1 *Investigating student writing with the BAWE corpus*

The book draws on the findings of a four-year study to investigate genres of student writing in higher education[1]. It provides an overview of the kind of writing British university students produce, showing the similarities and differences between writing assignments at different levels and across a range of disciplines. This information will be useful to researchers analysing the discourse of academic writing, to academics concerned with developing writing tasks at university level and to teachers who provide academic writing support to students, whether this is within the context of English for Academic Purposes (EAP) or in writing centres which largely cater for native speakers of English.

The book proposes a system of describing and distinguishing between different types of tertiary-level writing task. We identify and describe thirteen major types of assignment, each of which has a unique purpose and structure, but which is also subject to some variation in response to disciplinary requirements. Readers who devise academic writing tasks can use our descriptions of these assignment types as templates or as a stimulus for thought about the purpose and structure of the writing they expect their students to produce. Our descriptions may help them to distinguish between the different requirements of different writing tasks, and may also help them to make these distinctions clear to their students. Additionally, the descriptions can serve as a reference for writing teachers who are guiding their students towards more appropriate stylistic and organisational choices. The book describes the discourse features of successful assignments in terms of their underlying communicative purpose; successful assignments are those which achieve the intended purpose of the writing task, with due acknowledgement of disciplinary norms and expectations.

1.1 The educational context of university student writing

This book is written at a time of massive expansion in higher education. According to UNESCO (2008) about 138 million students were enrolled in tertiary education in 2005, an increase of 45 million university students worldwide since 1999. This rise has been partly due to population growth, and partly due to widening participation policies. Some countries have made great efforts to attract into higher education young people who have been academically disadvantaged, and to this end have encouraged universities to accept students without traditional university entry qualifications. In some countries a state university place is now guaranteed to all young people who have successfully completed secondary school.

Alongside widening participation there has been a huge rise in student mobility. Wächter (2008) cites UNESCO data indicating that the number of international students globally grew more than fourfold between 1975 and 2005, from 600,000 to 2.7 million. Most mobile students want to be taught in English, a language which they already know from their school studies, and which the international labour market requires. Countries where English is spoken as a first language are popular destinations for these students, but other countries have also gained a share of the international student market by adopting English as an educational lingua franca, for example Malaysia and Singapore (Sugimura, 2008), and non-English-speaking European countries (Wächter and Maiworm, 2008).

Thus, university students around the world are increasingly likely to be using English for their studies, although in many cases their pre-degree preparation will not have included extensive writing practice in English in the relevant genres. These students need to learn how to write well, because writing is the means by which they will construct disciplinary knowledge, the main means by which they will demonstrate their attainment for assessment purposes, and, in many cases, also the means by which they will communicate with professional colleagues in years to come.

However, although writing is probably the single most important skill necessary for academic success, and although we cannot assume that students will have acquired this skill before they begin their university studies, there is considerable confusion amongst students and writing instructors regarding the kinds of writing students are required to produce across disciplines and levels of study. Subject lecturers[2] often fail to make explicit the thinking behind the writing assignments they set, as Haggis (2006) points out, because traditionally student knowledge about genres has been acquired implicitly over time, via a process described by Turner as 'the pedagogy of osmosis' (2011: 21).

Moreover, academic writing programmes often proceed in ignorance of disciplinary genres, as Wardle (2009) discovered when she examined the 'pseudotransactional' assignments set in first year US university composition classes. Wardle rightly concludes that writing classes need to teach students about genres of student writing, and writing teachers need to be able to discern what the key features of these genres are.

1.2 What this book aims to do

This book is divided into two parts. Part I provides the context for our research, by explaining our methodologies and introducing the concepts of genre and genre family that are fundamental to our approach. This first chapter will describe the context in which our research took place, our data sources and our research methods. Chapter 2 explains how the genre classification was developed and introduces the thirteen genre families through their purpose, stages, genre networks, examples, characterisation in terms of multidimensional analysis, and distribution across levels of study and disciplinary groups.

Part II examines the social functions of university student writing, and the ways in which student writers develop and display various abilities through their writing. Genres and genre families are discussed individually, and Chapters 3, 4, 5 and 6 also each focus on a larger genre set by grouping together genre families which demonstrate similar educational purposes. These larger groupings help to distinguish some of the fundamental purposes of student writing, such as the ability to explain disciplinary concepts, to critically evaluate, to build sustained arguments, to carry out independent research projects and to prepare for professional practice. They also highlight writing requirements which are occasionally in conflict. Chapter 7 examines further functions of university student writing in those genres which enable writers to monitor their own personal development, and to practise writing for a readership outside their own specialism. The final chapter (Chapter 8) provides an overview of networks across genres and disciplines, and discusses the concept of academic register in relation to student writing. In this chapter we also suggest areas for further research, and provide details of how to access the BAWE corpus and other related resources.

1.3 Our starting point

The genres of student writing that we investigate in this book are those represented in the British Academic Written English (BAWE) corpus[3]. This collection presents a broad picture of British university

student writing at the beginning of the 21st century, thus providing insights into the prevailing teaching and learning practices, the priorities of departments, and the demands of disciplines and professional bodies.

We founded our investigation on the following assumptions, derived from our experience as teachers of academic writing, the findings of prior studies in the field and discussions with academic staff in many university departments in the UK and internationally.

1 *University students are required to produce a range of different genres of assessed written work, reflecting a range of different communicative purposes.*
We knew of a number of studies that had described writing tasks set at university level, or had analysed the written work produced by specified groups of university students, but we wanted to explore further the relationships between assignments produced for different disciplines and levels of study, and to compare and contrast different types of student writing on a much larger scale.

2 *The nomenclature used within university departments to specify different assignment types goes some way towards identifying and distinguishing these genres.*
We already knew that students produced assignments labelled as 'book reviews', 'business plans', 'case studies', 'film commentaries', 'lab reports' and so on. These titles are given to the students by their lecturers, and we assumed that they encapsulated information about the purpose of the assignment and its linguistic features, format and structure. On the other hand we were aware that some descriptors such as 'essay' or 'project' were used very loosely, and that different names were sometimes given to very similar assignments, whilst other dissimilar assignment types were sometimes known by the same name. Whilst drawing what insights we could from departmental usage, we wanted to identify more robust categories of assignment genres.

3 *Within broad discipline areas certain genres are favoured and others are produced only rarely, if at all.*
We were aware that it would be impossible to prove that a genre was completely absent from an academic discipline. This is because lecturers vary the assignments they set from year to year, and in some contexts they are encouraged to invent alternatives to old, familiar assignment types. We wanted, however, to create a clearer picture of the distribution of assignment types across the disciplines.

4 *The types of writing that university students are required to produce change as they progress through their course of study. Students are expected to conform increasingly to the norms of favoured genres, and may also be given generically different writing tasks at different stages of study.*
It seemed reasonable to assume that students would gain more and more technical expertise in their field, and that their writing might approximate more and more closely to published academic or professional workplace writing. In one way or another, we anticipated different expectations placed upon students in their first year at university and in subsequent years, and we wanted to see how these expectations affected the writing students produced.

5 *An overview of student writing in English at the beginning of the 21st century would not only reflect the educational context in which it was produced, but also resonate with accounts of university student writing internationally, produced in different contexts.*
This is to suggest that with the globalisation and internationalisation of higher education there is value in describing and explaining the genres of writing in one context, to inform any future comparisons and developments with other contexts that are removed in time or place.

We tested these assumptions, and tried to answer the questions we associated with them, by analysing assignment registers and genres in the light of discourse community perspectives. Central to our investigations was the creation of the BAWE corpus.

1.4 The contents of the BAWE corpus

A rationale for the creation of the BAWE corpus is presented in Nesi et al. (2005), and the process of its development is described in Alsop and Nesi (2009). Briefly, the corpus was designed so that roughly equal numbers of assignments could be collected from four levels of study (first year undergraduate to taught Masters level) and four disciplinary groupings. These 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. We only collected assignments that had already been positively assessed by subject tutors, because we wanted to ensure that they conformed to departmental expectations. When they were writing their assignments

the students would not have known that their work was going to become part of the corpus; at that time their only priority would have been to fulfil the task requirements and gain the best possible grade.

Assignments came from four different universities in England, which ensured access to a broad range of disciplines. Different departments operated rather different grading systems, but the pass mark in most was 40 per cent or above, and we only accepted assignments with grades of at least 60 per cent (or equivalent). The quality of these assignments was consistent with the award of an upper second class (2:1) or first class honours degree, and could be described in terms of 'merit' and 'distinction'.[4]

As far as possible, we collected equal numbers of assignments in each of the main disciplines we targeted, at each level of study. This ideal was difficult to achieve in practice, however, because of the much smaller numbers of students studying at Masters level in the Arts and Humanities, for example, and because of our requirement that all assignments should have reached a certain standard of proficiency. To facilitate data processing, only word-processed assignments were accepted for the corpus, excluding handwritten examination scripts, handwritten lab notebooks, assignments consisting solely of mathematical calculations and PowerPoint presentations assessed through oral delivery. As one assignment might include several essays or several lab reports (that is, several texts), a distinction was made between assignments which were submitted as one piece of work and texts which were analysed as genres. The final make-up of the corpus is illustrated in Table 1.1.

In addition to the assignments themselves, we collected information about the title of each assignment and its corresponding module, the department that set the assignment, and the grade that it had been given. At the end of the project some of this information was conflated, for example the assignment file headers identifying disciplinary rather than departmental provenance, as in some cases assignments from departments at more than one university contributed to the corpus holdings for a single discipline. Similarly, because of variation in the way assignments were graded, we simply divided them into those which had received a grade of between 60 per cent and 69 per cent, or its equivalent (a 'merit' grade, 'M') and those which had received a grade of 70 per cent or over, or its equivalent (a 'distinction' grade, 'D'). This distinction grade is comparable to an 'A' grade in the U.S. university system, while the merit grade is comparable to a 'B' grade, although the proportion of grades in each division varies.[5] The corpus contains almost equal numbers of distinction assignments (1,251) and merit assignments (1,402).

Table 1.1 Overview of BAWE corpus holdings

		Level 1	Level 2	Level 3	Level 4	Total
Arts and Humanities (AH)	assignments	239	228	160	78	**705**
Archaeology; Classics; Comparative	texts	255	229	160	80	**724**
American Studies; English; History;	words	468,353	583,617	427,942	234,206	**1,714,118**
Linguistics / English Language Studies; Philosophy; others						
Life Sciences (LS)	assignments	180	193	113	197	**683**
Agriculture; Biological Science; Food	texts	188	206	120	205	**719**
Science; Health; Medicine; Psychology	words	299,370	408,070	263,668	441,283	**1,412,391**
Physical Sciences (PS)	assignments	181	149	156	110	**596**
Architecture; Chemistry; Computer	texts	181	154	156	133	**624**
Science; Cybernetics / Electronic	words	300,989	314,331	426,431	339,605	**1,381,356**
Engineering; Engineering; Mathematics; Meteorology; Physics; Planning						
Social Sciences (SS)	assignments	207	197	166	207	777
Anthropology; Business; Economics;	texts	216	198	170	207	**791**
Hospitality, Leisure and Tourism;	words	371,473	475,668	447,950	704,039	**1,999,130**
Management; Law; Politics; Publishing; Sociology						
Total students		333	302	235	169	1039
Total assignments		807	767	595	592	2761
Total texts		840	787	606	625	2858
Total words		1,440,185	1,781,686	1,565,991	1,719,133	6,506,995

The discipline, level and grade of each assignment were important factors influencing our collection policy; we set a limit on the number of assignments at each level in each discipline, and we rejected assignments that had not achieved the required grade. However we also gathered other types of contextual information which did not affect our decision about whether or not to include an assignment in the corpus, such as the gender, year of birth and native speaker status of the contributor, and the number of years of UK secondary education he or she had received. Corpus findings concerning these factors must be treated with caution, because contributor features are not distributed equally across the corpus holdings. For example, anyone wishing to analyse the corpus from the perspective of gender should bear in mind that there are more female than male contributors, and their assignments are not entirely comparable in terms of discipline and disciplinary groupings. Likewise, anyone wishing to compare native and non-native speaker writing in the corpus should bear in mind that assignments contributed by speakers of languages other than English tend to be concentrated in the Social Sciences and at Masters level. Further details of the corpus contents in terms of the contributors' gender and first language are provided in Appendix 1.1.

1.5 Other sources of data

Throughout this book we will be drawing on corpus evidence as well as contextual data of various kinds gathered in connection with the project. This information was also considered when categorising assignments into genres and groups of similar genres, or genre families.

During the process of corpus compilation the students' own perceptions about the type of assignment they were submitting were recorded. We asked, for example, whether they thought their assignment was an essay, a lab report, a case study or some other kind of text. Students' responses were later compared with the way they had described their assignments within the text itself, and the way other contributors had described the same or similar tasks.

Although we did not have a prolonged engagement with each department, we were influenced by the ethnographic approach of Prior (1998), who used departmental documents and tutor representations of tasks to build a 'thick'[6] description of the contexts and processes of student writing. We referred to module descriptions from each of the target departments and explored departmental environments both informally and through semi-structured interviews with teaching staff and students.

Staff were selected for interviews because they were involved in teaching and in the assessment of student assignments. The interviews,

described in Nesi and Gardner (2006) and mentioned again in Chapter 2, centred on the following questions, adapted from a similar but smaller-scale investigation by Woodward-Kron (2002: 125):

- What role does assignment writing play in your department?
- What genres do you require your students to write?
- What are you looking for when you assess written work in different genres?
- How do your expectations of students' writing change during the course of the degree?

These methods were subsequently replicated by an undergraduate student researcher who interviewed 36 undergraduate students (Gardner and Powell, 2006). In this case the fact that the interviewer was a student encouraged the interviewees to respond more freely, and provided us with insights that we would not have obtained from talking to staff alone. The distribution of student interviews is reproduced in Appendix 1.2.

This contextual information was triangulated with textual information to inform our decisions regarding the classification of corpus holdings. The process of grouping similar genres together is described in Chapter 2; ultimately all texts in the corpus were assigned to one and only one genre family, making the description of large numbers of texts more manageable and facilitating comparisons across disciplines.

The distribution of the thirteen genre families across levels is shown in Table 1.2. The levels correspond essentially to first year, second

Table 1.2 Distribution of genre families by level

	Level 1	Level 2	Level 3	Level 4	Total
Case Study	26	30	35	103	194
Critique	78	79	68	97	322
Design Specification	24	19	35	15	93
Empathy Writing	10	3	18	5	36
Essay	416	360	267	194	1237
Exercise	28	28	31	27	114
Explanation	81	62	34	37	214
Literature Survey	10	6	9	10	35
Methodology Recount	120	127	49	65	361
Narrative Recount	18	19	21	17	75
Problem Question	12	19	6	3	40
Proposal	10	19	11	36	76
Research Report	7	16	22	16	61
Total	840	787	606	625	2858

year, third (or final) year and taught Masters level, although it should be noted that Masters level dissertations were not included in the corpus. Further tables throughout this book will show distribution in greater detail, in terms of disciplines and genres.

1.6 Methods of analysis

Throughout the book we refer to a variety of research techniques, choosing whichever method best reveals the character of a genre, or best distinguishes one genre family from another.

Our genre family classification system draws on the work of the Sydney School, which has been particularly influential in using Systemic Functional Linguistics (SFL) to identify and explain genres of secondary and primary school texts. Genre is widely regarded as 'the system of staged goal-oriented social processes through which social subjects in a given culture live their lives' (Martin, 1997: 13). We identify the educational purposes and stages that typify and distinguish genre families, for example Essays and Critiques (Chapter 4) and Research Reports and Literature Surveys (Chapter 5)[7]. Our genre families are different from those in the Sydney School classifications, however, partly because we aimed to develop them by grouping similar assignments, rather than imposing a classification developed for other contexts, and partly because we were also influenced by research on academic genres by Swales (1990) and in the field of academic literacies (Lea and Street, 2000).

Our examination of the linguistic features associated with the stages of genres draws on Systemic Functional Linguistics (Halliday and Matthiessen, 2004) to explore the prosodic nature of evaluation, and the functions of hyperNews (Martin, 1992) to make claims which help build an argument (Chapter 4). We also use the appraisal system developed by Martin (2000) and Martin and White (2005) to analyse evaluative resources in texts (Chapter 7). All these methods of analysis will be explained more fully in the relevant chapters.

We use the results of multidimensional analysis to help us characterise the genre families we describe. This analysis was conducted by Biber at the University of Northern Arizona, using five dimensions he identified (Biber, 1988). The BAWE corpus was tagged for 67 linguistic features, grouped into 16 grammatical / functional categories:

1 tense and aspect markers
2 place and time adverbials
3 pronouns and pro-verbs
4 questions
5 nominal forms
6 passives
7 stative forms
8 subordination features

(*cont.*)

9 prepositional phrases, adjectives and adverbs
10 lexical specificity (type–token ration and mean word length)
11 lexical classes such as downtoners, hedges, amplifiers and emphatics
12 modals
13 specialised verb classes such as 'public', 'private' and 'suasive' verbs
14 reduced forms and dispreferred structures such as split infinitives
15 coordination
16 negation

So, for example, place adverbials include *above* and *beside*, and time adverbials include *early, instantly* and *soon*. 'Public' verbs include *say, tell* and *explain*, 'private' verbs include *believe, think* and *know,* and 'suasive' verbs include *command, insist* and *propose*.

The texts in the corpus were compared across genre families, disciplinary groupings and levels of study, and scores along each dimension were allocated to each corpus subgroup. These scores characterise the register of the subgroup, and indicate tendencies towards information density, chronologically ordered narrative, deictic references to time and place, and so on. A summary of linguistic features in relation to the five dimensions is provided in Appendix 1.3, adapted from Biber et al. (2002). The dimensions are explained more fully below, with reference to Biber (1988) (also summarised in Conrad and Biber, 2001).

DIMENSION 1: *Involved versus informational*

This contrasts verbal and nominal styles. Biber found conversation to be extremely involved, with a score of 35, with high frequencies of present tense verbs, private verbs, first and second person pronouns, and contractions. At the opposite end of the scale general academic prose (published research in journals, books and reports from the Lancaster–Oslo/Bergen, or LOB, corpus) had a score of -15.

DIMENSION 2: *Narrative versus non-narrative*

This dimension is associated with past time narration. Biber found romance fiction to be heavily narrative, with a score of 7, because it contains many third person pronouns, past tense verbs, perfect aspect verbs, and public verbs such as *say* and *tell*. Academic prose, official documents and radio broadcasts were positioned at the opposite end of the scale, with scores below -2.

DIMENSION 3: *Elaborated versus situation-dependent*

Elaborated texts can be understood in contexts that are distant in time and place from the context in which they were originally produced. They identify referents explicitly, through features such as relative clause constructions, nominalisations, and time and place adverbials for temporal and locative reference. The official documents in Biber's study scored more than 7 on this dimension. Conversations and broadcasts, on the other hand, scored -4 or less, because when we listen to these sorts of texts we interpret what is being said in terms of where the speaker is, and what is happening at the time.

DIMENSION 4: *Persuasive*

This dimension identifies overtly argumentative texts, and is characterised by infinitives, suasive verbs such as *agree, ask, insist* and *recommend*, conditional subordination, split auxiliaries, and modals expressing prediction, necessity and possibility. The editorials and professional letters analysed by Biber scored 3 or more, while radio broadcasts scored below -4.

DIMENSION 5: *Non-impersonal versus abstract and impersonal*

Impersonal texts are characterised by passive constructions, conjuncts such as *thus* and *however*, and adverbial and postnominal clauses. Such features are typical of written as opposed to spoken texts. Biber found that general academic prose from the LOB corpus had high scores on this dimension (more than 5). Conversations had low scores (less than -3).

Tables 1.3 and 1.4 show scores for BAWE corpus texts across the four levels of study and across the four disciplinary groups. The entirely negative scores on the involved and narrative dimensions

Table 1.3 Dimension scores by level

Level	Involved	Narrative	Elaborated	Persuasive	Abstract and impersonal
1	–12.7	–2.7	5.1	–1.4	5.9
2	–13.9	–2.8	5.6	–1.4	6.2
3	–14.7	–3.0	5.7	–1.5	6.4
4	–17.2	–3.2	6.3	–2.0	5.5

Table 1.4 Dimension scores by disciplinary group

	Involved	Narrative	Elaborated	Persuasive	Abstract and impersonal
AH	–13.4	–2.1	5.7	–2.3	5.5
LS	–15.6	–3.0	5.7	–1.5	5.7
PS	–13.4	–3.7	4.4	–1.2	6.5
SS	–15.3	–3.0	6.5	–1.3	6.2

indicate a high informational focus and a low level of narration overall, but students' writing also becomes increasingly informational and elaborated as they progress through their degree programmes, and has progressively fewer narrative and persuasive features. Abstract impersonal features increase until Masters level; their decline at Level 4 has not yet been fully explained, but may be associated with the fact that Masters students contributed a greater number of case studies and proposals to the corpus, and these are some of the least abstract genre families (see Chapters 2 and 6).

Table 1.4 shows that texts in the Life Sciences (LS) are the most informational (that is, the least involved), and those in the Arts and Humanities (AH) have the greatest amount of narrative features. Physical Sciences (PS) have the fewest narrative features and are the most impersonal and persuasive. Texts in the Social Sciences (SS) are the most elaborated.

Dimension scores for the thirteen genre families are presented in Chapter 2, and the competing effects of disciplinary group and genre family are discussed in Chapter 9.

In addition to the results of multidimensional analysis, we refer throughout the book to data generated through the use of WordSmith Tools (Scott, 2010), and Sketch Engine (Kilgarriff et al., 2004). Details of the functions of Sketch Engine with special reference to the BAWE corpus are provided in Nesi and Thompson (2011).

WordSmith Tools and Sketch Engine were both used to create concordance lines. These provide contexts for corpus words and phrases throughout the book. Subcorpora of genres and genre families were manually prepared for use with WordSmith Tools. Sketch Engine enabled us to filter the corpus so that we could view concordance output from selected levels, genres and disciplines.

WordSmith and Sketch Engine were also used to create lists of keywords and lemmas, calculated by comparing their relative frequencies in a study corpus (all or part of the BAWE corpus) with those of a larger reference corpus (the British National Corpus or the entire

BAWE corpus). A word or lemma is considered positively key if its frequency in the study corpus is unusually high. WordSmith provides a 'keyness' score, measured by cross-tabulation and chi-square significance test (Scott, 2010). Sketch Engine provides a 'keyword score', using a statistic based on 'word W is N times as frequent in corpus X versus corpus Y'[8]. The keywords and lemmas listed in this book have very high scores, and the probability of the keyness being accidental is very low. The table in Appendix 1.4 shows the top key lemmas in the BAWE corpus compared with the British National Corpus, calculated using Sketch Engine.

Common word combinations, or 'clusters', were identified using WordSmith Tools Version 5. Clusters often reveal common underlying concepts and functions shared by groups of texts. Some writers use the term 'cluster' to refer to recognisable multi-word units which are identified by searching for strings containing a given 'seed term'. For example, a cluster search based on the seed term *of* might find multi-word units such as *in terms of, on account of, the context of* and so on. In this book, however, the term cluster is used for any frequent string of words; there is no seed term, and the only parameters are the length of the string and the minimum frequency. Thus, for us, the term is synonymous with 'n-gram' (see, for example, Gries, Newman and Shaoul, 2011) and 'lexical bundle' (see, for example, Biber, 2007).

All the words in the BAWE corpus have been annotated for part of speech, using the Constituent Likelihood Automatic Word-tagging System (CLAWS). The version we used identifies 137 part-of-speech categories and subcategories, for example singular and plural common, locative and temporal nouns, singular and plural proper nouns, and singular and plural units of measurement. The words in the BAWE corpus have also been annotated for semantic category, using the UCREL[9] Semantic Analysis System (USAS). This system groups words in terms of 21 thesaurus-style categories, developed on the basis of those in the *Longman Lexicon of Contemporary English* (McArthur, 1981). Details of the CLAWS system are provided in Garside and Smith (1997), and details of the USAS system are given in Archer et al. (2002).

Frequencies of words in some of the main USAS categories are provided in Appendix 1.5. Instructions on how to search for these features in Sketch Engine are provided in Nesi and Thompson (2011). Semantic analysis is used in Chapter 6, with reference to Case Studies, Design Specifications, Proposals and Problem Questions.

Basic statistics, such as the average word length of assignments, the average number of sentences per assignment and the average sentence length, were calculated from an Excel spreadsheet of the corpus

Table 1.5 BAWE corpus statistics

Average \ Level	1	2	3	4
words per assignment	1782	2323	2637	2903
sentences per assignment	75	95	108	122
paragraphs per assignment	21	29	34	40
words per sentence	24.8	25.6	25.5	24.6
sentences per paragraph	4.8	4.6	4.5	3.7
tables	0.6	0.7	1.0	1.0
figures	1.0	1.2	2.1	2.0
block quotes	0.5	1.1	0.8	0.7
formulae	2.0	5.2	7.3	1.8
lists	0.4	0.5	0.6	1.6

holdings. Table 1.5 provides statistics for each level of study produced by this means. These statistics clearly show how the average word length of assignments increases from Level 1 to 4, although average sentence and paragraph lengths do not increase. Assignments written at later levels of study contain more tables, figures and lists. The averages for formulae must be interpreted with caution, because there is very wide variation in their use. The range of formulae per assignment is 1–70 at Level 1 and 1–51 at Level 4, but at Levels 2 and 3 there are 13 assignments containing more than 100 formulae, and three containing more than 800.

Basic statistics for each genre family will be discussed in the appropriate chapters, where we will see meaningful differences between, for instance, Essays and Critiques, in terms of block quotes versus figures. These statistics help us characterise the genres and point to features that analyses of individual texts can miss.

1.7 Insights from the analysis of our data

In this book we recognise that there are tensions between the demands of various participants in the student writing process. While all probably subscribe to the view that the act of writing is a means of developing skills and constructing knowledge, many students are equally if not more concerned to satisfy their course demands and gain the grades they need in order to graduate. They may also seek self-expression and personal development through their writing. Departments, on the other hand, will view student writing as a form of quality control, visible to internal and / or external assessors during institutional, departmental and course reviews, and a key element in

the external examining system. Additionally, employers and professional bodies have their own requirements for student writing (see Chapters 6 and 7) which departments are usually keen to accommodate to improve the employability of their students.

We have also noted a possible conflict between the different requirements of different types of intended reader in some less traditional genres of student writing (see Chapters 6 and 7). Students may be expected to write ostensibly for a professional colleague or a non-expert, whilst at the same time meeting the academic assessment criteria of the department which assigns the grade, and addressing their own personal learning and self-development needs.

However, whilst acknowledging possible tension between these demands, our data does not suggest that there is conflict between the various participants in the communicative process. By and large employers, staff and students work together to enable each other's demands to be met. The very wide range of genres we have identified in use across the disciplines, and the enthusiasm of lecturers to innovate, bear testimony to a genuine desire on the part of staff to accommodate the needs of students, the discipline, the professions and industry. It seems to us that any problems with the assessment process are less likely to arise because of intransigence on the part of participants, and more likely to be due to failure to adequately explain the nature of the relevant assignment genres.

In this book we therefore aim to promote a better understanding of the diverse nature of writing in English university degree programmes. Our corpus does not represent every university discipline, and we do not provide detailed studies of every individual genre, but we do develop a framework for future researchers who might wish to make more detailed studies, using the BAWE corpus, another collection of student writing, or a combination of the two. We also hope that our genre family descriptions will be useful to those who teach academic writing for university study, especially by drawing attention to similarities and differences within families that have been obscured by departmental naming practices and were neglected in previous studies that have not been able to draw on such a large collection of textual evidence.

Notes

1. 'An investigation of genres of assessed writing in British higher education', funded by the UK Economic and Social Research Council, 2004–2008 (RES-000–23–0800) under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE],

University of Warwick), Paul Thompson (formerly of the Department of Applied Linguistics, University of Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes University).

2. Throughout this book the terms 'lecturer' and 'tutor' are used broadly to include professors, readers, teaching fellows and other academic tutors who set assignments for students.
3. One outcome of the project 'An investigation of genres of assessed writing in British higher education', see Note 1.
4. Although all the assignments in the BAWE corpus were given high marks, the writers were students, not experts, and the assignments were not edited to publication standard. For this reason the corpus does contain a few spelling and grammar mistakes.
5. In 2009/10, 14 per cent of all undergraduate degrees awarded in the UK were first class, i.e., included a majority of distinction grades (Higher Education Statistics Agency, 2011).
6. In ethnography, a 'thick' description is a description not only of people's behaviour, but also of its underlying meanings within their own culture.
7. We use lower case to refer to concepts in general (e.g., explanation) and upper case to refer to genre families (e.g. Explanation genres, Essay genres).
8. For an explanation of the statistic used to generate the score see http://trac.sketchengine.co.uk/wiki/SimpleMaths
9. UCREL is the University Centre for Computer Corpus Research on Language, a research centre at Lancaster University in the UK.

References

Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, 4(1), 71–83.

Archer, D., Wilson, A., & Rayson, P. (2002). *Introduction to the USAS category system*. University Centre for Computer Corpus Research on Language, Lancaster University, UK. See http://ucrel.lancs.ac.uk/usas/usas%20guide.pdf.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–86.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional analysis. *TESOL Quarterly*, 36(1), 9–49.

Conrad, S., & Biber, D. (2001). Multi-dimensional methodology and the dimensions of register variation in English. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies*. Harlow, Essex: Pearson Education.

Gardner, S., & Powell, L. (2006). An investigation of genres of assessed writing in British higher education: A Warwick–Reading–Oxford Brookes project.

Paper presented at the seminar 'Research, scholarship and practice in the area of Academic Literacies', University of Westminster, 30 June.

Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus annotation: Linguistic information from computer text corpora*. Longman, London, 102–21.

Gries, S.T., Newman, J., & Shaoul, C. (2011). N-grams and the clustering of registers. *Empirical Language Research*, 5(1).

Haggis, T. (2006). Pedagogies for diversity: Retaining critical challenge amidst fears of 'dumbing down'. *Studies in Higher Education 31(5)*, 521–35.

Halliday, M.A.K., & Matthiessen, C.M.I.M. (2004). *An introduction to functional grammar* (3rd edition). London: Arnold.

Higher Education Statistics Agency. (2011). *Product: SFR 153 higher education student enrolments and qualifications obtained at higher education institutions in the United Kingdom for the academic year 2009/10*. www.hesa.ac.uk.

Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, T. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), *Proceedings of Eleventh EURALEX International Congress*. Lorient, France: Université de Bretagne-Sud.

Lea, M., & Street, B. (2000). Student writing and staff feedback in higher education: An academic literacies approach. In M. Lea and B. Stierer (Eds.), *Student writing in higher education: New contexts*. Buckingham: The Society for Research into Higher Education and Open University Press.

Martin, J.R. (1992). *English text: System and structure*. Amsterdam: Benjamins.

Martin, J.R. (1997). Analysing genre: Functional parameters. In F. Christie & J.R. Martin (Eds.), *Genres and institutions: Social processes in the workplace and school*. London: Continuum.

Martin, J.R. (2000). Beyond exchange: APPRAISAL systems in English. In S. Hunston & G. Thompson (Eds.), *Evaluation in text*. Oxford: Oxford University Press, 142–75.

Martin, J.R., & White, P.R. (2005). *The language of evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan.

McArthur, T. (1981). *Longman lexicon of contemporary English*. Harlow: Longman.

Nesi, H., & Gardner, S. (2006). Variation in disciplinary culture: University tutors' views on assessed writing tasks. In R. Kiely, G. Clibbon, P. Rea-Dickins, & H. Woodfield (Eds.), *Language, culture and identity in applied linguistics* (British Studies in Applied Linguistics, vol 21). London: Equinox Publishing, 99–117.

Nesi, H., Gardner, S., Forsyth, R., Hindle, D., Wickens, P., Ebeling, S., Leedham, M., Thompson, P., & Heuboeck, A. (2005). Towards the compilation of a corpus of assessed student writing: An account of work in progress. In P. Danielsson & M. Wagenmakers (Eds.), *Proceedings from the corpus linguistics conference series*. Birmingham: University of Birmingham.

Nesi, H., & Thompson, P. (2011). *Using Sketch Engine with BAWE*. Available online at http://trac.sketchengine.co.uk/wiki/SharedResources.

Prior, P. (1998). *Writing / disciplinarity: A sociohistoric account of literate activity in the academy*. Mahwah, NJ: Lawrence Erlbaum.

Scott, M. (2010). *WordSmith Tools Version 5*. Oxford: Oxford University Press.

Sugimura, M. (2008). International student mobility and Asian higher education: Framework for global network. Paper presented at the Asia-Pacific Sub-regional Preparatory Conference for the 2009 World Conference on Higher Education, September 24–26, 2008, Macau, PR China. Available online at www.unescobkk.org/fileadmin/user_upload/apeid/workshops/macao08/papers/1-d-3.pdf.

Swales, J.M. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.

Turner, J. (2011). *Language in the academy: Cultural reflexivity and intercultural dynamics*. Bristol: Multilingual Matters.

UNESCO (2008). *Education for all. Global Monitoring Report 2008*. United Nations Education Scientific and Cultural Organisation. www.efareport.unesco.org.

Wächter, B. (2008). Teaching in English on the rise in European higher education. *International Higher Education*, 52(3).

Wächter, B., & Maiworm, F. (2008). *English-taught programmes in European higher education: The picture in 2007*. Bonn: Lemmens.

Wardle, E. (2009). 'Mutt Genres' and the goal of FYC: Can we help students write the genres of the university? *College Composition and Communication*, 60(4), 765–89.

Woodward-Kron, R. (2002). Critical analysis versus description? Examining the relationship in successful student writing. *Journal of English for Academic Purposes*, 1(2), 121–43.