



IS THERE ANY MEASURABLE LEARNING FROM HANDS-ON CONCORDANCING?

TOM COBB

City University of Hong Kong

This study attempts to identify a specific learning effect that can be unambiguously attributed to the use of concordance software by language learners. A base-level hypothesis for learning from concordances is proposed, that a computer concordance might simulate and potentially rationalize off-line vocabulary acquisition by presenting new words in several contexts. To test this idea, an experimental lexical tutor was developed to introduce new words to subjects, either through concordances or through other sources of lexical information. In a series of tests involving transfer of word knowledge to novel contexts, a small but consistent gain was found for words introduced through concordances. © 1997 Elsevier Science Ltd

INTRODUCTION

For more than a decade, corpus and concordance have been regularly described as one of the most promising ideas in computer-assisted language learning (Johns, 1986; Leech and Candlin, 1986; Johns and King, 1991; Hanson-Smith, 1993). Concordancing is a central idea in a proposed paradigm-shift from computer as magister to computer as pedagogue (Higgins, 1988), from a process-control model of language instruction to an information-resource model in which learners explore the language for themselves and the role of instruction is to provide tools and resources for doing so.

Oddly, however, the enthusiasm for hands-on concordancing has rarely resulted in attempts to test whether, how much, or under what conditions concordancing facilitates particular kinds or amounts of learning, particularly in comparison to traditional learning tools that are cheaper and more accessible. Even at the recent TALC96 (Teaching and Language Corpora) conference at Lancaster University, dedicated to "evaluating the claims made for the use of corpora in language instruction", none of the evaluations of hands-on activity took the form of a standard empirical study. For example, Aston (1996) reported a successful trial of the new 100 million-word British National Corpus and its SARA retrieval software with advanced language learners over ten sessions. But the research instrument was self-report, and the comparison with other learning tools suggested rather than demonstrated: "Compared with ... conventional reference instruments ... these learners reported greater success in finding solutions to problems of discourse interpretation and production" (p. 190). At some point, presumably, one would want to

confirm the learners' impressions empirically, for example comparing the success of two groups on some specified, quantified measure of learning, where one group solved language problems with conventional reference instruments (like dictionaries and grammar books) while another used corpora and concordances.

The only controlled experiment on student concordancing that I have been able to discover is a small off-line study by Stevens (1991) at Sultan Qaboos University in the Sultanate of Oman. Stevens' experimental task was to have students recall a known word to fill a gap in a text, which was either a single gapped sentence or a set of gapped concordance lines. Stevens predicted that learners would retrieve words from memory more successfully when cued by the concordance lines in spite of their chopped-off nature. His prediction was confirmed, so this was at least a proof in the limit for a facilitating effect of concordance data on some aspect of language processing and a hint of a possible role in learning.

Stevens' study appeared in a volume of learner concordancing studies assembled by Johns and King (1991), but none of the other studies in the volume ventured beyond merely describing students at various concordance activities in guided sessions. No theoretical underpinnings were explored, no falsifiable hypotheses formulated, no learning outcomes measured, no controlled comparisons attempted. When I asked Tim Johns whether he knew of any empirical study of any aspect of student concordancing other than Stevens', he replied that he did not (personal communication, 1994). Although he had "often proposed the idea to [his] graduate students", none had ever taken him up.

Some reasons for a lack of hard research can be ventured. One is that commercial concordance software does not generate user protocols, leaving informal observation the default research tool. Observation is unlikely to pinpoint exactly what a student is attempting to learn from a rich information resource like a concordance, even when an official task has been provided, and this makes it difficult to evaluate the success of the learning. Another reason is that a particularly fatal form of the internal-external paradox makes controlled studies of very novel learning media very difficult—learners must get used to a new medium over time, yet with time confounding of variables is almost inevitable, particularly in self-access settings.

THE PRESENT STUDY

The present study carries on from Stevens' (1991) study with Omani students at Sultan Qaboos University, using subjects and resources kindly provided by the same institution. The question to be answered is this: will the superiority of concordance information over a single sentence prevail, if (a) the information appears on a computer screen instead of on paper, and (b) the task is not to recall known words but to learn new ones?

The literature of vocabulary acquisition is virtually unanimous on the value of learning words through several contextual encounters, whether in a first language (Stahl and Fairbanks, 1986) or a second (Krashen, 1989; Nation, 1990). Learning a word from either a short definition or a single sentence context tends to produce inert lexical knowledge that does not facilitate the word's comprehension in a novel context, while learning a

word from several contexts, with or without a definition, tends to produce rich, transferable knowledge (Mezynski, 1983). A further question to be answered, then, is whether the several-contexts effect described in the reading literature occurs only when the contexts are naturally spaced, as they are in normal paper texts, or whether any important products of multicontextual learning, such as greater transferability to a novel text, are replicated when the contexts take the form of massed concordance lines. If they were, this would suggest a role for computers in rationalizing and shortening a learning process that, left to itself, is often protracted and haphazard (Nagy *et al.*, 1985).

Subjects

The subjects were first-year Arabic-speaking university students taking a year of intensive English in preparation for a full load of English-medium commerce subjects in a second year (such as accounting, marketing, and management information systems). Their English requirement was to achieve Band 4 on the Preliminary English Test, or PET (Cambridge, 1990), within three terms, or one academic year, a task many of them found difficult. Elementary task analysis (Cobb, 1995) identified some reasons for the difficulty, such as an incongruity between the PET's lexical base of 2387 words (the high-frequency band of the Cambridge Lexicon, Hindmarsh, 1980) and the students' average start-up vocabulary of less than 500 words (as established by Nation's (1990) Vocabulary Levels Test).

The students were aware of the vocabulary aspect of their problem; to say they were word-hungry would understate their interest in lexical acquisition. However, finding a commercial course that proposes to instruct language students in these high-frequency 2500 (or so) words of English is not simple (Cobb, 1994), in spite of a growing awareness that such a list exists (Willis, 1990: p. 46) and is something learners would benefit from knowing. Apparently the only commercial course that attempts comprehensive coverage of some version of the list is COBUILD, a three-book set normally worked through in one year and a half. With just one year to reach Band 4, these students needed some other way to get control of some significant portion of these words.

Suppose that learning 1500 new words would give these students a chance on the PET, quadrupling their vocabulary sizes from 500 to 2000 words. A study by Milton and Meara (1995) suggests how ambitious such a learning goal would be. Their study found that an average European secondary student learning a foreign language at school learned 275 new words per six-month term, or 550 per year—with the advantages of a cognate language and shared orthography. For these young Omanis, a minimum of 500 new words per term, not per year, was needed to bring the PET into range, and neither their first language nor the ambient culture was likely to be of much help.

The events described in this study took place half way through the subjects' first year, at a point when they were familiar with computers in general and CALL text-manipulation activities in particular, such as Millmore and Stevens' SUPERCLOZE (1990).

Materials: program design

The first challenge in any hands-on study is to get hands on and keep them there for a period of time. To this end, a suite of five familiar CALL-type activities grouped under the name PET•200, with a modified concordance as its main information source, was

designed and tested with more than 100 learners over an academic term in 1994. The software tutor was provided with a tracking routine that recorded all interactions.

All five activities access a 10 000-word corpus, which is simply 20 texts of about 500 words each assembled from the students' reading materials. The activities are driven by 12 wordlists of 20 words each, a total of 240 words over the term, or roughly 10% of the PET's 2387-word base. The 240 words were selected on the basis that they were unlikely to be known to the students, but likely to appear on a PET test, and occurred in the corpus at least four times. One 20-word alphabetical list per week was assigned for study in the computer lab and subsequent testing in the classroom. In the activities described below, the words are from "C-D" week.

The five activities move from easy to difficult, from word-level to text-level, and from reception to production. They present some form of concordance information at least three times for every word, in tasks where this information is needed to answer the tutor's questions. The tracking routine reveals that each learner viewed an average of 60 concordances per week, or 720 over the term.

PET•200's five activities are as follows:

Part 1: choosing a definition

The learner is presented with a small concordance of 4 to 7 lines, in KWIC format with the to-be-learned word at the centre, and uses this information to select a suitable short definition for the word from 1 correct and 3 randomly generated choices as in Fig. 1. The

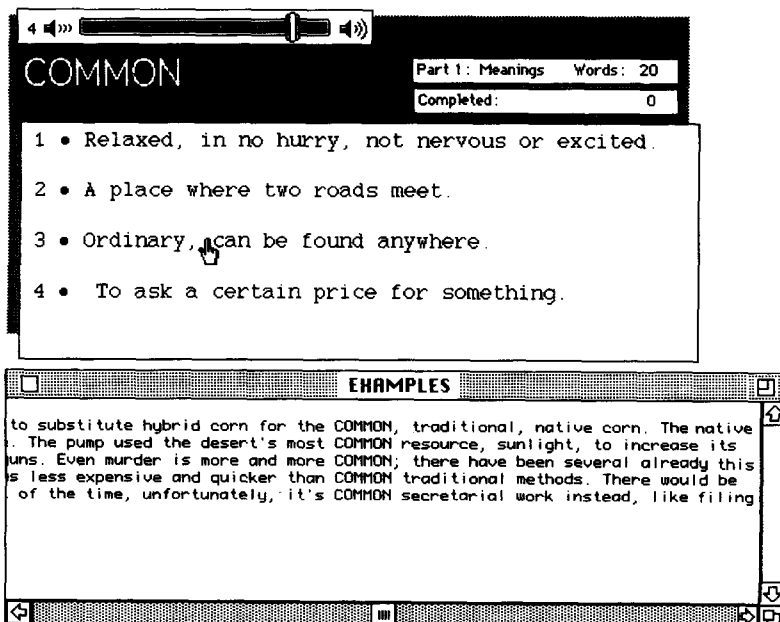


Fig. 1. Choosing a meaning.

definitions are tailored to the senses of the 240 words that happen to appear in the corpus, almost always the least marked, or most familiar, and hence most learnable sense (Kellerman, 1983).

The width of the context lines is not confined to the width of the concordance window. More context can be accessed by using the mouse on the slide control at the bottom of the window, or with the arrow keys. Also, a digitized soundbyte for each word can be heard by clicking the mouse on it. The 20 words cycle through in random order; if an incorrect choice is made, the word reappears later.

Part 2: finding words

After Part 1, the learner meets no further definitions. In Parts 2 to 5, the soundbytes and concordances, now with keywords masked, provide the basis for answers.

In Part 2, the 20 to-be-learned words again appear in random order. This time the task is to pull the target word out of a jumble of random letters, as in Fig. 2 (idea adapted from Meara, 1985). The learner drags the mouse across a string of letters, and on release finds out whether or not they make up the target word.

When the word is correctly identified, the concordance lines are filled in. As well as providing a measure of reinforcement, this visual change is designed to keep attention on the concordance window and discourage adoption of a trial-and-error strategy.

Part 3: spelling words

The 20 words once again cycle through in random order, and this time the learner is asked to type the correctly spelled word into the central space, cued by a soundbyte and a

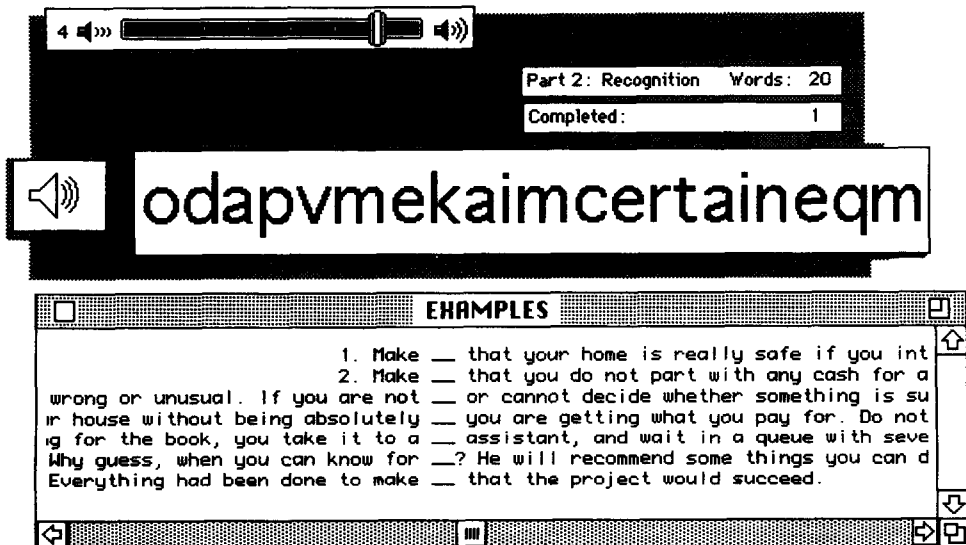


Fig. 2. Word recognition.

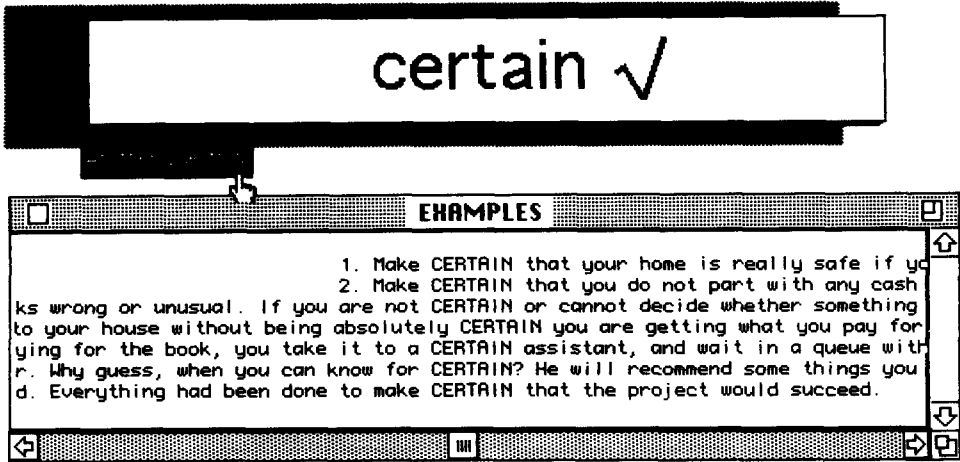


Fig. 3. Recognition feedback.

masked concordance (as in Fig. 4). A feature called GUIDESPELL helps learners shape their answers through incremental interaction. For example, if the target word is “certain” and a learner types “certin”, PET•200 responds by back-deleting to “cert” so that the learner can try again from there—as many times as necessary. Figure 4 shows the feedback following an attempt to enter “charge” as “chrg”. The tutor informs the learner that the string up to “ch” was correct, incidentally reminding a reader of unvowelled Arabic script that vowels are written in English.

Part 4: choosing words for new texts

After Part 3, soundbytes are no longer available; the activity focus changes from words to texts; and the cognitive focus changes from recall to transfer. In Fig. 5, PET•200 has gone into its corpus and found all the texts that contain a criterion number of “C” and “D” words, and masked these for the learner to replace.

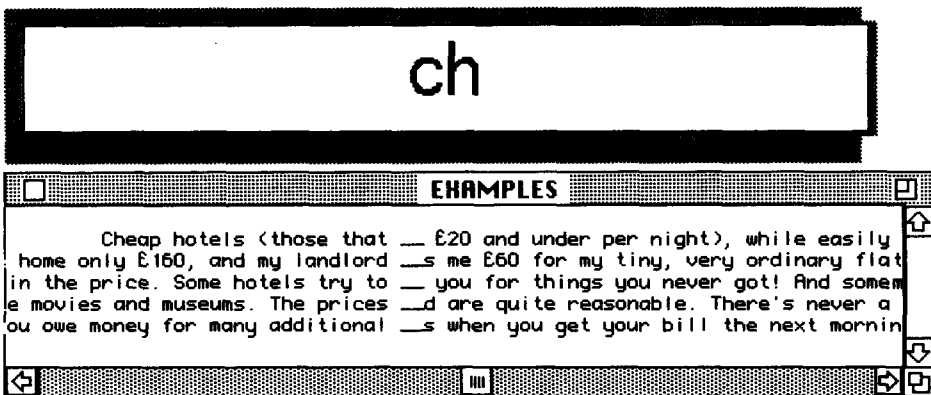


Fig. 4. Interactive spelling.

Part 4: Choose words for new texts 1 of 5 texts BLANKS = 7

Solar Pump

A solar pump was built in a small, poor, traditional desert village. The pump used the desert's most resource, sunlight, to increase its greatest necessity, water. In other words, it turned the desert into an advantage rather than a disadvantage.

Solar collectors were used to the sun's rays. Flat solar collectors were used instead of concave collectors, which are similar pieces of equipment but not exactly the same. Flat collectors can be stationary and do not have moving parts, which can be broken in sand storms, and thus they avoid unnecessary and repairs to expensive equipment. The system uses the 20 degree Centigrade temperature difference between solar collectors and the ground water to work a simple gas expansion engine which pumps water from under the ground. The pump could three times as much water as the old wells in the village - a major improvement for the whole village and all its people.

Some of the social effects of the new pumps were planned for. For

cash
 certain
 climate
 collect
 common
 damage
 deliver

Text from *Exploring Functions* (Oxford University Press, 1984).

Fig. 5. On-line transfer.

In Fig. 6, a learner has successfully replaced “common” and is about to grapple with “collect”. Various types of trail-marking help learners keep track of what they have done (used menu choices are italicized; successfully placed words are capitalized and underlined).

Predictably, the HELP available is a masked concordance of further examples of the needed word. A learner searching for “certain” might be cued by some other contexts of the word (see Fig. 7). Here again, there is a motivation for reading through the concordances.

Solar Pump

A solar pump was built in a small, poor, traditional desert village. The pump used the desert's most COMMON resource, sunlight, to increase its greatest necessity, water. In other words, it turned the desert into an advantage rather than a disadvantage.

Solar collectors were used to the sun's rays. Flat solar collectors were used instead of concave collectors, which are similar pieces of equipment but not exactly the same. Flat collectors can be stationary and do not have moving parts, which can be broken in sand storms, and thus they avoid unnecessary and repairs to expensive equipment. The system uses the 20 degree Centigrade temperature difference between solar collectors and the ground water to work a simple gas expansion engine which pumps water from under the ground. The pump could three times as much water as the old wells in the village - a major improvement for the whole village and all its people.

cash
 certain
 climate
 collect
common
 damage
 deliver

Fig. 6. Text gap-fill and feedback.

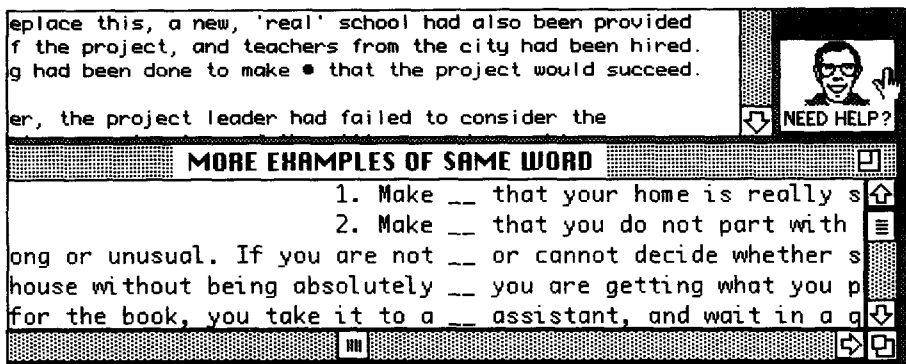


Fig. 7. Help from concordance.

Part 5: writing words for new texts

Part 5 is like Part 4, except that entry is by keyboard and words can be entered in any sequence. Word-selection is intelligent to the extent that if “day” or “deliver” is in the list of target words, then the tutor knows also to mask any plural, past-tense, or third-person “s” forms. GUIDESPELL is operative, enabling cumulative interactive reconstructions (as can be seen in the work under way in Fig. 8 on “days” and “delivered”).

Materials: designing for control

It was proposed above that a problem with concordancing research may be that in the time needed for learners to become accustomed to the medium, key learning variables are likely to undergo confounding. For example, in the present study, if PET•200 had been left in the computer lab for the 12-week run, it is unlikely that students assigned to a control group would have failed to use the tutor if they thought it would benefit them. Conversely, locking the doors and separating control and experimental groups for one or two sessions would have proven little except that the concordance was an unfamiliar medium.

A better way of establishing experimental control in CALL settings is to build two versions of a computer program and have all students use both. This idea, known as versioning, is

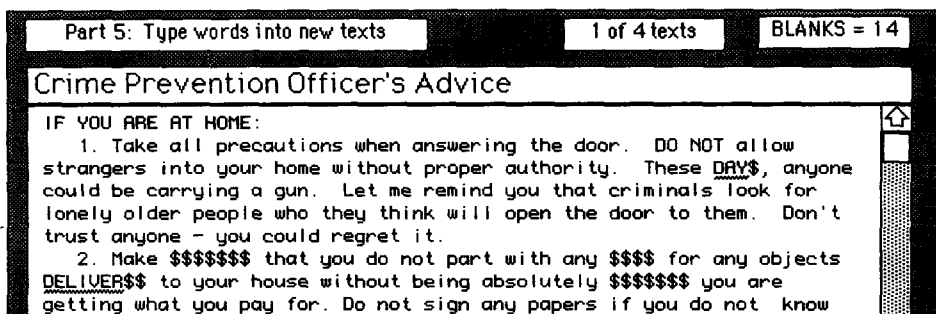


Fig. 8. Reading as writing.

discussed in Malone (1981). The two versions form a minimal-pair set that allows controlled comparisons to emerge longitudinally from a free-access resource.

After PET•200 had been pilot tested in the form already described, and the tracking system had confirmed that it could attract heavy and productive use, two versions of the tutor were then developed to run with new subjects on alternate weeks for 12 weeks. Version one was the experimental concordance version described above; version two was the same, but with example sentences and definitions where there had been concordances. For example, the initial activity in the no-concordance version is to choose a definition for a new word, cued not by multiple contexts in concordance format, but instead by a single complete sentence (as shown in Fig. 9). This difference between versions was intended to replicate Stevens' (1991) experimental distinction.

Then, for all activities after Part 1, the cue is the short definition, along with the digitized soundbyte in Parts 1 to 3. For example, in Part 3 the spelling activity is cued by the sound of the word and its definition (as shown in Fig. 10).

Everything about the two versions is identical except that the concordances are missing in the control version, so any difference in the weekly quiz results can be attributed to the presence of the concordance. It is worth pointing out in advance that with a distinction cut this fine, any gain for the concordance version is unlikely to be large, since a good deal of learning will probably take place with either version of the program. The words are met in several story-length texts in Parts 4 and 5 of either version; these texts are, of course, the source of the context lines that appear in the concordances. In other words, even in the

ANNOY

Part 1: Meanings Words: 20
Completed: 0

- 1 • To make someone feel angry.
- 2 • Describes something that makes you laugh or smile
- 3 • A small amount of anything.
- 4 • To say that someone is responsible for something bad that has happened.

EXAMPLES

But being woken up seemed to ANNOY him more than the damage to his car.

Fig. 9. Choose a definition, control version.

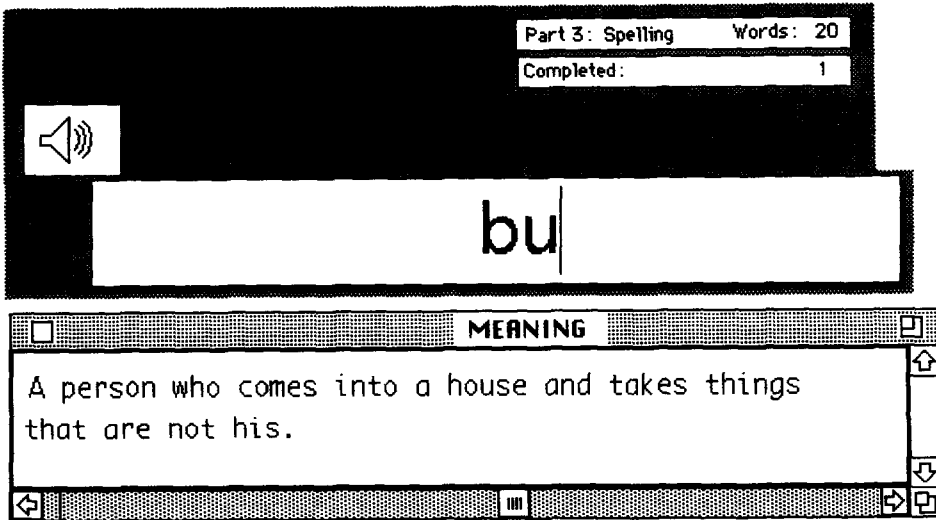


Fig. 10. Spelling, control version.

control version students have access to contextualized lexical information about the items, but not gathered together as lines of concordance. It is specifically the gathering-together feature that the comparison focuses on.

Measures

Subjects were subjected to several measures of word knowledge before, during, and after the 12-week run. They were pre-tested and post-tested with the Vocabulary Levels Test (Nation, 1990). They were given a questionnaire at the end of the term asking them to rate all their instructional materials, including specific CALL activities. They were quizzed weekly in the classroom on the words learned with PET•200. The quizzes involved two tasks, a spelling task included as a control measure, and an experimental task that had students fill gaps in a novel text with newly learned words. The lexis of the quiz texts was simplified as much as possible; the quizzes were all written prior to deciding whether to run the concordance or no concordance versions A-B-A-B or B-A-B-A fashion.

RESULTS

Just over 100 students used PET•200 half-way through their year of English studies. The endeavours of one intact group of 11 students were randomly selected for the analysis presented here.

Vocabulary levels test

The mean pre-test score for the experimental group on the 2000-level of Nation's Levels Test was 33.5% (SD 6.5), or 670 words, and their post-test score was 55% (SD 10.5), or 1100 words. This was a mean gain of 21.5% or 430 words in 3 months, far above the

European average (275 words per 6-month term). In other words, with a heavy emphasis on vocabulary in both computer lab and classroom, the students' vocabulary knowledge was growing roughly in line with PET requirements.

Materials questionnaire

PET•200 was consistently rated higher than all other published and in-house materials (mean 4.8 from 5, $n=107$), even edging out the usual area favourite—the grammar workbook.

Weekly quizzes

The weekly in-class vocabulary quiz scores reflected alternate concordance and no-concordance conditions of learning. Mean score on the six weeks without concordancing was 63.9% (SD 14.8), on the six weeks with concordancing 75.9% (SD 7.1), a mean concordance effect of 12% ($t=1.8$, $p<0.05$).

The pattern holds steady for all weekly pair-sets except one, as shown in Fig. 11. The exception occurred in week 10, which happened to be the final week of the fasting month of Ramadhan, often a period of atypical functioning in the area (see also Fig. 13).

In terms of individuals, 8 of the 11 students (73%) averaged higher scores on the text task when using the concordance version.

DISCUSSION

But was this gain caused by the subjects' use of concordance information? There are two reasons for believing so. First, as mentioned, the weekly quizzes included a spelling activity; if students had for some reason not been using PET•200 in the no-concordance weeks, then this should have produced a week-on, week-off pattern to their spelling scores as well as their novel-text scores. However, following a habituation phase these scores are steady, once again with a dip in week 10 (see Fig. 13).

Second, protocol files, as mentioned, recorded every interaction of every learner with PET•200. While these files do not record eye-movements, they do provide clues as to what subjects may have been doing while using each version of the tutor.

Table 1. Mean concordance effect by condition

	Concord No		Concord Yes	
	Week 1	40.9	Week 2	78.2
	Week 3	75.8	Week 4	78.8
	Week 5	65	Week 6	74.5
	Week 7	61	Week 8	65
	Week 9	83	Week 10	72.7
	Week 11	56.8	Week 12	86.4
Mean		63.9%		75.9%
Std Dev.		14.8		7.1

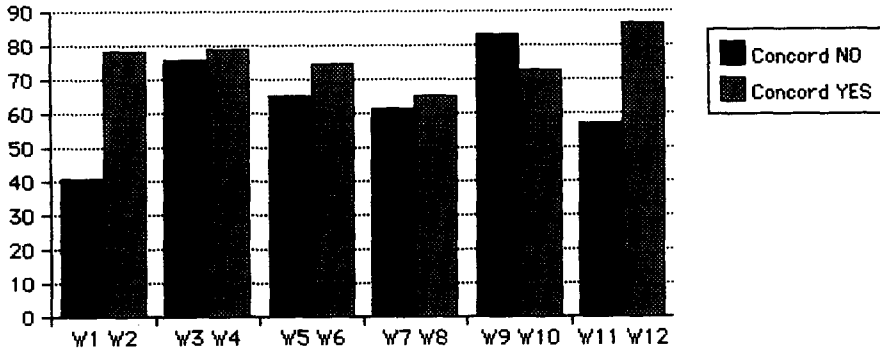


Fig. 11. Mean differences over 12 weeks.

The size of the protocol files directly reflects the number of interactions with the tutor, i.e. the number of its questions that subjects answered. So if the number of interactions was consistently lower when there were concordances to read, yet time-on-task was the same, this would suggest the extra time had been spent reading concordances. The protocol-file time logs show that the subjects spent an average of 10 h, or 600 min, using PET•200. The 600 min breakdown to 309.6 min using the no-concordance version, and 260.4 min using the concordance version represents a difference no greater than chance ($t=1.36$, $p>0.05$). But the mean size of the protocol files was 126.4 Kb (SD 49.5) in the no-concordance condition, dropping to 76 Kb (SD 44.9) in the concordance condition, a difference of about 40% ($t=2.38$, $p<0.05$).

In other words, the subjects were doing something in the concordance sessions that they were not doing in the no-concordance sessions (Fig. 14), something that helped them acquire 12% more transferable word-knowledge. It is hard not to conclude that it must have been reading concordances. It is also hard not to think that the gain would only increase as subjects became more familiar with the medium and more proficient in English.

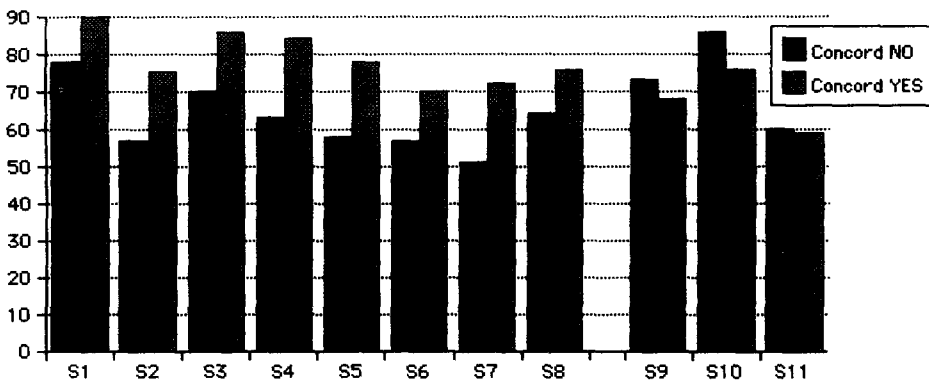


Fig. 12. Mean concordance effect by individuals.

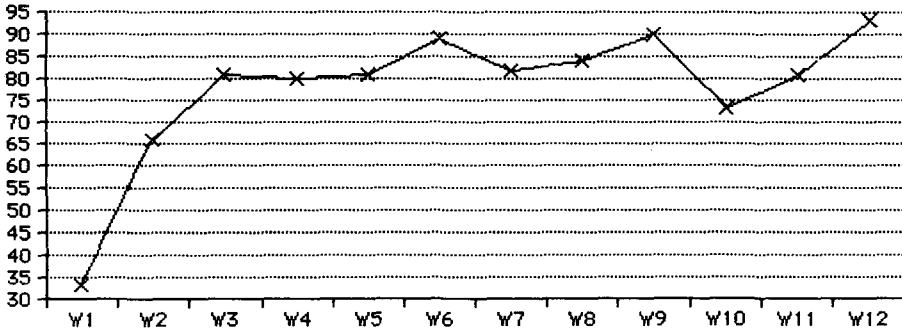


Fig. 13. Mean spelling scores over 12 weeks.

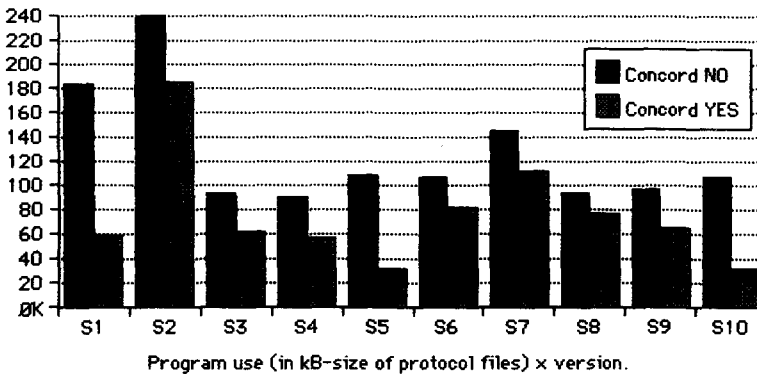


Fig. 14. A place for reading concordances.

CONCLUSION

Whether a concordance was available or not, subjects spent the same amount of time using PET•200 and got the same scores on the spelling quizzes. When a concordance was available, they answered 40% fewer of the tutor’s questions, but then achieved 12% higher scores on a novel-text task. The higher scores appear to result from the subjects’ efforts to use concordances to work out the meanings of new words.

Stevens’ (1991) off-line finding has thus been replicated on-line, over time, using new words, and in a pedagogically viable application. Further, Mezynski’s (1983) off-line finding has been broadly replicated on-line, in that multi-contextual learning whether from text or screen appears to facilitate the acquisition of transferable word knowledge. Further and more refined experiments are necessary to investigate this latter point more thoroughly.

Such experiments would be worth doing, because if important advantages of meeting words in several contexts could be shown to obtain whether the contexts were in natural

texts or on concordance screens, then concordance technology might help solve one of the toughest problems in language learning. In learning a second language, there is simply not the time, as there is in a first language, for rich, natural, multi-contextual lexical acquisition to take place. The usual prescription for this problem is that language learners should "read more" (Krashen, 1989), but it is doubtful that the necessary time actually exists for lexical growth through reading to occur to any useful extent. Long ago, J. B. Carroll (1964) expressed a wish that a way could be found to mimic the effects of natural contextual learning, except more efficiently; the way may be some version of concordancing.

Work is currently under way on an expanded lexical tutor, to be called PET•2000, which will access a more extensive corpus, raise the learning target, and explore more thoroughly the link between concordance and transfer. Also, the tutor's interface will be redesigned to profit from a suggestion raised above, that learners benefit less from answering a computer's questions than they do from having a computer answer theirs.

REFERENCES

- Aston, G. (1996) The British National Corpus as a language learner resource. In *Proceedings of Teaching and Language Corpora, 1996*, eds S. Botley, J. Glass, T. McEnery and A. Watson, pp. 178–191. University Centre for Computer Corpus Research on Language, Lancaster.
- Carroll, J. B. (1964) Words, meanings, and concepts. *Harvard Educational Review* 34, 178–202.
- Cobb, T. M. (1994) Which course prepares students for the PET? Research Report, Language Centre, Sultan Qaboos University, Oman.
- Cobb, T. M. (1995) Imported tests: analysing the task. Paper presented at TESOL Arabia, Al-Ain.
- Hanson-Smith, E. (1993) Dancing with concordances. *CÆLL Journal* 4(2), 40.
- Higgins, J. (1988) *Language, learners and computers: human intelligence and artificial unintelligence*. Longman, London.
- Hindmarsh, R. (1980) *Cambridge English Lexicon*. Cambridge University Press, Cambridge.
- Johns, T. and King, P. (eds) (1991) *Classroom concordancing: English Language Research Journal*, 4, Centre for English Language Studies, University of Birmingham.
- Johns, T. (1986) Micro-concord: a language learner's research tool. *System* 14(2), 151–162.
- Kellerman, E. (1983) Now you see it, now you don't. In *Language transfer in language learning*, eds S. Gass and L. Selinker, pp. 157–176. Newbury House, Rowley, MA.
- Krashen, S. D. (1989) We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis. *The Modern Language Journal* 73, 440–464.
- Leech, G. and Candlin, C. N. (1986) *Computers in English language teaching and research*. Longman, London.
- Malone, T. W. (1981) Toward a theory of intrinsically motivating discussion. *Cognitive Science* 4, 333–369.
- Meara, P. (1985) Lexical skills and CALL. In *Computers in English language teaching: a view from the classroom*, eds C. Brumfit, M. Phillips and P. Skehan, pp. 83–89. Pergamon Press, Oxford.
- Mezynski, K. (1983) Issues concerning the acquisition of knowledge: effects of vocabulary training on reading comprehension. *Review of Educational Research* 53(2), 253–279.
- Millmore, S. and Stevens, V. (1990) Super Cloze 2.0. Shareware available through CALL Interest Section, TESOL.
- Milton, J. and 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., Herman, P. A. and Anderson, R. C. (1985) Learning words from context. *Reading Research Quarterly* 20(2), 233–253.
- Nation, P. (1990) *Teaching and learning vocabulary*. Newbury House, New York.
- Stahl, S. A. and Fairbanks, M. M. (1986) The effects of vocabulary instruction: a model-based meta-analysis. *Review of Educational Research* 56(1), 72–110.

Stevens, V. (1991) Concordance-based vocabulary exercises: a viable alternative to gap-fillers. In *Classroom concordancing: English Language Research Journal*, 4, eds T. Johns and P. King, pp. 47–63. Centre for English Language Studies, University of Birmingham.

University of Cambridge (1990) *Preliminary English Test*. Local Examinations Syndicate: International examinations.

Willis, D. (1990) *The lexical syllabus*. Collins Cobuild, London.