2 reading does take place in the worlds of school and work, although probably with far larger infusions of top-down knowledge or other forms of guesswork than characterizes proficient L1 reading. And impossible only if we assume that input and practice must take place in L2 as it did in L1 - i.e., letting nature run its course, assigning to teaching a relatively minor, mainly facilitative role, and assigning to instructional design almost no role other than to ensure a wide range of text inputs for self-selection.

In L2, we apparently cannot rely on nature to do our teaching for us but must resign ourselves to mundane prospects like task and needs analysis, the establishment of feasible objectives, possibly different for different learners, and the usual processes of instructional design and materials selection - rather than pretending that reading will happen by itself as it appeared to in L1. But wait a minute: wasn't it L1 reading researchers who warned educators that reading, unlike speech, is an "unnatural act" that is inherently difficult and will leave many behind unless steps are taken to prevent it (Gough & Hillinger, 1980, cited in Adams, 1990)? Somehow, this was not one of the L1 ideas that got imported into L2 thinking.

To summarize, the chapter has so far argued that the computer has played a vital role in the task analysis of L2 reading. This role derives from the fact that only a computer can hold enough textual or other information in memory to disclose relevant patterns, whether in the vast expanses of the collective lexicon (e.g., coverage information) or the minutiae of individual word processing (e.g., reaction times). To the naked eye, either is as invisible as the role of bacteria in disease must have been 100 years ago. And just as computing has played a vital role in

defining the reading problem, so can it play a vital role in solving it. That is because the problem the computer solves for researchers is essentially the same one it can solve for prospective L2 readers – exposing the patterns buried in overwhelming data.

Part II: The role of computing in solving the problem

As discussed already, the obstacles to reading in a second language are predictable. But any problem that is predictable is in principle solvable. This part of the chapter deals with solutions to the problems outlined above, and proceeds with a problem and solution framework.

1. How can gaps in high-coverage lexical zones be prevented?

While any L2 reading course should offer large amounts and variety of reading, i.e. practice, it should also include some sort of direct, frequency based vocabulary component which will expose learners at the very least to the 2000 words that will reduce the dark spots in L2 texts to one word in five. This is not something that will happen by itself for all learners. It is common for L2 learners to know many words they will rarely see again yet still have gaps in the zones that provide high coverage in new texts (see Table 2, above). This situation can be prevented by making a level appropriate, frequency based vocabulary course a standard part of any L2 reading program, and at the same time making a pedagogical lexis course a standard part of any L2 teacher training program. In the writer's experience, neither of these is particularly common at present, certainly less common than the ubiquitous pedagogical grammar course (despite the fact that grammar has never been shown amenable to pedagogy while vocabulary clearly is).

A point to note regarding vocabulary courses is that commercial vocabulary courses are not necessarily frequency based; in fact, they are often devoted to "increasing your word power" precisely through the random pick-up of odd but interesting low frequency items that learners are already proficient in. A vocabulary course for learners could use a computer or not but, either way, if it is

done according to principles discussed above it is likely to be home made rather than store bought. An idea for a full computer version of such a course is described in Cobb (1999), in which learners used corpus materials to build their own dictionaries for a complete 1000-level frequency list and were tested using level-specific cloze passages (which can be built by anyone providing their own texts at Web reference [13]). Some principled, frequency based course books are now beginning to appear, including one developed for "mastering the AWL" (or Academic Word List, to be discussed below) by Schmitt and Schmitt (2005) that exemplifies many of the principles discussed above. A computer version of these materials providing maximum recycling and corpus back-up will be available by the time of publication at Web reference.

2. How can reading texts be chosen to maximize learners' skill development, vocabulary growth, and pleasure?

These benefits can be provided by matching texts to learners, i.e. by providing them with texts bearing 95% of words they can be reasonably expected already to know and only 5% to be handled as new items (looked up, worked out from context, etc.) While such proportions are clearly impossible to calculate with total precision, they can be approximated well enough by matching learners' Levels Test scores to texts that are one level in front of them, as indicated by running the text through the computer program Vocabprofile (Web reference [8]). For example, learners who are strong at the 1000 word level but weak at 2000 should read texts with 5% of their lexical items drawn from the 2000 level. The texts could be either found texts or modified texts. It would be an understatement to say that finding or writing such texts would be difficult without the help of a program like Vocabprofile.

3. Are there any shortcuts on the long climb to 95% coverage (12 thousand word families)?

The short answer is Yes, and again the shortcuts have been discovered through various types of computer text analysis and confirmed in empirical studies. Three of these will be outlined. First, Hirsch and Nation (1994) used Vocabprofile to determine how many words would be needed to read a specific kind of text, unsimplified fiction for young adults, with pleasure and with further vocabulary learning. They found that texts in this genre typically present 95% known-word conditions for learners with a vocabulary of only 2000 word families. This is

because of a distinct feature of this particular genre, namely a predominance of names and other proper nouns which reduces the density of novel lexis, since names are not normally words that require any learning. So these texts are an excellent source of consolidation reading for learners who know 2000 words, or acquisition reading for those who know 1000 but are weak at 2000. A limitation of this finding, of course, is that learners may have other reading goals.

Another approach and one of the most exciting research programs in L2 reading involves the creation of dedicated frequency lists for particular reading objectives. One of these is the already mentioned Academic Word List (AWL), which is intended for learners who are planning to undertake academic study through English. The AWL is a list of word families that are not included among the 2000 most frequent families and yet are reasonably frequent across a minimum number of specialist domains within a large corpus. It was constructed by first isolating all the post-2000 words of a large corpus with Vocabprofile, and then determining which of these occurred a minimum number of times in all of four academic domains (arts, law, commerce, and science). The resulting list comprises just 570 word families (words like *access, abandon,* and of course *academic*) but, when added to the 2000 list, makes a combined list of 2570 word families that typically provide a coverage of about 90% in academic texts (which is otherwise achieved somewhere beyond 5000 families, as shown in Figure 2).

Further shortcuts can be found beyond the AWL through the construction of frequency lists within particular domains. This was shown by Sutarsyah, Nation and Kennedy (1994) in the domain of economics. Briefly, an economics corpus was assembled, the 2000 and AWL lists subtracted out of it using Vocabprofile, and each item in the remainder evaluated for number and range of occurrences across the chapters of a substantial economics text, yielding an "economics list" in the vicinity of 500 word families. This word list, in conjunction with the 2000 list and AWL, typically raises coverage for texts *in this domain* to fully 95% in return for knowing just 3070 word families. The development of such technical lists is one of the most exciting areas of current research and development.

4. Notwithstanding shortcuts, 2000 or 3070 word families are still substantial learning tasks. Is there any way to speed up the acquisition of large numbers of new words?

Medium frequency words appear in texts only as often as they appear, and there is little that can be done to increase their numbers. But it is possible to increase the number of people looking for them, and then to share the findings. Something like this is often done by good teachers on their blackboards, but it can be done more effectively by learners themselves with a shared database on a computer network. The program GroupLex (Figure 4, Web reference [16]) is such a database. Hundreds of words can be entered into a simple database from the learners' reading materials, on whatever basis the group or teacher decides, and then sliced, diced, and recycled in several different ways. The program can encourage shared learning, domain specific learning, and can generate quizzes online or on paper. Hundreds of words per month can thus be accumulated and encountered, either in a single sentence contexts and definitions or in several contexts if this is encouraged. Learning results from an earlier version of this collaborative tool showed the learning rate increasing 40% over Milton and Meara's average (Cobb, 1999). Learning results from GroupLex and the story of its incorporation in a real language course are reported in Horst, Cobb & Nicolae (2005).

Figure 4: Group Lex: Learning words together

5. Rich word knowledge depends on meeting words in several contexts. Can contexts be multiplied by a computer in any learning-useful way?

The answer is Yes, through the use of an adapted concordance program. A concordance is a computer program that assembles all the contexts for a given word or phrase from different locations throughout a corpus, in the format of a central keyword and chopped off line endings (although with an option to request a broader contextualization; see Web reference [18]). There is a question, however, whether any "rich" learning benefits can be derived from this form of multi-contextualization. This was the question of a research study (already briefly mentioned) in which university bound learners in the Sultanate of Oman built personalized glossaries of the entire second thousand word list, in which testing had shown them to be deficient. Each word in the glossary had to accompanied by at least one clear example sentence found in a specially built corpus, and this entailed searching through several contexts for one that made sense to the learner. The corpus was a purpose-built collection of all the learners' course materials for one year. A control group performed a similar exercise but without the concordance work. The test of rich learning was to transfer several sets of learned words to gaps in level controlled, novel texts. The concordance users were 30% more successful than controls in this task, and therefore it appears that working with computer-generated contexts can make up at least to some extent for the lack of time to meet enough contexts in a more natural fashion. For a report on this research see Cobb, 1997; 1999; Cobb and Horst, 2001.

But how applicable is this finding to practical contexts? A purpose-built corpus while able to produce an interesting research result might be considered somewhat impractical. Building even a small special corpus is a rather large task. On the other hand, as a source of learning contexts, a natural corpus such as the Brown surely breaks the 95% known-word rule rather badly; many words are unlikely to be known in the inference base. Is there nothing in between?

A number of different sources of electronic text are being considered for their ability to provide comprehensible multicontexts for intermediate learners. One is a trial learning corpus currently being constructed in Montreal from a collection of

about 100 simplified readers. Such texts typically have more than 90% of their lexis in the 1000-word frequency zone, so virtually all contexts should be comprehensible for intermediate learners. Another idea involves providing concordances from within a single text or within the canon of a single writer, which should provide a consistency of style and substance that learners can gradually become accustomed to, so that as reading proceeds the contextual information becomes more and more clear.

For example, Figure 5 shows a screen from Lextutor's Resource-Assisted Reading version of Jack London's tale *Call of the Wild* (Web reference [19]), wherein a reader has clicked on the curious word *bristling* in Chapter 3 only to be reminded that there was already an occurrence of this word in Chapter 1 and that there will be several more in chapters to come. This information is provided without the reader leaving the text he or she is reading, i.e. without the text disappearing from view behind another window.

*Figure 5: Interactive story concordance – 1*
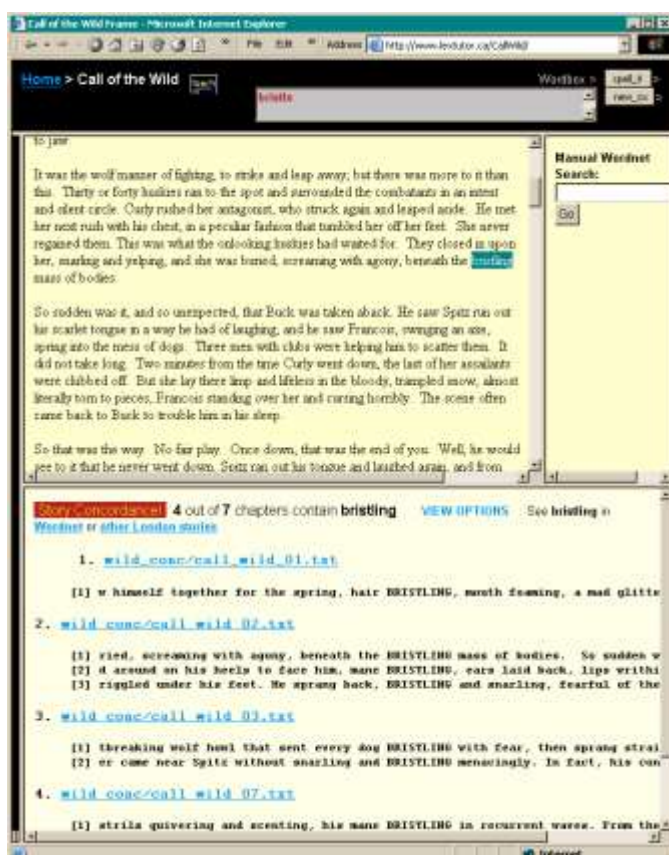
One of the concordance lines in Figure 5 features a further unknown word, *mane.* Learners who were curious about that word, or felt it might give a clue to the meaning of *bristling*, could just wait until it cropped up somewhere later in the text, when they might or might not remember why they were interested in the word, or they could just click on it in the concordance window itself, which is also a live window that produces concordances of its own. Another option is to drag the mouse across both words, *mane* and *bristling,* to see further occurrences of the phrase in the story, if any, and in a larger context (as shown in Figure 6).

Still more contexts for either words or phrases can be produced by clicking the link *Other London stories*; doing this will produce other uses of the same word or phrase in other works by the same author. Sure enough, this writer has employed this same expression in another story of his repertoire (see Figure 7). The point is that multiple contexts drawn from within a single canon should be more comprehensible than contexts drawn from random texts (like those of the Brown corpus), and moreover this comprehensibility should increase with familiarity.

Several screen shots have been provided so that readers can decide for themselves if the author-familiarity argument has any value. At what point 95% familiarity is achieved using this method is an empirical question.

Figure 6: Interactive story concordance - 2
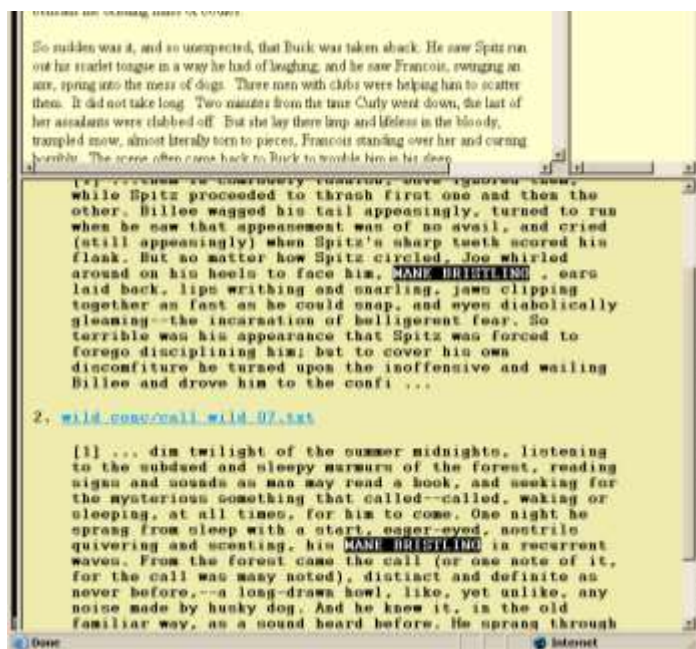
The fully wired version of this particular story, Jack London's *Call of the Wild*, can be tested at Web reference [19], and a similarly wired French story (de Maupassant's Boule de Suif, at Web reference [20]). *Call of the Wild* and *Boule de Suif* are of course just two randomly chosen stories, developed for purposes of demonstration and experimentation, and probably not all that useful for typical L2 readers ("bristling manes" are hardly to be met very often outside certain types of fiction). Teachers and learners would thus want to use these resources with texts of their own choosing, which they can do (for most of the resources described) at Web reference [21].

Is there any learning result from resource-assisted reading? A preliminary single-subject study by Cobb, Greaves and Horst (2001) showed 40% superior vocabulary acquisition from resource-assisted over unassisted reading, and of course vocabulary growth from reading is a good indication of text comprehension as well as an end in itself
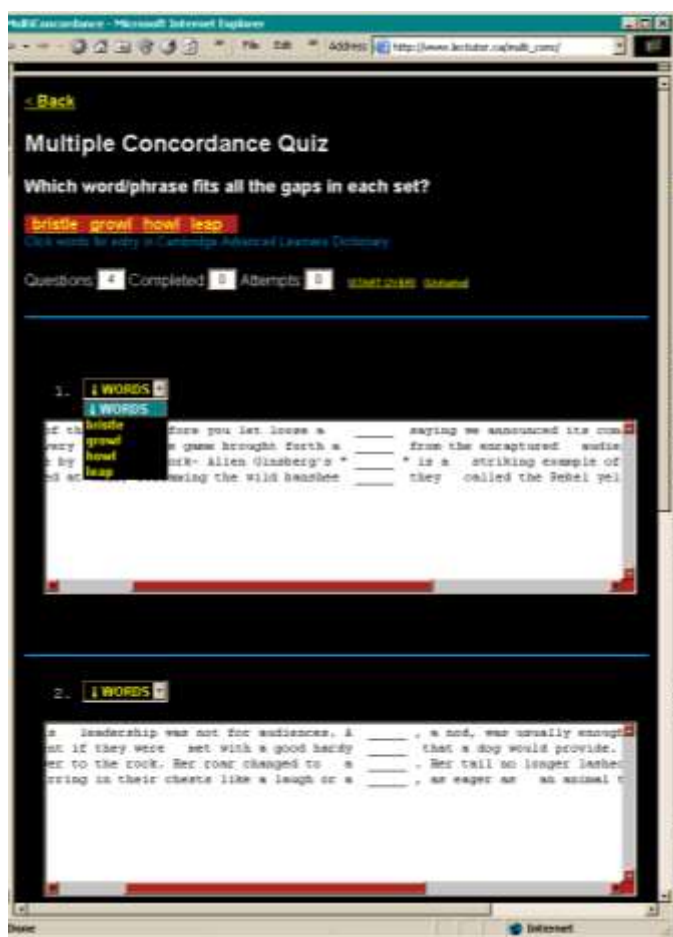
Figure 7: Interactive story concordance - 3

6. Working with familiar contexts is important at one stage of learning, but isn't the final goal to transfer word knowledge to novel contexts?

At some point, of course, any newly learned words will have to be applied in contexts beyond those of a particular writer. Transfer is the final step in most theories of learning, and transfer, as we know, is unlikely to happen unless provisioned in the design of instruction. How it is provisioned in the story concordance is as follows. When learners have identified an interesting word that they want to work on, they can record it for later attention simply by clicking on it with Alt-key held down. Doing this places the word in the silver box that can be seen at the top of Figures 5 or 7, and reading can proceed uninterrupted. Then, at some convenient point like the end of a chapter, all these words can be recycled in various ways – added to the learner's private database, pasted into Group Lex, or sent automatically to a program that generates novel-context quizzes using concordances from the Brown. This program is Multiconc (Web reference [18]). Figure 8 shows words gathered from a chapter in *Call of the Wild (bristle, howl, growl,* and *leap)* ready to be recycled in quite different sorts of contexts. The learner looks at the words around the gaps and tries to decide which of the targets will fit into each. Other options within MultiConc can be used to prepare learners for this activity.

Figure 8: Transfer of Call of the Wild words to novel contexts

Multiple Concordance Quiz

Which word/phrase fits all the gaps in each set?

bristle growl howl leap

Questions 4 Completed 0 Attempts 0

Conclusion

The vision presented here, then, is one where L2 learners undertake systematic vocabulary growth, read texts with known proportions of unknown lexis chosen in accordance with existing knowledge and learning goals, and maximize the vocabulary learning opportunities (recycling, recontextualization, transfer) within these texts through the use of technology. The ideas presented here barely scratch the surface of what can be done with computers to develop and encourage vocabulary growth in designed L2 reading development. As mentioned, the computer and the L2 reader are a natural match. Instances and teaching ideas could be multiplied, but by now the point is probably made that the computer's ability to take in, relate, and organize large spans of written language, whether single texts or 100-million word corpora, can compensate for many of the inherent difficulties of learning to read in a second language and in principle can play a very significant

role in the development of L2 reading instruction. And in this context the question of whether media can have any unique effects on learning seems to belong to a simpler time.

But will the computer ever play such a role in L2 reading in fact? It is not impossible, although a number of things are working against it. First, as Bernhardt (2005) recently pointed out, a faithful 20% of the variance in L2 reading can be attributed to L1 reading level, and "it is a rare L2 [intervention] study that appears to have an effect size large enough to overcome a 20% (or even a 10%) variance attributable to first language literacy" (p. 142). In other words, some L2 readers will always be weak for reasons neither learner nor teacher has any control over, and it is doubtful if any amount of vocabulary recycling or recontextualizing will produce a competent reader in L2 who was not a competent reader in L1. Thus a good-enough definition of success in L2 reading is likely to prevail, and designs for systematic, goal-oriented improvement such as those proposed here are likely to be seen as wishful thinking. Second, the institutional infrastructure for a thoroughgoing computational approach like the one suggested here is not simple to achieve. If 30 learners are going to spend two hours a week on Group Lex, and read 20 pages of text designed by a teacher with the help of Vocabprofile and linked to learning resources, for a long enough period to make any difference, – this will only happen, on other than a one-off basis, in the most supportive of institutions. (But such institutions do exist; for some names, see Appendix 1.) And third, while most learners are comfortable with screen reading, e-learning, and computing generally, many teachers are doubtful, resistant, or hostile to it. Any systematic use of computing in second language education probably awaits a generational change. All in all, the role envisaged for the computer in this chapter can probably only develop in conjunction with solutions to broader problems.

On the bright side, the best L2 reading researchers tend to endorse a strong role for computing in reading instruction, albeit often without specifying much detail. Koda (2005) concludes her recent book on reading research with the thought that "The enhanced capabilities of advanced computer technology … hold strong promise for major breakthroughs in L2 reading instruction" (p. 273). But the brightest spot on the bright side are some grassroots developments that do not normally make it into the research journals. Lextutor's records show that many individual teachers and even learners are bypassing their institutions and using the website to design their own principled instructional materials. Thousands of users throughout the day and night, from all over the world, log on to take Levels Tests,

run texts through Vocabprofile, build resource-assisted hypertexts on literally thousands of topics, reconfigure hypertexts as cloze passages, and run many, many, concordances for words and phrases of interest. If this is where the researchers, teachers and learners are heading, can the institutions be far behind?

Endnote

This chapter is based on two invited presentations from 2004, one at the reading symposium preceding TESOL 2004 in Long Beach, California, entitled *Computer solutions to classic L2 reading problems*, and the other at the English Language Institute of the University of Michigan entitled *The logical problem with acquiring second lexicons and how networked computing can solve it*.

References

Text references

Adams, M.J. (1990). *Beginning to Read: Thinking and Learning About Print.* Cambridge, MA: MIT Press.

Alderson, C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. Alderson & A. Urquhart (Eds.), *Reading in a Foreign Language* (pp. 1-24). London: Longman.

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Bernhardt, E. (2005). Progress & procrastination in second language reading. In M. McGroarty (Ed.), *Annual Review of Applied Linguistics, 25* (pp. 133-150). New York: Cambridge University Press.

Carrol, J., Davies, P., & Richman, B. (1971) The American Heritage Word Freuency Book. New York: Houghton Mifflin, Boston American Heritage.

Coxhead, A. (2000). A new academic word list. TESOL Quarterly 23 (2), 213-238.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System 25* (3), 301-315.

Cobb, T. (1997). Cognitive efficiency: Toward a revised theory of media. *Educational Technology Research & Development, 45* (4), 21-35.

Cobb, T. (1999). Applying constructivism: A test for the learner-as-scientist. *Educational Technology Research & Development, 47* (3), 15-33.

Cobb, T., Greaves, C., & Horst, M. (2001). Can the rate of lexical acquisition from reading be increased? An experiment in reading French with a suite of on-line resources. In P. Raymond & C. Cornaire, *Regards sur la didactique des langues secondes.* Montréal: Éditions logique.

Cobb, T. (In review). On chalkface & interface: A developer's take on the classroom-distance metacomparison. Response to Bernard et al. comparison of effectiveness of classroom vs. distance education (study and comment submitted to Erlbaum).

Cobb, T. (in preparation.) Toward a practical methodology of reaction time testing: How and why.

Cobb, T., & Horst, M. (2001). Reading academic English: Carrying learners across the lexical threshold. In J. Flowerdew & M. Peacock (Eds.) *The English for Academic Purposes Curriculum* (pp. 315-329). Cambridge: Cambridge University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34,* 213-238.

Clark, R. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53,* 445-459.

Clark, R. (Ed.) (2001). Learning from media: Arguments, analysis, evidence. Greenwich, CT: Information Age Publishing.

Ellis, N.C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition 24* (2), 143-188.

Gough, P. B., & Hillinger, M. L. (1980). Learning to read: An unnatural act. *Bulletin of the Orton Society, 30,* 179-190.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11,*207-223.

Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with a collaborative on-line database. *Language Learning & Technology, 9* (2), 90-110.

Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Review, 56* (2), 308-328.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8* (2), 689-696.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach.* New York: Cambridge University Press.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines.* Clevedon: Multilingual Maters.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word Frequencies in Written and Spoken English: based on the British National Corpus.* London: Longman.

London, J. (1903). *Call of the Wild*.
[Online: Accessed 20 Dec. 2005, http://www.archive.org/details/callw10].

Mezynski, K. (1983) Issues concerning the acquisition of knowledge: effects of vocabulary training on reading comprehension. *Review of Educational Research, 53* (2), 253-279.

Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics, 107/108,* 17-34.

Nagy, W.E., & Anderson, R.C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19,* 304-330.

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

Nation, 2001. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Ogden, C. (1930). *Basic English: A General Introduction with Rules and Grammar*. London: Kegan Paul, 1940.

Perfetti, C. (1985). *Reading Ability.* New York: Oxford University Press.

Perfetti, C. A., and Roth, S. (1981) Some of the interactive processes in reading and their role in reading skill. In A. Lesgold and C. Perfetti (Eds.), *Interactive Processes in Reading* (pp. 269-97), Hillsdale, NJ: Erlbaum Associates.

Schmitt, N. (2000). *Vocabulary in Language Teaching.* New York: Cambridge University Press.

Schmitt, N., & Schmitt, D. (2005). *Focus on vocabulary: Mastering the Academic Word List.* White Plains NY: Longman.

Schmitt, N. & Zimmermann, C. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36,* 145-171.

Segalowitz, S., Segalowitz, N., & Wood, A. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics, 19* (1), 53-68.

Segalowitz, S., Segalowitz, N. (1993). Skilled performance, practice, & the differentiation of speed up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics 14,* 369-385.

Sevier, M. (2004). Review of The Compleat Lexical Tutor. *TESL-EJ 8*(3). [Online: accessed Dec. 20, 2005, at http://www-writing.berkeley.edu/TESL-EJ/ej31/m2.html].

Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal, 25* (2), 34-50.

Thomas, T., & Collier, V. (2002). *A national study of school effectiveness for language minority students' achievement.* (No. 143). ERIC Document Reproduction Service: ED475048.

West, M. (1953). *A general service list of English words.* London: Longman, Green, & Co.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review,* 57(4), 541-572.

Web references

Note: Spaces in Web addresses represent underscores.
[1] Lextutor research page   http://www.lextutor.ca/research/
[2] Levels Tests   http://www.lextutor.ca/tests/
[3] Zipf Web page   http://www.nslij-genetics.org/wli/zipf/
[4] Lextutor freq. indexer   http://www.lextutor.ca/freq/
[5] Lancaster BNC lists
     http://www.comp.lancs.ac.uk/ucrel/bncfreq/lists/1_2_all_freq.txt
[6] Basic English Website   http://ogden.basic-english.org/basiceng.html
[7] General Service List   http://jbauman.com/aboutgsl.html
[8] Lextutor Vocabprofile   http://www.lextutor.ca/vp/
[9] Brown Corpus page http://helmer.aksis.uib.no/icame/brown/bcm.html
[10] Lextutor Range   http://www.lextutor.ca/range/
[11] Lextutor linked lists http://www.lextutor.ca/lists_learn/
[12] Lextutor RT builder http://www.lextutor.ca/rt/
[13] Lextutor cloze builder   http://www.lextutor.ca/cloze/vp_cloze/
[14] Lextutor Text Compare   http://www.lextutor.ca/text_lex_compare/
[15] Lextutor group lex   http://www.lextutor.ca/group_lex/demo/
[16] Lextutor concordancers   http://www.lextutor.ca/concordancers/
[17] Lextutor multiconcord   http://www.lextutor.ca/multi_conc/
[18] Lextutor Call of Wild   http://www.lextutor.ca/callwild

[19] Lextutor Boule de Suif   http://www.lextutor.ca/bouledesuif/
[20] Lextutor hypertext builder
                        http://www.lextutor.ca/hypertext/
[21] Lextutor cloze builder   http://www.lextutor.ca/cloze
[22] Lextutor text compare   http://www.lextutor.ca/text_lex_compare/

Appendix 1

These institutions support sustained use of some or all of the computational resources mentioned above.

- Group Lex is in regular use in the English Bridge Program, Simon Fraser University, Vancouver, Canada (contact Marti Sevier, <msevier@sfu.ca>).

- The ESL program at Seneca College, Toronto, Canada, makes extensive use of many of the tools mentioned in this chapter (contact Ross McCague <ross.mccague@senecac.on.ca>).

- The ESL programs in two Turkish universities make extensive use of Vocabprofile in their course development (Eastern Mediterranean University and Middle East Technical University, North Cyprus Campus (contact Steve Neufeld <steve@ingilnet.com>).