# Breadth and Depth of Lexical Acquisition with Hands-on Concordancing

## Tom Cobb

# Breadth and Depth of Lexical Acquisition with Hands-on Concordancing

Tom Cobb
Université du Québec à Montréal

## ABSTRACT

One of the biggest challenges in English for Academic Purposes is to help the students acquire the immense vocabulary they need in the short time available for their language instruction. This challenge has led course developers to choose between breadth (learning from word lists) and depth (learning through extensive reading). Both methods have distinct disadvantages. Computerized concordances can help resolve the breadth–depth paradox. In this paper, the author describes how students, in effect, become concordancers, using concordance and database software to create their own dictionaries of words to be learned. This method combines the benefits of list coverage with at least some of the benefits of lexical acquisition through natural reading. The method is further enhanced by computerized learning activities based on the principle of moving words through five stacks as they are reviewed and learned.

## 1. THE PROBLEM OF L2 WORD LEARNING

One of the biggest challenges in English for Academic Purposes is helping students acquire the vocabulary they need to begin reading in a subject area. Students typically need to know words measured in the thousands, not hundreds, but receive language instruction measured in months, not years. In this time-squeeze, vocabulary course developers have to 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 viewed as unresolvable using traditional learning technologies, but computer technology suggests new possibilities.

Correspondence: Tom Cobb. E-mail: tom@uquam.ca.

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 concordance 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 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.

But 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 which 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 L2 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 L2 vocabulary acquisition is a stark one, especially as the importance of vocabulary in language development, which was neglected in the early Chomskyan era, becomes more apparent (Meara, 1980). Over the years this problem has often been noted but has typically been 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, except 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, 1997a, for a discussion of cognitive efficiency as a basis for media development).

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. My first-year commerce students at Sultan Qaboos University in Oman arrive at the university with a receptive vocabulary size of about 1,000 words (as established by Nation's, 1990, Vocabulary Levels Test), while as mentioned they need more

like 3,500 to begin academic reading, 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?

These students are more than willing to commit to memory long lists of English words glossed with Arabic definitions, and indeed have already done so for many years in school. How can the students be routed instead through multiple contextual encounters with 2,500 words? The question is particularly difficult given that inadequate vocabulary and weak reading skills limit these students to a reading diet of about two or three pages a week.

## 2. CONCORDANCES IN PRINCIPLE

It has occurred to several instructional designers that the same concordance procedure that has been successful in identifying which words to learn might also be of use in learning 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, the Omani commerce students were 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 (http://www.liv.ac.uk/~ms2928/), 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.

But the figure also shows fairly clearly why a concordance might be of limited interest to low-level learners. The lexical information seems vast and confusing. Words appear in rich contexts, but many of the words in the contexts are themselves 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 people studying English for academic purposes are likely to have. The interesting information about the expression 'to sit on one's hands' displayed in Figure 1 has been obtained by requesting 'hands' subalphabetized 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.
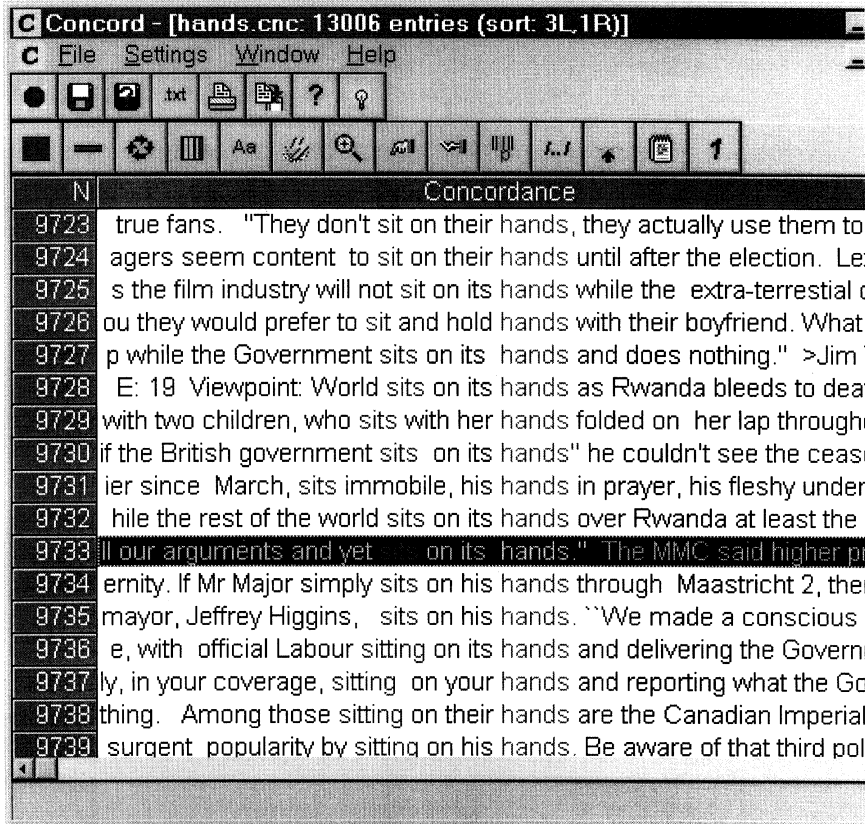
Figure 1.   *Wordsmith* screen showing how a concordance brings together lists and
                contexts.

On the other hand, this forbidding-looking interface may in principle offer
some opportunities for contextual word learning that are not present in other
more conventional text types. 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 typically get 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 suc-
cessful 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 principal 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 importantly, design features can help learners focus on basic questions of word meaning and offer them something to 'do' with the lexical information they gather.

## 3. COURSEWARE DESIGN AND IMPLEMENTATION

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 word list, before proceeding to their subject area studies. The students' objective was to use the program to raise their vocabulary level from about 1,000 to 2,000 words in a single academic session.

The useful fiction, following constructivist thinking (Cobb, in press), 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 the lexicography lab hour, each student looked through the relevant section of the word list, identifying the words that were unknown. There were of course too many words to look at in the hour without making choices, so that 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'.
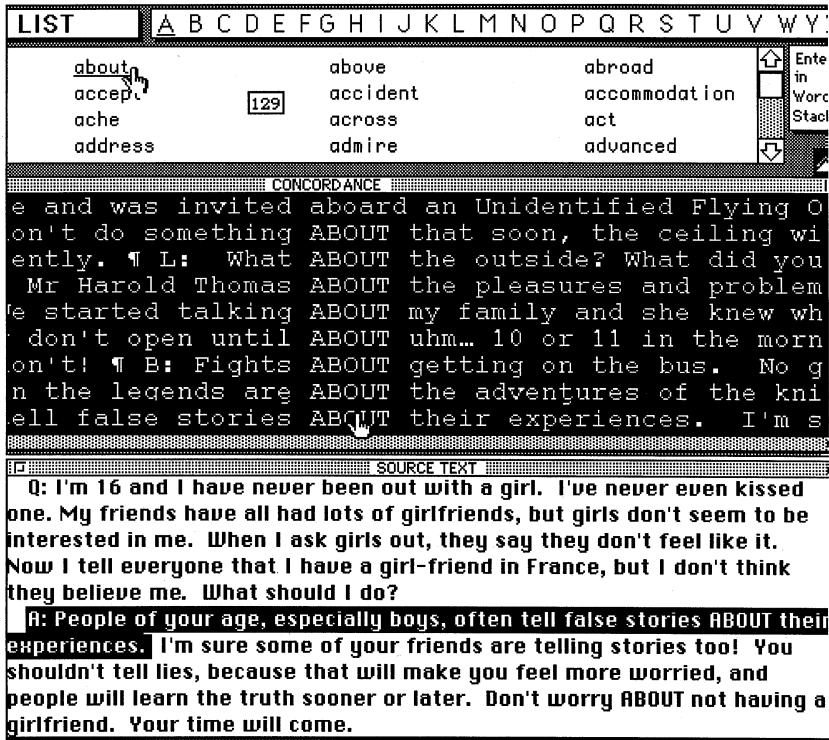
Figure 2.    PET•2000 interface.

When a word and one or more example contexts had been chosen, word and contexts were sent to the student's database on a floppy disk (Fig. 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 Arabic, and the day's cull of new words and accompanying examples could be printed up in an attractive-looking glossary (Fig. 4).

## 4. TESTING THE TUTOR

Students were assigned to learn 200 words a week for 12 weeks. Control groups used a word list and dictionary; experimental groups made their own dictionaries with the concordance and database software. Steps were taken to

Figure 3.　Personal word stack.

ensure equal time on task. Pre-post and weekly quizzes tested both experimental and control groups in both definitional knowledge as well as transfer of knowledge to a novel context (Fig. 5 shows the testing format).

## 5. RESULTS

In a year of testing, a clear trend emerged. Learning large numbers of words from a word list and a dictionary produced strong gains in definitional

PERSONAL WORD STACK ▪ LAILA ▪ 28/3/95                          p. 18

**EMPLOY**

يستخدم

They EMPLOY a consultant engineer to design it and prepare all the plans.

**EMPLOYEE**

مستخدم / أجير

The manager began to hire new EMPLOYEES who could put the idea into action.

**ENCOURAGE**

شجّع

They ENCOURAGED women to break the law.

**ENEMY**

خصم / عدو

His main ENEMY was the Sheriff of Nottingham, who was always trying to capture Robin but never managed to do it.

**ENGAGED**

(١) خاطيب / (٢) مشغول

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.

If you want to phone someone, it's often impossible because they're ENGAGED all the time.

**ENGINE**

محرك / ماكنة

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.

**ESCAPE**

يفلت من / يهرب

The thieves managed to get in and ESCAPE without setting off the security alarm by cutting off the electricity supply.

Figure 4.   Page from a student's personal glossary.

knowledge in the short term. However, this knowledge was not well retained, and students 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, short and long term.

More details on these tests including statistical criteria are available in Cobb (1996) or on Internet (at www.er.uqam.ca/nobel/r21270/cv/ webthesis.html). 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 Fig. 4), while only the concordance-lexicography groups made significant gains on the novel text measure (the right side).

Format for measuring two kinds of word learning.



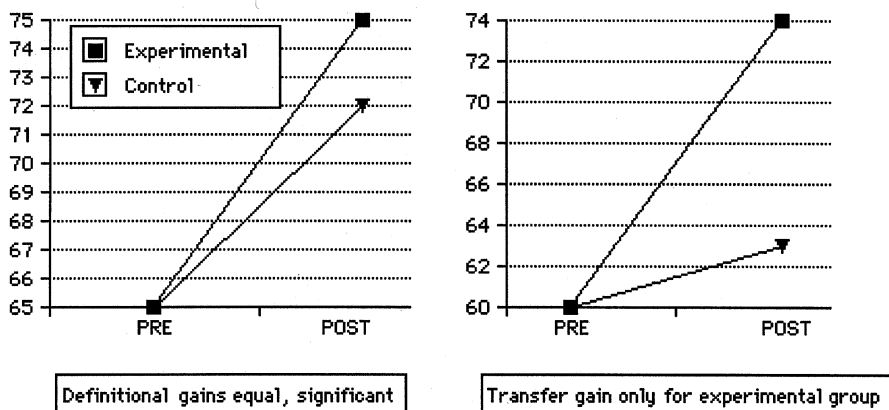Figure 5.   Format for measuring two kinds of word learning.



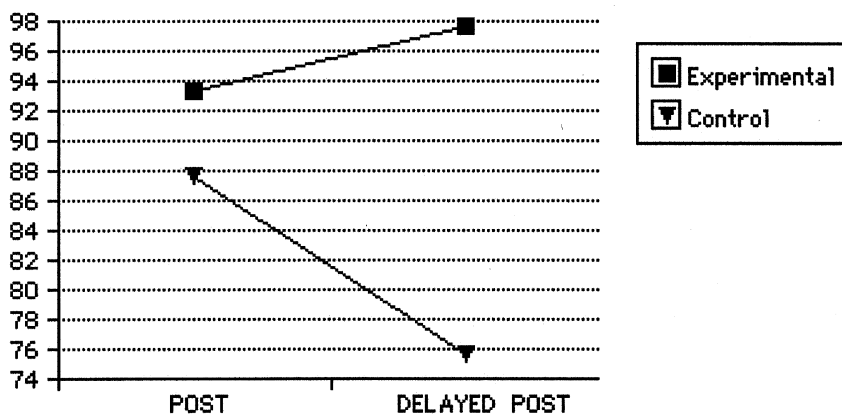Figure 6.   Static vs. transferable knowledge.

Figure 7.   Delayed post-test for definitional knowledge.

Further, the control groups' definitional knowledge did not last long, cer-
tainly not long enough to act as a stable substrate around which further learn-
ing could 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.

## 6. CONCLUSION

The corpus-based tutor, used as directed, seems to combine the benefits of
list coverage with at least some of the benefits of lexical acquisition through
natural reading (i.e., lasting and transferable word knowledge). Several hun-
dred students have now used PET•2000 at Sultan Qaboos University over two
years, and students regularly post-test at 2,500+ words within an academic
year.

## 7. FURTHER DEVELOPMENTS

As noted above, the target for reading in an academic discipline is not 2,500
but 3,500 words, and corpora and word lists will eventually be prepared to
extend the concordance approach to deal with a second tier of vocabulary. In
the meantime, development work is underway to further deepen learners'

experience with words and their contexts at the 2,500 level, particularly with regard to giving them more to 'do' with the words and contexts they have sent to their databases. For example, the students could use the contexts to cue recall of their words in some sort of flashcard activity.

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 some 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. There is no reason why 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. The four activities that move words up and down in the stacks are as follows.
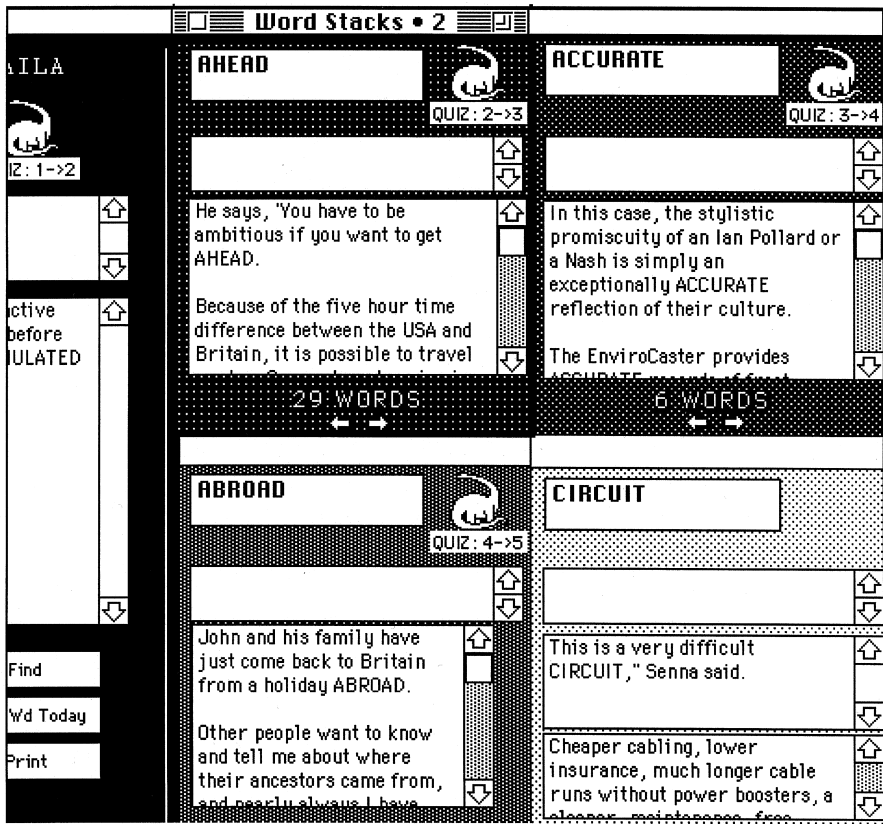
Figure 8.   Travelling through the stacks.

## From stack 1 to 2

The task here involves a simple reconstruction of a gapped sentence. The head-word and definition disappear, the entries are put in random order and a menu-entry button appears. The keyword is removed from each sentence, and replaced by the symbol '-•-'. Holding down the entry button brings up a menu of choices, as shown in 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.
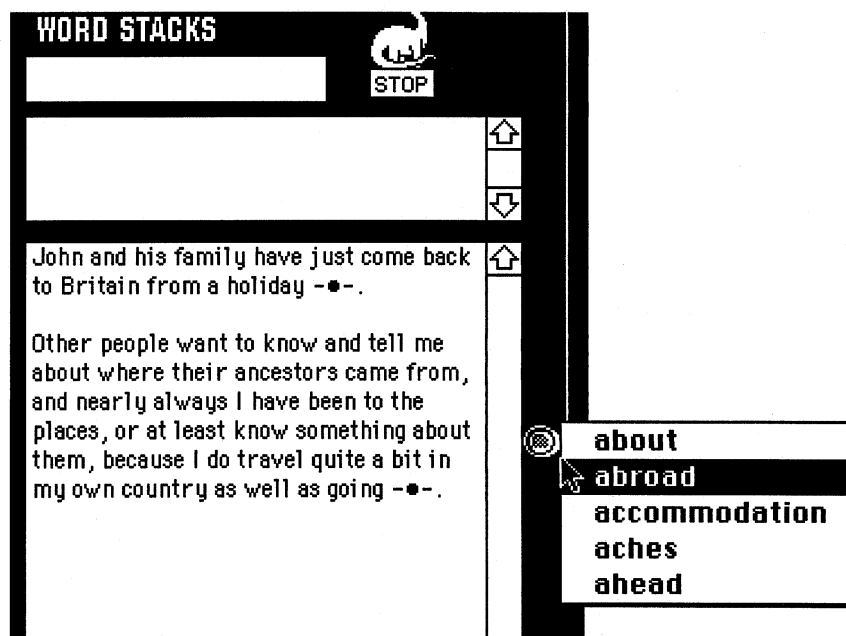
Figure 9.    Stack 1 to 2: Filling gaps in sentences chosen by learner.

### From stack 2 to 3

Here the task is to distinguish the target word from amidst a jumble of random letters, as in Figure 10, once again with a gapped context sentence as cue.

### From stack 3 to 4

Once again the target word is cued by a context but now the input 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 feed back to the information-gathering activities discussed earlier.
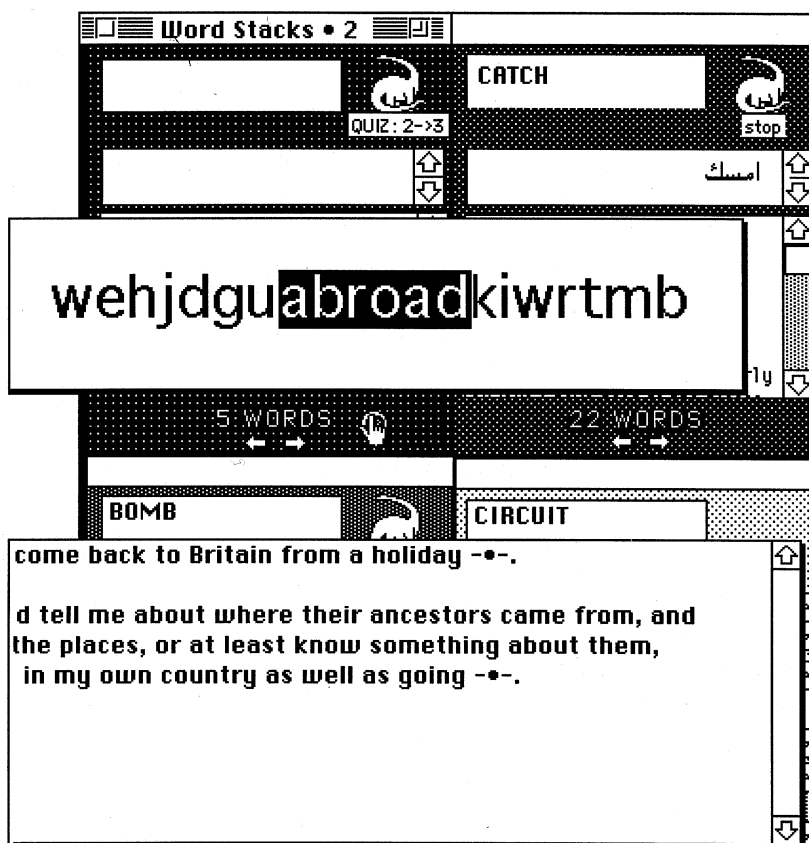
Figure 10. Stack 2 to 3: Distinguishing the target word from a jumble of letters.

**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? Unbeknown 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 (see Fig. 9).
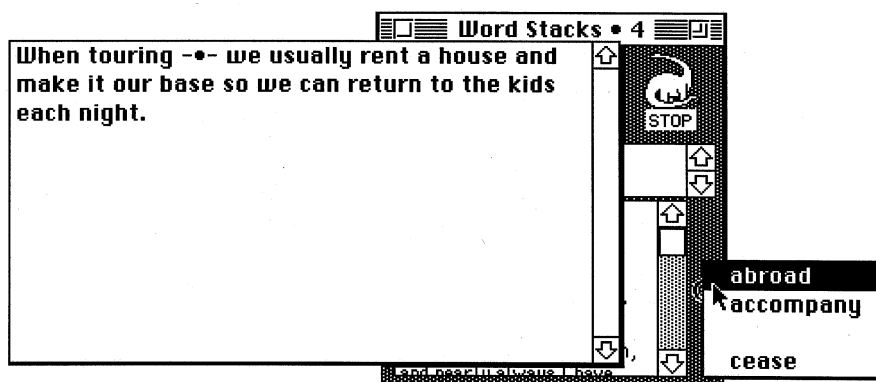
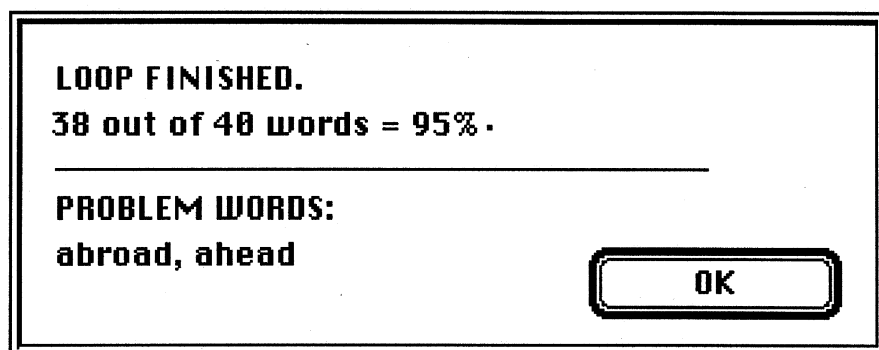Figure 11. Stack 4 to 5: Transferring abroad.



Figure 12. Stack feedback.

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

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 100 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.

## REFERENCES

Carroll, J.B. (1964) 'Words, meanings & 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.

Cobb, T.M. (1997a) 'Cognitive efficiency: Toward a revised theory of media', *Educational Technology Research & Development* 45 (4): 21–35.

Cobb, T. (1997b) 'Is there any measurable learning from hands-on concordancing?', *System* 25: 301–315.

Cobb, T.M. (in press) 'Applied constructivism: A test for the learner-as-scientist', *Educational Technology Research and Development*.

Hindmarsh, R. (1980) *Cambridge English Lexicon*. Cambridge: 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 & spelling by reading: Additional evidence for the input hypothesis', *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. *Wordsmith*. Computer program, accessible at http://www.liv.ac.uk/~ms2928/.

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 (2): 34–50.