



Higher Education Research and Development Society of Australasia, Inc

Higher education in a changing world

Research and development in higher education

Proceedings of the

28th HERDSA Annual Conference

3-6 July 2005

Sydney, Australia

Cargill, M. & Adams, R. (2005) Learning discipline-specific research English for a world stage: A self-access concordancing tool?, in *Higher education in a changing world, Proceedings of the 28th HERDSA Annual Conference, Sydney, 3-6 July 2005: pp 86.*

Published 2005 by the
Higher Education Research and Development Society of Australasia, Inc
PO Box 27, Milperra, NSW 2214, Australia
www.herdsa.org.au

ISSN: 0155 6223
ISBN: 0 908557 62 0

This research paper was reviewed using a double blind peer review process that meets DEEWR requirements. Two reviewers were appointed on the basis of their independence, expertise and experience and received the full paper devoid of the authors' names and institutions in order to ensure objectivity and anonymity. Where substantial differences existed between the two reviewers, a third reviewer was appointed. Papers were evaluated on the basis of originality, quality of academic merit, relevance to the conference theme and the standard of writing/presentation. Following review, this full paper was presented at the international conference.

Copyright© 2005 HERDSA and the authors. Apart from any fair dealing for the purposes of research or private study, criticism or review, as permitted under the Copyright, Design and Patent Act, 2005, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers at the address above.

Learning discipline-specific research English for a world stage: A self-access concordancing tool?

Margaret Cargill

The University of Adelaide, Australia
margaret.cargill@adelaide.edu.au

Ray Adams

The University of Adelaide, Australia
ray.adams@adelaide.edu.au

Abstract: *Researchers and postgraduate research students who use English as an additional language (EAL) often struggle to gain command of the highly discipline-specific English they need to write independently for successful publication in the international literature. A self-access language-learning tool that takes advantage of their existing facility with research processes and provides access to specific language elements they need to learn would be of benefit. A potential has been recognised in the literature on computer-aided language learning, but not widely taken up to date, for using Concordancing software for such a purpose. Concordancers allow a selected set of texts (corpus) to be searched for systematic evidence of how particular language items are used. This showcase session demonstrates a concordancing program that is simple and functional enough for novices to use on their own, and the steps in constructing a corpus from published journal articles. We also discuss outcomes of our initial trials of the package with EAL research students in the fields of applied linguistics and agricultural sciences. Session participants at the conference will be invited to formulate their own questions about English usage for on-the-spot investigation using the concordancer.*

Keywords: *corpus linguistics, concordancing, academic writing*

Introduction

International students enrol in a wide variety of academic disciplines at English language universities, and many are involved in high level research, requiring the writing of a thesis and often the publication of articles in English-language refereed journals. Once embarked on a career, researchers face continuing pressure to publish regularly, and to present at international conferences, likely also to be conducted in English. Students and researchers with English as an additional language (EAL) thus face not only the challenges of general and academic English, but those of the terminologies and structures inherent in the language of their chosen disciplines. The structures and patterns of such language usage have been researched for many years using concordancing, and this tool has also been used over recent decades for developing materials to assist learners in the observation of these structures and patterns (e.g. Tribble & Jones, 1997, Thurstun & Candlin 1993). The work of Johns (1986, 1991, 1994), Stevens (1995) and

others has focused the use of this software on the teaching of structures and patterns within academic English for EAL students. The development of concordancing software for use on personal computers has meant that the tool is now readily available and extremely fast in its output, which in its most common form presents each instance of the search term found within the corpus as a single line of text with the search term in the middle of the line (see Figure 1 for an example).

The above results for site 1 suggest that the longer extraction time is due to a high organic C content. We suggest that this resulted from an increase in the parents. Previous reports suggest that *S. perennis* is typically (Todd, 2001). RAPD markers, however, suggest that there may have been several planted treatments. This would suggest that the presence of decomposers in the field application. This would suggest possible detrimental effects (Schlesinger et al. (1990) suggest that any process that leads to

Figure 1: KWIC (key word in context) concordance output from a search for the word 'suggest' in a corpus of science research articles.

Concordancing has been described as a means of shifting the computer assisted language learning paradigm from one of a process-control model 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' (Cobb, 1997). This exploration provides data to serve as the basis for inductive learning of language (Todd, 2001), or data-driven learning. For example, the small number of examples in Figure 1 could suggest to a student that the most common structure after 'suggest' in science texts is a clause beginning with 'that'.

A well-known user of this approach is Tim Johns (1986, 1991), from the University of Birmingham, who used concordancing output in individual consultations with EAL students to show how words and phrases that they were having difficulty with were used in texts written by accomplished writers. He applied the term kibbitzing to the overall process, borrowing a word used to describe the practice of novice chess players watching over the shoulders of masters in the chess clubs of Europe. A collection of the outcomes of these sessions is available at <http://www.eisu.bham.ac.uk/Webmaterials/kibbitzers/> (see Kibbitzer 1 for a good demonstration of the use of concordancing output to solve a particular language query).

However, in spite of the work of a growing number of enthusiasts, the software has often remained in the hands of language researchers and educators as a means of developing teaching materials or of demonstrating the ways in which language is used in various contexts. The project reported here is investigating the feasibility and effectiveness of using a concordancing program as a tool for use by individual learners beyond the classroom, on their computer desktop as they compose or revise their texts. EAL researchers and research students have been targeted for the introduction of this approach for two reasons. Firstly, their facility with the processes of research and induction is likely to predispose them to being interested in using a language-learning approach based on research (Todd, 2001). The second reason is that, as discussed above, they have an acute, urgent and on-going need to learn accurate use of a specialised variety of English, which is exemplified in the articles published in journals from their sub-discipline area. A collection of such articles can therefore be used to construct a highly discipline-specific corpus to use with the concordancer. But will research students be able to manage concordancing software? Will they find it useful? How much training will they need to become proficient users? Will they be able to construct searches that will answer their uncertainties about how specific items of English are used? And will they make the time to construct their own corpus, or prefer to use a more general one if it is made available to them? These are the questions we set out to answer in our initial trails, with the aim of using the answers we obtained to direct the next stage of the project.

The next sections of the paper report on the identification of an appropriate concordancing package for our target users, and on issues related to constructing discipline-specific corpora of research English. We then present the methodology and results of the trials, and discuss the outcomes in terms of the future development of the approach.

Software

Generally, software in this field is not designed with students in mind. Concordancers are usually written for language researchers and often tend to be complex in their operation. They also often tend to use jargon known only to linguists and language specialists, potentially adding to the training necessary for students to grasp the use of the software. Web-based concordancers exist, but all seem to have severe limits — typically handling texts of about 1,500 words or less via a pasting operation, or limited to the use of sample texts provided on the web site. These would therefore be unsuitable for the intended purpose.

For this project investigation was limited to software written for the PC platform (as opposed to Macintosh) as this is the preferred platform of the University's Information Technology Services, and is therefore the most readily available platform among the target user group. Initially, promotional material and reviews were located to limit the field to those packages which could be used in the described context. Concentrating the search on packages that would minimise the secondary learning required, and would be operational within the intended operating environment, five packages were finally reviewed: Concordance, Corpus Presenter, WordSmith Tools, MonoConc Pro, and ConcApp. We summarise below details of each that are relevant to our search criteria.

Concordance

Concordance is published and sold directly by the author, R.J.C. Watt, from the University of Dundee. A trial version can be downloaded from http://www.rjcw.freemove.co.uk/concordance_software_download.htm, and purchase involves an online registration of the trial copy by credit card. As the title suggests, it is purely a concordancing tool, but provides a broad range of features and functions within the concordancing process. Of the software reviewed, it is one of the simplest to operate.

Corpus Presenter

This package was written by Raymond Hickey of Essen University, and is published and sold by John Benjamins Publishing Company. A trial, which is only a 'light' version of the original, can be downloaded from the author's site at http://www.uni-essen.de/~lan300/corpus_presenter.htm, but the full version must be purchased from the publisher. Corpus Presenter is possibly the largest package of its type, consisting of a 292-page book and a CD-ROM containing 27 separate, but integrated programs, as well as a sample corpus of Irish English. Consequently, this package could be overwhelming, and possibly presents the greatest challenge and training requirement for EAL students. It would, however be an excellent research tool for language specialists, as it has more text analysis functions available than other software in the field. As its title suggests, this package is designed for the production of documents derived from working with corpora. The concordancing functions seem to take a back seat to the many other features, and in this area the package is not as flexible and functional as others like Concordance or WordSmith Tools.

WordSmith Tools

WordSmith Tools was written by Mike Scott of the University of Liverpool. It is published and sold by Oxford University Press, and the trial program can be downloaded from their site at <http://www.oup.com/elt/global/catalogue/multimedia/demo/>. Purchase involves registration of the trial copy via

completion of an order form. WordSmith Tools has more features than Concordance, but is not as complex as Corpus Presenter. While some learning of the lexical tools would be required, most of the package is intuitively set out using standard Windows style. The author also provides a downloadable British National Corpus for use with WordSmith.

MonoConc Pro

MonoConc Pro is available for purchase from the authors/publishers at <http://www.athel.com/>. It has similar features to Concordance, and is similar in its simplicity of operation. Its interface is based on standard Windows and would require less training than Corpus Presenter or WordSmith Tools.

ConcApp

ConcApp was authored by Chris Greaves, coordinator of the Hong Kong Virtual Language Centre. It is available for free download from the VLC web site at <http://www.edict.com.hk/concordance/>. ConcApp proved to be a simple, lightweight program that would be easy to learn and use, while retaining a good range of options for the student. This package is also 'freeware', making it an ideal choice for students.

A note on MicroConcord

There is a large body of research available on the Internet which cites MicroConcord as the tool behind the investigations. However, Oxford University Press discontinued the product in 1997, and as it operated under the old DOS system it may not have been appropriate for our purposes.

After reviewing the above packages in the light of the intended purpose, ConcApp was chosen as the software to proceed with, because of its ready availability, lack of upfront cost and simplicity of use.

Corpora

The highly discipline-specific nature of journal articles and their ready availability in electronic format led us to investigate the possibility of using them as source texts for our corpora. Early in the development phase of the project advice was sought from the University's internal legal department in regard to the use of published and copyrighted texts for the intended purpose. The outcome was that the project was restricted to the use of material held or subscribed to by the University library, or contributed by academics if they held the copyright on the material contributed. The same conditions were deemed to apply as is the case for reproducing material for research purposes and it was arranged that the University's standard copyright notice would appear on computer screens when each corpus was used. There was also the restriction that the corpora developed in this way could only be made available to staff and students of the University of Adelaide.

Each corpus consisted of 5-8 journal articles. Preparing the articles for use in the corpora most commonly involved a copy-and-paste operation to convert text from Portable Document Format (PDF) to the ASCII text format required by the concordancing software. Care must be taken to exclude extraneous text such as headers and footers, and so columns or pages of text needed to be copied individually into a new file, which was then saved in .txt format. The texts were organised into separate folders labelled by discipline.

Trials: Procedure and results

For each group the basic features of the ConcApp software were demonstrated in a computer laboratory in which an on-screen demonstration could be projected to all computers. The demonstration covered the following elements:

- Opening the program and setting up a search
- Searching for an item as a word/phrase, as a prefix, as a suffix or as any string
- Accessing the full context of a line from the concordance output
- Sorting search output by left or right collocates (words appearing to the left or right)
- Examples of the types of questions that can be answered by concordance searches of the different types

The participants then had the opportunity to use the software with a corpus of journal articles from their broad research field, with instructor guidance available and printed instructions at hand. Members of the agricultural science group were asked to complete an evaluation questionnaire at the end of the session, and were contacted by email 8 months later to ascertain whether the software had been used again and whether action had been taken towards establishing a more discipline-specific corpus.

Applied Linguistics students

The first trial session was held with students of applied linguistics at both Masters and PhD levels. The intention was to demonstrate the software to a group with an interest in and knowledge of language, language teaching and language research, to assess their level of interest in the software and collect any ideas they may have for making it more useful and accessible for research students not primarily focussed on language issues. The sample corpora provided allowed for searches within the texts of more than one discipline, providing opportunities for the investigation of possible differences.

These students showed interest in using the program for their own purposes in developing teaching materials and researching discourse features and language structures. However, there were no suggestions put forward that were of benefit to the current research on the applicability of the approach for novice users not already experienced with language investigations.

Agricultural science students and researchers

The second trial was held with a group of EAL PhD students and researchers, predominantly from the discipline of agricultural science, but including others from science disciplines. The intention was to test the response of a sample from the intended target group: students and researchers who are not linguists or language teachers, are unfamiliar with concordancing and are seeing it work for the first time. This group focused on the capacity of the program as a self-help tool in further developing their own academic writing skills, as intended.

The evaluative responses of this group are summarised in Table 1. These results show that participants completed the session feeling confident that they could use the software on their own and that they understood the potential of the package. They were very interested in making their own corpus for future use, and thought that the package would be useful to researchers in their home countries. Less strong agreement was indicated with statements about needing more training or practice time, and about their ability to find answers to their own questions using the package as presented.

Questionnaire item	Mean score
The searches suggested in the handout helped me understand the potential of this package for identifying discipline-specific language usage	4.3
I could find answers to my own questions about language usage	3.8
I am confident to use ConcApp by myself now	4.0
If this software and this corpus were on my computer, I would use them to improve my written English	4.1
I am interested in making my own corpus to use with ConcApp	4.5
I need more training in the kinds of searches that might be useful for my language learning	3.7
I need more training in how to use ConcApp to search texts	3.3
I need more time to practise ConcApp with a trainer available to help me	3.1
This package (ConcApp plus corpus of articles) would be useful for researchers in my home country	4.1

Table 1: Mean scores (n=17) for evaluative items by agricultural science research students after a 90-minute introductory training session with ConcApp and a corpus of journal articles (1=disagree strongly, 5=agree strongly)

Results of the follow-up survey 8 months later were less clear-cut, but nevertheless provided valuable pointers for the future. The twelve participants who had provided email addresses were sent a message asking whether and how they had used ConcApp since the introductory session. Two messages were returned marked undeliverable. One of these had been addressed to a Chinese postdoctoral fellow whose contract has since expired, and we know from contact he initiated after the original session that he has downloaded the software and basic corpus and found them very useful (pers. comm. 2 June 2005). Of the remaining 10, only two replies were received, both answering 'no' to all questions. It can probably be assumed that the other answers would also have been negative. Thus almost all of the sample of students and researchers introduced to this software in May, and provided with training and instructions for installing it on their own computers, had not used it by the following January. Several factors can be suggested that may have contributed to this lack of uptake, and these are discussed below.

The timing of the introduction in terms of stage of candidature is suggested as a factor by a comment in one of the two replies received: "I think ConcApp is useful for research students, especially for international students. I'm going to say 'YES' for all questions, since I'm starting writing up my thesis very soon." For students in the sciences, stage of candidature is often closely related to the amount of writing they do, and the focus placed on it. Of the students who attended the session, five had just completed their formal proposals, and so eight months later would be heavily involved in experimental work, placing less emphasis on writing. The comment quoted above suggests that it may be important to pay attention to the timing of the introduction of the ConcApp approach, targeting students who may be embarking on a writing-intensive phase of their research. The specific timing would of course vary

with the discipline of the research and its publication conventions.

Another possible factor is the time and effort required to construct a truly discipline-specific corpus. The steps involved in converting a .pdf file to a text-only file are not difficult, but they are tedious and somewhat time-consuming. It may well be the case that, to encourage uptake of a new tool such as this, its use must be made as straightforward as possible in the first instance. In the next phase of our project we intend to approach post-graduate coordinators of Disciplines within the University, to canvass their interest in working with us to construct corpora that will cover the sub-disciplines operating in their area. These can then be made available on the University's password-protected MyUni site, along with the ConcApp software, and training provided at a Discipline level.

Another step that will be adopted is more intentional modelling of the use of ConcApp to solve particular language problems. Lecturers in the University's Integrated Bridging Program (IBP) for international postgraduate research students (Cargill 1996; Cargill, Cadman et al. 2001) will aim to use ConcApp to address one or more language errors in each individual consultation they conduct in 2005, as well as introducing the software early in the IBP teaching schedule. The students in this IBP will then be surveyed on a regular basis to see if their usage of ConcApp differs from that of other groups.

If the usefulness of this tool can be confirmed by future research, it will have implications for schools and university departments engaged in refining the writing skills of EAL students who are intending to study, or are studying, at English language universities around the world, and those assisting EAL researchers who must publish their research in English-language refereed journals. Pressures on these organisations and individuals to provide quality learning opportunities, while lessening the burdens of time and cost on teaching and consulting staff, is increasing, and concordancing as a self-help tool may well provide such opportunities.

References

- Cargill, M. (1996). An integrated bridging program for international postgraduate students. *Higher Education Research and Development* 15(2), 177-188.
- Cargill, M., Cadman, K. & McGowan, U. (2001). Postgraduate writing: using intersecting genres in a collaborative content-based program. In I. Leki (Ed.), *Case Studies in TESOL: Academic Writing Programs* (pp. 85-96). Alexandria, VA, USA: Teaching English to Speakers of Other Languages (TESOL).
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System* 25(3), 301-315.
- Johns, T. (1986). Micro-concord: a language-learner's research tool. *System* 14(2), 151-162.
- Johns, T. (1991). Should you be persuaded - two samples of data-driven learning materials. *English Language Research Journal, Birmingham University* 4: 1-16.
- Johns, T. (1994). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.) *Approaches to Pedagogic Grammar* (pp. 293-313). Cambridge, UK: Cambridge University Press.
- Stevens, V. (1995). Concordancing with Language Learners: Why? When? What? *CAELL Journal*, vol 6 no. 2, Summer 1995: 2-10.
- Thurstun, J. & Candlin, C. N. (1993). *Exploring Academic English: A workbook for student essay writing*. Sydney: NCELTR, Macquarie University.
- Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System* 29: 91-102.
- Tribble, C. & Jones, G. (1997). *Concordances in the Classroom*. Houston, TX, USA, Athelstan.

Copyright © 2005 Margaret Cargill & Ray Adams: The authors assign to HERDSA and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to HERDSA to publish this document in full on the World Wide Web (prime sites and mirrors) on CD-ROM and in printed form within the HERDSA 2005 conference proceedings. Any other usage is prohibited without the express permission of the authors.