# BALEAP PIM

**Uses of corpora in EAP**

*Palmer Building*  
*University of Reading*  
*15th November 2003*

## Programme

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.30-10.00</td>
<td>Registration &amp; refreshments</td>
</tr>
</tbody>
</table>
| 10.00-10.25| Paul Thompson (Room 1.09)  
*What use are corpora in the teaching of English for Academic Purposes?* |
| 10.30-11.00| Maggie Charles (Room 1.09)  
*Reporting Verbs: What Can We Learn From Corpus Data?*  |
| 11.05-11.35| Nigel Harwood (Room 1.09)  
*Corpus-based critical pragmatic EAP*  
Norbert Berger (Room 1.03)  
*Compiling and exploiting a learner corpus to improve academic writing*  |
| 11.35-12.05| David Stent (Room 1.09)  
*Methodological and procedural problems relating to an examination of how much Chinese students subscribe to an Anglo-American style of writing*  
Gerard Sharpling (Room 1.03)  
*The development of an electronic corpus of written assignments, and its use in the teaching of EAP*  |
| 12.05-1.10 | Lunch (concourse area)                                                  |
| 1.10-2.10  | Julie Hartill and David Lefevre (Room 1.09)  
*Practical applications of on-line corpora in a Science-Technology context*  
Tilly Harrison (Room 1.01)  
*Using on-line concordance lines in EAP*  |
| 2.15-2.45  | Fiona Barker (Room 1.09)  
*Learning from the learners: compiling and exploiting corpora of academic English*  
Olwyn Alexander (Room 1.03)  
*Singing from the same songsheet? Data commentary across the disciplines*  |
| 2.45-3.15  | Tea (concourse area)                                                    |
| 3.15-3.45  | Ken Hyland (Room 1.09)  
*Acknowledgements in graduate theses*  
Hilary Nesi (Room 1.03)  
*The use of numbering in British and American lectures: evidence from BASE and MICASE*  |
| 3.45-4.25  | Closing discussion (Room 1.09)                                          |
ABSTRACTS

What use are corpora in the teaching of English for Academic Purposes?
Paul Thompson
The University of Reading

Swales (2002) questions the relevance of corpus-based work to EAP, particularly with respect to its direct application to the classroom, and in the development of EAP materials. He posits a generally top-down approach to analysis taken by EAP practitioners, as compared to a bottom-up approach taken by corpus linguists, and argues that it is difficult for EAP practitioners, particularly materials writers, to integrate corpus-derived insights into their practice. Corpus investigation is perceived to be slow and erratic, and the results of dubious value.

In response to Swales' discussion, this paper reviews the uses of corpora in EAP, and assesses the contributions that corpus-based studies have made both to the general understanding of academic discourse, and to EAP teaching practice. Firstly, the various sources of data are described, divided into the categories of: academic subsection of general corpora, EGAP corpora, ESAP corpora; learner corpora; and corpora combined with extensive contextual information such as ethnographic studies. Relevant academic corpus projects such as the British Academic Spoken English (BASE) corpus are described, and issues involved in the direct application of corpus findings, and corpus analysis methodologies, to EAP classroom teaching, are then discussed. Finally, the results of a questionnaire sent to all BALEAP member institutions in 2002, and which asked for information about how corpora were used by each institution, are presented.

The paper concludes by arguing that corpus-based studies have already provided EAP practitioners with a wealth of insights into the language of academic discourse, and, to a lesser degree, into the diversity of textual practices between genres and disciplines. However, the direct application of corpus methodologies and corpus evidence to the classroom remains as yet limited and this is due in large part to a lack of familiarity amongst practitioners with 1) the methods of corpus analysis; and 2) with the criteria by which insights afforded by corpus-based linguistic research should be evaluated. In addition, many practitioners lack easy access to adequate corpus resources.

In order to develop the direct use of corpora in EAP teaching and course design, it is proposed that the following are required:

- empirical research into the usefulness (or not) of corpora in direct teaching
- improved channels for dissemination of practice
- better access to relevant corpora and corpus analysis tools, and an increased number of corpora created for use by practitioners rather than primarily for researchers
- training in corpus analysis techniques, and the evaluation of such methods, in EAP teacher development courses
**Reporting Verbs: What Can We Learn From Corpus Data?**

Maggie Charles  
Oxford University Language Centre

Reporting on research is a major area of concern for non-native-speaker academic writers and within this area, the use of reporting verbs is crucial. This paper draws upon two corpora of theses in politics and materials science in order to investigate one type of report clause, in which a finite reporting verb is followed by a complement clause beginning with *that*, e.g.

*Sakamoto et al (1994) showed that* a surfactant layer of bismuth has two effects on SSMBE of Ge/Si heterostructures... (dmabochn5)

*I argued above that* language and other cultural markers were important to Mazzini... (mpommmch3)

Although these verbs seem most often to be studied in relation to citation, as in example 1, results from the corpora show that their most frequent use is to report on the work of the thesis writer, as in example 2. In this talk I indicate which verbs occur in these ‘self-reports’ and identify some frequent patterns of use which are characteristic of the two corpora. In so doing I show how corpus data can reveal language patterns which are not necessarily identifiable through intuition alone and thus make a valuable contribution to the study of academic writing.

**Corpus-based critical pragmatic EAP**  
Nigel Harwood  
University of Essex

A key issue that the English for Academic Purposes (EAP) teacher/materials writer must confront is whether they wish to perpetuate the Anglo-American model of academic English. Will learners be expected to imitate this model unquestioningly? Or, at the other extreme, will learners be encouraged to allow their native languages’ rhetorical features to the fore—influencing their writing in English and resulting in a non-Anglo-American model?

Three approaches to the teaching of EAP are identified: the Critical, Pragmatic, and Critical Pragmatic approaches. Critical EAP is appealing because of its restive questioning of discourse norms, although it can seem reactionary at times. By focusing on the acquisition of the same dominant norms, Pragmatic EAP has a clear goal. However, its weakness is its failure to acknowledge difference in disciplinary community practices. Critical Pragmatism fuses Critical EAP’s focus on difference in the academy with Pragmatic EAP’s focus on access to the academy, with corpus data enabling the teacher to inform students of disciplinary tendencies in a systematic manner.

The Corpus-Based Critical Pragmatic approach is illustrated by activities for graduate students centring on the use of personal pronouns. The materials use quantitative and qualitative corpus evidence to raise learners’ awareness of academic writing conventions across the disciplinary spectrum, inviting students to speculate on the possible reasons for the differing practices between the hard and soft fields. The decision as to whether to
subscribe to the Anglo-American patterns rests with the learners, who also discuss the potential sanctions of flouting or conforming to the dominant conventions.

**Methodological and procedural problems relating to an examination of how much Chinese students subscribe to an Anglo-American style of writing**

David Stent
Portsmouth University

The paper presents an examination of a general research hypothesis: that the academic writing of Chinese students differs from that of Anglo-American students, with the following specific hypotheses:

1. That Chinese student writing is more indirect than Anglo-American student writing, tested out by examining the inclusion of a thesis statement, topic sentences and connectors in student essays.
2. That Chinese student writing is reader responsible, tested out by examining the inclusion of code glosses in student essays.
3. That Chinese student writing has a collective viewpoint with a lack of personal opinion, tested out by examining the inclusion of commitment markers in student texts.
4. That Chinese student writing demonstrates a gender difference, tested out by examining the findings of pre and post-test essays.

Participants in the study were students attending Portsmouth University during the academic year of 2002-3, including 40 Chinese students, 40 Greek students and 40 native-speaking English UK students.

The research design and data collection procedures involved were:

- A quantitative method using argumentative essays investigating the use of thesis and topic sentences and metadiscourse markers, a questionnaire given to all students involved in the study.
- The investigation also involved a qualitative method of investigation, using the questionnaire and individual interviews with each of the students.

The paper concludes with a consideration of possible pedagogical implications.

**Compiling and exploiting a learner corpus to improve academic writing**

Norbert G. Berger,
University of Exeter, & Karl Franzens University, Graz

This paper reports on work in progress to develop independent learning materials for students on pre-sessional and academic writing courses.

More and more students submit their drafts and final versions of essays and study skills projects in electronic form either through a website or as email attachments. So compiling a learner corpus is easy. The question is how we can use this corpus efficiently and effectively to design appropriate and useful remedial practice materials for students.
The report will start by explaining the purpose of the work done. Then I will demonstrate how we have modified standard software for marking essays and identify some research issues on the website as well as web-based materials in use. This will also involve a discussion of the classification system and the tagging process. We will then check some assumptions of most frequent error types based on the statistics generated and relate them to learners’ linguistic backgrounds. Finally, some implications for the design of independent learning and further research will be discussed.

Students on the pre-sessional summer course held at the ELC in Exeter contributed to the corpus. Besides, results from academic writing courses held at Karl Franzens University, Graz, Austria will be included to identify some similarities and differences.

The corpora are not representative but we can discuss some relevant implications for the courses these students study on. Besides it may be interesting to consider potential developments.

*The development of an electronic corpus of written assignments, and its use in the teaching of EAP*
Gerard Sharpling
University of Warwick

This paper reports on the development of a project at the University of Warwick to compile a corpus of good quality student writing, with particular focus on its potential use to inform the teaching of EAP. Although informal collections of student writing at university level exist, our corpus is copyright cleared, and is supplemented by contextual information, including contributor data, details of the courses for which the assignments were set, and the grades awarded for each assignment. The paper discusses the database’s rationale and shows how extracts from the database have been used to assist students in their understanding of issues relating to referencing and plagiarism.

*Practical Applications of On-line Corpora in a Science-Technology Context*
Julie Hartill & David Lefevre
Imperial College London

In the 2002–2003 academic year we initiated a corpus development project using a web browser interface in order to increase the relevance of the materials used in our classes, to increase students' motivation and to enhance the face validity of in-sessional classes in a science-technology context. This presentation will begin by demonstrating our in-house corpus builder software and then move on to discuss our uses to date of Web-based corpora to create classroom materials, presenting examples and discussing the rationale behind them. We will then introduce a further application of our corpus data; an on-line vocabulary learning tool. We will end by discussing some of the main issues that we have faced, some of which remain unresolved as the project has moved from the development phase to implementation.
Using on-line concordance lines in EAP
Tilly Harrison
University of Warwick

One of the major problems faced by students writing in English is that their awareness of natural-sounding collocates is often poor which results in texts which are grammatically and semantically 'correct' but somehow lexically strange. Monolingual dictionaries for learners are helpful in many cases but there are certain types of lexical error (collocation) where for many commonly used academic words even the best dictionaries give no answers. Errors of collocation can be frustrating to teachers as often the only explanation is 'That's just how we say it' which leaves the students with the impression that natural English is arbitrary and impossible to acquire. However, using readily available corpus-based tools, it is possible for teachers to back up their intuitive convictions about the frequency or infrequency of any given collocation, and students have a way of checking their collocations which does not simply rely on asking a native speaker.

The workshop will give several examples of how the BNC and Cobuild on-line can supplement and improve on dictionary information. The relative merits and uses of the two on-line corpora will also be discussed. Participants will have the chance to see these tools being used and input their own requests for concordance lines.

Learning from the learners: compiling and exploiting corpora of academic English
Fiona Barker
University of Cambridge ESOL Examinations

For over a decade Cambridge ESOL has been developing a corpus of learner writing derived from exam scripts. The Cambridge Learner Corpus (CLC) now includes 17 million words of general and business English from learners at most Common European Framework levels (A2-C2). The CLC will soon receive academic texts in the form of IELTS scripts which will enable more research to be carried out on the IELTS test, which is for many a crucial qualification for university entry and other high stakes purposes.

Alongside the written learner corpus there is a newer corpus of spoken learner language which already includes IELTS speaking tests. Some of the 8000 speaking tests available have been transcribed and several exploratory investigations have been completed including identifying functions of speech to aid question paper design and exploring candidate and examiner speech in various ways.

There are many possibilities for exploring and analysing the written and spoken IELTS texts contained in these corpora. This presentation will consider the nature of the IELTS data in these corpora, describe how they have already been used and propose future research directions.
Singing from the same songsheet? Data commentary across the disciplines
Olwyn Alexander
Heriot-Watt University

In written English Proficiency exams such as IELTS data in tables and graphs is used to generate simple descriptions of trends. Students are discouraged from accounting for these trends either by using the internal logic of the table or by referring to their knowledge of the world. Is this how data commentary works on university courses? I attempted to find out by asking lecturers and tutors from a range of disciplines to write a commentary for the same data set. This generated a small corpus of texts. In this talk I will present preliminary findings from this corpus with some suggestions for using these in class.

Acknowledgements in graduate theses
Ken Hyland
Institute of Education

While acknowledgements have been largely neglected in the EAP literature, they are almost universal in dissertation writing where they offer students a unique rhetorical space to convey their genuine gratitude for assistance and to promote a scholarly identity. This paper examines the content and structure of acknowledgements from 240 postgraduate theses written by ESL students in 6 disciplines. I argue that the unexamined nature of acknowledgements can disadvantage non-native writers and suggest that EAP teachers assist their students by raising awareness of the ways they can use this genre for rhetorical purposes.

The use of numbering in British and American lectures: evidence from BASE and MICASE
Hilary Nesi
University of Warwick

This paper explores lecturers' use of numbers to outline structure and content in lectures drawn from the British Academic Spoken English (BASE) corpus and the Michigan Corpus of Academic Spoken English (MICASE). University students need to be able to recognise predictive markers in order to take notes effectively, and in order to follow and create sustained lines of argument, but although there have been several studies of the way future content is signalled in academic writing (for example Tadros 1985, 1994), there has been little corresponding analysis of academic monologue. Flowerdew and Miller (1997) provide some evidence to suggest that lecturers are not very reliable in their use of enumeration, and may send out "false signals" by using markers which predict speech acts that are never actually produced. Flowerdew and Miller did not investigate this phenomenon in any detail, however, and drew all their examples of false signalling from just two lectures, delivered by the same person.

There are many examples of enumeration in BASE and MICASE, but British and American speakers seem to use numbering in rather different ways. In BASE, it often predicts the content of much longer sections of text, beyond adjacent clauses, whereas in MICASE there is not so much evidence of elaborate structuring.