

Giving learners something to do with concordance output

Tom Cobb

Dép. de linguistique et de didactique des langues
Université du Québec à Montréal
Canada

Introduction

Hands-on learner concordancing has been regularly described for more than 15 years as an interesting approach in computer assisted language learning. However, it is not clear that concordancing is ever used with any but advanced students (while the vast majority of language learners are intermediate), for more than a handful of guided sessions (which rarely lead to subsequent independent use), or that there is any demonstrated learning advantage of concordancing over cheaper means of exploring a second language (e.g., dictionaries, grammar books, TV or cassettes).

Believing that concordancing properly adapted to learner needs could indeed be a powerful learning technology, I have attempted to focus on a particular learning objective for concordances (rapid vocabulary expansion by meeting new words in multiple contexts; Cobb 1996), establish empirically the value of using concordances to meet this objective (Cobb, 1997a), and develop a learner interface and fiction (the learner as lexicographer; Cobb, 1999a) that would sustain extended independent use of concordances by intermediate and advanced learners alike. This paper will review this work with particular reference to the needs of learners of English for Academic Purposes (EAP), and demonstrate attempts now underway to make the whole system Web-worthy.

The problem of second-language word learning

One of the biggest challenges in EAP is helping students acquire the vocabulary they need to begin reading in a subject area. Students typically need words measured in the thousands, not hundreds, but receive language instruction measured in months, not years. In this time-squeeze, vocabulary course developers must choose between breadth (explicit learning of words on lists) and depth (implicit learning of words through extensive reading). But list-learning creates superficial knowledge, and acquisition through reading is too slow for the time available. This paradox has been seen as unresolvable using traditional learning technologies, but computer technology suggests new possibilities.

The advantages of word lists are many, particularly in the age of computational approaches to language. A corpus of subject-area texts can be assembled and "crunched" with a frequency program to determine which words a student needs to know to begin reading in the area. An interesting finding from corpus studies is that the vocabulary of a subject area is not be as large as it seems. Possibly as few as 3,500 words may be adequate preparation for independent reading in a discipline like economics (Sutarsyah et al, 1994). Such a number of words is in principle amenable to some form of direct instruction.

However, the disadvantages of word lists are also many. Giving lists to students has never been shown to be very effective. Lists send students running for their small, usually bilingual dictionaries, from which they construct fragile lexicons of one-to-one translation equivalents. These neither (a) improve their reading comprehension, even of texts employing the words they have worked on, or (b) serve as an adequate basis for future word learning (Miller & Gildea, 1987; Nesi & Meara, 1994). Large, well structured, richly interconnected and cross-referenced second-language lexicons appear to be acquired only

through meeting words in diverse natural contexts over lengthy periods of time, such as the ten or so leisurely, risk-free years of childhood (Mezynski, 1983; Stahl & Fairbanks, 1986).

The breadth-depth paradox in second-language vocabulary acquisition is a stark one, especially as the importance of vocabulary in language development, which was neglected in the Chomskyan era (Meara, 1980) becomes ever more apparent. Over the years this problem has often been noted but typically seen as insoluble. Long ago, Carroll (1964) expressed the wish that some form of vocabulary instruction could be found to mimic the effects of natural contextual learning, but more efficiently. More recently Krashen (1989) complained that "vocabulary teaching methods that attempt to do what reading does--give the student a complete knowledge of the word--are not efficient, and those that are efficient result in superficial knowledge" (p. 450). An "efficient" resolution of the paradox is something instructors might reasonably expect to find in some application of instructional technology (see Cobb, 1997b for a discussion of Cognitive Efficiency as a basis for media development).

Concordances in principle

It has occurred to several instructional designers that the same concordance procedure that has been successful in identifying which words learners should learn might also be of use in helping them learn the words. Some sort of concordance, which is a word list with contexts for each word, seems a likely first guess at a harmonization of depth and breadth. Accordingly, many language learners in the past 15 years have been invited to examine particular words with the aid of popular commercial corpus and concordance kits like Microconcord (Johns, 1986; Scott & Johns, 1993) or Wordsmith (Scott, 1996). In Figure 1 we see a screen from the Wordsmith Web page, where a user has just done a search through a collection of British

newspapers on the word "hands," showing fairly clearly how a concordance brings list and contexts together.

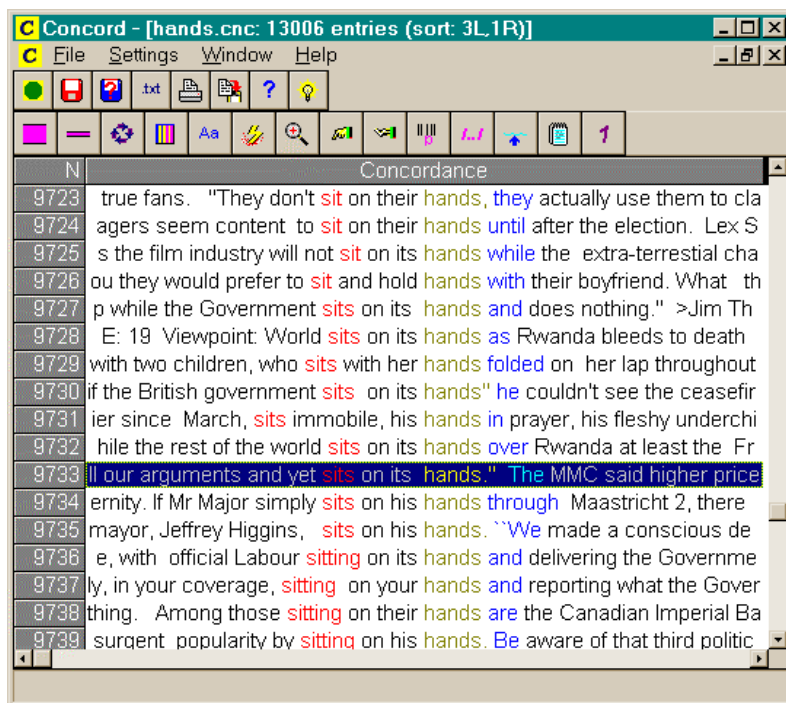


Figure 1

However, the figure also shows fairly clearly why a concordance might be of limited interest to intermediate level learners. The lexical information seems vast and confusing. Words appear in rich contexts, but many of the words in the contexts themselves are almost certainly unknown. The contexts are rich, varied and plentiful but they are also short, incomplete, and do not form a continuous storyline. The search procedure presupposes some well-focused questions on the part of the learner that not all students of English for academic purposes are likely to have.

The interesting information displayed in Figure 1 about the expression "to sit on one's hands" has been obtained by requesting "hands," sub-alphabetized by three words to the left of the search word and two to the right (as indicated in the bar at the top of the figure). And finally, if students made any sense of any of this information it is not clear what they should then "do" with it, other than try to remember it.

On the other hand, this forbidding-looking screenfull may in principle offer some opportunities for contextual word learning that are not present in other more conventional types of texts. First, the chopped-off lines may have advantages as well as disadvantages. Several studies including one by Mondria and Wit-de-Boer (1991) find that when learners are reading a full-length sequential text for meaning, they are typically caught up in the flow of discourse and fail to notice many of the new words they are encountering. Clearly, little flow is likely to be generated while reading concordance lines. Second, while meeting a word in several varied contexts is known to promote successful learning, even more successful learning is promoted by meeting words in varied situations in addition to varied contexts (Nitsch, 1978). A coherent text presents words in varied contexts but these tend to be limited to the few situations of principle concern to the writer, while a corpus is built from many texts and hence displays words in many more situations. Finally, the corpus and interface shown in Figure 1 are not the only ones possible. Learner corpora can be devised that limit the number of low frequency items on offer, and interfaces can be designed that presuppose less linguistic knowledge and curiosity on the part of the learner. Most important, design features can help learners focus on basic questions of word meaning and offer them something to "do" with the rich lexical information they gather.

Courseware design and implementation

The breadth-depth vocabulary problem is often most acute for academic learners in developing countries, who must use English as their medium of study but who do not use English in any other area of their lives. For example, my first-year commerce students at Sultan Qaboos University in Oman arrived at the University with a receptive vocabulary size of about 1000 words (as established by Nation's, 1990, Vocabulary Levels Test), while they needed more like 3,500 to begin academic reading, thus leaving 2,500 to be acquired in a year. Their situation is hardly atypical. Can a way be found to help such students learn something in the order of 2,500 words, fairly quickly, yet without sacrificing depth?

The first-year students' reading materials were typed and assembled into a learners' corpus, and a modified concordance interface was written to access this corpus. The interface was designed for extreme ease of use, and a frequency list of the 2,387 most common words of English (as determined by Hindmarsh, 1980) was built into it. Clicking on any word in the list produced a concordance of all the word's occurrences in the year's reading; clicking on a concordance line produced the source text with the searchword and its sentence highlighted. Figure 2 shows this interface, which was called PET•2000 in reference to the Cambridge Preliminary English Test (PET). Students were required to pass this test, which was based on the Hindmarsh wordlist, before proceeding to their subject area studies. The students' objective was to use the program to raise their vocabulary level from about 1000 to 2000 words in a single academic session.

The useful fiction, following constructivist thinking (Cobb, 1999a), was that the learners were lexicographers using concordance technology to build their own dictionaries. They were responsible for adding roughly 200 assigned words to

their cumulative dictionaries every week, and these words were tested in the classroom. In their "lexicography lab" hour, each student looked through the relevant section of the word list, identifying the words that were unknown. There were too many words to look at in the hour without making choices, so a non-optional metacognitive dimension was built into the activity. When a word was identified as unknown, the student used the concordance to search for an example sentence that made its meaning clear. Words in the contexts were sometimes themselves unknown, but with several contexts to choose from, students could use the computer to "negotiate comprehensible input."

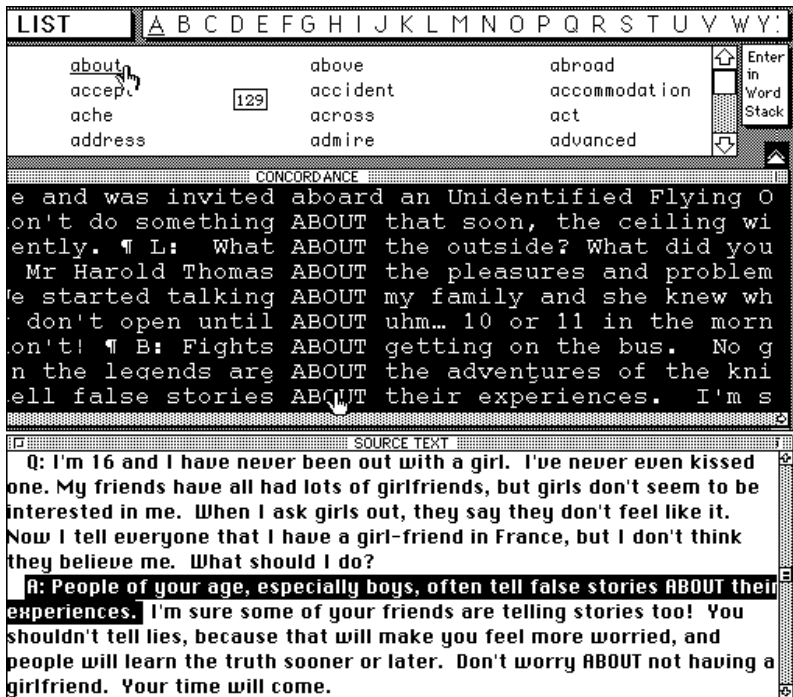


Figure 2: PET•2000 interface

When a word and one or more example contexts had been chosen, word and contexts were sent via one click to the student's database on a floppy disk (Figure 3). In the database, two things could be done with this information. There was a space for students to enter definitions if they wished, in English or another language (in this case, Arabic), and the day's cull of new words and accompanying examples could be printed up in an attractive-looking glossary (Figure 4).



Figure 3

PERSONAL WORD STACK ▪ LAILA ▪ 28/3/95		p. 18
<p><u>EMPLOY</u></p> <p style="text-align: right;">يستخدم</p> <p>They EMPLOY a consultant engineer to design it and prepare all the plans.</p>	<p><u>ENGAGED</u></p> <p style="text-align: right;">(١) خاطيب / (٢) مشغول</p> <p>Then the monster attacks and kills not only Frankenstein's friend but also his brother and the woman his brother is ENGAGED to, his brother's fiance Elizabeth.</p> <p>If you want to phone someone, it's often impossible because they're ENGAGED all the time.</p>	
<p><u>EMPLOYEE</u></p> <p style="text-align: right;">مستخدم / اجير</p> <p>The manager began to hire new EMPLOYEES who could put the idea into action.</p>	<p><u>ENGINE</u></p> <p style="text-align: right;">محرك / مائكة</p> <p>The ENGINES started to roar, the plane started to shake, and after a bumpy take-off we were suddenly up in the air over the water.</p>	
<p><u>ENCOURAGE</u></p> <p style="text-align: right;">شجع</p> <p>They ENCOURAGED women to break the law.</p>	<p><u>ESCAPE</u></p> <p style="text-align: right;">يفلت من / يهرب</p> <p>The thieves managed to get in and ESCAPE without setting off the security alarm by cutting off the electricity supply.</p>	
<p><u>ENEMY</u></p> <p style="text-align: right;">خصم / عدو</p> <p>His main ENEMY was the Sheriff of Nottingham, who was always trying to capture Robin but never managed to do it.</p>		

Figure 4

Testing the tutor

Students were assigned to learn 200 words per week for 12 weeks. Control groups used a wordlist and dictionary for this task; experimental groups made their own dictionaries with the concordance and database software. Steps were taken to ensure equal time on task. Pre-post and weekly quizzes tested experimental and control groups in both definitional knowledge as well as transfer of knowledge to a novel context (Figure 5 shows the testing format).

992 VOCABULARY QUIZ 2 (C)		Young Computer Criminals	
Part A. Recognition (1 each = 15 points)		Computers have changed our lives in many ways and made so many things easier to do. Most of us would probably agree that the computer is the most important invention of the 20th (1)..... By now most of us know how to use a computer and these days children learn how to use them at school. This sounds like a great idea. Computer games are good entertainment and educational programs are clearly very useful. But some children are able to do surprising things with computers. For example, in the United States some young computer users found a way to rob banks! These very (2)..... boys made a computer program that wrote thousands and thousands of different telephone numbers. When they (3)..... their program to the bank's telephone, the computer started dialing the numbers. The computer (4)..... dialing the phone numbers, day after day, until the bank's secret telephone number was finally found. Then the boys were able to open the files which (5)..... bank account information, and they started changing things. As you can imagine, you could easily make yourself very rich just by opening a new file and typing some numbers! Of course, bank (6)..... soon noticed that there were problems with their accounts and they started complaining to the bank manager. The police were brought in but it wasn't easy to (7)..... these criminals because no one expected the thieves to be children. In the end the police weren't sure what to do with the boys. They were too young to be sent to a (8)..... of law, and they were certainly too young to go to prison. The bank just had to find a better way to protect its files. Computer criminals have also (9)..... problems for credit card companies in the same way, and these companies have lost a lot of money. Finally, you may be interested to know that students who are good with computers have found a new way to (10)..... Again, by trying thousands of different numbers, they find their way into the computer network of their school and simply change their marks!	
Write the number of each word beside the correct definition.			
1. candle	___ conversation		
2. cabbage	___ green vegetable		
3. candy	___ teachers use it		
4. capital	___ important city		
5. chalk	___		
6. chat	___		
1. centre	___ two people or two things		
2. chapter	___ part of a book		
3. chimney	___ doctors work there		
4. church	___ Christian building		
5. clinic	___		
6. couple	___		
1. cry	___		
2. control	___ make unhappy sounds		
3. charge	___ make noise with hands		
4. cover	___ ask to pay; cost		
5. contact	___ have power over		
6. clap	___		
1. curious	___		
2. cruel	___ causing pain, hurt		
3. crazy	___ not sunny		
4. cloudy	___ always asking		
5. confident	___ questions		
6. convenient	___		
Part B. Cloze (1 each = 10 points)			
Choose from these words to complete the passage.			
continued	customers	catch	
century	court	congratulations	
cheat	clever	caused	
connected	curly	contained	

Figure 5: Format for measuring two kinds of word learning

Results

In a year of testing, a clear trend emerged. Learning large numbers of words from a wordlist and a dictionary produced strong gains in definitional knowledge in the short term. However, this knowledge was not well retained, and students using this method were not very successful at applying learned words to gaps in a novel text. But, searching through a corpus for clear examples of new words produced both definitional

knowledge and transfer of comprehension to novel texts in both the short and long terms.

More details on these tests including statistical criteria are available in Cobb (1996; 1999a). The main findings are summarized in the figures below. Figure 6 shows the result that was obtained over and over again in the testing sessions. Control and experimental groups both made substantial gains in terms of definitional knowledge (the left side of the test format in Figure 4), while only the concordance-lexicography groups made significant gains on the novel text measure (the right side).

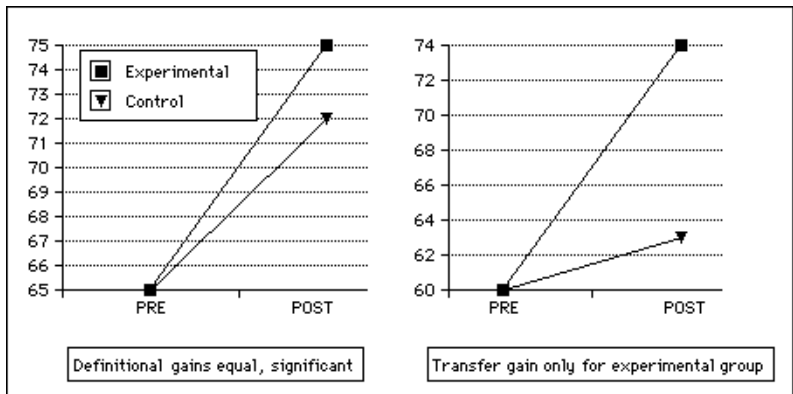


Figure 6

Further, the control groups definitional knowledge did not last long, certainly not long enough to act as a stable substrate around which further learning would form. Delayed retention tests consistently revealed that control groups did not retain their definitional knowledge, while the concordance groups if anything increased theirs with time, as shown in Figure 7.

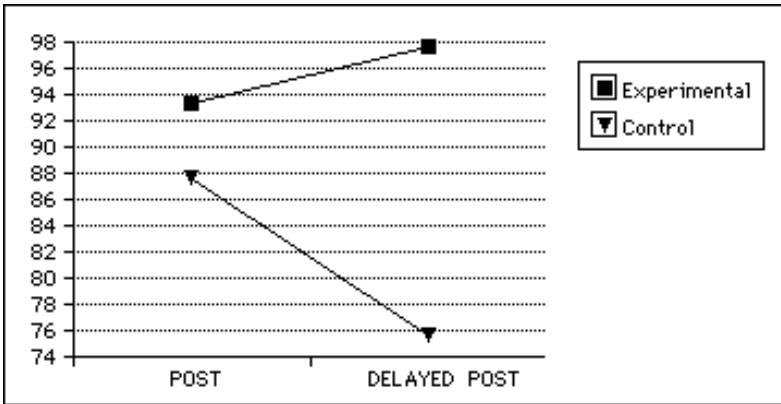


Figure 7: Delayed posttest for definitional knowledge

The corpus-based tutor, used as directed, seems to combine the benefits of list coverage with at least one important benefit of lexical acquisition through natural reading, i.e. lasting and transferable word knowledge. Several hundred students have now used PET•2000 at Sultan Qaboos University, and students regularly post-test at 2500+ words within an academic year.

Further developments

As already mentioned, the target to begin reading in an academic discipline is not 2500 but 3500 words, and accordingly corpora and wordlists are being prepared to extend the concordance approach to deal with a second tier of vocabulary. Also, as this corpus-based system is presently being used with francophone learners in Canada, wordlists and corpora are being developed that help these learners explore the Anglo-Saxon strand of the English lexicon that research has shown they are lacking (such as the word "lack"; Cobb, 1999b). At the same time, development work is under way to further deepen learners' experience with words and their contexts at whatever level or strand they are working on, particularly with regard to

giving them more to "do" with the words and contexts they have sent to their databases.

The *Hand Computer* concept

One promising idea for something more to do comes from a report by Mondria and Mondria-De Vries (1993) on using a *hand computer* for vocabulary practice. The hand computer is essentially a shoe-box divided in five compartments, bearing index cards with new words on one side and translations or short definitions on the other. Learners collect the words they want to remember, write out the cards, and then quiz themselves in their spare time. All words start out in compartment 1. To review the words, the learner shuffles the cards in a compartment and goes through them, looking at the English word and trying to recall the translation or definition, or vice-versa. If recall is successful, the card moves up one compartment, if not then down one compartment. The cards are recycled until they are all in compartment 5 (but of course new cards are entering the system all the time). Mondria and Mondria-De Vries present a convincing argument that this approach takes advantage of well-researched facts about optimal timing for the rehearsal of to-be-learned items.

However, the approach does not take good advantage of the finding that words are not optimally learned from definitions or translation equivalents, but rather from being met in multiple contextualizations. But there is no reason that Mondria's shoe box could not be computerized and attached to a concordance generating rich and varied contexts, so that the back of each card or electronic equivalent would present the learner not with definitions but contextualizations as cues.

Given that PET•2000 users have already collected in their databases the words they want to know and the contexts that make their meanings clear, an obvious further exploitation of these labours is to build some version of Mondria's five compartments into the database itself. On the student's database in Figure 3 a "Quiz" button is shown, which when

clicked unpacks the database into a set of five databases (called "stacks" since they are small Hypercard stacks). The object is to move all the words from Stack One to Stack Five through activities of increasing challenge.

In Figure 8, we see a portion of a student's screen with the five compartments or word stacks open. Words are at various stages in their journey from Stack 1 to Stack 5.

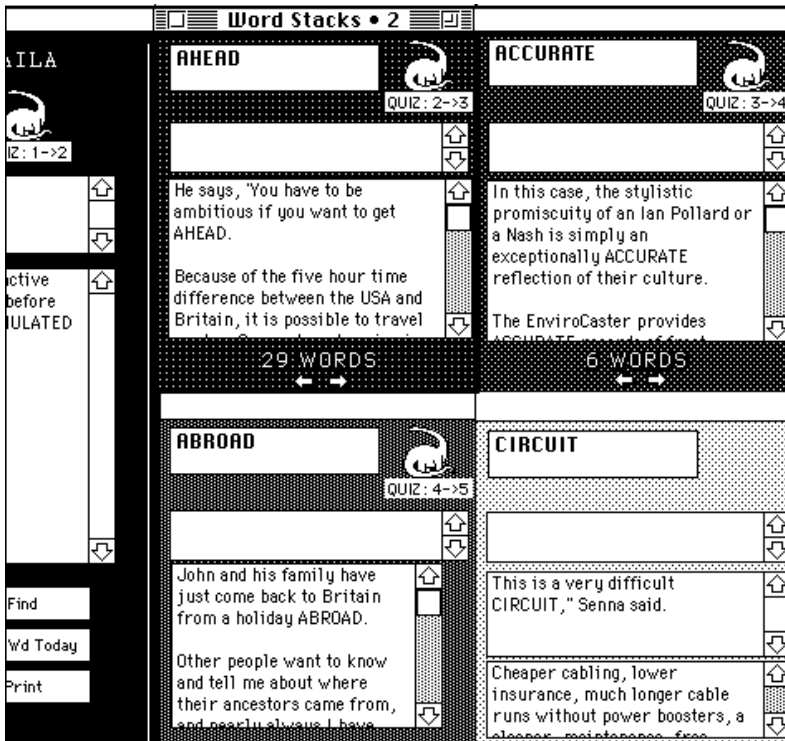


Figure 8: Traveling through the stacks

The four activities that move words up and down in the stacks are as follows.

From stack 1 to 2

The task here involves a simple reconstruction of a gapped sentence. The headword and definition disappear, the entries are put in random order, and a menu-entry button appears. The keyword is removed from each sentence, replaced by the symbol"-•-".

Holding down the entry button brings up a menu of choices, as shown in Figure 9.

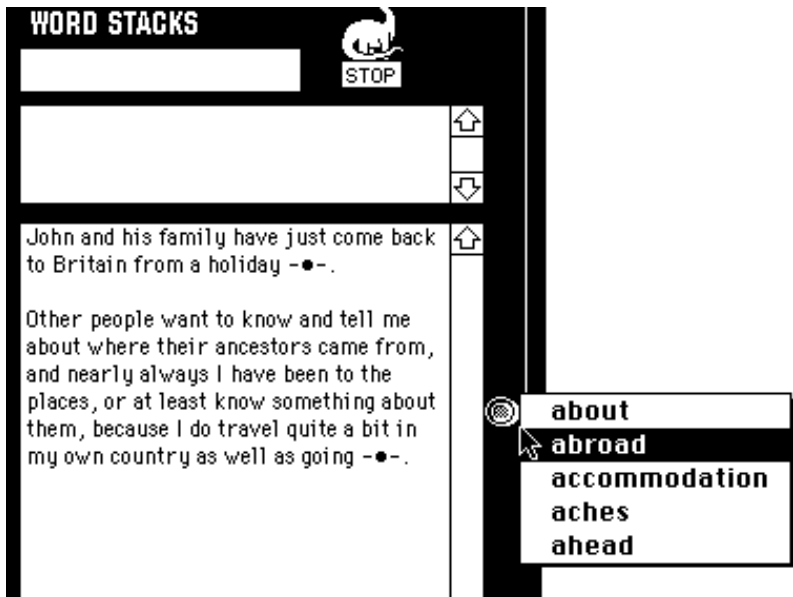


Figure 9

A correct entry sends the entire data structure (word, Arabic gloss, examples) up to the next stack; an incorrect entry sends it down to the previous stack. The idea, as set out by Mondria, is that the word in need of more practice gets it.

From stack 2 to 3

Here the task is to distinguish the hidden target word from amidst a jumble of random letters, as in Figure 10, once again with a gapped context sentence as cue (credit is due to Paul Meara for this idea).

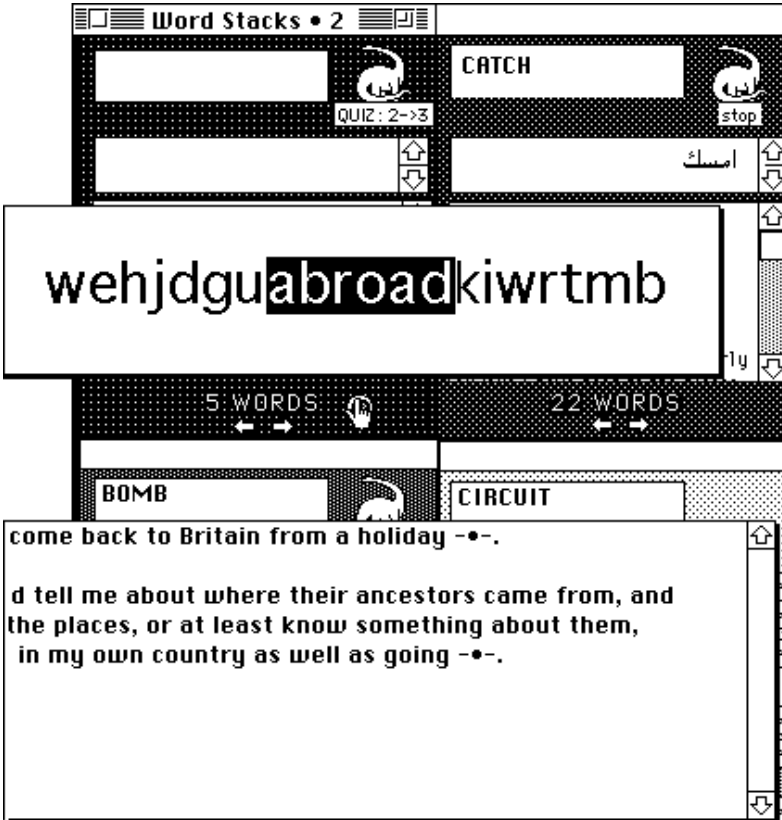


Figure 10

From stack 3 to 4

Once again the target word is cued by a context, but now also by a synthesized voice rendition of the word provided by Apple's Plaintalk technology. The input required of the learner is to spell the word correctly. A feature known as GUIDESPELL (Cobb, 1997b), allows the student to experiment with the spelling aided interactively by the computer.

In all these activities, the learner soon sees that recovering the word is easier if more than one example has been sent to the database, so some of this quiz activity should wash back to the information gathering activities discussed earlier.

From stack 4 to 5

Throughout the research and development sequence I have been describing, the test of rich word knowledge has been that the learner can supply the word to a gap in a novel context. This is the task in the fifth activity. Where does the novel context come from? Unbeknownst to the user, when a word and example were originally sent from the concordance to the database, another randomly chosen example of the word was sent along with it to hide in an invisible text field until needed. The ghost sentence rides with its data-set back and forth through the stacks. Now, on the move from Stack 4 to Stack 5 it appears, giving the student a novel context to transfer the word to.

In Figure 11, the learner is faced with a sentence requiring "abroad" that she has almost certainly never seen before (cf. Figure 9 above).

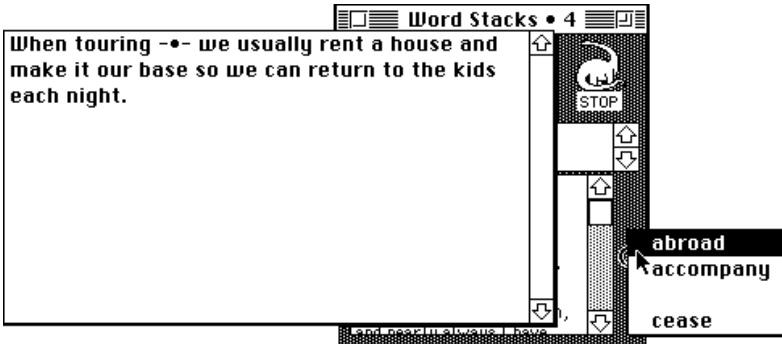


Figure 11: Transferring abroad

At the end of each stack, students get a score and are reminded of problem words, as shown in Figure 12.

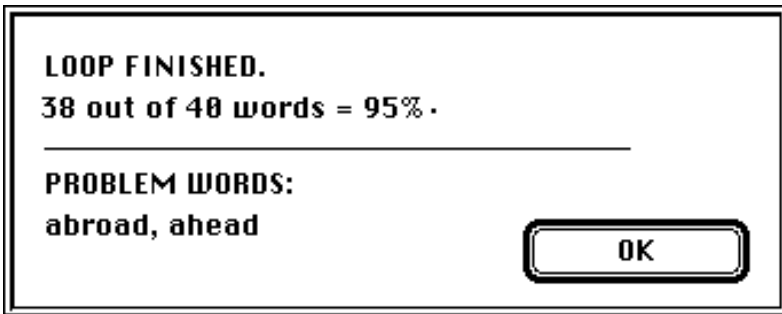


Figure 12: Stack feedback

Students can go back and forth between PET•2000 and their Personal Stacks as often as they like, and they can quit Stack activities without completing them. They can send 20 words from the concordance and then quiz themselves, or pile up 200 words from several sessions and practice them all later. Formal testing has not yet begun on this adaptation of Mondria's idea, and the interface may still be too cumbersome for use without teacher guidance.

The objective in all this work is to develop a complete set of corpus-based learning activities that will take learners through the stages of lexical growth from low intermediate up to functional reading within a discipline -- gaining broad word knowledge, in a short time, without sacrificing depth.

Launching PET•2000 onto the World Wide Web

While the theorizing, experimentation, and program code writing for this project were going on between 1994 and 1996, information technology was rapidly developing and the World Wide Web (WWW) was becoming educationally viable. Now, with the PET•2000 concept and technology shown to be of use to learners, there is no reason that it can not be made available to anyone who is interested wherever they may be, via the Internet.

An early link to the WWW (see Figure 13) was an addition to Stack 1 which, once learners had assembled examples from the concordance, allowed them to compare these to the information available in an online dictionary.



Figure 13: Linking to the Web

A web-worthy version of this tutorial system would be a comprehensive web-based lexical tutor where learners could do the following:

- test themselves to determine what frequency or other level they should be working at;
- work on any of several frequency or purpose-built vocabulary lists including eventually those of various academic or professional fields;
- meet the words in these lists in several contexts constrained to the vocabulary level they are working at;
- meet the words again in an online dictionary after they have met them in context;
- assemble contextual and definitional information in a downloadable file;
- test themselves regularly and receive advice to repeat or move forward.

The most difficult aspect of this adaptation will be to provide activities such as the five described above for further exposure to new words. However, several web-based authoring systems are being developed that make such flexibility conceivable (e.g., Hot Potatoes from the University of Victoria, British Columbia; Arneil, 1998).

With such a web-based tutor in place, learners anywhere will be able to supplement whatever vocabulary work they are doing in the classroom with a systematic exposure to the English vocabulary that is right for their needs and levels. Any number of learners can have concordance sessions with plenty "to do"!

References

- Arneil, S. (1998). *Hot Potatoes* (Development software for the WWW). [Computer software]. University of Victoria, B.C. Available: <http://web.uvic.ca/hrd/halfbaked/> [1999, October 7].
- Carroll, J.B. (1964). Words, meanings, and concepts. *Harvard Educational Review* 34, 178-202.
- Cobb, T.M. (1996). *From Concord to Lexicon: Development and test of a corpus-based lexical tutor*. Unpublished doctoral dissertation. Concordia University, Montreal. Available: <http://www.er.uqam.ca/nobel/r21270/webthesis/Thesis0.html> [1999, October 7].
- Cobb, T.M. (1997a). Is there any measurable learning from hands-on concordancing? *System* 25, 301-5.
- Cobb, T.M. (1997b). Cognitive efficiency: Toward a revised theory of media. *Educational Technology Research and Development* 45(4), 21-35.
- Cobb, T.M. (1999a). Applied constructivism: A test for the learner-as-scientist. *Educational Technology Research and Development* 47(3), 15-31.
- Cobb, T.M. (1999b). *Vocabulary Size Tests: One size fits all?* Available: <http://www.er.uqam.ca/nobel/r21270/cv/size.html> [1999, October 7].
- Hindmarsh, R. (1980). *Cambridge English Lexicon*. Cambridge University Press.
- Johns, T. (1986). Micro-concord: A language learner's research tool. *System* 14, 151-62.

Krashen, S.D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal* 73, 440-64.

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics: Abstracts* 13, 221-46.

Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research* 53, 253-79.

Miller, G.A. & Gildea, P.M. (1987). How children learn words. *Scientific American* 257(3), 94-9.

Mondria, J.-A. & Mondria-De Vries, S. (1993). Efficiently memorizing words with the help of word cards and 'hand computer': Theory and applications. *System* 22, 47-57.

Mondria, J.-A. & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics* 12, 249-67.

Nation, P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House.

Nesi, H. & Meara, P. (1994). Patterns of misinterpretation in the productive use of EFL dictionary definitions. *System* 22, 1-15.

Nitsch, K.E. (1978). *Structuring Decontextualized Forms of Knowledge*. Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN.

Scott, M. (1996). *Wordsmith*. [Computer program]. Available: <http://www.liv.ac.uk/~ms2928/> [1999, October 7].

Scott, M. & Johns, T. (1993). *Microconcord Manual: An introduction to the practices and principles of concordancing in language teaching*. Oxford: Oxford University Press.

Stahl, S.A. & Fairbanks, M.M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research* 56, 72-110.

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