Concordancing and the Teaching of the Vocabulary of Academic English

Jennifer Thurston and Christopher N. Candlin

Abstract—The project described in this article has used the concordancing program, Microconcord, with the Microconcord Corpus of Academic Texts (1993) to introduce students unfamiliar with the language of academic discourse to some of the most important, frequent and significant items of the vocabulary of academic English. The project has developed materials for classroom use and independent learning intended for native speakers of English as well as students of non-English speaking backgrounds. The materials deal in detail with frequently used words which are common to all fields of academic learning, not attempting to include specialized or technical vocabulary items associated with specific disciplines. Those working on the project are convinced of the value of concordancing in the development of teaching materials focusing on vocabulary and grammar and the line between them. This article describes the approach used and deals with the rationale behind that approach. © 1998 The American University. Published by Elsevier Science Ltd. All rights reserved

Keywords: concordancing, academic vocabulary, corpus-based teaching, grammar in context

Introduction

Since concordancing programs have become available to teachers and students, their possibilities have been seen as offering new and exciting directions for developing teaching materials, enabling students themselves to make direct discoveries about language (Johns 1991a; Tribble & Jones 1990) and as an aid to course design (Flowerdew 1993). This particular project* has used the concordancing program, Microconcord, and the Microconcord corpus of academic texts (1993) to develop teaching materials for independent study of the vocabulary of academic English.

Address correspondence to: Jennifer Thurston, NCELTR, Macquarie University, Sydney, NSW 2109, Australia. Tel: 61 2 9850 9936; Fax: 61 2 850 6129; E-mail: jthurston@nceltr.ncltr.mq.edu.au

*Sponsored by the Centre for Higher Education and Professional Development at Macquarie University as a Macquarie University Teaching Development Grant for the project entitled The Vocabulary of Academic English.
It was decided to develop corpus-based learning materials that would be of assistance to students from any discipline, focusing on lexical items shared across various disciplines. Li & Pemberton (1994: 184) point out that tertiary students do not necessarily find discipline-specific technical vocabulary difficult:

Rather, it is the vocabulary with a middle frequency of occurrence across texts of various disciplines that students find most problematic.

Nation (1990) refers to this range of vocabulary items as "academic vocabulary".

**Rationale**

The features of this project which require preliminary explanation are firstly, the decision to focus in detail on a restricted set of vocabulary items, and secondly the use of concordancing techniques to provide the student with intensive exposure to the use of these items.

In choosing the vocabulary items to be dealt with in this project, we began with the extensive University Word List cited in Nation (1990). Using this list, we developed categories of vocabulary items according to the various rhetorical purposes they can serve in academic writing. We then selected items according to frequency of use cited by Nation and our own perception (based on having marked many hundreds of student essays and in consultation with teachers of English for Academic Purposes) of the extent to which their investigation would be helpful to students. In this way, we created a list of about 150 examples (see Appendix A), attempting to group them according to purpose. We will refer to these purposes as rhetorical functions. Further selection was based on frequency counts provided by Microconcord, using the Microconcord Corpus of Academic Texts. The main rhetorical functions we identified, and the key words for each function on which the project finally focused were:

- Stating the topic of your writing
  - factor
  - issue
  - concept

- Referring to the research literature
  - evidence
  - research
  - source

- Reporting the research of others
  - according to
  - suggest
  - claim
• Expressing opinions tentatively
  may
  possible
  unlikely
  probably

• Explaining procedures undertaken in a study
  identification
  analysis
  criteria

• Drawing conclusions
  conclude
  summarize
  it is clear
  thus

Such a focus on particular rhetorical functions provides a purposeful basis for learning and potentially some structuring for the teaching of academic writing skills. The intensive focus on a limited number of vocabulary items is characteristic of concordance-based materials and is supported by the experience of Tim Johns (University of Birmingham, UK, personal communication). As with Li & Pemberton (1994), students, in his view, do not necessarily need to master a wide range of academic terms in order to write acceptable academic essays. They do, however, need to be competent users of a restricted set of “semi technical” vocabulary items. Pickard (1994: 218) suggested that, when preparing concordance-based material, there is

![A potential tension between the aims of wanting to expose students to a variety of vocabulary as a means of encouraging variety in their writing, and using a concordancer to search for a key word. Through selection of key, frequent words how does one encourage variety?]

Variety, however, is not necessarily an end to be pursued for its own sake in the teaching of academic writing, and Pickard goes on to point out that, by drawing students’ attention to collocates of the key word, concordance-based study has considerable potential for expanding student vocabulary while dealing in detail with selected items.

Academics from four Australian universities responded in a recent study (Bush et al. 1996) to questions about their expectations of student writing. Results indicated that accurate and appropriate use of academic vocabulary is considered to be extremely important, but there is much more concern that students convey their ideas clearly than that they attempt to rely on jargon.

These comments and studies supported our view that the most useful approach to helping students unfamiliar with the vocabulary required in academic writing would be to focus on intensive work on a few of the most
useful lexical items selected as typically realising each of the rhetorical functions listed.

These three or four selected words for each rhetorical function are all frequently used, most appearing more than once every 6000 words in the corpus. Words such as unlikely and forms associated with summary are not used quite so frequently in the corpus of professional, published work, (appearing with a frequency of 1/12 600 and 1/13 254 respectively) but were nonetheless included given their usefulness for student writers, particularly for dealing with modality and the creation of final statements. A range of grammatical forms of each word in question was included where appropriate.

Concordance-based materials offer the learner a rich experience of language (in this case, the language of academic English). Concordancing has been used in this project to present students with the opportunity to condense and intensify the process of learning through exposure to multiple examples of the same vocabulary item in context, and to promote awareness of collocational relationships. According to Nattinger (1988: 63),

guessing vocabulary in context is the most frequent way we discover the meaning of new words.

Johns (1991b) has argued that the central justification for using concordance-based materials is that they can help to develop this ability to guess the meaning and use of unknown words from context. In our materials, exposure to concordances for the purpose of discovering meaning focuses learner attention on the central importance of collocational relationships in connection with the key words. The broad objective of the materials is to develop the writing competence of students by promoting discovery of meaning and by making students aware of representative patterns of language use and to selected grammatical structures. This awareness is accompanied by guided opportunities for research, practice and improvisation.

At various stages in the preparation of this material, as with Flowerdew (1993: 240) who found “areas where concordancing has revealed a discrepancy between published materials and the specialist corpus,” we also encountered language use which questioned the standard patterns usually recommended to EAP students by teachers and grammar books. For example, it was found that the plural form, researches, occurred in the corpus on ten occasions, indicating that it is, in fact, accepted practice in published texts though usually not accepted by markers of student essays. Although we did not draw student attention to this specifically, we decided to modify questions about the use of this form so that students were not obliged to understand that it is never used in its plural form. Likewise, the standard advice to students to use the past tense of the reporting verb for author-prominent statements (see Weissberg & Buker 1990: 45, who advise students that “in these citations the simple past tense is used in the verb of
report") was not supported by the corpus, these reporting verbs being shown to be used more frequently in the present tense.

**Method**

The *Microconcord corpus of academic texts* contains approximately 1,016,000 words taken from academic texts (books, chapters and articles written by academics for academic audiences) in the fields of arts, belief and religion, applied science, science and social science. The concordancing program, *Microconcord* was used to establish the frequency of use of particular items in the corpus. Concordances of 100 characters were obtained for the student research activities, concordances of 400 characters for use in exercises and improvisation activities.

The material is being prepared for publication at this stage (in hard copy) and is primarily intended for independent study, although it is also appropriate for classroom and workshop use. Initial small-scale piloting in Australia showed that some students were puzzled by the cut-off sentences of the one-line concordances (see the Materials section for example) and daunted by the difficulty of the authentic academic texts. Since this mode of learning is quite new to most students, we included notes explaining that all the examples used are taken from authentic academic texts from a wide range of disciplines, with which students may not always be familiar. We endeavoured to emphasize that they are not expected to understand every word or every idea presented, but that the materials are intended to familiarize them with the use and meaning of the key words, helping them to develop insights into the collocations and grammatical structures with which the key words are associated.

**Materials**

Students are first presented with multiple examples of the vocabulary items in context. The student examines these one-line concordances, highlighting word groups surrounding the key word to discover how they are used, and then answers questions on their use in context. The presentation of concordances and these guided research tasks are followed by exercises enabling students to ensure that they are using the item appropriately. These tasks are then followed by creative and improvisational activities. Answer suggestions are provided for each unit. In the instructions for students, the term "meaning" is used to describe lexical and contextual significance, and the term "patterns of language" is used to describe collocations.

The learning pathway for students provides a chain of tasks or activities in which the successful completion of each task facilitates the completion of the succeeding activity, thus providing task continuity as defined by Candlin (1987) and Nunan (1989) and reflecting Johns's approach of Research-
Practice-Improvisation (Johns 1991b). This chain of activities, repeated for each key word, is as follows:

- LOOK at concordances for the key term and words surrounding it, thinking of meaning.
- FAMILIARIZE yourself with the patterns of language surrounding the key term by referring to the concordances as you complete the tasks.
- PRACTISE key terms without referring to the concordances.
- CREATE your own piece of writing using the terms studied to fulfil a particular function of academic writing.

For example, in the unit explaining procedures undertaken in a study, the rhetorical function is first introduced, vocabulary items which typically fulfil this purpose are mentioned and a number of sample texts which use these items are provided. The student is then presented with a full page of online concordances for familiarization with this item, as follows (see Fig. 1):

**UNIT 5A**

**GROUP ONE**

Study these concordances, underlining the central group of words which can stand alone, as has been done in the first example. Then answer the questions below. You may like to look at question 1 before you start. (Don't worry that these are cut-off sentences - just familiarize yourself with the key words).

**GROUP ONE**

- To make sense at puberty. The isolation and chemical identification of several sex hormones in the late 1920s In the first instance, a method lies not simply in the formal identification of a hormone, of its subgroups or a homotopon of hormones only after much controversy. Its occurrence in the production of muscle-specific proteins. Identification of the novel muscle protein has made possible the isolation of the discovery had led to the identification of two additional problems: 1. Do the principles that result...<list> 1. The isolation of 1. Greek literature, can nevertheless still be identified as distinctly and possibly definitive properties. 2. Many standards have been identified and utilized for recombination. 3. Instead, we have a second method identified by a series of regulatory causes identified by the receptor of the hormone. A later Muslim tradition identified the Enoch as the point from which Muhammad identified as either ribose or deoxyribose: for a time random. Forty years later these factors were supplied by 100 years or so before bacteria were identified as the cause of chromosome...<list>

**GROUP THREE**

- We do not rejoin. If we know enough and could identify all the individual animals alive, say, one hundred policies must be matched. If we can only identify the proper criteria correctly if we accept the therapist's role is to help the partners identify the problem that they face as a couple and that is out of bounds in politics. It tries to identify the reasons which lead people to embrace this. I want to identify one feature of that relationship which seems to separate effects. Research was essential to identify the ill-effects attributable to each substance relative to that tradition. They hope to identify a coherent body of ideas which places them so we don't have to identify the chemical constituents of fungi and discover.

**FACILITATE**

1) Which of the following statements do you think are true? Tick your answer in the box.

- **TV IDENTIFY** involves naming  true false
- **TV IDENTIFY** involves describing  true false
- **TV IDENTIFY** involves imagining  true false
- **TV IDENTIFY** involves deciding what something is  true false
- **TV IDENTIFY** involves recognizing  true false
- **TV IDENTIFY** involves criticizing  true false

Figure 1.
Students are encouraged to become sensitive to the patterns of language frequently surrounding the key word. The following exercises are designed to develop awareness of the appropriate preposition and the linking of the key word with its anaphoric referent (by the creation of lists) (see Fig. 2).

This section is followed by exercises using complete sentences taken from the corpus, and then by exercises involving students in their own improvised writing using the key word. As the unit progresses, including the study of two or three other words, the exercises become cumulative, including all items studied, and allowing the introduction of more interesting, innovative exercises using concordances. Concordancing provides opportunities to create innovative tasks for students, but gap-filling exercises continue to be useful because, as Pickard (1994) points out, they are easy to prepare and students have a sense of familiarity with this format. The following are some examples of representative exercises within the units of the materials (see Fig. 3).

Johns (1991b) calls exercises of the above type "gapping using multiple contexts". They involve forming and testing hypotheses while noticing meaning and collocation (see Figs 4 and 5).

Gapping and matching exercises of the above types are familiar to most students and have been included, partly to provide this sense of familiarity given the novel nature of the materials with which they are associated. Concordancing allows a materials developer enormous possibilities for the creation of such exercises from authentic texts (see Fig. 6).

2) Which preposition most frequently follows the word IDENTIFICATION?

identification

3) Look at the second group of concordances with the key word IDENTIFIED. Notice that many of these use part of the verb TO BE (is, are, was, were, been, be) as auxiliaries. Circle the concepts/objects identified in these concordances and write them down. Decide whether the singular or plural form of TO BE should be used.

... the facts............ (is/are) identified
... (is/are) identified
... (is/are) identified
... (is/are) identified
... (is/are) identified
... (is/are) identified
... (is/are) identified
... (is/are) identified
... (is/are) identified
... (is/are) identified
... (is/are) identified
... (is/are) identified

4) Now look at the concordances in Group 2 which do not use part of the auxiliary TO BE and circle the objects identified. Who or what are the identifiers in these concordances? What are the concepts or objects they identify?

<table>
<thead>
<tr>
<th>IDENTIFIER</th>
<th>OBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>we</td>
<td>identified two characteristics.</td>
</tr>
<tr>
<td>identified</td>
<td></td>
</tr>
<tr>
<td>identified</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.
In this final group of exercises, students are provided with prompts and outlines to facilitate their own use of the vocabulary studied. Piloting indicates that a number of students find these exercises more demanding than the preceding exercise types.

Ability on the part of the teacher to select appropriate, representative concordances improves with experience. A useful technique is to use the delete key on all concordances of a particular word, looking at the words and word groups on each side of the key word which can be shown in alphabetical order, and then inserting and saving concordances which provide a representative sample of frequently used contexts.

The word issue/s, for example, appeared once every 3019 words (338 times), eliciting nine densely packed pages of one-line concordances. As a basis for selecting concordances for student research that fit onto one page and provide a representative sample of the use of the word in context, the words preceding and following the key word were examined. It was found that some adjectives (e.g. environmental, moral, political) appeared with noticeable frequency and so needed to be included, and that frequent postposed prepositions (e.g. for, of, in) also needed to be brought to the attention of the students. The exercises developed for students to familiarize them-
Decide which verbs or adjectives to use in the blanks in the following examples. Choose from the lists below. (The following are extension exercises, not necessarily related to the concordances)

ADJECTIVES: legal, mental, modern, definite, new
VERBS: to defend, to form

- Our ............... concept of history, however rationalized and secularized it may be, still rests on the concept of historical time which was inaugurated by Christianity.

- The capacity ............... a concept of one's own identity and a process of comparison allowing an estimation of one's esteem among one's fellows that affects one's view of oneself are central features of humanity.

- According to this definition, time has no substantive existence but is a ............... concept or means of measurement - a point of view that strikes us today as being remarkably modern.

- One may even say that the whole purpose of this book is ............... a concept of political freedom.

- If this is the case, rather than attempt to invent a ............... conceptual framework, there is much to be said for turning to marriage as the one, already existing ............... concept which has the obvious potential for expansion so as to provide the institutional framework for such a union.

- As S. G. F. Brandon has pointed out, the great popularity of the cult of Osiris meant, in effect, the adoption by the Egyptians of a ............... concept of time; although this may not have been consciously recognized.

Draw lines matching parts of the sentence on the left hand side with the appropriate part on the right.

Plaget does not believe that concepts factor governing the doctor-patient relationship.

The other single most important of God and Nature used to evoke a poetic response in the nineteenth-century reader, the modern poets rely on sex and myth to produce the same effect.

The issue here is whether provocation should remain a qualified defence to murder and, if so, how far it should extend.

Another unifying principle is said to be the take their origin from linguistic concept of trust. This, it is said, is the key structures.

Where the concepts issue in the early years of the war.

With the conduct of hostilities the paramount factor responsible for raising the concern for both the USA and Britain, civil standard of care was the recognition of aviation remained a minor venereology as a clinical specialty in its own right.

Figure 4.

selves with frequent patterns suggested their own further questions. These led the student to explore, on the basis of the primary source of the language itself, with no mention of rules or need to resort to pedagogic grammar, the contexts in which the different prepositions are used. This section of the materials is provided in Fig. 7.

The possibilities of leading students to grammar-in-context discoveries of this type using learning materials and with direct student access to the computer, have been investigated by Johns (1991b: 31). He promotes the idea of "the teacher abandoning the role of expert and taking on that of research organiser", allowing students access to a computer equipped with
December 1987: 8

1. unreasonably see in these actions the...EVIDENCE...of communal meal-taking in man and its im
2. experiences of human beings as valid...for studies of consciousness. Contemporary
3. who like myself had conducted field...in Africa, I undertook a survey of existi
4. st clearly be expanding international...and monitoring efforts. But this should n
5. helpful, however, in a search for the...and significance of our social organizati
6. boration in mammals comes from recent...by Olwyn Rose (1977) on the dwarf mongooses
7. d countries unable to borrow from any...without the IMF's 'seal of approval' must u
8. are as N20 

Figure 5.

Figure 6.

a concordancing program and a relevant corpus as the primary source of information, language itself. A similar proposal was presented in Leech & Candlin (1986: xvi) where in their introduction, the editors advocated a classroom with online links to just such a corpus resource:

We need, in short, comprehensible input tailored to our learners' needs and abilities... we need... classroom access to language databases, lexicographic and grammatical corpora, oriented to learners' interlanguages and displayed in terms that learners (not only lexicographers and grammarians) can understand... What of the process tasks to be made available as problem-solving exercises for learners to apply to text? Two characteristics will predominate: they will have to involve learners in solving problems and experimenting with language learning, and they will need to be differentiated in terms of offering
alternative routes, varying levels of demand and attainment, and alternative possibilities of solution. In short, they will need to mirror the cognitive requirements of language learning.

Conclusion

Teachers and students who have piloted the materials have commented that they provide a helpful, very different and innovative approach to vocabulary learning. They confirm our opinion that the words chosen are those that students need as basic tools for academic writing. In a real sense, materials of this type provide more than just vocabulary learning. They provide a rich experience of the language with insights into collocations and contextualized grammatical structures linked to opportunities to develop students' analytical abilities.

Although students have overwhelmingly indicated that they find all exercises “very helpful” or “helpful”, and teachers involved in piloting (in Australasia, Canada and Spain) have commented, for example, that “the material has provided structure for my whole course”, some student reservations should also be mentioned. To their comment that “there are many exercises for just three words” we direct them to the point that although the focus is on a restricted number of words, learning extends far beyond the
particular items around which the material is based. The rich context of language within which these vocabulary items are studied provides considerable opportunities for students to broaden their lexical and grammatical awareness.

Over-exposure to concordance lines can conceivably tire students if teaching of this type depends solely on deduction from concordance lines. We have, therefore, provided a variety of different activities, all based on concordanced texts but demanding different problem-solving techniques from the student.

There is some evidence from piloting that some students may benefit more from the material if it is presented in a teacher-mediated workshop format, while others may find it valuable for independent study.

(Revised version received 22 May 1997)

REFERENCES


Jennifer Thurstun is a Lecturer in English for Academic Purposes at Macquarie University, Australia. Her research interests include language maintenance, academic writing and the practical application of concordancing.

Christopher N. Candlin is Professor of Linguistics and Executive Director of the National Centre for English Language Teaching and Research at Macquarie University, Sydney, and is currently Visiting Chair Professor of Applied Linguistics at the City University of Hong Kong. His research interests include discourse analysis and pragmatics, academic writing and second language acquisition in workplace contexts.

**Appendix**

The University Word List cited in Nation (1990) provided the basis for the development of categories. Items were selected according to frequency of use and our perception of the extent to which their investigation would be helpful to students. Some vocabulary items of frequent occurrence in the Microconcord Corpus of Academic Texts were added. Our list after initial selection, grouped according to rhetorical functions, was as follows:

**Stating the Topic of Your Writing**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Proposition</th>
<th>To Predict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesise</td>
<td>To Consist</td>
<td>Topic</td>
</tr>
<tr>
<td>Issue</td>
<td>Thesis</td>
<td></td>
</tr>
<tr>
<td>Premise</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**General Vocabulary of Research**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Controversy</th>
<th>Homogeneous</th>
<th>Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>Debate</td>
<td>Implicit</td>
<td>Source</td>
</tr>
<tr>
<td>Anecdotal</td>
<td>Empirical</td>
<td>Issue</td>
<td>Study</td>
</tr>
<tr>
<td>Anomaly</td>
<td>Evidence</td>
<td>Norm</td>
<td>Subsequent</td>
</tr>
<tr>
<td>Assimilate</td>
<td>Exception</td>
<td>Pertinent</td>
<td>Survey</td>
</tr>
<tr>
<td>Cogent</td>
<td>Explicit</td>
<td>Principle</td>
<td>Tentative</td>
</tr>
<tr>
<td>Conform</td>
<td>Feasible</td>
<td>Relevance</td>
<td>Theory</td>
</tr>
<tr>
<td>Consequent</td>
<td>Findings</td>
<td></td>
<td>To Equate</td>
</tr>
<tr>
<td>Trend</td>
<td>Valid</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Verbs of Reporting**

to add  
to consider  
to dispute  
to question  
to affirm  
to contend  
to emphasise  
to recommend  
to agree  
to contradict  
to enumerate  
to refute  
to argue  
to demonstrate  
to explain  
to reject  
to assert  
to deny  
to imply  
to remark  
to assume  
to describe  
to infer  
to report  
to challenge  
to determine  
to maintain  
to repudiate  
to claim  
to disagree  
to point out  
to say  
to clarify  
to discern  
to postulate  
to suggest  
to comment  
to discuss  
to query  
to stipulate  
to view

**Using Tentativeness**

may  possible  probably  unlikely

**Linking Ideas Correctly**

also  however  since  whereas  
among  moreover  so  while  
as  on the other hand  therefore

**Indicating Writers’ References to the Work of Other People**

according to  to cite  to claim  to quote  
to refer to  to suggest

**Drawing Conclusions**

conclusion  It is clear that  therefore  to signify  
evident  outcome  thus  to verify  
It can be seen that  summary  to reveal

**Words with Different Meanings in Different Contexts**

appreciate  complex  constant  manifest  
appetite  conceive  construct  objective  
attribute  concentrate  function  potential

**Explaining Processes Undertaken in a Study**

administer  data  to elicit  to interrelate  
apalyse  process  to establish  to investigate  
conduct  to denote  to evaluate  to modify  
correlate  to derive  to identify  undertake  
criterion  to design  to integrate