

Issues in the development of the British Academic Written English (BAWE) corpus

Alsop, Sian and Nesi, H.

Published PDF deposited in CURVE March 2013

Original citation & hyperlink:

Alsop, Sian and Nesi, H. (2009) Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, volume 4 (1): 71-83

<http://dx.doi.org/10.3366/E1749503209000227>

Publisher: Edinburgh University Press

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the published PDF of the journal article.

CURVE is the Institutional Repository for Coventry University

<http://curve.coventry.ac.uk/open>

Issues in the development of the British Academic Written English (BAWE) corpus

Sian Alsop¹ and Hilary Nesi²

Abstract

The British Academic Written English (BAWE) corpus is a collection of texts produced by undergraduate and Master's students in a wide range of disciplines, for assessment as part of taught degree programmes undertaken in the UK. The majority of the contributors to the corpus are mother tongue speakers of English, but, in order to be included in the corpus, each assignment had to be judged proficient by assessors in the contributor's discipline, regardless of the writer's mother tongue. The corpus contains, therefore, only texts that have met departmental requirements for the given level of study.

University writing programmes are typically aimed at undergraduate and Master's students, and it would be useful for writing tutors to know more about student assignment genres and the linguistic features of successful writing at undergraduate and Master's level. However, most large-scale descriptive studies of academic writing focus on published or publicly accessible texts, or learner essays on general academic topics, probably because there are practical difficulties associated with collecting large amounts of well-documented student output. This paper charts the experience of collecting data for the BAWE corpus, highlighting the problems we encountered and the solutions we chose, with a view to facilitating the task of future developers of academic student writing corpora.

1. Introduction

The British Academic Written English (BAWE) corpus³ was designed to fill a gap in current corpus resources by complementing academic writing

¹ Freelance.

² Faculty of Business, Environment and Society, Coventry University, Priory Street, Coventry, CV1 5FB, United Kingdom.

Correspondence to: Hilary Nesi, *e-mail:* h.nesi@coventry.ac.uk

³ BAWE was created as part of the project *An Investigation of Genres of Assessed Writing in British Higher Education*, conducted at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics (previously known as CELTE), Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800).

collections representing published and/or publicly available texts, on the one hand, and non-discipline-specific learner output, on the other.

Most academic corpora tend to be made up of professionally edited and expertly written text. The PERC Corpus of Professional English, for example, (Rayson *et al.*, 2005; Noguchi *et al.*, 2006) will consist of academic journals, trade magazines, textbooks and web pages. Similarly the TOEFL 2000 Spoken and Written Academic Language Corpus is composed of sources such as textbooks and course packs. Although the TOEFL 2000 claims to represent 'the full range of spoken and written registers used at US universities' (Biber *et al.*, 2002: 11) it fails to include any texts produced by students.

Until now, most publicly available corpora of student writing have consisted of general essays produced in the writing class or under examination conditions, and have been designed primarily to monitor non-native-speaker errors and the processes of language acquisition, rather than the development of academic literacy skills and disciplinary knowledge. Well-known examples are the International Corpus of Learner English (see Granger *et al.*, 2002) and the Louvain Corpus of Native English Essays (see Granger and Tyson, 1996). Discipline-specific student writing, produced for assessment within the student's own department, has tended to be collected for individual scholarly purposes rather than as part of formal corpus-building projects; see, for example, Woodward-Kron (2004), Moore and Morton (2005), North (2005) and Samraj (2004, 2008), all of whom refer to small private collections of student assignments.

The BAWE corpus is intended to enable the identification and description of student writing genres across disciplines and at different stages of academic development. It currently appears to be the only formally planned and archived corpus of its kind, although the Michigan Corpus of Upper-level Student Papers (MICUSP; Ädel, 2006a, 2006b) is now nearing completion (at around two million words) and a similar corpus at Portland State University is in the early stages of development (Conrad and Albers, 2008).

The main BAWE corpus was designed with reference to a pilot corpus constructed between May 2001 and November 2002.⁴ The pilot corpus contained 499 assignments from eighteen university departments, and has been referred to in several studies (e.g., McKenny, 2005a, 2005b). Its relative success highlighted the need for a larger corpus with more contextual annotation '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' (Nesi *et al.*, 2004: 443). However, the pilot corpus also illustrated the difficulty of collecting a representative selection of work from a shifting student population, who produced varying amounts of writing at various stages in the academic year, and who had relatively

⁴ This was funded by the University of Warwick Research and Teaching Development Fund.

little incentive to co-operate with our research agenda. This paper overviews the strategies and processes we employed in an attempt to achieve greater balance and representation in the full-scale project. It is hoped that future corpus compilers can learn from the mistakes and the solutions recorded here.

2. Corpus design

Assignments for the main BAWE corpus were collected at Warwick University, Reading University, Oxford Brookes University and, towards the end of the project, at Coventry University. Most assignments were collected in 2005 and 2006, with a few late additions in 2007.

Of the various sampling methods we considered, simple random sampling would have been the most statistically valid way of achieving representation, had it been possible to identify the full range of assignments produced within each of the participating universities, and to acquire a proper sample from this resource pool. Unfortunately, we had no real means of assessing the volume or nature of assignments that would be at our disposal, and we knew from our experience with the pilot corpus that if we invited all students to submit all their work until the desired corpus size had been reached, we would have a very unrepresentative sample, produced only by those students who found it convenient to call into the project offices, and/or those who had particular sympathy with the project aims.

We decided, therefore, to use a matrix of four disciplinary groupings and four levels of study to plan the corpus structure, forming sixteen cells of approximately equal size. The groupings (Arts and Humanities, Life Sciences, Physical Sciences and Social Sciences) were intended to facilitate comparison with two influential corpora of academic spoken English: the Michigan Corpus of Spoken Academic English (MICASE) and the British Academic Spoken English (BASE) corpus. Although neither of these corpora represent exactly the same range of disciplines, they divide their holdings in much the same way: BASE groupings have the same names as BAWE groupings, and MICASE groupings are similar (Biological and Health Sciences, Humanities and Arts, Physical Sciences and Engineering, and Social Sciences and Education). MICUSP (see Introduction), the Michigan academic student writing corpus that is still under development, uses the same divisions as MICASE.

Four levels were used to identify the stage in the degree programme that a student had reached at the time of writing the assignment. Most undergraduate courses in British universities last for three years, and most taught Master's courses last for one year. In the case of a four year undergraduate course, year three is often spent abroad or in a job placement located outside the university. Some four year courses, however, consist of three years of undergraduate level study and a final year at Master's level. Assignments written in the fourth year of university registration were thus categorised according to the weight attributed to them by the department, as

either level three (if year three had been an intercalatory year) or level four (Master's level). Information about levels was provided by contributors in the final stages of submission. Master's dissertations (theses) were excluded from the corpus, on the grounds that they were much longer than most of the other assignments the students wrote, and often underwent a process of drafting, redrafting and review which made them more like published texts.

We initially intended to represent as many disciplines as possible across the three universities, but we discovered that departments varied in cooperativeness, the size of the student enrolment, particularly at Master's level, and the extent to which written work was produced in digital form. Our final sampling scheme listed the seven potentially most productive disciplines in each disciplinary grouping, plus an 'other' category in which we would place a few assignments that belonged to other disciplines or were surplus to requirements. Modules are not a perfect match with disciplines – economics departments, for example, deliver modules in mathematics – but, for the purposes of this project, we treated every assignment produced for every module taught by staff belonging to the same department as belonging to the same discipline. Initially, each discipline was represented by a single department in a single university, but later in the project we mixed contributions to the same discipline from different universities, in order to make up numbers.

Table 1 represents the sampling scheme, with a target of 3,500 assignments. Not all the cells were completely filled in the end, as will be seen later. In each grouping, six disciplines were considered 'core' and one 'peripheral' (in brackets). Two disciplines were treated somewhat differently from the rest: engineering was double weighted because of the size and diversity of the engineering department, and all the assignments in medicine were counted as level four, because the Medical School was graduate entry.

Assignments were required to be written in English, and had to be submitted electronically. They also had to meet a certain proficiency standard, as judged by the students' subject tutors. Both formative and summative assignments⁵ were accepted for the corpus, provided that department staff had awarded them a mark equivalent to 60 percent or more. We asked contributors to provide proof of their grades in the form of, for example, a mark sheet or transcript. Where proof was not available, we randomly selected cases to confirm with departmental staff; we also checked the grade of any submitted assignments that seemed to us to fall below the required level of proficiency.

Grading practices varied across departments and modules, and we did not want to use numerical grades as statistical data. Thus the file header information we created for each assignment identifies it as either 'distinction' (D), if given a grade equivalent to 70 percent or above, or 'merit' (M), if

⁵ Nesi *et al.* (2005: 10) make the distinction between 'formative assessment, which will not contribute to the final grade for the module, and summative assessment, which will determine the grade for the module'.

<i>Disciplinary Grouping</i>	<i>Subject</i>	<i>Per Yr (1, 2, 3/4, 4/M)</i>	<i>Total</i>
<i>Arts and Humanities</i>	Linguistics	32	128
	Classics	32	128
	Comparative American Studies	32	128
	English Studies	32	128
	History	32	128
	Philosophy	32	128
	(Archaeology)	16	64
<i>Life Sciences</i>	Agriculture	32	128
	Biological Sciences/ Biochemistry	32	128
	Food Science and Technology	32	128
	Health and Social Care	32	128
	Plant Biosciences	32	128
	Psychology	32	128
	(Medicine)	64	64
<i>Physical Sciences</i>	Architecture	32	128
	Chemistry	32	128
	Computer Science	32	128
	Cybernetics and Electronic Engineering	32	128
	Engineering	64	256
	Physics	32	128
	Sciences	16	128
<i>Social Sciences</i>	Business	32	128
	Economics	32	128
	Hospitality, Leisure and Tourism Management	32	128
	Law	32	128
	Politics	32	128
	Sociology	32	128
	(Anthropology)	16	64
<i>Other</i>	Other	43	172
			3,500

Table 1: Our plan for the BAWE corpus

graded between 60 percent and 69 percent. Assignments for the Medical School tended to be classed as ‘excellent’ or ‘satisfactory’ rather than receiving a percentage grade; most of the assignments we accepted from medicine were in the ‘excellent’ category (identified as D), but a few were ‘satisfactory’ (identified as M).

Certain other criteria also governed the acceptance of assignments. To avoid over-representation by individual writers it was initially agreed that contributors could submit no more than three assignments from a single level, no more than ten in total across four levels, no more than two from any single module, and up to five at Master’s level. However, this proved to be

too limiting to attract sufficient numbers of assignments, especially in those disciplines where student enrolment was quite low; so we changed the rules to allow individuals to submit up to five assignments from any single year (including Master's level), no more than ten in total at undergraduate level, and no more than three from any single module. Assignments contributed by the same student, but in different disciplines, constituted a different count of ten, so students undertaking joint honours courses and modules outside their home department had the opportunity to submit more than ten assignments. In practice, however, only forty-seven contributors out of a total of 627 submitted ten assignments, and only five submitted more than ten (the highest number received from a single contributor was eighteen).

As a rough-and-ready means of ensuring a wide variety of genres in the corpus, contributors were asked to identify the *type* their assignment belonged to, from a choice of 'case-study', 'essay', 'exercise', 'notes', 'presentation', 'report', 'review' and 'specified other', and no more than ten assignments of the same self-reported type were accepted from any single module. Text types often failed to match the labels students themselves used metatextually, however; for example, an assignment identified as an 'essay' might begin with the words 'In this report', or an assignment identified as a 'report' might begin with the words 'In this essay'. We decided, therefore, not to include the contributors' *type* labels in the header information provided for each file in the final corpus, and assigned files 'genre family' labels instead, derived from our own examination of the entire dataset (see, for example, Nesi, 2008).

3. Collection practice

The first stage in the collection process involved the electronic submission of a single document, usually as a Microsoft Office, plain text or PDF file. At the outset of the project this process was carried out entirely through e-mail correspondence. The student and the research assistant at each university would have several exchanges in order to clarify issues relating to the assignment and the process. This proved time-consuming, however, and it soon became clear that students were losing their initial enthusiasm due to the time and effort involved. Indeed, many potential assignments were lost to the corpus in this way, before the final stage in the process had been completed. Also, the mass submissions that often occurred following a particularly successful advertising strategy could not be dealt with efficiently through the e-mail system. We therefore established an online system which allowed students from any of the universities involved to attach multiple assignments to a single submission page at the University of Warwick. The majority of the required contextual information could be provided here by means of a drop-down menu, and students could access information concerning the project and its aims, view assignment quotas (which were regularly updated as submissions were accepted), and receive automated confirmation of receipt

of their assignments. The system was also helpful at the processing stage, since it saved the details of all the assignments and contributors.

Once it had been established that an assignment met our submission criteria, we required certain contextual information to complete the process. For example, although assignments from speakers of languages other than English were accepted, all contributors were required to state their first language and number of years of secondary education in the UK. This information was eventually transferred to the file header, so that corpus users can use it to filter assignments if they wish. Contextual information valid for all the assignments submitted by the same contributor was supplied at the online submission stage, and was then transferred to a database which automatically created a copyright disclaimer form for each assignment. In the interests of efficiency, the student only filled in a single information page for all of his or her submissions at the online stage, however; so any fields that varied according to the assignment (such as 'module tutor' or 'grade') were either completed by the research assistant when the assignment was anonymised (as this information was frequently provided within the body of the assignment), or manually by the student at the final stage. Contact details and names of contributors were gathered during the collection process in case we needed to check contextual information, but were destroyed at the end of the project.

The assignments attached to the submission page were given an author and assignment reference, and stored electronically as 'awaiting signature'. The student was then notified by e-mail to come to sign the copyright forms and receive payment. The funding budget allowed a small sum to be offered as an incentive for participation.

On completion of the final stage in the process, assignments were transferred to the 'completed' category and were added to the corpus. Following this, the files were encoded and marked-up for analysis, as described in Ebeling and Heuboeck (2007), and in the BAWE corpus manual.

4. Advertising strategies

Dissemination of information concerning the project was approached in a variety of ways. The target departments were contacted (usually in person) and asked to forward an e-mail about the project to all relevant students with a link to the online information and submission page. An e-mail with departmental endorsement was assumed to hold more weight than an e-mail directly from the BAWE team. After this the BAWE research assistants attended lectures for modules within each year group with the highest student attendance, and gave a brief overview of the project, reinforcing the information in the e-mail and providing a human point of contact. Information posters with detachable contact information were also displayed in areas frequented by target students, and stalls were set up in university social spaces to give students an opportunity to ask questions about the project and to increase its visibility.

This first wave of advertising tended to be most successful for disciplines that demanded a high number of written assignments per year—especially Business Studies, Psychology and Law. In these disciplines students were more likely (and able) to submit the maximum number of assignments, making it more profitable for them to take the trouble to go through the submission process. The first wave of advertising was also most successful in attracting contributors from levels one and two.

Students at the end of their studies were more difficult to reach through the university e-mail system. There is a much shorter period for the collection of assignments at Master's level and also in the final year of the undergraduate degree, especially taking into account that many assignments are not written (or graded) until late in the academic year. Moreover, because students' university e-mail accounts are closed soon after graduation, it was very difficult to contact students after the end of the summer term. Once they had graduated, students tended to lose interest in the project, anyway, and were unlikely to return to the university to submit their work. The collection of assignments from Master's students was particularly problematic because Master's students are less plentiful than undergraduates, and in many disciplines Master's programmes do not constitute a seamless continuation from undergraduate study. Master's students therefore failed to submit final year undergraduate assignments, whereas final year undergraduate contributors often submitted assignments they had written in previous years. Many Master's students had taken their first degrees overseas, while British Master's students were often taking a career break, and had lost their undergraduate assignments long ago.

The second stage of disseminating information, which began roughly halfway through the project, had a more focussed approach. In addition to continued mass-mailing through departments, e-mails were sent directly from the BAWE account to targeted potential contributors; departmental secretaries were also asked to hand out fliers when graded assignments were returned to students. BAWE 'open-afternoons', advertised in person in targeted departments, provided information and computer access so that assignments could be submitted and processed on the spot, and contributors could receive immediate payment. Student involvement in sports associations and interest groups also offered a means of collecting multiple assignments, and in an attempt to reach recent graduates, notices were placed in the Warwick Graduate Association Newsletter, and on various online graduate forums. However, although our open afternoons attracted a large number of submissions, the strategy of approaching clubs was only partly successful, and there was relatively little response from the graduates we attempted to contact.

Following our two advertising campaigns, the problem of unbalanced collection across year groups was less severe, but still a matter of concern. In addition, it became clear that certain disciplines in the physical sciences were significantly under-represented in the corpus. A lower rate of contribution from these areas had been anticipated from the start of the project, following the pilot corpus experience. Preparatory interviews

<i>Disciplinary group</i>		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>Total</i>
<i>Arts and Humanities</i>	students	101	83	61	23	268
	assignments	239	228	160	78	705
	texts	259	231	161	83	734
	words	468,353	583,617	427,942	234,206	1,714,118
<i>Life Sciences</i>	students	74	71	42	46	233
	assignments	180	193	113	197	683
	texts	191	208	119	203	721
	words	299,370	408,070	263,668	441,283	1,412,391
<i>Physical Sciences</i>	students	73	60	56	36	225
	assignments	181	149	156	110	596
	texts	186	156	169	129	640
	words	300,989	314,331	426,431	339,605	1,381,356
<i>Social Sciences</i>	students	85	88	75	62	313 ¹
	assignments	207	197	162	202	777 ²
	texts	218	202	169	204	802 ³
	words	371,473	475,668	440,674	688,921	1,999,130 ⁴
<i>Total students</i>		333	302	234	167	1039 ¹
<i>Total assignments</i>		807	767	591	587	2761 ²
<i>Total texts</i>		854	797	618	619	2897 ³
<i>Total words</i>		1,440,185	1,781,686	1,558,715	1,704,015	6,506,995 ⁴

¹ Includes 3 of unknown level.

² Includes 9 of unknown level.

³ Includes 9 of unknown level.

⁴ Includes 22,394 in texts of unknown level.

Table 2: Numbers of students, assignments, texts and words by disciplinary grouping and level

with academic staff had also highlighted the fact that relatively few written assignments are set for physics, chemistry and mathematics. This shortfall demanded that collection of assignments from the physical sciences be opened up to students in similar departments at all three universities, whereas, originally, each discipline had been assigned to a single university.⁶ Shared effort across the universities, and the late introduction of a fourth university (Coventry), boosted the number of contributions and helped to fill some previously empty cells. As a final incentive to contributors, we promoted a BAWE ‘sale’, increasing payment in under-represented areas from three to five pounds per assignment.

⁶ As a matter of policy, information about the source university for each assignment is not provided in the file header or in the accompanying BAWE documentation.

<i>Disciplinary group</i>	<i>Discipline</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>Total</i>
<i>Arts and Humanities</i>	Archaeology	23	21	15	17	76
	Classics	33	27	15	7	82
	Comparative American Studies	29	26	13	6	74
	English	35	35	28	8	106
	History	30	32	31	3	96
	Linguistics	27	31	24	33	115
	Other	19	22	9	0	50
	Philosophy	43	34	25	4	106
	<i>Total</i>	239	228	160	78	705
<i>Life Sciences</i>	Agriculture	35	35	30	34	134
	Biological Sciences	52	50	26	41	169
	Food Sciences	26	36	32	30	124
	Health	35	33	12	1	81
	Medicine	0	0	0	80	80
	Psychology	32	39	13	11	95
	<i>Total</i>	180	193	113	197	683
<i>Physical Sciences</i>	Architecture	2	4	2	1	9
	Chemistry	23	24	29	13	89
	Computer Science	34	13	30	10	87
	Cybernetics and Electronics	4	4	13	7	28
	Engineering	59	71	54	54	238
	Mathematics	8	5	12	8	33
	Meteorology	6	9	0	14	29
	Other	0	1	0	0	1
	Physics	37	14	14	3	68
	Planning	8	4	2	0	14
	<i>Total</i>	181	149	156	110	596
<i>Social Sciences</i>	Anthropology	14	12	6	17	49
	Business	32	33	31	50	146
	Economics	30	30	23	13	96
	Hospitality, Leisure and TM	14	21	29	29	93
	Law	37	37	31	28	134*
	Other	0	2	3	4	9
	Politics	37	33	15	25	110
	Publishing	11	4	0	15	30
	Sociology	32	25	24	21	110 [†]
	<i>Total</i>	207	197	162	202	777 [‡]
<i>Overall Total</i>	807	767	591	587	2,761 [‡]	

* Includes 1 of unknown year.

[†] Includes 8 of unknown year.

[‡] Includes 9 of unknown year.

Table 3: Number of assignments by discipline and year

The final corpus is under target, but not significantly so, as can be seen from Tables 2 and 3. Distribution is not perfectly balanced – this would have been almost impossible to achieve – but is sufficient to inform comparative studies across levels and disciplinary groupings. The greatest inequality is in year 4 (Master’s), where there was greatest variation in student enrolment figures in different disciplines. It should be noted that although the majority of submissions consisted of one piece of writing, with just one introduction and conclusion, in some cases contributors submitted two or more independent pieces of writing as one ‘compound’ assignment, (for example, a series of different lab reports or case studies, or a research report together with a critique), because they formed part of one unit of assessment and had received a single grade. In Table 2, separate figures are given for the number of texts collected and the number of assignments collected.

5. Conclusions

Various methods were employed to gather assignments for the BAWE project, some with better results than others. We regret, now, that we did not take note of students’ information sources for the project systematically. One way of monitoring the success of advertising campaigns might have been to provide slightly different e-mail addresses on advertisements at different times and in different locations. Also, a simple ‘Where did you hear about the project?’ option on the online submission form or copyright disclaimers could have provided a further record of strategy effectiveness.

Data collection proved to be more problematic than originally anticipated. One recurring obstacle was the time lag between the moment at which assignments were conditionally accepted into the corpus (based on quotas at that time) and the point at which the student came to the office to sign the permission form and accept payment (by which time the assignment was sometimes no longer required). Full automation of the submission process, including authorisation and payment, was impossible due to the administrative costs of setting up multiple one-off bank transfers; however, we did manage to finalise some submissions through postal correspondence with students who were absent or studying at a distance.

In terms of future corpus construction, our experiences suggest a need for alternative collection strategies to supplement or replace reliance on financial motivation. Conrad and Albers (2008) discussed the idea of giving students extra credit for submitted work. Compulsory submission might also be a possibility in certain contexts, provided that necessary contextual information could also be collected.

The BAWE corpus is now freely available to researchers, and has been deposited in the Oxford Text Archive, together with accompanying documentation. It is listed as resource number 2539.⁷

⁷ See: <http://www.ota.ox.ac.uk/headers/2539.xml>

References

- Ädel, A. 2006a. 'Citation practices across the disciplines: the case of proficient student writing', Paper Presented at AELFE 5. Zaragoza, Spain.
- Ädel, A. 2006b. 'Metadiscourse in MICUSP: reflexive phraseology in a new corpus of student writing', Paper Presented at ICAME 27. Helsinki, Finland.
- Biber, D., S. Conrad, R. Reppen, P. Byrd and M. Helt. 2002. 'Speaking and writing in the university: a multidimensional comparison', *TESOL Quarterly* 36, pp. 9–48.
- Conrad, S. and S. Albers. 2008. 'A new corpus of student academic writing', Paper Presented at the American Association for Corpus Linguistics Conference, 13–15 March. Utah: Brigham Young University.
- Ebeling, S.O. and A. Heuboeck. 2007. 'Encoding document information in a corpus of student writing: the experience of the British Academic Written English (BAWE) corpus', *Corpora* 2 (2), pp. 241–56.
- Gardner, S. and J. Holmes. Forthcoming. 'From section headings to assignment macrostructures in undergraduate student writing' in E. Swain (ed.) *Thresholds and Potentialities of Systemic Functional Linguistics: Applications to Other Disciplines, Specialised Discourses and Languages Other than English*. Trieste: Edizioni Universitarie Trieste (EUT).
- Granger, S., E. Dagneaux and F. Meunier. 2002. *The International Corpus of Learner English/Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Granger, S. and S. Tyson. 1996. 'Connector usage in the English essay writing of native and non-native EFL speakers of English', *World Englishes* 15, pp. 19–29.
- McKenny, J. 2005a. 'Stance and spin in academic writing' in L. Lagerwerf, W. Spooren and L. Degand (eds) *Determination of Information and Tenor in Texts: Multidisciplinary Approaches to Discourse*, pp. 115–37. Amsterdam: Stichting Neerlandistiek VU; Münster: Nodus Publikationen.
- McKenny, J. 2005b. 'Content analysis of dogmatism compared with corpus analysis of epistemic stance in student essays', *Information Design Journal* 13 (1), pp. 40–9.
- Moore, T. and J. Morton. 2005. 'Dimensions of difference: a comparison of university writing and IELTS writing', *Journal of English for Academic Purposes* 4 (1), pp. 43–66.

- Nesi, H. 2008. 'BAWE: an introduction to a new resource' in A. Frankenberg-Garcia, T. Rkibi, M. Braga da Cruz, R. Carvalho, C. Direito and D. Santos-Rosa (eds) *Proceedings of the Eighth Teaching and Language Corpora Conference*, pp. 239–46. Lisbon, Portugal: ISLA.
- Nesi, H. and S. Gardner. 2006. 'Variation in disciplinary culture: university tutors' views on assessed writing tasks' in R. Kiely, P. Rea-Dickins, H. Woodfield and G. Clibbon (eds) *Language, Culture and Identity in Applied Linguistics*, pp. 97–117. *British Studies in Applied Linguistics*, Volume 21. London: Equinox Publishing.
- Nesi, H., S. Gardner, R. Forsyth, D. Hindle, P. Wickens, S. Ebeling, M. Leedham, P. Thompson and A. Heuboeck. 2005. 'Towards the compilation of a corpus of assessed student writing: an account of work in progress', *Proceedings of the Corpus Linguistics Conference Series* 1 (1). Available online at: www.corpus.bham.ac.uk/PCLC.
- Nesi, H., G. Sharpling and L. Ganobcsik-Williams. 2004. 'The design, development and purpose of a corpus of British student writing', *Computers and Composition* 21, pp. 439–50.
- Noguchi, J., T. Orr and Y. Tono. 2006. 'Using a dedicated corpus to identify features of professional English usage: what do "we" do in science journal articles?' in A. Wilson, D. Archer and P. Rayson (eds) *Corpus Linguistics Around the World*, pp. 155–66. Amsterdam: Rodopi.
- North, S. 2005. 'Different values, different skills? A comparison of essay writing by students from arts and science backgrounds', *Studies in Higher Education* 30 (5), pp. 517–33.
- Rayson, P., Y. Tono, Y. Morita, M. Hoshino, T. Nakamura, H. Aizawa and R. Watanabe. 2005. 'Building a corpus of professional English', Poster Presented at the *Corpus Linguistics 2005 Conference*, 14–17 July. Birmingham, UK. See also: <http://www.perc21.org/menu.html>
- Samraj, B. 2004. 'Discourse features of the student-produced academic research paper: variations across disciplinary courses', *Journal of English for Academic Purposes* 3 (1), pp. 5–22.
- Samraj, B. 2008. 'A discourse analysis of master's theses across disciplines with a focus on introductions', *Journal of English for Academic Purposes* 7 (1), pp. 55–67.
- Woodward-Kron, R. 2004. "'Discourse communities" and "writing apprenticeship": an investigation of these concepts in undergraduate Education students' writing', *Journal of English for Academic Purposes* 3 (2), pp. 139–61.