Title of project

**A comparison of computerized and traditional techniques for learning academic vocabulary**

Type of grant application
**Priority Research Grant**

Applicants

**Marlise Horst**, Concordia University, Montreal
Principal Investigator

**Diane Schmitt**, Nottingham Trent University
Co-investigator

**Tom Cobb**, Université de Québec a Montréal, dép de linguistique et de didactique des langues
Co-investigator

**Detailed proposal**

The problem
In many countries around the world, university students must read and write in English in order to achieve educational and professional goals. But at the same time, they often arrive at the university inadequately prepared for studying in English. For such learners, reading materials designed for native speakers is laborious and slow, and writing academic essays and reports in the second language is even more challenging. Preparing students who are short on the basic proficiency needed to succeed in university courses and shorter still on time to address the deficit poses urgent questions for English for Academic Purposes course design: How can instruction be tailored to meet these students' needs most effectively and efficiently? And given the wide disparity in needs, can computer technology provide a solution?

A useful answer to the question of what academic learners of English would do well to study has already been provided by technology. Second language (L2) vocabulary acquisition researchers have drawn on large computerized corpora of academic texts and the tools of corpus linguistics to identify the words that occur most frequently in academic texts across a wide variety of disciplines (e.g. Xue & Nation, 1984). This efficiency-based approach has resulted in the the Academic Word List (Coxhead, 2000), a set of 570 word families that includes words such as *focus*, *phenomenon* and *perceive*. Research shows that adding knowledge of this subtechnical vocabulary to a basic English lexicon of 2000 frequent words results in high levels of coverage of the words that occur in academic texts, such that learners are well positioned to comprehend course readings (Nation, 1990, 2001; Sutarysyah, Nation & Kennedy, 1994) and to express abstract concepts in writing (Cobb & Horst, in press). Study of the AWL is now in important part

of the curriculum in EAP courses around the world; recent textbooks (Obenda, 2004; Schmitt, in preparation) testify to the enthusiasm for this approach.

Technology may also be able to provide an answer to the question of how students can learn AWL vocabulary efficiently. One of the most promising tools for computer assisted vocabulary learning so far is concordancing (Johns, 1986). From a theoretical perspective, putting concordancing tools in the hands of learners is consistent with the current interest in data-driven learning (Hunston, 2002). From a practical perspective, a concordancer linked to a useful corpus is a rich resource that is difficult to replicate in traditional classrooms with chalk and paper. Good classroom teachers may add a sentence example when they present the meaning of a word like *perceive,* and perhaps also ask students to find an additional example in a learners' dictionary or encourage them to note uses as they encounter the word again in their reading. But this cannot match the concordancer's ability to search millions of words of text for instances of a target word, locate many examples, and display them to learners in a variety of presentation and quiz formats format -- all in a matter of seconds.

However, a conclusive experimental case for the efficicacy of concordance-based activities in learning L2 vocabulary has yet to be made. The few studies that have investigated this type of activity and identified a learning advantage usually involve small groups of subjects in supportive, resource-rich environments where learners have easy access to computers. Measures used provide a limited picture of what students can do with newly acquired word knowledge, and comparisons to outcomes in groups where learners studied the same material using other methods are usually lacking. Finally, we are not aware of any research that speaks to resource-poor environments where teachers have some (often limited) access to computers and the Internet but most of their students do not. Since this is a reality confronting many teachers in EFL contexts worldwide, this is an important issue to investigate. The research we propose is designed to address these theoretical and practical challenges.

## Theoretical and research background

What are the theoretical arguments for an instructional technique that involves university students in examining concordance lines for a word, attempting to infer its meaning, and then checking the guess? If we look to L1 research vocabulary acquisition, a prominent perspective is that of Nagy, Herman and Anderson (1985) who advocate an input-based model. They argue that an adult-size mental lexicon is largely acquired through reading; form-meaning associations are built up gradually as readers meet new words in context and infer their meanings. A strong version of this position has been extended to L2 vocabulary acquisition by Krashen (1989, 1993) and others. Nagy (1997) argues that many varied contextual exposures are critical for building the complex associative network that enables L2 learners to access meanings as words are met again in ever novel contexts. Since even fairly common words may occur only a few times per million running words of text (Schmitt, 2000), this clearly entails a great deal of reading -- something that many L2 learners may be reluctant to do. Reading context lines gathered by a concordancer instead may be able to offer a convenient shortcut through this process.

A more moderate theoretical position is held by L2 vocabulary acquisition researchers such as Nation (1990, 2001), Paribakht and Wesche (1997) and others, who see direct vocabulary instruction as a valuable complement to the process of learning words through contextual encounters. Notable in the quest to understand characterisitics of tasks that enhance the effects of contextual learning is the work of Hulstijn and Laufer. Drawing on learning theory and the construct of depth of processing (Craik & Lockheart, 1972) in particular, they have posited the Involvement Load Hypothesis (Hulstijn & Laufer, 2001; Laufer & Hulstijn, 2001) which contends that learners are most likely to acquire and retain vocabulary if learning tasks feature high levels of three operationalizable components: *need, search* and *evaluation*. If we apply this paradigm to the task of inferring meanings from concordance lines presented to L2 learners as an activity on paper, it is clear that *need* (a motivational construct) is low since the task is teacher imposed rather than student initiated; but if learners initiate their own concordance quests (as advocated by Cobb, 2000 and Johns 1994), this factor is increased. Thus an on-line concordance quest scores high on the *search* component, especially if the learner also verifies the correctness of a guess in an on-line dictionary. And, if the learner looks for fit between the dictionary definition and the uses of a target word in concordance lines, then *evaluation* comes into play as well.

Finally, asking learners to complete concordancing exercises that highlight vocabulary they have met previously in course readings (the activity type we prepose) is consistent with currently held views on the role of form-focused activities within a larger context of language instruction that is meaning-focused. There is also congruency with Chapelle's "meaning focus", "authenticity", and "positive impact" criteria for designing research informed computer activities (2001, pp. 55-57) .

Thus there are strong theoretical arguments for encouraging learners to do computerized concordancing tasks, but what do experimental studies say about the efficacy of this technique for acquiring new L2 vocabulary knowledge? Several studies have shown that language learners can use concordance examples to good effect for error correction (Gaskell, 2002; Ng & Burton, 2001), and translation (Bowker, 1999). The question of lexical acquisition has been also been addressed in a studies by Cobb and Horst (Cobb 1997, 1999; Cobb & Horst, 2001). Overall, this research points to an advantage for concordancing over other computerized vocabulary activities; for instance, Cobb (1999) found that concordance users were more able to transfer newly acquired word knowedge to novel contexts than definitions users. However, this work was conducted with small groups in resource-rich settings, and did not investigate comparable off-line learning activities. There is clearly a need to build on these promising research precedents in carefully designed experiments with larger numbers of participants in more varied settings. The research we propose will address these experimental design issues and investigate the following questions:

1. When the quality of two instructional activity types is similar but one draws on computerized resources and the other does not, is there an advantage for the computer-mediated activity?

2. When the type of instructional activity is the same but the mode of delivery is different, (on-line vs paper worksheet), is there an advantage for computer delivery?

3. How does availabilty of computers impact learning results?

## Methodology

The proposed experimentation involves engaging learners in vocabulary study tasks in three conditions: concordance-based activities on paper, concordance-based activities on-line, and definitions-based activities on paper. Following Cobb's (1997) versioning design, all learners will study words in the three conditions. The context for the study is EAP courses at universities in Canada and Britain, where students will use a textbook (Schmitt, in preparation) to study AWL vocabulary as part of their regular coursework. The experimental activities are a supplement to these course materials to aid students in preparation for regular quizzes. In Phase 2 of the research, we would like to test the materials and support activities with EAP students at a university in a non-English speaking setting where students have limited access to computers. Additional funding will be sought to to support this plan.

### Participants

Four EAP classes of approximately 25 students each will be recruited at each university, amounting to a total of 8 groups and approximately 200 students in Phase 1. Similar numbers will be involved in Phase 2. The researchers will use a measure of receptive vocabulary size (Schmitt, Schmitt & Clapham, 2001) to identify groups of students whose scores on the AWL section of the test indicate that they stand to benefit from studying AWL vocabulary. Scores on this measure will also allow us to identify comparable groups of learners at the two universities. A large proportion of the students in Canada and Britain will be of non-Latin language background (i.e Arabic or Mandarin Chinese speakers); the largely Greco-Latin AWL vocabulary is especially challenging for these learners.

### Materials

The textbook to be used in the participant EAP groups is divided into eight units; each unit presents reading passages that feature about 70 AWL items along with activities designed to support learning of the target words.

The experimental activities that offer additional support consist of concordance-based activities on paper, concordance-based activities on-line, and definitions-based activities on paper. The central task in the concordance-based activities presents learners with several lines of lines of text, each featuring a different use of the targeted word (see example below. In the paper concordancing condition, students can use dictionaries to check the correctness of guesses; the on-line condition provides a computer link to an on-line dictionary. Activities in the definitions-based condition (on paper) will include techniques recommended by Sökmen (1997) such as linking target words to synonyms and definitions and building semantic maps.

Which Unit 5 word fits in all of the spaces?

The viewer ___s a big-screen TV image floating in space.

The enemy had not ___ed them and such was the darkness that they could not see each other.

His secret discussions about the plans are ___ed by the journalists to be unjust.
Ne Win, Burma's ruler since 1962, ___ed her as a threat to his leadership.

The cold was intense. The survivors suffered cruelly, but they scarcely ___ ed it.

(Answer = perceive)

An important aspect of the proposed research is the development of a purpose-built corpus of modified academic texts on which concordance searches for AWL items can be based. In our work, a common complaint has been that students find concordance lines derived from searches of authentic texts (e.g. the Bown corpus) too lexically dense to comprehend (Cobb, 1997; Horst, Cobb & Nicolae, under review; Horst & Plomer, 2004). In the exercise above, concordance lines have been edited slightly to consist entirely of words in one of the following categories: the 2000 most frequent word families of English (West, 1953), the AWL (Coxhead, 2000), or lexically transparent proper names.

Testing

In order to test the quantity and quality of word learning in a valid manner, a measure that includes both definitional and contextualized tasks will be administered at the beginning and end of the session. This pre-post instrument will sample the entire 570-item list and emphasize receptive knowledge (e.g. recognizing synonyms or appropriate contextual uses). Students will also take in-class quizzes as part of their regular coursework. To assess learners' ability to put AWL knowledge into active use, samples of essays produced at four intervals will be collected and analyzed using lexical frequency profiling tools (see Cobb's *Vocabprofile* available at http://www.lextutor.ca/ and Laufer & Nation, 1995); students' performance in content courses will also be recorded. Participants will be asked to complete survey questions about access to the internet, their use of the on-line study resources, and other study strategies.

Procedures and data analysis

Over the course of the EAP course, students will work their way through the eight units of the AWL textbook, with each unit supported by activities presented in one of the three conditions. Thus activities that support the first unit will be concordance-based and on paper, those accompanying the second unit will be concordance-based and available for independent work on-line, those accompanying the third unit will be definitions-based and on paper, while the fourth unit will have no accompanying support activities (to provide a baseline for comparisons). This sequence will be repeated for the remaining four units. The receptive knowledge measures administered at the beginning and end of the course will contain equal proportions of AWL items studied in each condition. Students will take a similar measure as a delayed post test one month after finishing the course.

Answering the first question about instruction with and without computer-mediation involves comparing pre-post gains on knowledge of words studied in the concordance-based paper and definitions-based paper conditions. The second question that addresses the issue of how activities are delivered can be answered by comparing pre-post gains on words that were studied in the two corcordancing conditions: one teacher-produced and on paper, and the other student-initiated and on-line. Survey responses will be used to exclude data from participants who did not have access to computers. To answer the third question about the impact of computer availability, we will look for a relationship between reported availabilty levels and test performance on items that were supported by on-line activities.

## Implications of the research

The research is positioned to provide clear information about both learning and cost effectiveness. It is specifically designed to test whether one teacher with a single computer can offer her students the benefits of data-driven learning with effects comparable those achieved by students in settings where computer access is commonplace. The study materials developed in the research present teachers and students with research-informed approaches to learning L2 vocabulary. Teachers and learners of academic English who are interested in data driven learning but frustrated by the lexical density of material in most available corpora will find the learner corpus produced for the project a useful resource.