Student papers across the curriculum: designing and developing a corpus of British student writing

Nesi, H. , Sharpling, G. and Ganobcsik-Williams, L.

Abstract

This paper reports on a collaborative project, currently being carried out by the Centre for English Language Teacher Education and the Warwick Writing Programme at the University of Warwick, England, to compile a multimillion-word corpus of student writing. Since May 2001, we have collected samples of proficient written coursework produced by students at all levels and in a range of disciplines. We believe this student writing collection will eventually provide an invaluable database for use by researchers and writing teachers, enabling them to identify and describe, in a systematic way, the characteristics of assigned work across disciplines and levels of study. Our corpus is confined to shorter assignments assessed within departments—the most common form of student writing, but unpublished and therefore generally unavailable to researchers. This paper describes the project, and explains the rationale for developing the corpus. It also considers the corpus' potential role as a resource for research and teaching within and across subject disciplines.

1. Introduction: The role of corpora in academic writing research

Corpora play an important role in language research today. The largest multimillion-word databanks of general English enable researchers to describe a text's linguistic features with far greater accuracy than was possible in pre-computer days. Such corpora aim to represent a wide range of spoken and written genres, including fiction, journalism, and academic publications, and they inform the design of virtually all recent major English dictionaries and grammar reference books. Though entries in the weighty Longman Grammar of Spoken and Written English (1999), for example, were derived from analysis of the Longman Spoken and Written English Corpus (over forty million words), the New Oxford Dictionary of English (2001) compilers drew upon the 100 million-word British National Corpus.

Several large English corpora are now available for purchase or subscription,[1] and newcomers to corpus analysis can also try simple corpus queries online for free. At the British National Corpus website, for instance, concordancing software will search the second edition of the corpus for any given key word or phrase, and provide up to 50 contexts.[2]

Corpora are particularly useful for establishing the frequency and co-occurrence of lexical and grammatical features, and it is obvious that lexicographers and grammarians, who seek to describe an entire language system, or significant parts of it, need to work with a large and representative collection of texts, not all corpus research is of this nature, however. Specialized corpora of only a few thousand words may be used to investigate features typical of specific genres or fields, although large corpora may be used to corroborate findings from the qualitative analysis of a much smaller body of texts.[3]

Recent studies have drawn on specifically compiled corpora to identify and examine generic features of academic writing, such as discourse structure, rhetorical strategies, citation practices, and lexical choice. These corpora, however, tend to consist solely of professionally edited and expertly written text, perhaps because expert text is in the public domain and therefore relatively easy to obtain. The Test of English as a Foreign Language (TOEFL) 2000 Spoken and Written Academic Language Corpus, for example, which claims to represent "the full range of spoken and written registers used at United States universities" (Biber et al., 2002, p. 11), contains textbooks, course packs, university web pages, and similar expert sources, but no examples of student writing.

Expert text is, of course, an important area for research, as it constitutes a model to which student writers ultimately aspire, and which they will repeatedly encounter in their program of reading. Novice writers do not, however, begin by writing for publication, or for a readership of strangers. Their early attempts at academic writing are more likely to be assessed texts produced in the context of a course of study. Although there are undoubtedly generic similarities between the student assignment and the published paper, there are also significant differences in their communicative purposes and rhetorical features.

Far more academic writing is produced for assessment purposes than for publication purposes, but because of the lack of a suitable corpus, research into the generic features of published academic writing vastly outweighs research into the generic features of assessed student writing. Moreover, those few corpus-based studies of assessed student writing that have been undertaken tend to focus on the work of native speaking students in the final stages of study, at the end of a postgraduate program,[4] or on writers in the process of acquiring English as a second language[5] (ESL) rather than on the proficient writing of native or near native speakers during the course of their academic development.

Clearly, the lack of suitable corpus data has not prevented researchers from investigating this type of writing. There have been many excellent studies of assignment genres and student writing processes, which have not required access to a large, representative body of student work. Questionnaires, case studies, observations, diaries, focus groups, and interviews with student writers and subject specialists can provide valuable insights, as can the close analysis of small samples of text produced by individual writers. Nevertheless, we believe that a corpus is needed to support our work as researchers, to corroborate findings from small qualitative studies, to triangulate data collected by other means, and to provide strong quantitative insights into student writers' use of grammar, lexis, and discourse patterns across the disciplines.

2. The University of Warwick student writing corpus project
The purpose of the corpus project at the University of Warwick was to collect and index as many samples of proficient, academic writing as possible during an 18-month period. As such, the research has developed within a British academic context, although contributors to the corpus come from all over the world.

The corpus is not currently accessible via the World Wide Web. It can, however, be made available to writing teachers and researchers upon request.6 Student papers are stored in word document form and as text files, and are grouped according to the disciplines and faculties at Warwick University: humanities, social studies, and sciences. Researchers can search the corpus using any standard concordancing program such as *WordSmith Tools* (Scott, 1996). In the future, additional information will be added in coded form to each text file, so it will be easier to identify and search text subgroups both within and across disciplines.

Between May 2001 and November 2002, we employed a research assistant to collect assignments from students each Wednesday afternoon throughout the term. The administrative costs of collecting the students' work and paying the students for their contributions during this time were met through a research award from our university's research and teaching development fund. There was no set limit to the number of assignments we planned to collect, though it was, at first, hoped that approximately one thousand samples of writing would be amassed during the 18-month project. The current corpus holdings are actually smaller than this, but substantial funding has recently been obtained from the Economic and Social Research Council in the UK. This will allow the database to expand to many times its present size. Nevertheless, it is felt that the current holdings, obtained during the initial 18-month period, in themselves form a self-contained set of papers that merit the attention of researchers and can also be drawn on by teachers to devise writing tasks for their students.

It was intended that the writing samples collected should be indicative of the standards set by departments for each phase of academic study. For this reason, we decided to only accept assignments that had already been assessed and had received a mark of at least 65% or an II i, equivalent to a B+ or above on the North American grading scale. There are many different kinds of writing that students are engaged in across disciplines, and itwas hoped that the corpus might reflect something of this diversity. However, the kind of writing most widely submitted was the essay-type assignment. The length of the assignments collected was important, as we did not want to collect (and pay for) very short pieces of writing. Nor, at this stage, were we interested in dissertations or theses, as these are often edited by supervisors, and are more akin to published academic writing. For this reason we stipulated that submissions should be between one thousand and five thousand words in length.

To provide an incentive to students to submit their writing, we advertised for contributions through fliers and posters placed around campus. The fliers and posters provided brief details of the project and directed students to our website for further information. Early attempts were made to use other advertising media, such as a campus-based radio station and an events magazine published by the students' union, but these generated less interest, and so their use was subsequently abandoned. We discovered that the best way to attract contributors was via departmental contacts personally known to the researchers, in part because this kind of advertising could be timed to coincide with the return of specific assignments. Assignments are returned at different points in the academic session according to discipline and program of study, and we found it useful to be knowledgeable about this schedule. A further incentive to contributors was that in exchange for each assignment submitted in electronic form, we offered a small cash sum (£1 per assignment, later raised to £2). Though modest, this payment became an incentive for students to submit as many assignments as possible.

There were two main administrative processes that had to be followed for each submission. First, contributors were asked to bring hard copies of their assignment, because these gave proof of the marks awarded, and also to bring an electronic version of the text. Contributors then completed a paper-based submission form, detailing their gender, first language, department, course of study, assignment title and the year of study the assignment was written, as well as the name of the tutor who set the writing task. Importantly, the form also asked students to transfer copyright of their work over to the research project. These paper-based forms are retained permanently on file. The students did not put their name on the form, but a code was placed on the form that matched up with the input code placed on the database spreadsheet.

The second part of the administrative procedure involved direct use of the computer itself. In order to preserve the students' writing, the files were temporarily stored on the computer, though they were also backed up on the university's server in case of potential system failure. Periodically, the files were saved to CD-ROM. Contextual details about the contributor and his or her work were stored on an Excel spreadsheet. For each assignment, writer information from the previously detailed paper-based submission form was entered onto the spreadsheet.

To help preserve anonymity, each text file saved to the database was listed under an ID code rather than the contributors name. This code consisted of a series of letters indicating the contributor's department (for example, ET for the Centre for English Language Teacher Education [CELTE], PS for psychology, and HI for history), a four digit number and the first two letters from the contributor's first and last name. Assignments submitted by the same student had the same name code, but different numbers. For example, if Mary Evans submitted three assignments in history, and these were the third, fourth, and fifth assignments to be submitted in that particular discipline, then the assignments would be saved as HI0003, HI0004 and HI0005, followed by the input code "maev."

During the 18 months of the project, it was relatively easy for a clerical assistant to maintain the database on a weekly basis by simply saving new papers to disk according to faculty, backing them up on the university server, adding contextual information for each new assignment to the spreadsheet, and keeping the paperwork updated. These procedures were neither highly sophisticated, nor especially technical in nature. However, in order to take the project to its next stage of development and enable large-scale linguistic analysis we need to add coded information to the files. This is particularly the case as the corpus is likely to grow substantially from its present size and will include papers from other institutions. Files converted to text format will be encoded so contextual and statistical information can be included, and so structural and presentational information available in the original format will not be lost.

3. Current corpus holdings

Thus far, 70 students have contributed a total of 499 assignments to our database. As illustrated in Table 1, written assignments held in the corpus represent 18 departments, with the largest contributions coming from psychology (22%) and history (21%). For the purposes of this paper we have followed University of Warwick faculty divisions and have separated the holdings into humanities, social studies, and science. It can be seen that the majority of contributions are from the humanities, with significantly lower percentages from science and social studies—submissions from chemistry, physics, computer science, mathematics, and statistics are entirely lacking. This imbalance is perhaps a reflection of the fact that students in the humanities and social sciences produce more written work than science students. Nevertheless, our aim is to create better representation across disciplines, with an equal quantity of corpus holdings for the less prolific subject areas. We also note that a number of humanities and social sciences departments are currently unrepresented in our corpus, for example education, classics, film and television studies, American studies, and cultural studies.

Humanities	Social Studies	Science
English Literature (58)	Economics (11)	Biological Sciences (10)
French (12)	CELTE (49)	Engineering (18)
History (106)	Business (17)	Psychology (110)
History of Art (11)	Philosophy (7)	
Italian Studies (3)	Politics (18)	
Theatre Studies (19)	Law (20)	
	Sociology (21)	
	Women's Studies (9)	
Total 209 (42%)	Total 152 (30%)	Total 138 (28%)

Table 1: Breakdown of corpus contributions by faculty and discipline

Table 2 lists contributions by the year of study the assignment was completed and assessed. Most assignments are from the first year of undergraduate study, with progressively fewer from second-year and third-year courses. This distribution reflects the fact that most current third year students have a collection of writing assignments from their previous years at university upon which to draw, and second-year students will have essays to contribute from their first year as well their current year. Students on postgraduate courses have also contributed assignments from their undergraduate study.

220 (44%)	First-Year Undergraduate
96 (19%)	Second-Year Undergraduate
73 (14.5%)	Third-Year Undergraduate
98 (19.5%)	Postgraduate (Master's Level or Diploma)
14 (3%)	Other (mainly foreign exchange students without specific year status)

Table 2: Breakdown of corpus contributions by year of study

Postgraduate master's or diploma students contributed a surprising number of assignments—many of these (50% of the total for this category), however, were submitted by students in the CELTE, the host department for our study. Our aim now will be to gather a more representative collection of second-year, third-year, and postgraduate assignments to facilitate comparison between stages of academic development.

A further issue of representation is that of language. Although 27% of our contributors did not list English as their first language, we hesitate to attach too great a significance to this. The University of Warwick is a multicultural, multilingual environment, and in their departments students are assessed on merit, without regard for their language background. Many international student contributors were native-like in their speech and writing, the submission form being the only indicator of their non-native speaker status. The corpus provides opportunities to investigate the possible influence of mother tongue, but we assume that all contributors are proficient users of English, given that their assignments have been awarded high grades.

4. Current research based on the corpus

As has already been stated, our corpus is still under development, but promises to provide useful research data. We plan to join forces with other universities to collect assignments in a wider range of discipline areas than the University ofWarwick affords.With the aid of external grant funding, we intend to eventually collect, file, and tag five million words of proficient, assessed work, for deposit in a text archive accessible to researchers everywhere. However, even in its present form, the database is useful for researchers. Studies of citation practices (Nesi, 2002a, 2002b), and writer's stance (McKenny, 2003) are underway. The database has also formed part of a wider study on the value of corpora in designing teaching materials (Nesi, 2002c) and has recently been matched as a control for a learner corpus produced by international students at Simon Fraser University in Canada.

The database is not, however, intended solely for academic professionals researching student writing. As a teaching resource, staff working on the University of Warwick's pre-sessional, insessional, and one-year pre-university foundation programs for non-native speakers of English have used the corpus. The assignments serve as models of the kind of writing required at degree level, and provide invaluable insights into departmental assessment procedures and practices that are of value to both teachers and learners.

Samples of writing drawn from the database have already been used directly in writing instruction. In a recent series of language support classes for one-year international postgraduate students at Warwick held from October to December 2003, a number of short extracts were drawn from the database to demonstrate a variety of writing features that were being discussed in class. These writing features included levels of formality, use of the first person pronoun, the way sources are incorporated into the text, and use of citations. In the language support classes, students were put into small groups and encouraged to discuss how writers in different subject areas approach these features of writing differently, and as a result, to consolidate their own approach to such issues. For example, the students found that there was virtually no use of the first person pronoun in psychology papers, but students following teacher education courses used the first person pronoun more frequently, because it lent greater interactivity to their own writing.

Initial trialling of this kind of teaching material in the classroom situation has been encouraging. We intend to further develop such materials into a volume in the *EASE* series of CD-ROMS, designed to develop academic skills in English (Kelly, Nesi, & Revell, 2000; Kelly, Nesi, & Richards, 2004). By using extracts from the database for a CD-ROM, students will be able to work independently of the teacher, though continuing to engage interactively with material drawn from the corpus. For example, if the students are working on their own (without the instructor) on an exercise where they are asked to identify communicative

purpose in a series of extracts, they can be given instant feedback by clicking on "done", to see whether their judgment of the writing is correct. The feedback can be quite detailed, giving information about the specific rhetorical features of a given extract. It can thus raise the students' consciousness regarding specific aspects of composing texts.

5.   Rationale for the University of Warwick database

Although student writing has become a focus for study and teaching in United Kingdom universities only in recent years, the British educational system has traditionally relied on students' ability to produce extended pieces of writing. At the secondary school level, General Certificate of Secondary Education (GCSE) and A-Level examinations[7] in most subjects require students to develop an argument or state a case (typically in the form of an essay or other genre of analytical paper), rather than to respond through testing formats such as multiple choice, true/false, or fill-in-the-blank. At university, essays and other types of extended writing are viewed as integral to student learning and function as a primary method for ongoing, in-course assessment as well as final examination. At the University of Warwick, for instance, most humanities and social science students write papers between fifteen hundred and five thousand words throughout the course of their undergraduate study, and final examinations commonly take the form of essay-response. Furthermore, most departments, including the sciences, require or offer the option of writing a five thousand to ten thousand word final-year undergraduate dissertation or report, in which content knowledge as well as argumentative ability are meant to be displayed.[8]

Although writing is valued as such in British universities, it has tended to be viewed as a transparent medium; thus, providing students with explicit criteria and guidance on written assignments has not been a priority. A key problem that writing teachers often experience, and which may be redressed by a well-designed corpus, is that there is still too little known about the writing requirements of individual university departments (Kusel, 1992). In many departments, for instance, the essay predominates, but other types of assessed writing should not be ignored. A recent survey of three subject areas in British universities (education, English, and engineering) recorded 64 varieties of writing, including tasks as diverse as business plans, website texts, written material to support visual work, classroom observation notes and lesson plans for student teaching, and text that forms the basis for scientific posters (Ganobcsik-Williams, 2001).

The lack of departmental practice knowledge is exacerbated by the fact that so few writing course books available in the United Kingdom use student-generated writing as models. Those intended for a second language writer audience, for instance, often draw on generalized, semi-authentic and stylized academic texts (e.g., Jordan, 1990). Alternatively, they focus on the published research paper (e.g., Swales & Feak, 1994, 2000) or on journalistic discourse (e.g., Trzeciak & Mackay, 1994). Few course books differentiate sufficiently between published and unpublished academic writing, despite their obvious differences in readership and communicative purpose. Even in cases where student writing has been directly incorporated into teaching materials (e.g., Oshima & Hogue, 1999), the context from which these samples are drawn and their relevance to the discourse community of the student are not always apparent or adequately explained.

A corpus of student writing can serve as an awareness-raising tool for writing teachers in universities, schools, and colleges, and can thus accord with an academic literacies approach to writing instruction (Jones, Turner & Street, 1999; Lea & Stierer, 2000). The term "academic literacy" was coined in the research of Norman Fairclough (1992), and it refers to a whole range of skills—including reading and writing skills—that participants in academic communities must acquire to succeed (Hyland, 2000).

Useful as it is, however, for teachers to identify different genres of academic writing (see Bloor, 1996), it is also necessary to recognize that "variation is found not only from genre to

genre, but also within genres" (Hewings & Hewings, 2001, p. 71). Ann Hewings and Martin Hewings argued that there is a danger of stereotyping writing within a particular discipline in a crude and unrealistic way. Identifying the types of assessed work set in an engineering department, for example, should be only the first stage in the process of designing a writing syllabus. This stage should be followed by a consideration of more subtle variations between different courses within the department, and variation in writing produced at different stages of academic study.

Importantly, too, a corpus of student writing such as ours provides data to enable writing developers to chart the growth patterns in student texts across years of study in a subject discipline, and, on a larger scale, to judge if the maturation process is similar across disciplines. Just as student writing changes from A-Level to university (Scott, 2001), students write differently—and have different expectations placed upon their writing—in their first year at university than in subsequent years. Preliminary findings from a longitudinal study of the undergraduate writing at Harvard University (Sommers, 2000), for example, suggested that although the writing of first-year students tends to be descriptive, in that it uses research sources to affirm students' own ideas and to demonstrate student comprehension, the writing of second-year students begins to move beyond summary to interpretation, and develops a more "questioning disposition" toward sources. These findings are corroborated by responses of staff surveyed in British universities and higher education colleges, who noted a shift from descriptive to analytical writing, and an increase in confidence and critical evaluation of source materials between first-year writing and the writing of degree finalists (Ganobcsik-Williams, 2001). Our database of assessed student writing has the potential to chart growth patterns such as whether students' arguments became more complex as their education advanced, whether students learned to integrate material from different sources in formulating conclusions, and whether students' vocabulary became more specialized and precise.

6. Applications of corpora in process writing research

The publication of Janet Emig's seminal report, *The Composing Processes of Twelfth Graders* (1971), inaugurated the process research movement within composition studies in North America. By the early 1980s, this new awareness of writing as "recursive" rather than "linear", and the importance of attending to composing processes as well as written products, was hailed as a paradigm shift in writing research and pedagogy (Perl, 1994). Over the last 30 years, investigations into how people learn to write have taken many forms, from empirical to anthropological, and as a result of this ongoing research scholars have discovered the need to broaden and contextualize their understanding of composing processes. Since the late 1980s, studies have tried to incorporate factors such as the view of writing as a social rather than individual act, the effects of writing contexts, and the relevance of the choices and assumptions that shape research designs (Perl, 1994).

One characteristic of writing corpora is focus on product, and this might be seen as a limitation for the use of corpora in academic writing research into composing processes. Whereas the assignments in most writing databanks are assessed products, a significant line of enquiry for writing developers is writing processes: We are interested in how to teach students to compose, and in learning at which points during the composing stages feedback intervention is most effective. Hence, from a process research point of view, it would be ideal if writing corpora could be designed to include, for instance, assignment guidelines, student plans and drafts, and teachers' comments. Such corpora would be of direct use for rhetorical models of research and teaching, in which knowing the audience, context, and teacher's expectations for a given assignment is key.

We should bear in mind, however, that teacher expectations are not always made explicit to students. Indeed, feedback from teachers tends to focus on content rather than on language and organization, and student writers do not always adopt traceable drafting techniques. For

this reason, we may never succeed in acquiring a complete picture of the writing process as it occurs in university departments, rather than in the writing class. Nevertheless, simple collections of written product can still play a useful role in process research. Scholars can make use of writing corpora to form initial impressions across a spectrum—or alternatively, within a specific cohort—of student papers, and then explore these impressions more fully through process-oriented research. For example, having noted students' use of particular transitional constructions when integrating quotations into essays, a researcher could hypothesize about possible reasons for these features, and test these hypotheses by observing students as they write, or examining a series of drafts. A large collection of student writing could also be used to prove quantitatively that a given feature was widespread, confined to certain groups of writers, or representative of the writing in certain disciplines. This finding could then be investigated more deeply in a small-scale qualitative study. Thus, although corpora alone do not provide a complete picture of writing processes or of the range of writing activities in a university, they can provide a basis for generating and grounding more than one type of writing research.

7. Conclusion

In the United States, where composition and rhetoric is an established discipline, corpus studies have not been a hallmark of research on student writing. In Great Britain, "Academic Writing" is a young field, and writing practitioners and researchers work out of a range of subject areas, including linguistics, English language teaching, education, literary studies, and student and staff development.[9] Perhaps because of the diversity of approaches and methodologies these writing developers bring to the study of writing, it is not surprising that our project has arisen from the British context: Our corpus database has grown out of our collaboration between an EAP (English for Academic Purposes) lecturer in the Centre for English Language Teacher Education (Gerard Sharpling), a corpus linguist (Hilary Nesi), and an American compositionist in the Warwick Writing Programme in the Department of English (Lisa Ganobcsik-Williams). Although the Warwick corpus of assessed student writing reflects the educational context and hybrid research culture of academic writing in Great Britain, it is our belief that the database can serve as a model and a research tool for researchers within, but also beyond, the United Kingdom.

Notes

1. See, the British National Corpus resources page <http://www.natcorp.ox.ac.uk/corpora.html>.
2. See the British National Corpus page <http://sara.natcorp.ox.ac.uk/lookup.html>.
3. Caldas-Coulthard (1993, p. 198), for example, critically examined gender representation in a limited number of media sources, referring to the huge COBUILD corpus "to back up and confirm my findings."
4. Dudley-Evans (1986) examined the rhetorical moves in sections of a collection of MSc dissertations, while Thompson (2000) and Thompson and Tribble (2001) studied citation patterns in a small corpus of doctoral theses.
5. The International Corpus of Learner English (ICLE), for example, is a two million word collection of argumentative essays produced by advanced learners of English.
6. Researchers who are interested in accessing the data should contact the authors.
7. GCSE exams are generally taken at age 16, at the end of the first (compulsory) period of schooling. A- or Advanced-Level exams are taken as a pre-requisite for entry to higher education.
8. Some academic departments require lengthier writing tasks for final year undergraduates. The Department of Computer Science, for instance, sets a Group Project Report "typically of length 30,000 words (excluding appendices)", <http://www.dcs.warwick.ac.uk/undergraduate/modules/cs407.html>.
9. For a sense of this variety see the membership of Britain's Writing Development in Higher Education (WDHE) listserv and participants in WDHE conferences: <http://www.jiscmail.ac.uk/lists/writing-dev-he.html>.

References

Biber, Douglas, Conrad, Susan, Reppen, Randi, Byrd, P., & Helt, Marie. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, *36*, 9–48.

Bloor, Meriel. (1996). Academic writing in computer science: A comparison of genres. In Eija Ventola & Anna Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp. 59–88). Amsterdam and Philadelphia: John Benjamins.

Caldas-Coulthard, Carmen. (1993). From discourse analysis to critical discourse analysis: The differential representation of women and men speaking in written news. In John Sinclair, Michael Hoey, & Gwyneth Fox (Eds.), *Techniques of description: Spoken and written discourse* (pp. 196–208). London and New York: Routledge.

Dudley-Evans, Tony. (1986). Genre analysis: An investigation of the introduction and discussion sections of MSc dissertations. In Malcolm Coulthard (Ed.), *Talking about text* (pp. 128–145). Birmingham, AL: University of Birmingham Press.

Emig, Janet. (1971). *The composing processes of twelfth graders*. Urbana, IL: NCTE.

Fairclough, Norman. (1992). *Critical language awareness*. London and New York: Longman.

Ganobcsik-Williams, Lisa. (2001, March). Teaching writing at university: A survey of staff perspectives. In *Paper presented at Teaching writing in higher education: An international symposium*.

Hewings, Ann, & Hewings, Martin. (2001). Approaches to the study of disciplinary variation in academic writing: Implications for syllabus design. In David Hall & Ann Hewings (Eds.), *Innovation in language teaching: A reader* (pp. 71–86). London and New York: Routledge.

Hyland, Ken. (2000). *Disciplinary discourses: Social Interaction in academic writing*. London: Longman Pearson Education.

Jones, Christopher, Turner, Joan, & Street, Brian. (1999). *Students writing in the university: Cultural and epistemological issues*. Amsterdam: John Benjamins.

Jordan, Robert. (1990). *Academic writing course*. Harlow, UK: Longman.

Kelly, Tim, Nesi, Hilary, & Revell, Rod. (2000). *EASE volume one: Listening to lectures*. Coventry, UK: CELTE, University of Warwick [CD-ROM].

Kelly, Tim, Nesi, Hilary, Sharpling, Gerard, & Richards, Lynnette. (2003). *EASE volume two: Seminar skills—oral presentations*. Coventry, UK: CELTE, University of Warwick [CD-ROM].

Kusel, Paul. (1992). Rhetorical approaches to the study and composition of academic essays. *System*, *20*, 457–470.

Lea, Mary, & Stierer, Barry. (2000). *Student writing in higher education: New contexts*. Buckingham, UK: Open University Press.

Leki, Ilona. (1995). *Academic writing: Exploring processes and strategies*. New York: St. Martin's Press.

McKenny, John. (2003). A double take on writer's stance: Content analysis of dogmatism compared with corpus analysis of student essays. In Luuk Lagerwerf, Wilbert Spooren, & Liesbeth Degand (Eds.), *Determination of information and tenor in texts: Multidisciplinary approaches to discourse* (pp. 93–103). Amsterdam: Stichting Neerlandistiek, VU & Munster: Nodus Publikationen.

Nesi, Hilary. (2002a). Citation in the electronic age: Academic norms and current problems. In paper presented at the *36th International IATEFL Conference*.

Nesi, Hilary. (2002b). Current problems with citation practices. In paper presented at the *35th BAAL Annual Meeting*.

Nesi, Hilary. (2002c). BASE, BAWE and EASE: How insights from academic corpora have informed the design of an EAP series. In paper presented at the *First Inter-Varietal Applied Corpus Studies (IVACS) Group International Conference*.

Oshima, Alice, & Hogue, Ann. (1999). *Writing academic English*. New York: Addison Wesley Longman.

Perl, Sondra (Ed.). (1994). *Landmark essays on writing process* (Vol. 7). Davis, CA: Hermagoras Press.

Scott, Mary. (2001, March). Students in transition: Writing about literature at A level and at university. In paper presented at *Teaching writing in higher education: An international symposium*.

Scott, Mike. (1996). *WordSmith Tools*. Oxford: Oxford University Press.

Sommers, Nancy. (2000, July). Something more, something deeper: The experience of undergraduate writing. In paper presented at the *51st Annual Conference on College Composition and Communication*.

Swales, John, & Feak, Christine. (1994). *Academic writing for graduate students*. Ann Arbor, MI: Michigan University Press.

Swales, John, & Feak, Christine. (2000). *English in today's research world: A writing guide*. Ann Arbor, MI:University of Michigan Press.

Thompson, Paul. (2000). Citation practices in PhD theses. In Lou Burnard & Tony McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 91–101). Frankfurt, Germany: Peter Lang.

Thompson, Paul, & Tribble, Christopher. (2001). Looking at citations: Using corpora in English for academic purposes. *Language Learning and Technology*, 5, 91–105.