Edited by
Maggie Charles, Diane Pecorari
and Susan Hunston

Academic Writing

At the Interface of
Corpus and Discourse

Academic Writing

This page intentionally left blank

Academic Writing

At the Interface of
Corpus and Discourse

Edited by

Maggie Charles, Diane Pecorari
and Susan Hunston

Contents

Part III   Focus on Learner Discourses

Contributors

Marina Bondi is Professor of English and Director of the Department of Studies on Language, Text and Translation at the University of Modena and Reggio Emilia, Italy. She has worked on various aspects of discourse analysis and EAP, with special reference to argumentation, metadiscourse and evaluative language. Her recent work centres on language variation across genres, disciplines and cultures through the analysis of small specialized corpora. On this issue she has published various articles and co-edited a number of volumes: most recently, *Academic Discourse across Disciplines* with Ken Hyland and *Managing Interaction in Professional Discourse: Intercultural and Interdiscoursal Perspectives* with Julia Bamford.

Maggie Charles is a Tutor in English for Academic Studies at Oxford University Language Centre, where she specializes in teaching academic writing to graduates. Her research interests include the pedagogical applications of corpus linguistics, the study of stance/evaluation and discipline-specific discourse. Her work has appeared in, among others, *Applied Linguistics, Journal of English for Academic Purposes, English for Specific Purposes* and *System.* She is also a consultant on academic writing for the *Oxford Advanced Learner's Dictionary.*

Caroline Coffin is a Reader in Applied Linguistics at the Centre for Language and Communication at the Open University, UK. Over the last few years she has been using both corpus linguistics and systemic functional linguistics to investigate disciplinary discourse including computer mediated dialogue and different forms of argumentation. Published books include *Applied Linguistics Methods: A Reader* (Routledge, 2009, with Lillis and O'Halloran), *Historical Discourse: The Language of Time, Cause and Evaluation* (Continuum, 2006), *Applying English Grammar: Functional and Corpus Approaches* (Hodder-Arnold, 2004, with Hewings and O'Halloran) and *Teaching Writing in Higher Education* (Routledge, 2003 with Curry, Goodman, Hewings, Lillis, & Swann).

John Flowerdew is Professor of Applied Linguistics at the University of Leeds, UK. He was formerly a professor at City University, Hong Kong. His research interests include discourse analysis, critical discourse analysis, English for specific purposes, corpus linguistics, curriculum theory and practice and

qualitative research methodologies. His approach is problem-driven and he uses a range of methodologies, from the micro-analysis of texts (including corpus-based analysis) to more ethnographic approaches to data collection and analysis. He has published widely, including 12 books and over 75 articles, especially on academic discourse and political discourse.

Richard W. Forest is Assistant Professor in the Department of English Language and Literature at Central Michigan University. His research interests include English for Academic Purposes, evaluation in text, and usage-based models of grammar and discourse. His current research focuses on the nature and distribution of abstract nouns in various genres (lectures, textbooks, and journal articles) and disciplines (from the natural and social sciences) in academic English.

S. Amina Gardezi lives in Pakistan. She has her Masters in English Language Studies and Methods from the University of Warwick UK. Prior to this she studied in the prestigious Kinnaird College for Women, Lahore, Pakistan. At Kinnaird she obtained her Bachelors in Economics and Journalism and Masters in Applied Linguistics. While at Warwick she concentrated on Discourse Analysis and researched in the area of writing. She is presently working at Beaconhouse National University, Lahore and conducts teacher training workshops and is also engaged in consultancy work in the field of education. She also teaches part time at her old alma mater, Kinnaird College.

Sheena Gardner is Reader in Educational Linguistics in the Language, Discourse and Society Academic group and MOSAIC Centre for Research on Multilingualism, School of Education, University of Birmingham. Her academic writing interests stem from her English language teaching and course development work in EAP/ESP contexts as well as North American Writing Programme contexts, and her teacher education work with EFL, ESL/EAL and ESP teachers internationally. Her research centres on the linguistic analysis of educational texts, as illustrated in her many publications and her contributions to the genre analysis of the BAWE corpus of assessed university student writing (ESRC RES-000-23-0800).

Sylviane Granger is Professor of English Language and Linguistics at the Université catholique de Louvain in Belgium. She is the director of the Centre for English Corpus Linguistics, where research activity is focused on the compilation and exploitation of learner corpora and bilingual corpora. In 1990 she launched the *International Corpus of Learner English* project, a learner corpus initiative which involves a large number of universities internationally. She has written numerous articles and edited several volumes on these topics and gives frequent invited talks and seminars to stimulate learner corpus research and to promote its application to ELT materials design and development.

Ann Hewings is a Senior Lecturer in the Centre for Language and Communication at the Open University, UK. She has previously taught English

in Europe, Asia, and Australia from primary to tertiary levels. She worked for a number of years on the COBUILD project, researching and contributing to dictionaries and other English language reference material. Her research focus is academic writing in disciplinary contexts, by both professional academics and university students. She has recently been involved in projects investigating student writing in electronic environments, and non-Anglophone context scholars writing for publication in English.

Jasper Holmes's PhD thesis was in Word Grammar, which he did with Dick Hudson at University College, London. He worked for many years in education research and computer linguistics before joining the BAWE project in 2006. In 2008 he became a Fellow at Coventry University.

Susan Hunston is Professor of English Language at the University of Birmingham, UK. She specializes in corpus linguistics and discourse analysis. She is author of *Corpora in Applied Linguistics* (2002, CUP), co-author of *Pattern Grammar: a Corpus-driven Approach to the Lexical Grammar of English* (1999, Benjamins) and co-editor of *Evaluation in Text: Authorial Stance and the Construction of Discourse* (2000, OUP) and *System and Corpus: Exploring the Connections* (2005, Equinox). She has also published numerous articles on the expression of stance or evaluation, on the use of corpora to describe the grammar and lexis of English, and on the interface between corpus and discourse studies.

Ken Hyland is Professor of Applied Linguistics and Director of the Centre for Applied English Studies at the University of Hong Kong having recently moved from the University of London. He has taught applied linguistics and EAP for over 30 years in Asia, Australasia and the UK and has published over 130 articles and 14 books on language education and academic writing. Most recent publications are *Academic Discourse* (Continuum, 2009), *EAP: An Advanced Resource Book* (Routledge, 2006), and *Academic Evaluation* (edited with Giuliana Diani, Palgrave, 2009). He was founding co-editor of the *Journal of English for Academic Purposes* and is now co-editor of *Applied Linguistics.*

Suganthi John is a lecturer at the Centre for Academic and Professional English, College of Arts and Law at the University of Birmingham. Her research is in the area of academic discourse and she is particularly interested in the linguistic representations of the academic identity, the role of the revision process in writing and the nature of research writing at postgraduate level. She is also interested in applying research insights to the development of teaching materials for academic writing courses. Her teaching is primarily in the field of English for academic and professional purposes.

Dimitra Koutsantoni holds a PhD in English (Applied Linguistics) from the University of Birmingham (2003). She has researched and taught academic writing at the Universities of Warwick, Aston and Bedfordshire, UK. She has published in the areas of genre and discourse analysis, contrastive

rhetoric, e-learning/blended learning, and intercultural pragmatics, while she has also written policy papers and reports on the internationalization of higher education. She is the author of the 2007 monograph: *Academic Literacies: Understanding Disciplinary Communities' Cultures and Rhetoric* (Peter Lang).

Hilary Nesi joined Coventry University as Professor in English Language in October 2007, having worked for twenty years at the University of Warwick. She led the AHRB-funded project to create the *BASE* corpus of British Academic Spoken English (2001–2005) and also the ESRC-funded project which involved the creation of the *BAWE* corpus of British Academic Written English (2004–2007). She was chief academic consultant for the *Essential Academic Skills in English* (*EASE*) series, based around video from the BASE corpus, and she is currently involved in several projects relating to the use of English in international higher education.

Sarah North is a senior lecturer at the Open University, involved in the development of courses in applied linguistics. She previously worked in ELT in the UK and overseas, including Tanzania, China, India, Malaysia and Mexico. Her research interests centre on English for academic purposes, language and literacy, and computer-mediated communication. Recent publications include chapters in *A Companion to English Language Studies* (Maybin and Swann, 2009), *IELTS Collected Papers: Research in Speaking and Writing Assessment* (Taylor & Falvey, 2007), and *Language and Literacy: Functional Approaches* (Whittaker, O'Donnell, & McCabe, 2006), as well as journal articles.

Magali Paquot is a postdoctoral researcher at the Centre for English Corpus Linguistics of the Université catholique de Louvain in Belgium and a fellow of the Belgian National Fund for Scientific Research (FNRS). Her research interests include EAP vocabulary, phraseology and corpus-based analysis of L1 transfer in Second Language Acquisition. In 2008 she launched the *Varieties of English for Specific Purposes database (VESPA)* project, which aims at collecting a large corpus of learner texts in a wide range of disciplines and genres in academic settings.

Diane Pecorari is a senior lecturer in the School of Education, Culture and Communication at Mälardalen University in Sweden. Her research interests include discourse analysis, second-language writing and the acquisition of academic literacy. Her publications include *Academic Writing and Plagiarism: A Linguistic Analysis* (Continuum, 2008), as well as articles in the *Journal of Second Language Writing* and *English for Specific Purposes*. She is a member of the editorial board of *English for Specific Purposes*.

Philip Shaw has taught English and linguistics in Thailand, Germany, Britain and Denmark, and is currently a Professor in the English Department of

Stockholm University. His interests include varieties of English round the world, English for academic and business purposes, and reading and vocabulary acquisition in a second language. He is co-author of *World Englishes: An Introduction* and recently co-edited a special issue of the *Nordic Journal of English Studies* on EAP.

John M. Swales is Professor Emeritus of Linguistics at the University of Michigan, where he was also Director of the English Language Institute from 1985 to 2001. In 2006 a retirement conference was held in his honour in Ann Arbor, the transcripts from which are now available at the MICASE website. In 2009, the University of Michigan Press will publish two small textbooks, one on writing abstracts, the other on writing literature reviews, both co-authored with Chris Feak, as well as an autobiographical volume entitled *Incidents in an Educational Life: A Memoir (of Sorts).* John Swales continues to serve on the editorial boards of several journals.

Ramona Tang is an Assistant Professor in the English Language and Literature Academic Group at the National Institute of Education, Singapore, where she teaches a range of applied linguistics, writing, and teaching methodology courses at undergraduate and postgraduate level to pre-service as well as in-service teachers. She developed and now manages NIE's academic discourse skills course which is taken by all undergraduates at the university, and she is also the programme coordinator of NIE's Postgraduate Diploma in English Language Teaching programme for university EFL lecturers from China. Her research interests include academic literacies, discourse analysis, and teaching practices and platforms for learning in higher education.

This page intentionally left blank

Introduction: Exploring the Interface between Corpus Linguistics and Discourse Analysis

Maggie Charles, Diane Pecorari and Susan Hunston

This volume explores the interaction between two traditions of investigating written academic prose that might broadly be called 'discourse analysis' and 'corpus linguistics'. The two traditions have much in common. Both take selected examples of naturally occurring discourse as their starting point. Both attempt to identify recurring patterns in those examples. Both relate their findings to the social, intellectual or ideological contexts in which the discourse plays a role. The priorities of the two approaches do tend to diverge, however. Discourse analysis prioritizes whole texts and their cultural context, identifying patterns that extend across sentences and paragraphs. Corpus linguistics tends to use techniques that decontextualize individual texts and focuses on recurrent patternings of small-scale items such as words and phrases. In the opinion of many researchers, however, the availability of both approaches offers possibilities of enriched analysis, combining not only complementary methodologies but also alternative theoretical approaches. Increasingly, then, the two priorities are meshed in practice.

This introduction considers a further similarity between discourse analysis and corpus linguistics: neither is a monolithic whole comprising a single theoretical model and a single method of analysis. Within both traditions there is diversity; this diversity is reflected in the chapters in this book and its background is explored here.

1 Discourse Approaches

Discourse approaches have traditionally been seen as 'top-down' (Swales, 2002) in that they are concerned with whole, individual texts including the social conditions of their production and reception. At least two common themes can be identified in discourse-based analyses of academic writing, and the various approaches can be distinguished by the ways they interpret and

balance these themes. One is the concern with writing as social practice, considering not only the effect on the text of its social role but also the role of the writer and the text in creating the contexts in which they occur. The second theme is the recurring pattern of meanings in texts of similar types, often described as 'generic elements' or 'moves'.

Since Swales (1981) originally put forward his CARS (Create a Research Space) model, which provides a move analysis of introductions to research articles (RAs), there have been numerous investigations of this type, with perhaps the most studied academic genres being the RA and the thesis. Swales himself has refined his analysis of RA introductions (1990, 2004) and other parts of this genre have been analysed (e.g. Brett, 1994; Holmes, 1997; Samraj, 2005; Yang & Allison, 2003). There has also been growing interest in the part-genres of theses (e.g. Bunton, 2002, 2005; Dudley-Evans, 1986; Kwan, 2006; Samraj, 2008). While early work on genre analysis tended to employ discourse methods alone, more recent studies, such as those by Kwan and Bunton, make use of corpora and the associated more quantitative techniques.

The second type of 'discourse-based' research includes studies that prioritize the social context surrounding the production of academic writing. This work draws on the theories of Halliday (1978) and Giddens (1984) and interprets academic writing as a social practice. Early examples of this type of research, which show how expert texts are shaped by the disciplinary community's expectations and requirements, include Bazerman (1988), Berkenkotter and Huckin (1995) and Myers (1990). More recently, the social dimension of student writing has been the specific focus of attention in research such as the volume from the London-based 'new literacies group' edited by Jones et al. (1999), Casanave's (2002) examination of academic literacy practices in higher education and Pecorari's (2008) work on plagiarism.

Work in the Systemic Functional Linguistics (SFL) tradition combines a focus on genre and social context with fine-grained linguistic analysis, drawing on early contributions by Halliday to the development of academic writing (e.g. 1987/1993), by Hasan on genres and their structural elements (e.g. 1984/1996), and by Martin and others on the contexts and types of writing by school children (e.g. 1985). Typical of this type of research are papers in the recent volume devoted to student writing, edited by Ravelli and Ellis (2004). Thus, for example, Hood (2004) examines how undergraduates use evaluative resources to construct their stance, while A. Hewings (2004) shows how attention to the choice of Theme can help students understand the rhetorical organization of geography essays. Also within this theoretical framework, Coffin's book-length treatment of historical discourse (2006) shows how the genres of history can be distinguished by means of detailed SFL analyses.

2　Corpus Approaches

Corpus-based studies have traditionally been less concerned with whole texts or with the social context and have thus been characterized as working from the 'bottom up' (Swales, 2002), that is, they typically examine large amounts of data from many texts and provide frequency and distributional information about surface features of the language used. Corpus research has played a key role in distinguishing the overall characteristics of academic prose by means of multi-dimensional analysis, pioneered by Biber and co-workers (e.g. Biber, 1988, 2006b; Biber, Conrad, Reppen, Byrd, & Helt, 2002; Biber, Johansson, Leech, Conrad, & Finegan, 1999).

Large-scale quantitative studies of phraseological tendencies are particularly suited to contrastive work and specific disciplines and genres have been the focus of attention (e.g. papers on writing in the volume on disciplinary discourse edited by Hyland and Bondi, 2006). Of particular importance in this area is the work of Hyland (e.g. 1998, 2000, 2001, 2005, 2008), who has examined a large number of individual linguistic features and shown how they vary systematically according to discipline and/or genre. Much of his work is notable for combining a corpus approach with information about the social context derived from interviews with members of the relevant discourse communities.

Other corpus studies have focused on lexico-grammatical features such as introductory *it* patterns (Groom, 2005; M. Hewings & A. Hewings, 2002), nouns (Charles, 2007; J. Flowerdew, 2003), personal pronouns (Harwood, 2005a, 2005b), and *if*-conditionals (Carter-Thomas & Rowley-Jolivet, 2008). Research on citation (e.g. Charles, 2006a; Hyland, 1999a; Thompson, 2005; Thompson & Tribble, 2001) and stance or evaluation have proved to be areas in which corpus methods can be fruitfully applied to discourse-level concerns (e.g. Biber, 2006a; Charles, 2006b; Hyland, 1999b; Shaw, 2003; and papers from Tognini-Bonelli & Del Lungo Camiciotti, 2005).

Contrastive corpus work has also been able to establish differences between L1 and L2 production. Work in this area is particularly associated with the research of Granger and co-workers, who set up the International Corpus of Learner English, a large corpus of essays written by students from many different L1 backgrounds (Granger, Dagneaux, Meunier, & Paquot, 2009). Studies based on this data have shown that there is systematic variation in L2 production according to students' L1, thus paving the way for the production of more closely targeted pedagogic materials (e.g. papers from Granger, 1998 and Granger, Hung, & Petch-Tyson, 2002).

The focus on surface features can give corpus approaches an 'ad hoc' feel, and it can appear that research of this kind lacks the consistent theoretical underpinning of much discourse-based work. This is particularly true where the research is 'corpus-driven' and deliberately avoids predicting what will be

found to be of importance in a set of texts from what is known of their social context. What unites much of this work, however, is a concern with what Sinclair and Coulthard (1975) called 'latent patterning' – the recurrence across many texts of detailed features of expression that play a crucial, though often hidden, role in the construal of academic knowledge.

3   Integrative Approaches

Although we have presented discourse and corpus approaches separately, as we have indicated, there has been some adoption of corpus techniques within certain areas of discourse analysis and a corresponding acknowledgement of the importance of discourse-level concerns within some corpus-based work. This gradual coming-together has led to considerable interpenetration of the two approaches. In this section, we briefly mention some studies that actively seek an integration of corpus and discourse approaches in the methods they use and/or the aims they pursue.

As far back as 1998, L. Flowerdew drew attention to the potential of corpus linguistics to contribute to the analysis of discourse and called for the development of techniques for use at the discoursal level, which, she argued, could lead to improved exploitation of corpus findings for pedagogical purposes. An illustration of such an approach is provided by her (2008) investigation of the Problem-Solution pattern in technical reports written by professionals and students. She uses keyword analysis to characterize the move structure of this genre and to differentiate between expert and apprentice writing.

Conrad (2002) provides an overview of corpus linguistic approaches that can be used to examine discourse phenomena and distinguishes four types: (1) studies which examine a feature of language in use; (2) studies of the realizations of a function of language; (3) studies of a variety of language and (4) studies that trace the occurrence of a linguistic feature throughout a text. She gives examples of each type and argues that corpus techniques have much to contribute to research on discourse. In a similar vein, Partington (2004) sees corpus and discourse methods as being complementary and points to the need for what he calls 'corpus-assisted discourse study' (19), while Baker (2006) illustrates how corpus methods can usefully be applied in discourse analysis.

Two recent volumes are devoted specifically to the interface between discourse and corpus research. In a collection that covers several different contexts of use, Ädel and Reppen (2008: 2) argue for 'the viability of corpus-based research and corpus-assisted tools for discourse studies'. Chapters by Bondi (2008) on emphatics in history and economics RAs and by Sanderson (2008) on the use of pronouns in RAs in five humanities disciplines exemplify the application of corpus techniques to investigating written academic discourse.

Biber, Connor and Upton (2007) contrast two approaches to the genre analysis of biochemistry and biology RAs. In the first, top-down analysis (Biber, Connor, Kanoksilapatham, & Upton, 2007), a corpus is used as the basis for the manual coding of discourse units according to their function in the text. This study, then, takes an approach similar to that envisaged by Partington (2004), in which the corpus functions as a subsidiary aid to the discourse analysis. In the second approach (Biber, Connor, Jones, & Upton, 2007), the unit of analysis is the 'vocabulary-based discourse unit' (VBDU), which is identified from the bottom up, on the basis of the automatic segmentation of the text according to linguistic rather than functional criteria. Here the analysis is more firmly corpus-based, as the units are identified solely through the corpus techniques of multi-dimensional and cluster analysis. Biber, Connor and Upton find considerable, though not complete, overlap between these two types of analysis, and make the point that the advantage of the automatic determination of discourse units is its ability to be scaled up to provide genre analyses from corpora of any size.

4   Academic Writing at the Interface: A Continuum of Approaches

So far, then, we have distinguished research using corpus linguistic techniques from that based on discourse analysis and have seen how some recent work seeks an integration of the two. From this point on we would prefer to regard discourse and corpus approaches not as opposing ideas, but as constituting a continuum from top-down (more discourse-analytic) to bottom-up (more corpus-based) along which we can situate individual studies.

This volume reflects that view. The writer of each chapter carries out their analysis according to their own methodological perspective and in so doing, establishes their own individual position on the interface between corpus and discourse approaches. Thus the studies by John and Tang dealing with student identity and authority can be situated towards the discourse end of the continuum, while the genre analysis by Koutsantoni provides an example of a 'corpus-assisted discourse study'. Representing the integrative approach are the papers by Bondi, Charles, Shaw and Hewings, Coffin and North, all of which move between analysis of language features and consideration of their context of use. The chapters by Hyland and J. Flowerdew and Forest exemplify integration in two slightly different ways: the former includes information on the social context by using interview data to supplement the corpus analysis, while the latter combines the corpus technique of keyword analysis with a more discourse-based move analysis. The remaining chapters, by Holmes and Nesi, Granger and Pacquot, Gardezi and Nesi, Gardner and Holmes and Pecorari can be placed towards the corpus end of the continuum. These can best be

described as 'discourse-informed corpus studies' in that the discourse applications take second place to their primary concern, which is with the phraseological patterns found in their corpora.

By focusing on the interface between corpus and discourse approaches, this volume has clear and direct pedagogical implications. The research reported here is driven not just by the desire to investigate features of academic writing, but more importantly by the need to bring those discoveries to the classroom and writing centre. It is the multi-level concern with detailed linguistic analysis, whole texts and their context of use that enables this volume to offer research that can be used and built upon both by academics and practising teachers.

The volume closes with an Afterword by John Swales. Swales himself is the acknowledged pioneer of genre approaches to discourse studies in general and to EAP in particular. More recently he has been closely associated with the MICASE academic corpus project at the University of Michigan. He is therefore well placed to see the advantages of exploring the interface of discourse and corpus studies. His paper ends with something of a challenge, in that he suggests that it is more fruitful for the broad sweep of discourse approaches to be supplemented by the detail of corpus studies than vice versa. This is a statement that deserves a far longer response than space permits here, but it rightly raises questions to which we hope the book itself offers a number of answers.

5   Themes and Connections within this Volume

The three-part organization given in the table of contents reflects one constellation of themes addressed in this volume. The chapters in Part I are concerned with genre and disciplinary discourses, while those in Part II have a common focus on interpersonal discourses. In the final part, the emphasis is on learner discourses, and on the ways that research findings in applied linguistics can contribute to the learning process. This division reflects our understanding of the ways in which the chapters that make up this volume can inform each other, but it is not, of course, the only possible grouping. It is therefore worth highlighting a number of other themes which form productive connections between the chapters, thus providing alternative pathways for readers to take through the collection.

The needs and concerns of writers depend greatly, of course, on the position they occupy within the academic community. Eight of the chapters in this book (3, 6, 7 and 9–13) deal with undergraduate writing, three (1, 8 and 14) with postgraduate work and five (2, 4, 5, 10 and 11) with the texts of expert writers.

Two of the chapters are concerned with the academic processes themselves. In Chapter 7, Hewings, Coffin and North look at e-conferencing on distance

courses, and shed light on the role such conferences play in the process of participants orienting themselves to the course and to each other. In Chapter 14, John examines an aspect of the writing process, revision.

Chapters 3, 4–6, 8 and 10–12 all share a concern with the investigation of lexico-grammatical features, while others take up structural and rhetorical moves and patterns (1, 2, 7, 9) and the ways they are signalled (13).

Attention to disciplinary differences and linguistic features which occur within specific disciplines is a common component in research on academic writing, and that is also reflected in the chapters of this volume. Disciplines examined include applied linguistics (1, 14); biology (5); business (2); economics (12); English studies (9, 11); health and social care (7); materials science (8); physics (3); and politics (8, 12), while multi-disciplinary comparisons are carried out in Chapters 3, 6, 8, 11 and 13.

This collection of papers, then, is illustrative of the many interlocking concerns that motivate research on written academic discourse and the wide range of approaches currently applied to its analysis. In so doing, it indicates the richness of description that can be achieved by exploiting these multiple perspectives. As synonyms of the term *interface*, the *Oxford English Dictionary* offers *meeting-point, common ground* and *dialogue*. It is hoped that the volume provides a useful meeting-point for these multiple themes and that it can contribute, in particular, to the further development of common ground and dialogue between corpus linguistics and discourse analysis.

References

Ädel, A., & Reppen, R. (Eds). (2008). *Corpora and discourse: The challenges of different settings.* Amsterdam: Benjamins.

Baker, P. (2006). *Using corpora in discourse analysis.* London: Continuum.

Bazerman, C. (1988). *Shaping written knowledge.* Madison: University of Wisconsin Press.

Berkenkotter, C., & Huckin, T. (1995). *Genre knowledge in disciplinary communication: Cognition/culture/power.* Hillsdale, NJ: Erlbaum.

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

—(2006a). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*, 97–116.

—(2006b). *University language.* Amsterdam: Benjamins.

Biber, D., Connor U., Jones, J., & Upton, T. (2007). Vocabulary-based discourse units in biology research articles. In D. Biber, U. Connor & T. Upton, *Discourse on the move* (pp. 175–212). Amsterdam: Benjamins.

Biber D., Connor U., Kanoksilapatham, B., & Upton, T. (2007). Rhetorical moves in biochemistry research articles. In D. Biber, U. Connor & T. Upton, *Discourse on the move* (pp. 73–120). Amsterdam: Benjamins.

Biber, D., Connor, U., & Upton, T. (2007). *Discourse on the move.* Amsterdam: Benjamins.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly, 36*, 9–48.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow: Longman.

Bondi, M. (2008). Emphatics in academic discourse. In A. Ädel & R. Reppen (Eds), *Corpora and discourse* (pp. 31–55). Amsterdam: Benjamins.

Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for Specific Purposes, 13*, 47–59.

Bunton, D. (2002). Generic moves in PhD thesis introductions. In J. Flowerdew (Ed.), *Academic discourse* (pp. 57–75). London: Longman.

—(2005). The structure of PhD conclusion chapters. *Journal of English for Academic Purposes, 4*, 207–224.

Carter-Thomas, S., & Rowley-Jolivet, E. (2008). *If*-conditionals in medical discourse: From theory to disciplinary practice. *Journal of English for Academic Purposes, 7*, 191–205.

Casanave, C.P. (2002). *Writing games: Multicultural case studies of academic literacy practices in higher education.* Mahwah, NJ: Erlbaum.

Charles, M. (2006a). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*, 310–331.

—(2006b). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics, 27*, 492–518.

—(2007). *Argument or Evidence*? Disciplinary variation in the use of the noun *that* pattern in stance construction. *English for Specific Purposes, 26*, 203–218.

Coffin, C. (2006). *Historical discourse: The language of time, cause and evaluation.* London: Continuum.

Conrad, S. (2002). Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics, 22*, 75–95.

Dudley-Evans, T. (1986). Genre analysis: An investigation of the introduction and discussion sections of MSc dissertations. In M. Coulthard (Ed.), *Talking about text: Studies presented to David Brazil on his retirement* (pp. 128–145). Birmingham: ELR.

Flowerdew, J. (2003). Signalling nouns in discourse. *English for Specific Purposes, 22*, 329–346.

Flowerdew, L. (1998). Corpus linguistic techniques applied to textlinguistics. *System, 26*, 541–552.

—(2008). *Corpus-based analyses of the problem-solution pattern.* Amsterdam: Benjamins.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Cambridge: Polity Press.

Granger, S. (Ed.). (1998). *Learner English on computer.* Harlow: Addison Wesley Longman.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *The international corpus of learner English. Version 2. Handbook and CD-ROM.* Louvain-la-Neuve: Presses universitaires de Louvain.

Granger, S., Hung, J., & Petch-Tyson, S. (Eds). (2002). *Computer learner corpora, second language acquisition, and foreign language teaching*. Amsterdam: Benjamins.

Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes, 4*, 257–277.

Halliday, M.A.K. (1978). *Language as social semiotic*. London: Arnold.

—(1987/1993). Language and the order of nature. In N. Fabb et al. (Eds), *The linguistics of writing*. Manchester: Manchester University Press. Reprinted in M.A.K. Halliday & J.R. Martin, *Writing science: Literacy and discursive power* (pp. 106–123). London: Falmer.

Harwood, N. (2005a). 'Nowhere has anyone attempted . . . In this article I aim to do just that': A corpus-based study of self-promotional *I* and *we* in academic writing across four disciplines. *Journal of Pragmatics, 37*, 1207–1231.

—(2005b). 'We do not seem to have a theory . . . the theory I present here attempts to fill this gap': Inclusive and exclusive pronouns in academic writing. *Applied Linguistics, 26*, 343–375.

Hasan, R. (1984/1996). The nursery tale as genre. *Nottingham Linguistic Circular*, 13. Revised version in R. Hasan, *Ways of saying: Ways of meaning* (pp. 51–72). London: Cassell.

Hewings, A. (2004). Developing discipline-specific academic writing: An analysis of undergraduate geography essays. In L. Ravelli & R. Ellis (Eds), *Analysing academic writing: Contextualized frameworks* (pp. 131–152). London: Continuum.

Hewings, M., & Hewings, A. (2002). 'It is interesting to note that . . .': A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes, 21*, 367–383.

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes, 16*, 321–337.

Hood, S. (2004). Managing attitude in undergraduate academic writing: A focus on the introductions to research reports. In L. Ravelli & R. Ellis (Eds), *Analysing academic writing: Contextualized frameworks* (pp. 24–44). London: Continuum.

Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam: Benjamins.

—(1999a). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics, 20*, 341–367.

—(1999b). Disciplinary discourses: Writer stance in research articles. In C.N. Candlin & K. Hyland (Eds), *Writing: Texts, processes and practices* (pp. 99–121). London: Longman.

—(2000). *Disciplinary discourses: Social interactions in academic writing*. Harlow: Longman.

—(2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes, 20*, 207–226.

—(2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

—(2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4–21.

Hyland, K., & Bondi, M. (Eds). (2006). *Academic discourse across disciplines*. New York: Peter Lang.

Jones, C., Turner, J., & Street, B. (Eds). (1999). *Students writing in the university*. Amsterdam: Benjamins.

Kwan, B.S.C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes, 25*, 30–55.

Martin, J.R. (1985). *Factual writing: Exploring and challenging social reality.* Deakin: Deakin University Press.

Myers, G. (1990). *Writing biology: Texts in the social construction of scientific knowledge.* Madison: University of Wisconsin Press.

Partington, A. (2004). Corpora and discourse, a most congruous beast. In A. Partington, J. Morley, & L. Haarman (Eds), *Corpora and discourse* (pp. 11–20). Bern: Peter Lang.

Pecorari, D. (2008). *Academic writing and plagiarism: A linguistic analysis.* London: Continuum.

Ravelli, L., & Ellis, R. (Eds). (2004). *Analysing academic writing: Contextualized frameworks.* London: Continuum.

Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes, 24*, 141–156.

—(2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes, 7*, 55–67.

Sanderson, T. (2008). Interaction, identity and culture in academic writing. In A. Ädel & R. Reppen (Eds), *Corpora and discourse* (pp. 58–92). Amsterdam: Benjamins.

Shaw, P. (2003). Evaluation and promotion across languages. *Journal of English for Academic Purposes, 2*, 343–357.

Sinclair, J., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils.* Oxford: Oxford University Press.

Swales, J.M. (1981). *Aspects of article introductions.* Birmingham: University of Aston Language Studies Unit.

—(1990). *Genre analysis: English in academic and research settings.* Cambridge: Cambridge University Press.

—(2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J. Flowerdew (Ed.), *Academic discourse* (pp. 150–164). London: Longman.

—(2004). *Research genres.* Cambridge: Cambridge University Press.

Thompson, P. (2005). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes, 4*, 307–323.

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

Tognini-Bonelli, E., & Del Lungo Camiciotti, G. (Eds). (2005). *Strategies in academic discourse.* Amsterdam: Benjamins.

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes, 22*, 365–385.

Part I

Focus on Genre and Disciplinary Discourses

This page intentionally left blank

Introduction to Part I

Maggie Charles

The chapters that constitute Part I of this volume provide a good illustration of the different levels of analysis that can be the focus of attention in research on written academic discourse. As Part I progresses, the chapters become more specific, moving from studies at the generic level to those that focus primarily on the phraseology of a single discipline.

The section begins with two chapters that analyse genre. Flowerdew and Forest integrate corpus and discourse approaches in order to examine the PhD literature review in applied linguistics. They first carry out a manual analysis of the moves and steps in this genre and then use the corpus technique of determining key keywords as a means of linking moves and steps to their typical lexico-grammatical realizations. In particular, they investigate the phraseology and communicative functions associated with the noun *research,* thus bridging the gap between generic and lexico-grammatical analyses. As they point out, it is through addressing the need for reliable analyses at both levels that the production of more useful pedagogic materials can be achieved.

Koutsantoni also performs a move analysis, but uses a more discourse-based, top-down methodology. She draws on a corpus of cases of support from successful funding applications in the discipline of business/management in order to provide an analysis of the under-researched, but high-stakes genre of grant proposals. She stresses the socially constructed nature of this genre and how it forms part of a genre set which includes, for example, research articles and the grant proposal guidelines provided by funding bodies. Koutsantoni's work has immediate applications in research staff development and she suggests a pedagogical approach which would involve writers in carrying out their own analyses of grant proposals and comparing them with other genres in the set.

The next three chapters in Part I focus on disciplinary discourses. The study by Holmes and Nesi deals with novice academic writing and is based on the corpus of British Academic Written English (BAWE), which consists of university assignments that have been awarded good grades. Using keyword analysis of verbal and mental processes, this chapter first contrasts two 'pure' disciplines, one 'hard' (physics) and one 'soft' (history), and then compares the results with those from three applied disciplines (engineering, medicine and hospitality, leisure and tourism management). Holmes and Nesi link the differences found between these disciplines to differences in their epistemology and

argue that pedagogical materials need to provide discipline-specific examples of both lexis and the grammar of the clause.

Both Bondi and Pecorari focus on the expert writing of research articles in a single discipline, investigating aspects of the phraseology of history and biology respectively. Bondi's chapter moves between corpus and discourse approaches, analysing not only frequent clusters, but also their co-text. She first examines 'chrononyms', nominalizations of time periods, for example, *the First World War*, and shows how they function both to set an event in time and to provide an interpretation or evaluation of the period. She finds that these chrononyms often form part of longer adverbial expressions that refer to transitional periods by relating one event to another, landmark event (e.g. *in the wake of the Terror*). Bondi argues that such phraseological sequences blend the notions of time and causality, thereby reflecting the epistemology of the discipline.

Pecorari takes a rather more corpus-based approach, using n-gram analysis to identify lexical bundles in a highly specialized corpus of biology research articles dealing with the topic of *Candida albicans*. She finds both considerably more and considerably longer bundles than previously reported in the literature. Although it is to be expected that texts on a given topic will share a large stock of multi-word units, Pecorari also shows how the lexical bundles indicate conventionalized discoursal moves within this topic area of biology. In pedagogical terms, she makes the point that instruction at postgraduate and postdoctoral level could benefit from being more specifically targeted, not just to the discipline, but to the topic area in which the student is working.

An underlying theme that runs through all the studies in Part I concerns the specificity of the discourse examined and its associated phraseology. It is by now well-accepted that academic discourse varies according to discipline and genre. However the findings from these studies suggest that this variation may perhaps be more extensive and deep-seated than hitherto realized. These studies, therefore, have profound implications for pedagogical practices, suggesting that the more specific and homogeneous the corpus analysed, the more reliable the pedagogical descriptions are likely to be and the greater the likelihood that these descriptions can form the basis of useful teaching materials for the relevant group of students.

Chapter 1

Schematic Structure and Lexico-Grammatical Realization in Corpus-based Genre Analysis: The Case of *Research* in the PhD Literature Review[1]

John Flowerdew and Richard W. Forest

1 Introduction

Within discourse analysis much research has been conducted on academic genres. A considerable amount of this research, following in the tradition of Swales (1990) and Bhatia (1993), has focused on schematic structure, the sequential patterning of communicative acts within genres into moves and steps. This work has proved to be of great value in developing pedagogic models for application to the training of apprentice scholars in their writing (see e.g. Paltridge & Starfield, 2007; Swales & Feak, 1994, 2000). More recently, research has started to investigate the potential relationship between moves and steps and their typical lexico-grammatical realizations (L. Flowerdew, 1998, 2008; Upton & Connor, 2001). This has seen the introduction of corpus techniques into what was previously a text linguistic enterprise. Using a corpus of PhD thesis literature reviews (LRs) the present study builds on this corpus-based genre analysis. It takes the notion of key keyness of lexical items – how some words are more (or less) frequent in many texts across a corpus (Scott & Tribble, 2006) – to select a key keyword in the LR corpus, *research*, for investigation. Using corpus techniques, the various uses of this word in the different moves and steps of the LRs are revealed, identifying its typical collocations and associated communicative functions. The study thus shows how key keywords can be used as the point of departure for establishing form–function relations in corpus-based genre analysis. The implications for both genre research and pedagogical application are highlighted.

2 Literature Review

2.1 Move analysis and the CARS model

Leading work in genre analysis focused on schematic structure was carried out by Swales (1990) and, a little later, Bhatia (1993). Swales and Bhatia both define

genre as a type of purposeful communicative event which is readily identified by what they refer to as its discourse community (those people who regularly engage in it). Genres are staged events (Bhatia, 1993; Swales, 1990; see also Martin, 1992), that is, they develop through a sequence of what Swales (1990) and Bhatia (1993), following earlier work by Sinclair and Coulthard (1975) on classroom discourse, call *moves*. This sequence, referred to as a genre's schematic structure, may vary between different instances of a genre. Some moves may be optional, some may occur in different orders, some may be embedded in others, and some may be repeated (Swales, 1990; see also Hasan, 1977, 1979, 1989; Ventola, 1987). However, a prototypical schematic structure will be recognizable in terms of the most typical realization pattern, as identified by the discourse community (Swales, 1990: 58).

Perhaps the best known model of generic staging, and one highly relevant for this study, is Swales's (1990: 141) CARS ('Create a Research Space') structure, which he posited for academic research article introductions. This model consists of three moves, each with its component 'steps', as follows:

Move 1. Establishing a territory
    Step 1. Claiming centrality   and/or
    Step 2. Making topic generalization   and/or
    Step 3. Reviewing items of previous research

Move 2. Establishing a niche
    Step 1A. Counter-claiming   or
    Step 1B. Indicating a gap   or
    Step 1C. Question-raising   or
    Step 1D. Continuing a tradition

Move 3. Occupying the niche
    Step 1A. Outlining purposes   or
    Step 1B. Announcing present research
    Step 2. Announcing principal findings
    Step 3. Indicating Research Article structure

The model has been widely studied since its original formulation and, while there have been other studies which emphasize the recursive nature of the moves (Bunton, 2002) and different types of variation across disciplines and contexts (Ahmad, 1997; Burgess, 2002; Melander, 1998), the three-move structure has been shown to be fairly stable. Among other work on research article introductions, that of Lewin, Fine, and Young (2001) is worthy of mention. This work emphasized the obligatory-optional aspect of the steps in the moves, a feature which is employed in the model used in the

present study for PhD LRs. Also relevant here is the work of Bunton (2002). Based on Swales's CARS model, Bunton (2002) developed a model for the introductions to PhD theses. His model employs the same basic three-move structure for the introduction to research articles of establishing a territory, establishing a niche and occupying a niche, but also establishes a range of new elements.

Most relevant, though, for the present study is the work of Kwan (2006). Given that introductions and LRs have been claimed to belong to the same genre, Kwan set out to study potential similarities and differences between them. Based on the same corpus as the one used in the present study (J. Flowerdew, 2004), Kwan conducted a move analysis to show how the LR of the PhD thesis in Applied Linguistics, while exhibiting a fairly predictable schematic structure similar to that of the research article and PhD thesis introduction, at the same time '[is] not structurally entirely the same' (p. 52).

Kwan identified a schematic structure for the PhD thesis in Applied Linguistics, as shown in Table 1.1. Following Bhatia, Kwan uses the term 'strategy' in place of Swales's 'step' (Kwan, 2006: 34) for the sub-components of the moves. As we will be applying Kwan's model in the present study, we will also follow this convention.

Table 1.1　A move structure for LRs

Move 1 Establishing one part of the territory of one's own research by:	
Strategy A[a]	surveying the non-research-related phenomena or knowledge claims
Strategy B[a]	claiming centrality
Strategy C	surveying the research-related phenomena
Move 2 Creating a research niche (in response to Move 1) by:	
Strategy A	counter-claiming
Strategy B	gap-indicating
Strategy C	asserting confirmative claims about knowledge or research practices surveyed
Strategy D	asserting the relevancy of the surveyed claims to one's own research
Strategy E	abstracting or synthesizing knowledge claims to establish a theoretical position or a theoretical framework
Move 3 (optional) Occupying the research niche by announcing:	
Strategy A	research aims, focuses, research questions or hypotheses[b]
Strategy B	theoretical positions/theoretical frameworks[b]
Strategy C	research design/processes[b]
Strategy D	interpretations of terminology used in the thesis[b]

[a] Strategy 1B tends to precede Strategy 1A when the two co-occur.
[b] Sub-strategy: justifying or claiming contributions.

Source: Reproduced from Kwan, 2006: 51.

2.2   Generic moves and linguistic realization

In his original formulation of genre theory, Swales (1990) claimed that as well as having a prototypical schematic structure, the various communicative functions (moves and steps) of a genre exhibit typical conventionalized verbalization patterns (realizations) which, like the schematic structure, are recognized as such by the discourse community. The following are examples of authentic realizations drawn from academic articles of the first step of the first move of Swales's research article introductions, 'claiming centrality', as presented by Swales (1990: 144):

Recently, there has been a spate of interest in how to . . .

In recent years, applied researchers have become increasingly interested in . . .

The possibility . . . has generated interest in . . .

Recently, there has been wide interest in . . .

The time development . . . is a classic problem in fluid mechanics.

The explication of the relationship between . . . is a classic problem of . . .

Many investigators have recently turned to . . .

It is important to stress that there is no one-to-one relation between move or step and realization pattern (unless a genre is extremely conventionalized, such as vows at a wedding, or the oath at a public swearing in). An important question, however, is to what extent it is possible to predict likely realization patterns of moves and steps such as these. Since the publication of Swales's original work, relatively little has been done on this area. There has been much more interest in analysing schematic structure than in investigating linguistic realization. This is curious, because, if we take those patterns cited by Swales, we might intuitively feel that there is a quite high probability of such patterns occurring as realizations of the given rhetorical move. For example, immediately noticeable is the recurrence of the adverbial 'recently/in recent years', the noun 'interest', and the noun phrase 'classic problem' in more than one instance.

The probable main reason why there has been relatively little work in this area of linguistic realization of generic moves and steps may be methodological. If systematic relations between form and function are to be identified, it is extremely time-consuming to examine the large number of examples of individual moves and steps which would be required. Fortunately, however, with the advent of computerized text analysis software, the opportunity to do such work is now available.

There are a number of studies in the literature related to such an approach. As a first example, Connor and Upton at the University of Indiana, along with

various collaborators, have taken an interest in the application of corpus tech-
niques to genre analysis (Connor, Precht, & Upton, 2002; Upton & Connor,
2001; Upton, 2002). Of most interest here is their 2001 study (Upton & Connor,
2001), where the two researchers examined politeness strategies in the two
moves in a corpus of letters of application. Other work in this area is that of
L. Flowerdew (1998, 2008). Connor et al. indeed cite L. Flowerdew (1998) as
showing the way in this field. L. Flowerdew has argued as follows:

> Another suggestion, which I believe would have wide pedagogical applica-
> tions, is more exploitation of the tagging function of existing software on
> the market. As Leech (1991) remarks, most of the work on text annotation
> (tagging) has been done at the grammatical (word class) or syntactic (pars-
> ing) level. Very little has been done on the semantic or pragmatic discourse
> level to date. For example, text could be tagged manually to indicate the
> generic 'move structures' such as background, scope, purpose in the intro-
> ductory sections of a report. (549)

L. Flowerdew has herself done some empirical work in this area. One study is
a recent paper (L. Flowerdew, 2008), where she again suggests a combination
of genre-based and corpus-based analysis, exemplifying the approach with an
investigation of the Problem–Solution pattern in professional reports.[2] She
shows in her analysis how words such as *problem* and *impact* are typically used
in the realization of the Means-Purpose relation in the Problem-Solution
pattern.

In a recent book, Biber, Connor, and Upton (2007) present a number of
studies which investigate generic structure, focusing on fund-raising genres
and research articles. Two approaches are exemplified: traditional functional
analysis of move structure (by hand) and multi-dimensional analysis, using
vocabulary-based discourse units (VBDUs). In the latter approach, the VBDUs
are identified automatically by comparing 50-word 'windows', or segments, of
text. The windows are opened one word at a time and, at each point, the two
open windows are compared. This occurs throughout the whole text under
examination. VBDU beginning and end points are identified where win-
dows maximally diverge. This approach, as Biber et al. (2007) make clear, is
a complementary way to establish move boundaries. The boundaries that are
established do not coincide with the functional boundaries identified by hand
in traditional functional analysis. While VBDU analysis is not a functional
approach, nevertheless, the distinct sets of words which are used in each of the
individual units might be operationalized in some sort of pedagogic applica-
tion, although this is not developed in Biber et al. (2007).

Of more interest to the present study, in one of the studies reported in Biber
et al. (2007, chapter 5) a rather different approach is used and keywords are
analysed to show how they occur typically in the *appeal* section of fund-raising

letters and in research articles. This approach is close to the sort of analysis conducted in the present study, where we will look at how one keyword (or, rather, key keyword [see below]) may relate to different discourse moves.

3   Keywords and Key Keywords

The terms keyword and key keyword have already been used. At this stage it is appropriate to make clear the meaning we are attributing to these two terms. A keyword is a word that is particularly frequent (a 'positive' keyword) or infrequent (a 'negative' keyword) in a corpus in comparison to its frequency in a reference corpus. In this study, we will use the British National Corpus as our reference corpus. A key keyword is simply a keyword which is key in multiple texts in a corpus. Key keyword identification has a number of uses, but as far as the present study is concerned, its purpose is to ensure that no single text in the corpus has a disproportionate effect on the analysis. Key keyness thus helps to reduce the likelihood that the resulting keyword lists are overly influenced by any particular author's idiolect or by text-internal patterns of repetition.

Key and key keyword identification procedures are useful tools for sorting corpus data in meaningful ways, identifying words which are likely to be worthy of more detailed study. Corpus linguists are often faced with the problem of how to select from a very large pool of data particular features worthy of follow up analysis: the language corpora of today generally provide far more data than can be reasonably summarized in any single study. Keywords and key keywords provide one method of applying *selectivity* (Scott, 2007: 125) to the corpus, allowing the study to focus on the most significant patterns for analysis. The utility of findings for teacher education and classroom application in particular is dependent on finding ways to tease out noteworthy patterns from the large volume of data provided by even a small language corpus.

4   The Study

4.1   Corpora

This study is based on a 379,397-word corpus of 20 PhD literature review chapters in applied linguistics created by J. Flowerdew (2004). All of the chapters were written by native speakers of English who have also published internationally and can hence be assumed to be good models. The corpus is a small one by today's standards (for justifications of the uses of small corpora, see e.g. J. Flowerdew, 1993a; L. Flowerdew, 2008; Ghadessy, Henry, & Roseberry, 2001). There are two justifications for the use of a corpus of this size in this study. First, there is a practical justification. With the type of work undertaken here,

involving, on the one hand, the hand-coding of moves and steps and, on the other hand, a fine-grained analysis of the semantic/pragmatic values of the many exemplars of the keyword(s) identified, there is a limit to the amount of data that can be handled. Second, and more importantly, there is a theoretical justification. Because the 20 texts which make up the corpus are all of the same genre and all belong to the same discipline, in spite of the small size of the corpus, sufficient data is generated for the purpose of the analysis.

The corpus exists in two forms: one in which it retains its original structure as a corpus of 20 literature review chapters, and a second in which it has been segmented into moves and strategies according to Kwan's (2006) model of the schematic structure of PhD literature reviews. The segmentation process for the identification of moves and strategies is discussed in detail in Kwan (2006).

Keyword and key keyword analysis requires a reference corpus. As previously stated, this study uses the British National Corpus (BNC) for this purpose. All keyword and key keyword generation procedures in this study have been carried out using Scott's (2004) WordSmith Tools 4 package. The BNC wordlist used in this study is a wordlist for the whole of the BNC World edition (100 million words) and was retrieved from the WordSmith Tools website. The BNC is a useful reference corpus for a number of reasons. It is readily available to a large number of linguists, its contents have been thoroughly documented (Lee, 2001), and its use in numerous corpus-based studies means that it provides a recognizable common ground for keyword and key keyword comparisons.

4.2 Procedure

Keywords and key keywords were identified using the following procedure. First, a wordlist was generated for each of the 20 literature review chapters in the corpus. Following this, the Keywords function of WordSmith Tools was used to compare each of these wordlists to the BNC World wordlist. This procedure identifies those words which are markedly frequent or infrequent in the literature review chapters compared against their general frequency in the BNC. Keywords were calculated using the default statistical test and p value set in WordSmith: Dunning's log likelihood test (Dunning, 1993) to determine keyness and a minimum p value of 0.000001.

Using the resulting keyword lists, key keywords were then identified. WordSmith Tools was used to create a database of all keywords which occur in multiple texts in the corpus, recording the number of texts in which a given word is key. This information was exported to Microsoft Excel for ease of analysis. The key keyword database presents an overview of which keywords are most widely distributed in the corpus as a whole.

Table 1.2 shows the highest ranked key keywords in our corpus. Only those words which are key in 60 per cent or more of the texts have been included. As Scott and Tribble note (2006: 78), there is no standard cut-off point for

the number of texts in which a word should be key to be included in a study. Our fairly strict cut-off point for Table 1.2 is made possible by the specialized nature of the corpus.[3]

Numerals are key in all 20 literature reviews in the corpus. Concordancing confirms that this is primarily due to the frequency of in-text citations. The importance of discipline to word choice is also readily observable from the results. All 20 of the literature reviews are from PhD theses in applied linguistics, and this fact is reflected in a number of key keywords: *language*,

Table 1.2   Key keywords in the literature review corpus

Rank	Key keyword	Texts	% of Texts	Overall word frequency	Standardized frequency per 1,000[a]
1	#	20	100	15,840	25.7
2	language	19	95	1,512	2.5
3	of	17	85	16,994	27.6
4	non	16	80	274	0.4
5	context	15	75	547	0.9
6	studies	15	75	535	0.9
7	analysis	14	70	604	1.0
8	linguistic	14	70	330	0.5
9	research	14	70	900	1.5
10	study	14	70	978	1.6
11	text	14	70	1,128	1.8
12	texts	14	70	680	1.1
13	chapter	13	65	193	0.3
14	communicative	13	65	182	0.3
15	discourse	13	65	617	1.0
16	example	13	65	580	0.9
17	knowledge	13	65	684	1.1
18	students	13	65	1,081	1.8
19	approach	12	60	532	0.9
20	approaches	12	60	193	0.3
21	discussion	12	60	225	0.4
22	English	12	60	826	1.3
23	strategies	12	60	475	0.8

[a] Frequency counts per thousand words have been standardized following the procedures outlined in Appendix B of Biber (2006). The formula is 'normed # of word types = # of word types / square root of corpus size' (Biber, 2006: 252–257). This adjustment is necessary because the sub-corpora are of varying sizes, and word type frequency is sensitive to corpus size. Just as type/token ratios are not directly comparable across corpora of different sizes, and must be normalized, so are word counts per thousand or million words if comparing corpora of different sizes. We have adopted Biber's procedure for its relative simplicity. It allows more accurate comparisons to be made between sub-corpora of varying sizes, reducing the distorting effect of smaller corpora tending to produce higher word per thousand counts than larger corpora. All 'word per thousand' counts in this chapter are normalized in this manner. The resulting tables provide a more conservative comparison of word frequencies per thousand than would raw counts. We have opted for the more conservative approach. Raw counts of word frequencies per thousand can be easily derived from the data available in the chapter, if desired.

non- (primarily from *non-native* SPEAKER), *context, linguistic, text, texts, commu-nicative, discourse, students, English* and *strategies.* Roughly half of the top key keywords included in Table 1.2 seem primarily to reflect discipline-specific patterns. While the chapters were drawn from a variety of areas within applied linguistics, a core of shared disciplinary vocabulary is in evidence.

In addition, there are other key keywords which are less clearly discipline-bound. These have the potential to be of broader interest, and include *of, stud-ies, analysis, research, study, chapter, example, knowledge, approach, approaches* and *discussion.* These words are candidates for a tentative pool of words likely to be characteristic of the genre of literature review. It is from this pool of terms that we have selected terms for more in-depth investigation.

We will discuss three related key keywords from Table 1.2 in this study, using the key keyword *research* as the starting point. Our study is meant to be illustrative of the kinds of findings which a procedure such as this can throw up, and it is part of an ongoing project which includes analyses of the other top key keywords in the corpus. While *research* is not itself the most frequent or most highly ranked key keyword of interest, it is both frequent (with 900 occurrences) and widely distributed (key in 70% of the texts) and is illustrative of the kinds of results that the procedure can produce. In addition, we will draw attention to some charac-teristics of the lemma STUDY, which occurs twice in the Table 1.2: the word form *study* (978 tokens) is key in 70 per cent of the texts and the word form *studies* (535 tokens) in 75 per cent. A clear picture of the relationship between *research* and schematic structure can only be provided by taking into account the division of labour between *research, study* and *studies* across moves and steps in the corpus.

4.3 Findings

4.3.1 *The corpus sorted into moves and strategies*

Our study uses the move and strategy labels from Table 1.1 (e.g. Strategy 2A: Counterclaiming, Strategy 2B: Gap-indicating, etc.), with the following clarifications[4]:

1. Text assigned to a 'miscellaneous' or 'uncategorized' category within a move is identified by an X. Thus, uncategorized text belonging to Move 1 is identified as 1X, and to Move 2 as 2X.
2. Our study uses an unmodified version of Kwan's (2006) corpus, which breaks Moves 1 and 2 down into sub-strategies but treats Move 3 as a single unit. Kwan's (2006) coding scheme for the corpus treats all items in Move 3 under a single label ('3') rather than a Move + Strategy label. Thus, while the alphanumeric code '2B' indicates the strategy labelled 'Gap-indicating', the numeral '3' is used for all sub-strategies involved in 'occupying a niche'. We have chosen to retain this structure so that the findings can be readily compared with previous work on the corpus.

Table 1.3   Corpus of moves and strategies

Move and strategy	Size of sub-corpus
Move 1	
1A Surveying non-research claims	154,281
1B Claiming centrality	20,301
1C Surveying the research	68,900
1X Uncategorized	20,294
Move 2	
2A Counter-claiming	78,959
2B Gap-indicating	5,833
2C Confirmative claims	8,230
2D Asserting relevancy	4,396
2E Abstracting/synthesizing	794
2X Uncategorized	1,411
Move 3	
3 Occupying the niche	15,998
TOTAL CORPUS	**379,397**

The structure of the corpus when sorted into moves and strategies is shown in Table 1.3.

4.3.2   Schematic structure and the case of research

As was shown in Table 1.2, the key keyword *research* occurs 900 times in the corpus and is key in 70 per cent of the LR texts. While the word *research* has the potential to be a noun or a verb, in all 900 occurrences of the word in our corpus it occurs as a noun. Table 1.4 shows the frequency of *research* in each move and strategy in the corpus.

A visual representation of the relative frequency of *research* across the different strategies is shown in Figure 1.1.

Strategy 2B ('Gap-indicating') is of particular interest: *research* shows a dramatic spike in frequency in this strategy to a rate of 7.4 occurrences per thousand words. This spike is approximately five times its overall rate of occurrence in the corpus (1.5 occurrences per thousand words). We will investigate the reasons for this jump in frequency below. Compared with its overall rate of occurrence, it is also relatively frequent in Strategy 1B ('Claiming centrality'), with 3.2 hits per thousand words, followed by 2C ('Confirmative claims') and 2D ('Relevancy'). It is comparatively infrequent in 1A ('Surveying non-research related knowledge claims'), at 0.8 hits per thousand. While there are no instances of *research* in strategy 2E ('Abstraction/ synthesis'), little import can be attributed to this fact given the small size of that particular strategy.

Table 1.4   Key keyword *research* by move and strategy

Move and Strategy	Total	Standardized frequency per 1,000
Move 1		
1A Surveying non-research claims	197	0.8
1B Claiming centrality	107	3.2
1C Surveying the research	184	1.6
1X Uncategorized	56	1.7
Move 2		
2A Counter-claiming	184	1.4
2B Gap-indicating	70	7.4
2C Confirmative claims	35	2.6
2D Asserting relevancy	17	2.4
2E Abstracting/synthesizing	0	0.0
2X Uncategorized	3	1.3
Move 3		
3 Occupying the niche	47	1.8
TOTAL CORPUS	**900**	**1.5**
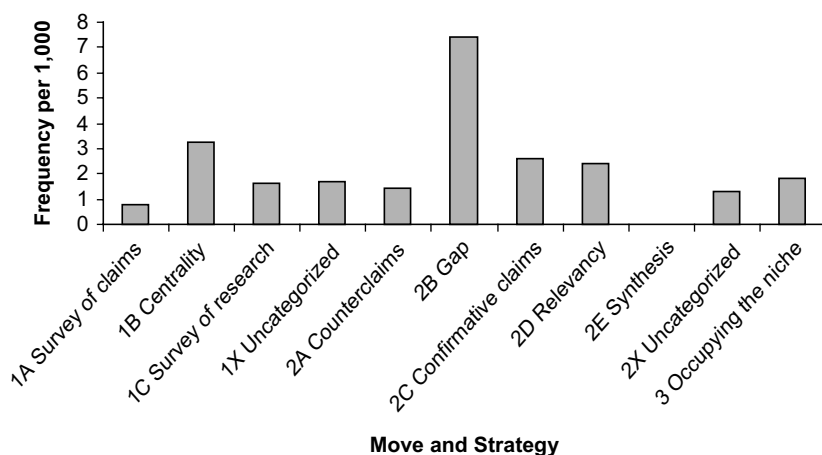
FIGURE 1.1   Key keyword frequencies: *research*

4.3.3   *Overall collocational behaviour of research*

Before looking at Strategy 2B in detail, it is important to get a general picture of the overall lexical behaviour of *research* in the corpus. It regularly collocates with words that specify a type or domain of research (be it a discipline, a research methodology or an object of research) and orient the PhD thesis with respect to it. This tendency is not restricted to any particular move or strategy,

but rather is spread throughout the corpus, reflecting the overall concern of the LR to situate the thesis with respect to the field and to prior work. One typical pattern is of a noun phrase with *research* as its head, taking a noun or adjective as a premodifier. This premodifying noun or adjective serves to specify the domain or approach in question. The most common examples can be placed in two broad semantic groups. Representative collocations and their frequencies in pre-modifiers are as follows:

> **Identifying a discipline, domain, field, or sub-field**: *(L1/L2) writing research* (15), *reading research* (11), *educational research* (10), *linguistic research* (10), *contrastive rhetoric research* (9), *SLA research* (9), *L2 research* (5)

> **Identifying an approach or methodology**: *qualitative research* (12), *questionnaire-based research* (11), *action research* (11), *empirical research* (7), *experimental research* (7), *naturalistic research* (7), *test-based research* (6), *observation-based research* (5), *language research* (5)

Many of these frequent collocations are technical or semi-technical terms. These terms serve multiple functions in the LR. They situate the research with respect to its place in the academic environment, aiding the prospective PhD candidate in building an identity with respect to the numerous sub-fields and specialty areas within the discipline as a whole. They establish boundaries, showing both what areas and approaches the researcher will engage with as well as indirectly marking off domains and approaches which will not be addressed. They narrow down the scope of the study to a few particular domains, a fact that is reflected in the distribution of the above collocations. For while *research* is a key keyword in many texts, and the types of collocations exhibited above are also spread across LRs, each individual collocation is largely representative of one author's concerns. Most of the particular collocations presented above are restricted to a particular LR. Two-thirds of the instances of *writing research* come from a single LR, while all but one instance of *reading research* come from another. And so it goes even for those collocations that appear on the surface to be fairly general in scope: every instance of *naturalistic research* comes from one LR, and all but two of *qualitative research* from yet another. But the basic pattern of pre-modifier + research is constant: it is within this common lexico-grammatical framework that particular kinds of research are distinguished from others.

The exception is the collocation *empirical research*, which occurs in seven of the LRs. In all but one of these instances, *empirical research* is mentioned as part of Strategy 2B in order to indicate a gap: *minimal, little, some, small-scale*, and *limited empirical research* has been done, while *further* empirical research *is needed*. In all these instances, *empirical* is used in a primarily evaluative sense – that research be *empirical* is a highly desirable feature, and the lack of empirical research in an area is a clear gap for the PhD thesis to fill.

This brings us to a third pattern of pre-modification into which *research* tends to fall: general statements about the 'extent' or 'scope' of prior work. Examples featuring pre-modifiers include *further research* (15), *little research* (11), *other research* (10), *some research* (9), *much research* (8), *more research* (5), *such research* (5). The lexico-grammatical pattern of 'EXTENT/SCOPE' + *research* in part reflects the demand that academic research provide novel or newsworthy findings with respect to current knowledge in the field. The two most frequent of these patterns, *further research* and *little research* are particularly interesting in this respect: they cluster strongly in Strategy 2B, to which we will now turn.

4.3.4   Strategy 2B

As noted in the discussion of Figure 1.1, *research* is five times as frequent in Strategy 2B: 'Gap-indicating' as it is in the corpus overall. A closer look at the 70 instances of *research* in this strategy reveals the important role it plays in explicitly indicating the gap itself. This is done in two broad ways, each of which can be identified with a canonical wording:

1. *There has been little research / little research has been done*
2. *Further research is needed / called for*

(1) **Little research has been done** – a claim is made that there is <u>a paucity of research</u> in a given approach, domain, or discipline, using a given method, and/ or carried out with a given purpose:

- And <u>little research</u> has been carried out with 'SUBJECT POPULATION'.
- However . . . there has been <u>little research</u> into 'DOMAIN/TOPIC'.
- Aside from 'PROPER NAME'S' work, there has been <u>little research</u> into 'DOMAIN/TOPIC'.
- While there has been <u>little research</u> with 'SUBJECT POPULATION 1', there has been even less, if any, with 'SUBJECT POPULATION 2'.
- NP is clearly important but there is <u>little research</u> in 'NAME OF APPROACH' to . . .
- There is <u>little research</u> on 'DOMAIN/TOPIC' however, and . . .
- There is <u>little research</u> on 'DOMAIN/TOPIC'.
- Relatively <u>little research</u> has been done in 'NAME OF DISCIPLINE'.
- and so it is perhaps surprising that <u>little research</u> has been conducted into 'DOMAIN/TOPIC'.
- There is very <u>little research</u> on 'DOMAIN/TOPIC'.
- There has been <u>minimal research</u> into 'DOMAIN/TOPIC'.
- But <u>minimal empirical research</u> . . . has been carried out to date within 'NAME OF DISCIPLINE'.

A gestalt of the common realizations suggests the following frames:

- There is little research in/within 'NAME OF DISCIPLINE/AREA' into/in/on 'DOMAIN/TOPIC.'
- Little research has been done/conducted/carried out in 'DOMAIN/TOPIC'.

It is also interesting to note that of the 11 instances of the collocation *little research* in the corpus, 10 are found in Strategy 2B. This suggests that this particular collocation is a strong indicator that the gap-indicating function is being expressed. This is a useful specific point for use in materials design and pedagogy: the collocation *little research* is a strong indication that a gap is being identified. Writers should be aware of the use of this phrase for explicit gap indications as well as the details of its realization: in addition to noting the tendency of *little research* to occur in the above identified frames, it is also perhaps worth adding that ***a little*** *research* does not occur in the corpus, and that the collocation *a little* is not a fitting alternative.

(2) ***Further research*** **is needed**: a claim is made that there is a <u>need for *research*</u> on a given topic or question, in a given approach, domain, or discipline, using a given method, and/or carried out with a given purpose.

- Further **research** is called for.
- <u>Further **research**</u> is needed in this area.
- . . . there is still a call for <u>further **research**</u> in this area.
- . . . a great deal of <u>further empirical **research**</u> is needed.
- The need for <u>further **research**</u> into 'DOMAIN/TOPIC' has been increasingly recognized . . .
- . . . <u>further **research**</u> is called for in two important areas when applying 'NAME OF METHODOLOGY'.
- There is an increasing recognition that <u>further **research**</u> is needed into 'DOMAIN/TOPIC'.
- Therefore, <u>further **research**</u> investigating the what as well as the how . . . could have positive implications.
- . . . which suggest areas in 'DOMAIN/TOPIC' where <u>further **research**</u> would be useful.

As was the case with *little research*, the string *further research* is particularly common in this strategy: in this case, 10 of its 15 occurrences occur in this category. Most occurrences of *further research* and nearly all of *little research* are limited to this strategy, in spite of the fact that Strategy 2B accounts for only 1.5 per cent of the total running words of the corpus. This suggests that these collocations are particularly strong signals of gap-indicating moves.

   Of note also is that both of the 'gap indicating' patterns above tend to colligate with present perfect verb phrases, often in the passive. This is exemplified

by the colligation of *little research* with *has been,* which appears both preceding it, as in the existential *there has been little research*, as well as following it, as in *little research has been done/carried out/conducted*. The use of the passive here makes sense in light of the purposes of the strategy, as the main goal of Strategy 2B is to indicate that little or no work has been done. Assuming the writer's claim is correct, there is no agent available. Of course, most of these claims are present in the form of *little research*, not *no research*, ensuring that the knowledge claims made are hedged.

Further research also colligates with passive verb phrases, as well as indications of necessity for the further research in question. In this case, the missing agent in strategy 2B instances of *further research* will likely be the writer herself or himself, who will step in and fill this gap in current knowledge in the form of the PhD thesis.

The relationship between *little research* and *further research* is also interesting in light of the fact that the word *research* is a non-count noun. A review of the examples from the corpus suggests that when *little research* is used, the 'non-count' semantics of the word are being exploited for strategic purposes. By stating that *little research* has been done, the writer is able to signal that there is a gap without specifically needing to challenge or criticize any particular researcher or study. The absence of work in the area is presented as a general fact rather than a specific failing of a particular researcher. This allows the writer to avoid threatening the face of other experts in the field (a field which the prospective PhD candidate hopes to join) while still emphasizing the importance of the gap: if *so little research* has been carried out on a particular subject or in a particular area, the need for this work is all the more pressing.

The flexibility of the non-count semantics of *research* is again open to strategic use. When talking about the *need for further research*, this *further research* is not really general and faceless at all. It is, in fact, quite specific: it is the work being carried out by the author of the text itself. The word *research* thus fills both non-specific and specific roles without changing form.

This is further evidenced with reference to one of the other key keywords in the corpus, the plural word form *studies*. Interestingly, this word form shares some of the distributional as well as the discourse semantic features of *research*, and these uses could be taught together. Its distribution is shown in Figure 1.2. Like *research*, it is particularly frequent in the gap-indicating strategy (2B), occurring 4.0 times per thousand words (its overall frequency in the corpus being 0.9 per thousand).

And it is used in a similar way, to indicate a relative lack of research in an area:

- Few **studies** have attempted . . .
- There were relatively few empirical **studies** . . .
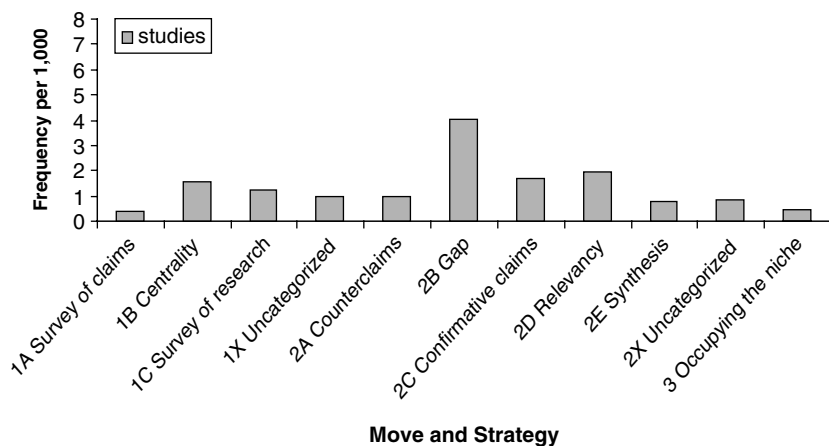- Relatively few empirical **studies** have involved . . .

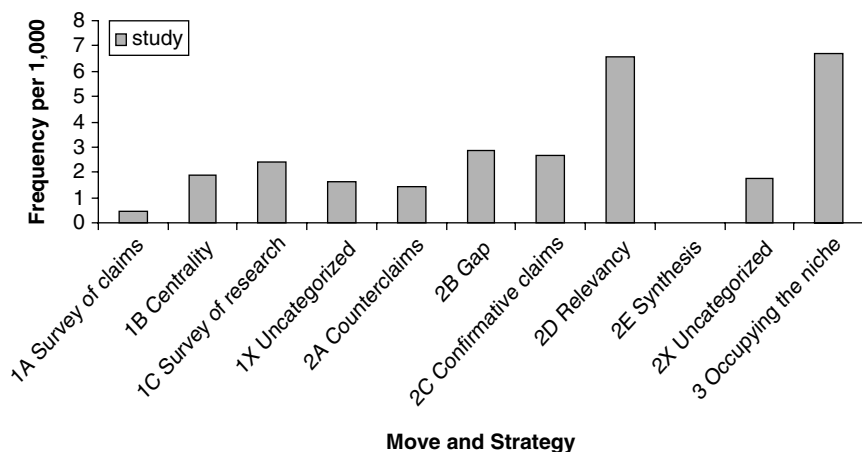**Figure 1.2**   Key keyword frequencies: *studies*

- There are very few published **studies** . . .
- These **studies** do not address the issue of . . .
- Aside from one or two **studies** . . .
- Only a limited number of **studies** have been undertaken . . .
- **Studies** are limited . . .
- Only a few **studies** have attempted . . .
- The **studies** remain largely non-replicable . . .
- No published research **studies** have been found . . .
- None of the **studies** . . . report how
- Only a few **studies** have considered . . .

Interestingly, *studies* is rarely used to indicate a need for future work, with only 2 instances of this use in the corpus:

- there is a clear need for . . . **studies**
- the need for future **studies** . . .

If our hypothesis about *research* is correct, this should not be surprising: while the plural form *studies* is an effective way to signal a general lack of research without criticizing or challenging specific researchers or studies, it lacks the flexibility of *research* to signal that the current study will fill the need for more research. Our corpus suggests that the singular form *study* is not available to fill this role, either, as it tends to occur in very different contexts from the plural.

   As Figure 1.3 shows, the singular *study* (overall frequency 1.6 per thousand) is noticeably frequent in Strategies 2D (6.6 per thousand) and 3A (6.7 per

FIGURE 1.3   Key keyword frequencies: *study*

thousand). Treating the singular and plural of STUDY as variant forms of a single lemma would obscure an interesting fact about the primary role each word form plays in the schematic structure of PhD LRs. *Study* is primarily used to manage 'ownership' or 'affiliation' with a particular investigation or research report: its most frequent 2–3 word collocations include *this study* (210), *the present study* (68), PROPER NAME's *study* (43), *their study* (35), *the current study* (33), *his study* (18), *another study* (11), and *her study* (11). Interestingly, while *this study* can be used to reference a study which has been recently introduced in the discourse, it is most often used to refer to the author's own study. Together, *this study*, *the present study*, and *the current study* are the most frequent self-referential uses in the corpus. In contrast, *this analysis* (10) and *this research* (30) are relatively rare collocations, particularly as references to the author's own study: *this analysis* never occurs as a reference to the author's own study in our corpus, and *this research* is primarily used cohesively, in reference to a recently mentioned research project in the text rather than in reference to the author's own work.

This use of *study* to refer to the author's own work and to track its relationship to specific projects carried out by other researchers is an interesting overall fact, but one made more significant in view of where it is most frequent in the schematic structure of the LRs. As shown in Figure 1.3, the singular *study* is most frequent in those strategies where the author is making the most explicitly self-referential moves – Strategy 2D: 'asserting the relevancy of the surveyed claims to one's own research' and Move 3: 'Occupying the research niche'.

5   Discussion

As examples examined in this study show, there is a close relationship between lexical choice and schematic structure in the corpus. The key keywords *research*, *studies* and *study* share in a kind of 'division of labour' when the LR author negotiates a place for the PhD thesis within its wider disciplinary environment, with *research* showing the most flexible set of uses among the three. Throughout the corpus, *research* is often specified with respect to an academic domain or methodology. Once specified, the specific type or method of research under discussion is then defined or evaluated – defined or circumscribed to position the LR with respect to other research in the field and outside it, and evaluated with regard to which research methodologies or approaches are problematic (those used by other researchers) and which are effective (those used in the research of the current thesis). More general but still specific types of *research* act as a background or environment against which specific analytical approaches and specific methods can be judged sufficient or insufficient, and from which specific analytic approaches and methods can be adopted. Prospective authors of thesis LRs need to negotiate a place within a discipline but without adopting as acceptable all the varied approaches to the study of the topic that have been explored. In order to achieve this, control over typical collocations associated with particular schematic structures in the text is an essential tool.

6   Conclusion

This study has shown the complex interactions occurring between the key keyword *research* (along with its related key keywords *studies* and *study*) and its complex range of uses. The study has demonstrated the potential of key keyword analysis in investigating the relation between moves and steps and their linguistic realizations in the chosen genre, the PhD literature review. The study is presented as a contribution to genre theory in showing the potential and value of investigating realization patterns as well as schematic structure in genre analysis. At the same time, and importantly for this volume, the study demonstrates the power of corpus techniques in such analysis.

If in general terms this study demonstrates how corpus techniques can be applied to genre analysis, in the use of a relatively small corpus, it also shows that, while, *other things being equal*, bigger is better in corpus linguistics; small corpora, in the right circumstances, can be as valuable to work with as much larger corpora. The fine-grained collocational and semantic/pragmatic analysis as well as the labour-intensive move and step analysis involved in this study could not realistically have been carried out with a much larger corpus, but the small size of the corpus does not undermine the findings.

In the conclusion to her cited study Kwan (2006: 51) states that '[t]he multi-chaptering, multi-thematic sectioning, and the highly recursive and complex move structures in the LRs in this study reveal that the genre is a sophisticated rhetorical exposition that serves to delineate the complex conceptual and theoretical contours of a thesis'. The findings of the present study reinforce this statement, demonstrating the subtle meanings and collocations of lexis across moves and steps to be equally, if not even more, complex than the elements of schematic structure. Kwan also states in her conclusion that

> [t]he sophistication that is involved in the construction of the LR is probably a result of the complex nature of the writer's research topic, the objects that are studied, the disciplines to which the writer belongs, the need that the writer feels to demonstrate their extensive knowledge of – and critical thinking about – the field, and in some cases the long history of the field.

All of these factors no doubt also feed into the complex choices that are involved in the selection of preferred linguistic realizations for individual moves and steps, including, importantly, keywords such as *research*.

Turning now to the potential pedagogic applications of this study, what was stated above on the subtle contextual influences on key keywords can be conveyed to learners; consciousness-raising for students of the subtleties of lexical choices and their interactions between moves and steps in generic structures is an obvious application of the findings of the study. As has also already been stated, Kwan (2006) concludes that there are notable structural differences between LRs and introductions in research writing. Given these different configurations of moves and steps in the two related genres, it is likely that realization patterns will also vary. This reinforces Kwan's suggestion that writing instructors would be advised to use authentic LRs in their teaching and present them in their entirety. In this way, learners will appreciate the complexity and subtle choices that are involved in the use of lexical items in specific moves and steps of specific genres. This, of course presupposes specific purpose teaching. However, there is no reason why specific purpose examples cannot be used for consciousness-raising purposes in more general pedagogic contexts, for education as opposed to training (see J. Flowerdew, 1993b).

At a less general level, the results of this study and similar ones which might follow it might be incorporated into research writing manuals such as those by Swales and Feak (1994, 2000) and Paltridge and Starfield (2007). The analysis presented in the findings section of this paper provides a useful metalanguage for discussing the subtleties of the use of the particular keyword *research*. Here, canonical realization patterns of the different moves and steps, as presented in this study, might be directly used for the purpose of exemplification. As well as example realization patterns in manuals, in classroom contexts, students may be given key keywords and asked to do their own 'data driven learning'

(Johns, 1991, 1994, 2002; Lee & Swales, 2006). In such work, it must be borne in mind that this study was conducted in one disciplinary area and that there are likely to be field-specific variations. However, students can be guided to perform the sort of analysis conducted here (or presented with such analyses) for their particular field. The present study can thus be considered to be a starting point for further work by both applied linguists and apprentice writers in English for specific or more general purposes.

Of course, there are limitations to this small-scale study. Further work suggests itself in investigating other key keywords besides *research*: other words high on the list of key keywords in the corpus are content words such as *analysis* and *approach*, discipline specific words and morphemes such as *language, non-* (mainly from *non-native speaker*) and *discourse*, grammatical words such as *of,* and textual signalling nouns such as *chapter* and *discussion*. Given the findings of Kwan concerning the differences between introductions and LRs, a study comparing realization patterns in these two genres suggests itself. Similar work, of course, could be conducted with other genres.

Notes

[1] The research reported in this paper was partially funded by City University of Hong Kong, Faculty of Humanities and Social Sciences Exploratory Grant No. 9360101.

[2] Swales (1990: 18), incidentally, identified the problem-solution pattern (Hoey, 1983) as influential in his original formulation of genre theory.

[3] For comparison, Scott and Tribble (2006: 78–79) use a cut-off point of 5 per cent in their discussion of the key keywords of the BNC, and the highest ranked key keyword in their analysis (*you*) is key in only 30 per cent of the texts in the BNC. Our corpus has much higher percentages due to its specialized nature. It is a design feature of the BNC that it should be heterogeneous, while the purpose of the PhD literature review corpus is to represent a particular genre in a single field.

[4] We would like to thank Becky Kwan for her help in clarifying these aspects of the coding scheme.

References

Ahmad, U.K. (1997). Research article introductions in Malay: Rhetoric in an emerging research community. In A. Duszak (Ed.), *Culture and styles of academic discourse* (pp. 273–303). Berlin: Mouton de Gruyter.

Bhatia, V.K. (1993). *Analysing genre: Language use in professional settings.* Harlow: Longman.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* Amsterdam: Benjamins.

Biber, D., Connor, U., & Upton, T.A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure.* Amsterdam: Benjamins.

Bunton, D. (2002). Generic moves in Ph.D. thesis introductions. In J. Flowerdew (Ed.), *Academic discourse* (pp. 57–75). London: Longman.

Burgess, S. (2002). Packed houses and intimate gatherings: Audience and rhetorical structure. In J. Flowerdew (Ed.), *Academic discourse* (pp. 196–215). London: Longman.

Connor, U., Precht, K., & Upton, T. (2002). Business English: Learner data from Belgium, Finland and the U.S. In S. Granger, J. Hung, & S. Petch-Tyson (Eds), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 175–194). Amsterdam: Benjamins.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19,* 61–74.

Flowerdew, J. (1993a). Concordancing as a tool in course design. *System, 21,* 231–244.

—(1993b). A process, or educational, approach to the teaching of professional genres. *ELT Journal, 47,* 305–316.

—(2004). *Corpus of Applied Linguistics PhD literature reviews.* Hong Kong: City University of Hong Kong.

Flowerdew, L. (1998). Corpus linguistic techniques applied to text linguistics. *System, 26,* 541–552.

—(2008). Determining discourse-based moves in professional reports. In A. Ädel & R. Reppen (Eds), *Corpora and discourse: The challenges of different settings* (pp. 117–131). Amsterdam: Benjamins.

Ghadessy, M., Henry, A., & Roseberry, R.L. (Eds). (2001). *Small corpus studies and ELT: Theory and practice.* Amsterdam: Benjamins.

Hasan, R. (1977). Text in the systemic functional model. In W.U. Dressler (Ed.), *Current trends in text linguistics* (pp. 228–246). Berlin: Walter de Gruyter.

—(1979). On the notion of text. In J.S. Petöfi (Ed.), *Text versus sentence: Basic questions in text linguistics* (pp. 369–390). Hamburg: Helmut Buske.

—(1989). The structure of a text. In M.A.K. Halliday & R. Hasan, *Language, context, and text: Aspects of language in a social-semiotic perspective* (pp. 52–69). Oxford: Oxford University Press.

Hoey, M. (1983). *On the surface of discourse.* London: Allen and Unwin.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning materials. In T. Johns & P. King (Eds), *Classroom concordancing* (pp. 1–16). Birmingham: ELR.

—(1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 27–45). New York: Cambridge University Press.

—(2002). Data-driven learning: The perpetual challenge. In B. Kettermann & G. Marko (Eds), *Teaching and learning by doing corpus analysis: Proceedings of the fourth international conference on teaching and language corpora* (pp. 107–117). Amsterdam: Rodopi.

Kwan, B.S.C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes, 25*, 30–55.

Lee, D.Y.W. (2001). Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology, 5*, 37–72.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25,* 56–75.

Lewin, B.A., Fine, J., & Young, L. (2001). *Expository discourse: A genre-based approach to social science research texts.* New York: Continuum.

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

Melander, B. (1998). Culture or genres? Issues in the interpretation of cross-cultural differences in scientific articles. In I. Fortanet, S. Posteguillo, J.C. Palmer, & J.F. Coll (Eds), *Genre studies in English for academic purposes* (pp. 211–226). Castello de la Plana: Universitat Jaume I.

Paltridge, B., & Starfield, S. (2007). *Thesis and dissertation writing in a second language: A handbook for supervisors.* London: Routledge.

Scott, M. (2004). *WordSmith Tools 4.0* [Computer software]. Oxford: Oxford University Press.

—(2007). *Oxford WordSmith Tools 4.0* [Computer software]. Oxford: Oxford University Press. Retrieved 16 July 2008 from www.lexically.net/downloads/version4/WordSmith.pdf.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education.* Amsterdam: Benjamins.

Sinclair, J. McH. & Coulthard, R.M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils.* Oxford: Oxford University Press.

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

Swales, J.M., & Feak, C.B. (1994). *Academic writing for graduate students: A course for nonnative speakers of English.* Ann Arbor, MI: University of Michigan Press.

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

Upton, T.A. (2002). Understanding direct mail letters as a genre. *International Journal of Corpus Linguistics, 7,* 65–85.

Upton, T.A., & Connor, U. (2001). Using computerized corpus analysis to investigate the textlinguistic discourse moves of a genre. *English for Specific Purposes, 20,* 313–329.

Ventola, E. (1987). *The structure of social interaction: A systemic approach to the semiotics of service encounters.* London: Frances Pinter.

Persuading Sponsors and Securing Funding: Rhetorical Patterns in Grant Proposals

Dimitra Koutsantoni

1   Introduction

Discourse analysis research has often used corpus-based approaches to explore lexical, grammatical and text organizational patterns in discourse. Specialized, genre-based corpora are frequently used in genre analysis research, which is concerned with genres' rhetorical patterns and has pedagogical implications. The study of academic genres has focused almost exclusively on published genres, primarily the research article (RA) and the doctoral thesis (e.g. Brett, 1994; Bunton, 2002, 2005; Holmes, 1997; Kwan, 2006; Posteguillo, 1999; Samraj, 2002; Swales, 1990, 2004 among others), neglecting genres not available in the public domain, such as the research grant proposal (GP). However, such neglected genres have central positions in academic discourse communities and their study can have important pedagogical implications for research staff development.

The GP is the basis by which funding applications are judged and grants are allocated. A good track record of obtaining funding is one of the criteria for recruitment of researchers in universities and research funding allocation by sponsors. The recently created European Research Council judges applicants by their 'funding IDs', their track record of securing external research income. In England, research income informs funding allocations by the Higher Education Funding Council of England (HEFCE); it is a metric for the formula allocation of the Higher Education Innovation Fund (HEIF), and is fast becoming one of the two main metrics in the new Research Evaluation Framework (REF-formerly RAE; HEFCE, 2007).[1]

The GP is therefore a 'high-stakes genre crucial to researchers' work' (Tardy, 2003), and the ability to write proposals that can persuade sponsors of the timeliness of a piece of research and secure funding is an essential skill for researchers. Success in winning grants is cumulative; a track

record of grant application success, in combination with a top publication record, can considerably boost chances of winning a grant. In contrast to RAs' peer review process, GP review is not blind. The proposer's reputation, track record and perceived ability to deliver the research programme significantly affect reviewers' feedback, affirming the genre's promotional nature.

We know very little about the structure of this 'occluded' genre (Swales, 1996). Relevant work includes Myer's report on the social construction of two biologists' proposals (1990), Tardy's (2003) analysis of the genre system of the research proposal, and analyses of communicative moves in EU and US GPs by Connor and Mauranen (1999) and Connor (2000) respectively. Addressing the paucity of research on GPs in general and in the UK context in particular, this study aims to offer further insight into the rhetorical organization of this research genre and the ways it is affected by other genres. Focusing on the UK context in particular, the present study is a genre analysis of a corpus of 14 'cases of support' of recent GPs submitted to and funded by two of the UK's major funders, the Economic and Social Science Research Council (ESRC), and Engineering and Physical Sciences Research Council (EPSRC). The ESRC and EPSRC are among the biggest UK research sponsors, making over £100 and £740 million per year available respectively to sponsor research in the disciplines within their remit.[2] The proposals were submitted by researchers from the wider field of Management/ Business Studies covering areas as diverse as Management, Finance, and Actuarial Science and Insurance. The 'case for support' is the section where researchers demonstrate the need for their proposed research. The analysis explores this section's similarities in terms of rhetorical structure with sections of the RA (introductions and methods), and the applicability of Connor and Mauranen's (1999) moves model in this GP corpus. A moves-analytic approach is a useful way of investigating persuasive discourse as a move is a specific part of the text which achieves a particular purpose within the text and contributes to the overall purpose of the genre (Henry & Roseberry, 2001). As grant writing is not an isolated activity but requires familiarity with a number of genre sets (Tardy, 2003), the moves identified are compared with the Research Councils' (RCs) GP guidelines. Given the increasingly globalized research environment, where competition for funding is fierce, the study is highly topical and timely. The universities in the United Kingdom are challenged to demonstrate their research income generating potential in order to secure government funding (REF and HEIF) and achieve financial sustainability. The study's findings therefore have implications for the training of researchers in GP writing while their applicability could be extended to other disciplines and other research funding contexts.

2   The Grant Proposal Genre and the Genre System of Research Funding

The GP is an essentially promotional genre; its objective is to persuade sponsors of the timeliness of a programme of research and secure their funding. Researchers 'sell' their ideas and their expertise to sponsors, make them see the innovation and value in their 'product' and persuade them to 'buy' it. As GPs involve public money, the stakes are high and gatekeeping requirements are very strict, making the chances of winning grant money notoriously slim. Proposals are judged on multiple levels: they must first be approved by research council administrators with regard to their fit with the RC's remit and adherence to its presentation and associated documentation requirements. They are then evaluated by reviewers (academic and practitioner experts) with regard to their scientific value, methodological soundness, innovativeness and potential impact. They are finally graded by assessors (generalists) who base their decisions on referees' comments (ESRC applications), while the final decision lies with peer review panels. GPs therefore need to adhere to a number of discourse communities' expectations.

As Connor (2000) and Tardy (2003) point out, the GP genre does not exist in isolation but is part of a system of interacting genres and discourse communities. Examples of these interacting genres include grant-writing guidelines, sponsors' mission statements, strategic and delivery plans, funding scheme specifications, online application forms, reviewers' assessment forms and reviewers' guidelines. For example, sponsors' mission statements form the basis for choosing a sponsor and assessing whether the proposed research programme falls within its remit, and reviewers' assessment forms and guidelines for reviewers indicate the assessment criteria. These genres do not originate within one discourse community, but are the result of interaction and dialogue among a number of communities. For instance, a sponsor's strategic plan is the result of a dialogue between government departments, businesses associations, professional bodies, professional associations and academics, and responds to their research needs. Assessment criteria address both purely academic and applicability aspects of the research and involve both academic and non-academic beneficiaries.

3   Previous Studies on Grant Proposals

Work on the rhetorical structure of GPs is extremely limited and rather dated with no studies specific to the UK context. Extant work includes studies

by Connor and Mauranen (1999) and Connor (2000). Working within the moves-analysis approach, these studies explored the rhetorical structure of GPs and the ways this is defined by sponsors' expectations and other interacting genres. Connor and Mauranen analysed 34 proposals from Finnish universities and research institutes, written between 1992–94 for four different EU programmes funding scientific/technological research (industrial and material technologies, and environmental issues). They identified ten moves, some of which bear similarities to RA introductions, while others are specific to the GP genre. On the whole, the moves follow EU guidelines on proposal writing for the programmes in question. They were: *Territory* (the context of the research); *Gap* (indication of a knowledge gap in the territory); *Goal* (statement of aims and objectives); *Means* (specification of how the goal will be achieved); *Reporting previous research*; *Achievements* (anticipated results, findings and outcomes); *Benefits* (usefulness of anticipated outcomes); *Competency Claims* (researchers' track records and research environment); *Importance Claim* (importance or topicality of the project); and *Compliance Claim* (relevance of the project to the sponsor's objectives), which was found to be specific to EU programmes.

Connor applied the above model in her 2000 study of 14 GPs written by five humanities and science researchers for US government funders (National Science Foundation, National Institutes of Health, National Endowment for the Humanities) and one private sponsor. GPs were in the disciplines of biology, chemistry, English, geology and history. While she was able to identify most of the moves, Connor found that the compliance move was absent as this was not required by the funding bodies in question. Instead those sponsors required an *institutional commitment* claim. Connor additionally found that the competency claim could be either a claim on the researcher's personal competency or the institution's capability and expertise in the area of the proposed research. Finally, she observed that not all moves were obligatory but often occurred in cycles, and there was variation in their functions: for instance, a self-citation could be interpreted as a competency claim. Connor supported her analysis with interviews with the applicants, who confirmed her moves model and talked about the ways the moves used responded to granting agencies' guidelines and comments from peers.

The motivation for the present study is the paucity of research on the rhetorical organization of GPs, the virtual lack of such studies in the UK context and the discipline of business/management, and the changes in the global funding environment which limit the applicability of the existing research. This study focuses on the current UK context and two of its major Governmental funders which have broadly similar expectations and criteria. The UK context is characterized by a focus on innovation, enterprise and applicability of research to business and the community, guided by the Knowledge and Technology Transfer agenda. Governmental funders expect to see research programmes

that engage with users and have the potential to lead to technology transfer and commercial exploitation. The research councils' detailed guidelines, strategic plans and associated policy initiatives are available on their websites, and applicants are expected to consult and in address them when preparing their applications.

The focus of this study is the work of researchers within the discipline of business/management in UK universities, whose persuasive style may differ from those in Finnish and US universities and in different disciplines. Cross-disciplinary and cross-cultural variation in academic writing has been widely reported, with rhetorical and persuasive strategies varying depending on disciplinary methods of enquiry and norms of social interaction and culture-specific intellectual styles (see Koutsantoni, 2007 for an overview).

4   Data

The main data for the analysis comprises cases for support and the RCs' GP writing guidelines. The study additionally draws on genres such as online application forms, RC mission statements and delivery/strategic plans.

4.1   Cases for support

Applications to UK RCs are submitted electronically, with the 'case for support' submitted as an attachment.[3] The analysis includes the cases for support from all successful funding applications submitted by researchers at one UK business school, between 2002 and 2006 (3 EPSRC and 11 ESRC, totalling 14). The proposals were held on file in the School's research development office, and as they are not in the public domain, the authors' consent to analyse and refer to their proposals was sought prior to undertaking the study. All the authors gave their consent but some requested anonymity and removal of sensitive data.[4] Following their wishes, the proposals have been anonymized and encoded as GP1, GP2 etc.

4.2   Research council grant proposal writing guidelines

The ESRC expects: (1) an introduction; (2) proposed methodology; (3) expected outcomes; and (4) expected impact.[5] The EPSRC's expectations include: (1) background; (2) methodology; (3) relevance to beneficiaries; and (4) dissemination and exploitation.[6] Their detailed expectations in each section are shown in Table 2.1.

Table 2.1 ESRC and EPSRC grant proposal guidelines

ESRC	EPSRC
Introduction ■ Aims and objectives ■ Main work on which the research will draw ■ Research questions	Background ■ Topic of research ■ Past and current work in the subject area in the UK and abroad
Methodology ■ Description of methods and innovation ■ Data description and data collection procedures with reasons for choice ■ Methods for analysis and reasons for their choice	Methodology ■ Aims and objectives ■ Methodology and justification for choice ■ Specific expertise of researchers ■ Timeliness and novelty ■ Description of programme of work, time lines and project management
Outcomes (articles, papers, datasets, events) ■ Academic ■ User orientated	Dissemination and exploitation ■ Dissemination and technology transfer routes ■ Exploitation of results
Impact	Relevance to beneficiaries ■ Potential impact ■ Beneficiaries ■ Collaborations with beneficiaries

5 Method

The analysis follows the moves-analytic approach, with a move defined as a functional unit used for some identifiable rhetorical purpose (Connor & Mauranen, 1999) and explores the applicability of Connor and Mauranen's (1999) and Connor's (2000) moves models for GPs. These researchers have shown that some moves in GPs correspond to the structure of RA introductions as described by Swales (1990). The detail RCs expect on the methodology of the proposed research suggests that this section might resemble the Method section of RAs. The study thus also investigates the similarity of moves with moves in introductions and methods sections of RAs.

For introductions, the study uses Swales's (2004) revised CARS model, which consists of: Move (1), *establishing the territory*, realized via topic generalizations of increased specificity; Move (2), *establishing a niche*, realized via references to literature, indicating gaps/adding to what is known and presenting positive justification and Move (3), *presenting the present work*. This move is realized via such steps as announcing present research; presenting research questions or hypotheses; definitional clarifications; summarizing methods; announcing principal outcomes; and stating the value of present research.

For methods, this study uses Lim's (2006) model for management RA methods sections, which includes three main moves: (1) *describing data collection*

procedures; (2) *delineating procedures for measuring variables*; and (3) *elucidating data analysis procedures*. Move (1) is realized via describing the sample; describing the sample technique or criterion; justifying the data collection procedure. Move (2) is realized via presenting an overview of the design; explaining methods for measuring variables; justifying the methods of measuring variables (by highlighting the advantages of using the sample and by showing its representativity). Finally, move (3) includes the following steps: relating data analysis procedures; justifying them (by citing previous research methods and by highlighting acceptability of these methods); and previewing results.

6   Results

The analysis shows that proposers divide their proposals into sections that correspond to RCs' guidelines for GPs, as shown in Table 2.2. Deviation is observed in some EPSRC cases for support which include an Introduction section separate from the Background section. In addition, the majority of cases for support have separate sections with the aims and objectives of the study when these are not required as a separate section by either RC (93%). The background and methodology often extend over more than one section, based on the various sub-topics studied and separating data/data analysis/methods. The majority of proposals (71%) have a dedicated section for dissemination and outputs, while over half of the proposals (57%) have a section dedicated to relevance to beneficiaries/user engagement.[7] EPSRC applicants are on the whole more faithful to the headings suggested by the RC, while ESRC applicants are more atypical in their use of headings. The division into sections facilitates the identification of moves, as each section corresponds to a move.

The analysis confirms the partial applicability of Swales's (2004), Lim's (2006) and Connor and Mauranen's (1999) models of moves/steps, as seen in Table 2.3.

Table 2.2   Sub-section headings in the cases for support

Section	No. of occurrences in ESRC GPs	No. of occurrences in EPSRC GPs	Total number of occurrences	% GPs with this feature
Introduction	6	2	8	57
Aims and objectives	10	3	13	93
Background	10	3	13	93
Methodology	10	3	13	93
Relevance to beneficiaries/user engagement	5	3	8	57
Dissemination/outputs	7	3	10	71
Total number of cases for support	14			

Table 2.3 Moves and steps in the cases for support

Moves and steps	Occurrences in ESRC GPs	Occurrences in EPSRC GPs	Total occurrences	% GPs with this feature
Introduction (establishing the territory)	5	3	8	57
Topic generalizations	5	3	8	57
Focus/object/methods of research	6	1	7	50
Review of studies and gap/ limitation	3	1	4	29
Summary of method	3	1	4	29
Summary of data	3	0	3	21
Positive justification of research	4	0	4	29
Background (establishing a niche)	10	3	13	93
Topic generalizations	10	3	13	93
Reviews of previous studies and gap/limitation/shortcoming	10	3	13	93
Reviews of previous studies and extension of previous work	4	1	5	36
Reviews of previous studies and statement of problem	2	2	4	29
Statement of need/justification	4	1	5	36
Aims and objectives (presenting current research/goal)	11	3	14	100
Description of research	2	0	2	14
Aim of research	4	3	7	50
Research questions or hypotheses or objectives	8	2	10	71
Stating value of research	3	2	5	36
Methodology (means)	11	3	14	100
Data collection procedures				
Describing the sample	6	1	7	50
Recounting steps in data collection	0	0	0	0
Justifying data collection procedures by:				
highlighting advantages of using the sample in comparison to other samples AND/OR	5	0	5	36
showing representativity of the sample	7	1	8	57
Procedures for measuring variables				
Presenting overview of the design	7	3	10	71
Explaining methods of measuring variables	7	2	9	64

Table 2.3 Continued

Moves and steps	Occurrences in ESRC GPs	Occurrences in EPSRC GPs	Total occurrences	% GPs with this feature
Justifying methods by:				
highlighting acceptability of the methods with ref to lit/ self citation AND/OR	9	2	11	79
highlighting originality of the methods (as compared to existing ones)	4	2	6	43
Elucidating data analysis procedures				
Relating data analysis procedures	4	1	5	36
Justifying data analysis procedures	1	1	2	14
Previewing results and their significance	7	2	9	64
Outputs and Dissemination/ exploitation (achievement)	7	3	10	71
Journals	6	3	9	64
Practitioner orientated publications	5	2	7	50
Conferences	4	1	5	36
Workshops	3	1	4	29
Press releases/media publicity	3	1	4	29
Tools (business models, skills development, guidelines)	0	1	1	7
Presentations to commercial partners	0	1	1	7
Relevance to Beneficiaries/user engagement (benefits/ importance claim)	8	3	12	86
User groups and impact	4	3	8	57
Policy implications	5	0	5	36
Compliance with RC mission	1	0	1	7
Competency claims	4	2	6	43
Collaborations	3	2	5	36
Links with previous projects (by the PIs)	2	0	2	14
Researchers' expertise	1	0	1	7
Links with other projects	1	0	1	7
Research environment	1	0	1	7
Time plans	2	1	3	21
Project management	0	2	2	14
Ethical statement	1	0	1	7
Total number of cases for support	14			

7   Discussion

7.1   Introductions in the cases for support

Fifty-seven per cent of the cases for support have an introduction in which researchers place their proposed research in the current policy/regulatory context and demonstrate their awareness of any issues or problems that need to be addressed. The majority of introductions are limited to an outline of the territory of the research and its focus, resembling the move of *establishing the territory* in RA introductions. Fewer preview studies, indicate gaps/problems, preview method and data, and justify the research, combining moves (1) and (2) of Swales's CARS model (see Example 1).

Example 1

Commercialization of academic inventions is currently central to the innovation policy agenda of most developed nations as it contributes to the global competitiveness (. . .)	Topic generalization
However, the Lambert review of business-university collaboration (Lambert 2003) suggested that spinning out is overemphasized in the UK while licensing might be a better way forward for UK universities	Problem
This research proposes to explore the reasons universities opt for spin outs instead of licensing. According to real options theory, technological and market uncertainty may affect the way academic inventions will be commercialized. Using this theory, the proposed research aims to explain the (. . .)	Aim
Empirically, the best way to control for institutional differences in technology transfer policy (and focus on uncertainty) is to research a sample of inventions from a single institution (. . .) We have gained access to a proprietary dataset of (. . .). This is a unique dataset as it provides a large and longitudinal sample for our study.	Data and method and justification for choice
The results will be generalizable to other developed countries such as the UK (GP14)	Significance

7.2   Background section in the cases for support

Ninety-three per cent of the cases for support include a background section introducing the topics the proposed research addresses, reviewing previous studies and indicating gaps, limitations and shortcomings, thus carving a niche for the proposed research. This is the section where researchers demonstrate the topicality of their research, place the research in context and demonstrate their knowledge of issues relevant to academic and practitioner communities. These sections are in effect literature reviews following a cyclical pattern of moves and steps similar to RA introductions, with each cycle dealing with a different sub-topic within the overarching topic of the proposal. In EPSRC cases for support, this section functions as an introduction too and as Kwan (2006) has shown for PhD theses, these two sections can be very similar in terms of moves. The moves in this section correspond to Connor and Mauranen's (1999) *reporting previous research* and *gap* moves.

Example 2

High involvement, lean production and family-friendly practices are viewed by academics and policy makers as critical ways of improving both the individual's performance and well-being and the organization's efficiency. Moreover they are often presented as related sets of practices.	Topic generalization
Yet much of the research has concentrated on one or other type, and on the effects on organizational performance to the exclusion of other outcomes, often with a limited methodology.	Limitations of previous research
There is need for better studies of organizational performance and more research on (a) the nature of these practices – the extent to which they tend to be used together and, if so, whether this reflects an underlying approach to management; and (b) their effects on employees, and in particular their well-being, job satisfaction, and commitment. (GP7)	Statement of need for research

7.3   Aims and objectives section in the cases for support

All of the cases for support include this move, with 93 per cent of them having it as a distinct section. Here, researchers aim to show the innovation in their approach and the significance of their research: the *goal* of the research (Connor & Mauranen, 1999). Steps here are similar to the *presenting present*

work move of Swales's CARS model and include: announcing the focus of the research and presenting research questions, aims, objectives or hypotheses.

Example 3

The project has a dual focus – (a) to develop (. . .), and (b) to explore the (. . .). It is concerned with the 'valuation' of innovation processes in two senses (. . .)	Focus
It has four objectives focused on innovation (. . .) (GP3)	Objectives

7.4   Methodology section in the cases for support

All of the cases for support detail the methodology to be used, as this is a key RC requirement. Methodology sections include three moves: (1) data descriptions; (2) methods for measuring variables; and (3) data analysis procedures, similar to the management RAs Lim (2006) analysed. These sections can also follow a cyclical pattern, as researchers endeavour to justify their methods, and show the innovation in their method and the representativity of their datasets. While this sub-section broadly corresponds to Connor and Mauranen's (1999) *means* move, it is clearly not a single-step move, and is much more complex and detailed than they describe.

7.4.1   Data description

The description of data move includes a step in which researchers describe the sample (50% of cases of support included this), and a step in which researchers justify the dataset used by either highlighting the advantages of using this particular dataset in comparison to others (35%) or by showing its representativity (57%). Lim's (2006) *recounting steps in data collection* step was not encountered at all, presumably as researchers would be able to recount the steps in data collection after completion of the research and not beforehand.

Example 4

The distinctive features of the data on which this study will be based are as follows:	Highlighting advantages of the data sample
(1) A series on the official rate is used. Unlike Hofmann and Mizen (2003, 2004), the actual rate will be used, rather than the average of base rates for the 'big four banks'. (. . .)	
(2) The data will cover a wide range of retail products, including (. . .). Including all the key retail products will provide a more complete picture of the pass through to retail rates. (GP1)	Highlighting representativity of the sample

7.4.2 Procedures for measuring variables

This move includes steps such as presenting an overview of the design (71%), explaining methods of measuring variables (64%), and justifying methods by either highlighting acceptability of the methods with references to the literature or self-citation (79%) or highlighting originality of the methods (43%).

Example 5

We will use a longitudinal research design since we are concerned with data collected over time.	Presenting overview of the design
Longitudinal designs are unique, innovative and in great demand as they capture the relationships over a period of time and not in a 'snapshot' moment as normal cross-sectional surveys.	Justifying method
Dependent Variable: Firm Formation In each year a patent could be licensed or not and the license could be issued to an established firm, or a new firm. Similar to Shane (2001b) we define firm formation (spin out) as occurring in a given year, if the invention was licensed to a for-profit firm that did not exist as a legal entity in the previous year. To account for the timing of commercialization we will count the number of days from the invention disclosure date to invention commercialization date. (. . .) (GP14)	Explaining methods for measuring variables and justifying them by reference to the literature

7.4.3 Elucidating data analysis procedures

This move comprises steps such as relating data analysis procedures (36%), justifying data analysis procedures (14%) and previewing results and their significance (64%). This move is less frequent than the previous two, as at this stage researchers cannot be expected to offer detailed accounts of data analysis procedures nor preview results, as is the case with RA authors.

Example 6

We propose refining the approach of Lewis et al. (2002) by specifying a four- or five-state semi-Markov model (compared to their three-state model which excludes duration dependence).	Relating data analysis procedures
Further, the transition probabilities will be estimated from linked panel data of labour market activity. Hence, we will be able to extract as much information as possible while considering possible measurement bias. (GP6)	Justifying data analysis procedures

7.5   Outputs/dissemination and exploitation sections in the cases for support

This move occurs in 71 per cent of the cases for support, always as a distinct sub-section, and corresponds to Connor and Mauranen's *achievement* move. In this move, researchers outline their dissemination plan, including journal publications, conference presentations, media publicity events and workshops. Researchers are expected to demonstrate that their research is going to reach a wide range of audiences and user groups, including policy makers. The ESRC, in particular, ascribes a great deal of importance to what it terms 'communication and user engagement' (reflected in the dedicated section on this aspect of a research programme on its online application form)[8] and points out the importance of project marketing and dissemination of research findings through the web, media and events, as well as publication in journals.

The analysis, however, showed that researchers still primarily propose to disseminate their research via academic papers (64%), followed by publications in practitioner journals and periodicals (50%). Conference presentations are the third most popular option (36%), while workshops/events and press releases/ media publicity were mentioned by few researchers.

Example 7

The results of the research will be published in peer reviewed academic journals (e.g. . . .).	Academic papers
The work will be presented at suitable conferences (e.g. the X Conference).	Conference presentations
The results will also be publicized in newspapers and specialist magazines, thus reaching a wider public audience. The Press Office at X Business School has a very good track record in publicizing research. For example, the results of X were reported in five national newspapers and several trade magazines.	Media
We will also organize workshops for interested parties (e.g. health authorities and relevant Government departments). (GP9)	Workshops

7.6   Relevance to beneficiaries/user engagement sections in the cases for support

This move occurs in 86 per cent of the cases for support (57% as a separate section) and corresponds to Connor and Mauranen's *benefits* and *importance claim* moves. While developing this move, researchers need to show the ability to think beyond the immediate technical part of their proposal to assess

the potential economic and social impact of their research, identify potential users and beneficiaries of the research and ways to engage with them. Fifty-seven per cent of the cases of support discuss the relevance of research for academic and non-academic user groups and the wider impact of the research in science, economy and society. Less commonly, this section discusses the policy and regulatory implications of the research (36% of cases of support) or its compliance with the research council mission and its remit (7%).

Example 8

The resultant new methodologies for scheduling machinery, people and orders in relation to batching of jobs and the management of orders in the supply chain, should benefit the academic community, in Operational Research and Operations Management. The company, X, expects to gain insight into how their microbiology laboratory can be run, especially during surges in demand. Specific algorithmic scheduling solutions are, however, considered to be a bonus for (. . .).	User groups
This project could ultimately benefit all automated microbiology laboratories, and has potential impact on the food manufacturing (and hence retailers and consumers) since they are dependent upon the speed of results from microbiology testing for release times of food produced. (GP11)	Impact

The focus of this section is defined by the remit of the RC and its strategic plan, namely, technology transfer and commercialization of research,[9] engagement with stakeholders and the public and the translation of research into some social and economic benefit, and influencing of policy and regulation.

7.7   Competency claims in the cases for support

Competency claims were identified in 43 per cent of the cases for support. They included references to the expertise of the research team (7%),[10] the supportiveness of the research environment (7%), links with other projects (7%), links with previous projects undertaken by the principal investigator (14%) and commercial collaborators (36%).

Example 9

We will also build on previous projects such as X's project and ESRC studentship of X as well as X's ESRC biotechnology projects. (GP4)	Link to other projects

Example 10

We have assembled a consortium of industrial   Commercial
parents for this research project, covering   collaborations
different perspectives in the market and bringing
a wide range of experience and expertise to the
project (. . .) (GP10)

7.8   Timeplans, project management and ethical statement sections in the cases for support

Researchers can indicate their timeplan on the online application form (for ESRC) or attach it as a Gantt chart, and in my sample only three cases for support included one (two of which were by the same principal investigator). Only two cases for support discussed project management (both EPSRC projects, as this is a requirement of EPSRC). Finally, only one case for support included an ethical statement, as again this is covered in the ESRC's online application form (EPSRC's online application form does not include this section).

8   Summary and Pedagogical Implications

This chapter has presented a genre analysis of a corpus of cases for support from 14 successful funding applications, submitted to two of the UK's major research councils. Informed by a social constructionist perspective, the study argues that GP writing does not take place in isolation but is part of a cluster of genres, which are in dialogic relationship with the discourse communities that use them. It explores ways the GP is influenced by this network of interacting genres that form the genre system of research funding (Tardy, 2003), comprising RCs' GP guidelines, delivery and strategic plans, and online application forms.

The analysis confirms the broad applicability of Connor and Mauranen's (1999) and Connor's (2000) moves models for GPs, as all of the moves they identified were also found in this study's dataset. However, some moves are more complex than they describe and comprise steps resembling the ones employed in sections of RAs. The *territory*, *reporting previous research* and *gap* moves are not separate from one another but co-occur in the overall background move, as Swales (2004) suggests for RA introductions. The *goal* move is similar to the *presenting current work* move of RA introductions. The *means* move comprises several sub-moves and steps, as Lim (2006) found in RA methods sections. There is a clear similarity in moves in certain sections of RAs and GPs, arguably due to the fact that both genres go through a process of review and negotiation of their claims, and as RAs are often written accounts of externally funded research.

The differences noted between past models and the present model can be attributed to disciplinary differences in developing a convincing argument as well as to specific intellectual and persuasive styles of national cultures. Cross-disciplinary variation in the use of moves has been reported by Posteguillo (1999), Samraj (2002) and Anthony (1999) for RA introductions, Lim (2006) and Kanoksilapatham (2005) for RA methods, Holmes (1997) and Brett (1994) for RA results and discussion sections. It has been attributed to differences in the audience and goals of disciplinary communities, methods of investigation and agreed methods of presenting results. Cross-cultural studies of academic argumentation have identified discernible cultural patterns of writing, of organizing discourse, and of interacting with members of the community. Differences have been attributed to culture-specific intellectual styles, to size and structure of academic communities, and to educational systems and literacy practices (see Koutsantoni, 2007). Cross-cultural variation in the employment of Swales's CARS model in RAs, for instance, has been reported by Ahmad (1997), Shaw (2003), Martin (2003) and Martín-Martín and Burgess (2004) in Malay, Spanish and Danish contexts. A detailed investigation of such disciplinary and cultural variables is, however, beyond the scope of this study.

The analysis indicates that the rhetorical pattern of cases for support is greatly influenced by a number of other interacting genres and responds to the expectations of a number of discourse communities. Their structure is generally defined by RCs' guidelines which correspond to review assessment criteria. The authors of the analysed GPs demonstrated familiarity with a number of different genre sets when putting forward their proposals for a programme of research. Their success in securing funding shows that the GPs met sponsors' specifications as these were outlined in their guidelines, and that they were in line with the RCs' remits and funding priorities as these were specified in their strategic plans. The RCs' online application forms showed authors which aspects of their proposal are seen as important by the sponsors. Their previous experience as referees and their knowledge of reviewers' guidelines and assessment forms, together with their nomination of reviewers and informed guess of other potential reviewers helped them meet reviewers' expectations. By researching previous work and identifying gaps and shortcomings, they showed the innovation in their approach, in their method and datasets and demonstrated the need for their proposed research programme. Proposers also demonstrated knowledge of the political, social and economic context, ability to assess the potential impact of their research, identify potential users and beneficiaries of the research and ways to engage with them.

The findings of this study have pedagogical implications for research staff development and training in GP writing and in the processes of funding and evaluation of research, which are arguably key skills that researchers need to develop. In the United Kingdom, major investment is made in research staff development following the Roberts Report (2002), which recommended that

researchers at doctoral and post-doctoral level should be given the opportunity to develop a range of research and transferable skills in order to improve the quality of research, increase their prospects of employment and enhance the reputation of universities. Dedicated funds (the 'Roberts money') are made available for developing research staff development opportunities. Similar policy initiatives exist at European level: the EU Charter for Researchers (2005) is a representative example. It defines roles and responsibilities of researchers and their employers, and aims at ensuring that the relationship between these parties contributes to successful performance, generation of knowledge and career development of researchers.

This chapter suggests that for research staff development courses to be most effective, they need to be socioculturally oriented, aiming to raise researchers' awareness of the social construction of the GP genre and its interactivity with other genre sets. Researcher development courses should provide researchers with an understanding of the values and social practices of the discourse communities in which these genre sets are produced, the dialogic relationship between communities and genres, and the social forces that contribute to their formation. Courses should help researchers see that GP writing is not an isolated activity that merely involves writing a document, but requires a thorough understanding of the wider socio-economic context, and the ways this defines research priorities and public money allocations. Teaching materials should involve researchers in actual genre analysis of GPs in terms of structure and examination of these against funders' guidelines and reviewing criteria. Researchers should have access to examples of both successful and unsuccessful proposals for comparison, as well as to reviewers' comments on both. Such a genre-based approach for the teaching of EAP in general is proposed in Koutsantoni (2007).

The findings of this study can inform the development of a moves model for GPs targeted at UK RCs, while the applicability of the model could be tested in other disciplines and research funding contexts. Research is also necessary into the linguistic realization of GP authors' evaluations, claims and judgements, and the ways these are managed interpersonally within the moves identified. This combined with analysis of reviewers' commentary on both successful and unsuccessful proposals could enrich our knowledge of the rhetoric that persuades sponsors of the timeliness and merit of proposed research and of the social factors that make this rhetoric persuasive.

Acknowledgements

Menezes and Dr Ben Rickyazen. I would also like to extend my thanks to their co-investigators.

Notes

[1] The other one is number of citations.

[2] www.esrcsocietytoday.ac.uk/ESRCInfoCentre/about/

[3] The ESRC requires cases for support to be up to 6–12 pages long depending on the amount requested. The EPSRC requires that cases for support include: the research team's track record (2 pages) and the description of the research and its context (6 pages). The analysis focuses on the second part of the EPSRC case for support.

[4] Those authors were happy to be acknowledged by name but did not wish to be cited by name in the body of paper.

[5] www.esrc.ac.uk/ESRCInfoCentre/Support/research_award_holders/FAQs2/index2.aspx?ComponentId=5079&SourcePageId=5441

[6] www.epsrc.ac.uk/ResearchFunding/HowToApply/ProposalDocumentation.htm

[7] Some of the proposals deviate from this pattern combining sections under one (e.g. one comprises project aims and a project summary including background, data, methods, etc; another has a section titled: research design, methodology, user engagement, influence and dissemination).

[8] ESRC has developed a toolkit to help applicants define their communication/dissemination strategy and focus their thinking of who the end users of the research are and what constitute the best ways to reach them. www.esrc.ac.uk/ESRCInfoCentre/CTK/default.aspx?ComponentId=25076&SourcePageId=19165

[9] www.esrcsocietytoday.ac.uk/ESRCInfoCentre/about/delivery_plan/priorities_and_funding/index.aspx?ComponentId=9508&SourcePageId=13007

[10] The limited occurrence of statements of researchers' expertise is not surprising in EPSRC applications, as a separate two pages on researchers' track record and the attachment of the principal investigator's CV are required. ESRC applications do not require a track record but CVs of the principal investigator and all named researchers as an attachment.

References

Ahmad, U.K. (1997). Research article introductions in Malay: Rhetoric in an emerging research community. In A. Duszak (Ed.), *Culture and styles of academic discourse* (pp. 273–303). Berlin: Mouton de Gruyter.

Anthony, L. (1999). Writing research article introductions in software engineering. *IEEE Transactions on Professional Communication, 42*, 38–45.

Brett, P. (1994). A genre analysis of the results sections of sociology articles. *English for Specific Purposes, 13*, 47–59.

Bunton, D. (2002). Generic moves in PhD thesis introductions. In J. Flowerdew (Ed.), *Academic discourse* (pp. 57–75). London: Longman.

—(2005). The structure of PhD conclusion chapters. *Journal of English for Academic Purposes, 4*, 207–224.

Connor, U. (2000). Variation in rhetorical moves in grant proposals of US humanists and scientists. *Text, 20,* 1–28.

Connor, U., & Mauranen, A. (1999). Linguistic analyses of grant proposals: European Union research grants. *English for Specific Purposes, 18*, 47–62.

EPSRC Strategic Plan (2006). Retrieved 28 March 2008, from: www.epsrc.ac.uk/CMSWeb/Downloads/Publications/Corporate/EPSRCSP06.pdf

ESRC Communications Tool. Retrieved 28 March 2008, from: www.esrc.ac.uk/ESRCInfoCentre/CTK/default.aspx?ComponentId=25076&SourcePageId=19165

ESRC Strategic Plan (2005–2010). Retrieved 28 March 2008, from: www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/Strategic_Plan_2005-10_tcm6-12995.pdf

European Commission (2005). *European charter for researchers.* Retrieved 28 March 2007, from: www.grad.ac.uk/downloads/documents/Reports/European%20reports/European%20Charter%20and%20code%20of%20conduct%202005.pdf

HEFCE (2007). *Research excellence framework: Consultation on the assessment and funding of higher education research post-2008.* Retrieved 28 March 2008, from: www.hefce.ac.uk/pubs/hefce/2007/07_34/

Henry, A., & Roseberry, R. (2001). A narrow-angled corpus analysis of moves and strategies of the genre 'letter of application'. *English for Specific Purposes, 20*, 153–167.

Holmes, R. (1997). Genre analysis and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes, 16*, 321–337.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes, 24*, 269–292.

Koutsantoni, D. (2007). *Developing academic literacies: Understanding disciplinary communities' culture and rhetoric.* Oxford: Peter Lang.

Kwan, B.S.C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes, 25*, 30–55.

Lim, J. (2006). Methods sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes, 25*, 282–309.

Martin, P.M. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes, 22*, 25–43.

Martín-Martín, P., & Burgess, S. (2004). The rhetorical management of academic criticism in research article abstracts. *Text, 24*, 171–195.

Myers, G. (1990). *Writing biology: Texts in the social construction of scientific knowledge.* Madison: University of Wisconsin Press.

Posteguillo, S. (1999). The schematic structure of computer science research articles. *English for Specific Purposes, 18*, 139–160.

Roberts, G. (2002). *SET for success.* HM Treasury. Retrieved 28 March 2008, from: www.hm-treasury.gov.uk/documents/enterprise_and_productivity/research_and_enterprise/ent_res_roberts.cfm

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes, 21*, 1–17.

Shaw, P. (2003). Evaluation and promotion across languages. *Journal of English for Academic Purposes, 2*, 343–357.

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

—(1996). Occluded genres in the academy: The case of the submission letter. In E. Ventola & A. Mauranen (Eds), *Academic writing: Intercultural and textual issues* (pp. 45–58). Amsterdam: Benjamins.

—(2004). *Research genres.* Cambridge: Cambridge University Press.

Tardy, C. (2003). A genre system view of the funding of academic research. *Written Communication, 20*, 7–36.

Chapter 3

Verbal and Mental Processes in Academic Disciplines

Jasper Holmes and Hilary Nesi

1   Introduction

Interview surveys have shown that university lecturers from different disciplines look for different attributes in the writing of their students, and describe and evaluate academic activities in discipline-specific ways (Lea & Street, 2000; Nesi & Gardner, 2006). This chapter aims to identify some important disciplinary differences at clause level in student assignments which have been awarded good grades, and have therefore met lecturers' expectations, at least to some extent. Using keyword analysis, which is a corpus linguistic technique, followed by closer analysis of clause meanings in context, it considers the ways in which student writers position themselves as members of their discourse community, as purveyors of hard or soft and pure or applied knowledge. The identification of discipline-specific clausal features helps to further our understanding of the way disciplinary knowledge is conceptualized and expressed, and may also inform the design of discipline-specific writing programmes for novice academic writers.

2   Process Types in Academic Writing

According to the Systemic Functional Linguistics (SFL) model (Halliday, 1994), the world of experience has three forms of representation, each realized by two Process types. 'Outer' experience is represented as actions or events, and is realized in Material and Behavioural Processes. 'Inner' experience is represented as reaction and reflection, and is realized in Mental and Verbal Processes. The third form of representation, 'generalization', is the relationship between these experiences, and is realized in Relational and Existential Processes. Mental and Verbal Processes usually involve at least one animate participant, a 'Senser' or a 'Sayer' and human agency in these types of Processes is usually recoverable from the text, even if it is disguised by the use of metaphor, as in 'This dissertation considers an alternative view' (John, this volume), where the Mental Process is attributable to the writer. In Material Processes, on the other hand, the 'Actor'

may be animate, inanimate or abstract and agency may be hidden by means of a passive construction, while in Relational Processes human activity can be entirely disguised through nominalization (Halliday, 1994: 352–353), which metaphorically transforms congruently worded clauses such as 'he argued' or 'he composed a piece of music' into nominal groups such as 'the argument' or 'the musical composition', which are then related to new entities in clauses such as 'the argument is valid' or 'the composition was in binary form'.

Transitivity mechanisms allow writers to adjust their presence or distance in the text according to their communicative goals, as noted by John (this volume) who discusses the effect of Process on writer visibility, and Tang (this volume) who examines writers' means of self-expression. Because different transitivity choices achieve different communicative effects, we expect the distribution of Processes to vary according to domain, genre and context. John (this volume), for example, found Material and Relational Processes to be the most frequent in the methodology sections of MA dissertations in the field of Applied Linguistics, while Martínez (2001) found that Material Processes dominated in the Method sections of scientific research articles whereas Relational Processes dominated in the Results and Discussion sections. Love (1993) noted the high frequency of Relational and Existential Processes in geology textbooks, while Babaii and Ansary (2005) reported more Relational and Existential Process types in physics book reviews than in sociology and literature, and more Material Process types in sociology and literature book reviews than in physics.

This study concentrates on students' lexical choices relating to Verbal and Mental Processes. As these are processes which involve human agency it was assumed that they could shed some light on the students' sense of scholarly identity. Martínez (2001: 241) found that the Verbal and Mental Processes in research article introductions helped writers achieve their goal of 'contextualizing . . . previous research, reviewing theories, ideas and previous findings'. In John's examples of MA dissertations (this volume) Mental Processes were often found to involve subjective interpretation and result in greater visibility for the writer, whereas the Sayer in Verbal Processes was usually a cited authority behind which the student writer could hide. It is reasonable to assume that Verbal and Mental Processes will play similar roles in the types of university assignment we will examine, given that students often have to introduce and review others' theories and findings, even if, prior to the dissertation, their ultimate objective is often to demonstrate their acquisition of academic knowledge, rather than to occupy a research niche in the manner described by Swales (1990).

3   Classification of Academic Knowledge

A standard typology for the classification of academic knowledge distinguishes between 'pure' or 'applied' and 'hard' or 'soft' disciplines (Becher & Trowler,

2001). In general, the natural sciences and mathematics are classed as hard-pure, the science-based professions such as engineering are classed as hard-applied, the humanities are classed as soft-pure and the social professions such as education and law are classed as soft-applied.

Knowledge in hard-pure disciplines is quantitative and tends to develop steadily and cumulatively; new findings derive linearly from an existing body of knowledge. Soft-pure knowledge, on the other hand, is qualitative and new developments in these disciplines tend to derive from the combination and recombination of existing work and results (Becher, 1989: 13; Becher & Trowler, 2001: 39). This accords with Hyland's (2000: 37–40) claim that writers in soft disciplines use more, and more varied, reporting verbs than writers in hard disciplines, because they need to support their arguments with references to other researchers whose works are known and respected. In the hard disciplines causal and logical relationships are relatively easily established from observations and quantitative data, and so there is less need to cite the opinions of others. Hyland (2000: 28) also finds that writers in the soft disciplines use more 'discourse act' reporting verbs such as ASCRIBE, DISCUSS and STATE, whereas writers in the hard disciplines prefer to use 'research act' reporting verbs such as OBSERVE, DISCOVER and CALCULATE.

The distinction between pure and applied disciplines depends on the extent to which the discipline is concerned with theory, or practice. Applied knowledge builds on theory, but is ultimately practical; it is concerned with 'knowing how' as opposed to 'knowing that' (Becher, 1989: 15). Scholars in the science-based professions aim to produce products and techniques, and those in the social professions aim to produce protocols and procedures. Applied methods operate in the real world rather than under experimental conditions where variables can be carefully controlled, and for this reason they always entail some qualitative judgement, even when the discipline is science-based.

All this suggests that Verbal and Mental Processes across the discipline types are likely to vary, especially in the pure fields where the distinction between soft and hard knowledge is greatest. Students' use of the lexical items associated with these Processes should indicate the extent to which they are positioned within one of the four quadrants into which hard, soft, pure and applied disciplines fall.

4　Keywords and Keyness

The four disciplinary areas can be examined using WordSmith Tools software (Scott, 2004), which enables identification of keywords (KWs). Scott and Tribble (2006: 56) explain 'keyness' as 'what the text "boils down to" . . . once we have steamed off the verbiage, the adornment, the blah blah blah'. The technique for identification of KWs described by Scott (1997) and Scott and

Tribble (2006) requires both a reference corpus and one, or more than one, study corpus, often a subset of the larger reference corpus. Keyness is obtained by statistical comparison of word frequencies in these two types of corpora; the standard default setting for WordSmith Tools requires a minimum of three occurrences of each keyword in the study corpus, and a Log Likelihood statistic (Dunning, 1993) with a p value of 0.000001. A word is deemed to be positively key if its frequency in the study corpus is unusually high and negatively key if its frequency in the study corpus is unusually low.

Scott and Tribble (2006: 59–69) illustrate the process of keyword analysis using Romeo and Juliet as a study corpus and the entire collection of Shakespeare's plays as a reference corpus. Their analysis shows that KWs which occur significantly more frequently in the study corpus than in the reference corpus reflect important themes specific to Romeo and Juliet. A keyword analysis, then, provides an opportunity to examine the typical uses of lexical items associated with Verbal and Mental Processes.

5   Method

For this study our reference corpus was the entire British Academic Written English (BAWE) corpus,[1] a 6.5 million word collection of student assignments which have been awarded high grades when assessed as part of degree coursework at three British universities. The corpus holdings are distributed fairly evenly over four disciplinary groups (Arts and Humanities, Life Sciences, Physical Sciences, Social Sciences) and four levels of study (from first year undergraduate to taught masters level) (see Alsop & Nesi, 2009; Ebeling & Heuboeck, 2007; Gardner & Holmes, this volume; Nesi, 2008). The main study corpora were two subsets of the BAWE corpus: 96 history assignments (309,761 words) and 68 physics assignments (196,487 words). At a second stage of analysis, KWs in these two 'pure' disciplines were compared with further subsets of the BAWE corpus in 'applied' disciplines: engineering (238 assignments, 599,687 words), medicine (80 assignments, 214,226 words), and hospitality, leisure and tourism management (HLTM) (93 assignments, 296,709 words). Becher and Trowler (2001: 39) note that it is not always straightforward to determine a priori whether a particular discipline is pure or applied, since different researchers and different university departments give different emphasis to different aspects of their field. However, for the purposes of this study HLTM, Medicine and Engineering were selected as representative of the applied disciplines, based on the kinds of assignments submitted to the BAWE corpus by students in these disciplines, and also on interview data gathered in the early stages of the project 'An Investigation of Genres of Assessed Writing in British Higher Education' (see Nesi & Gardner, 2006).

While undertaking our analysis, we took into account that the term 'Process' has two senses in SFL: it can refer to what goes on in the whole clause, and it can refer to 'that part of the proposition encoded in the Verbal Group' (Bloor & Bloor, 1995: 110). In our study we considered both the Verbal Group and all clausal elements that suggested the Processes of 'saying' and of 'internal cognition'. Berber Sardinha's (2000) formula for predicting the number of KWs in a corpus was found to yield around 1,500 KWs for our history corpus alone, and over 500 KWs even if we used a very stringent level of significance (p<0.0000000001). This number was clearly too great for the detailed analysis we had in mind, and we therefore restricted our data to those 'saying' and 'internal cognition' words identified in WordNet (Fellbaum, 1998).

First, a list of word forms was extracted from WordNet containing all the hyponyms and troponyms of relevant senses such as 'argument', 'belief', 'cognition', 'opinion', 'say', 'state' and 'tell'. This initial list was shortened by removing some words that are not used epistemologically (e.g. BARK, GROWL, HISS) and some which did not appear in our chosen corpus (e.g. AVER, DISAVOW, OPINE). The reduced list of 122 words was then expanded by adding all inflectional forms for each lemma, resulting in the word forms shown in Table 3.1.

Table 3.1 Verbal and mental process words, including inflected forms

abstract	confirmation	explanation/s
account/s	consider/s/ing/ed	fact/s
acknowledg/e/es/ing/ed	consideration/s	falsification
add/s/ing/ed	contradict/s/ing/ed	falsif/y/ies/ying/ied
advanc/e/es/ing/ed	contradiction/s	find/s/ing/found
announc/e/es/ing/ed	criticis/e/es/ing/ed	findings
announcement/s	criticism/s	grounds
answer/s/ing/ed	criticiz/e/es/ing/ed	identif/y/ies/ying/ied
apparent/ly	declaration/s	indicat/e/es/ing/ed
argu/e/es/ing/ed	declar/e/es/ing/ed	indication/s
argument/s	defend/s/ing/ed	infer/s/ring/red
ascertain/s/ing/ed	defense	inference/s
assert/s/ing/ed	definitely	information
assertion/s	demonstrat/e/es/ing/ed	justif/y/ies/ying/ied
assum/e/es/ing/ed	demonstration/s	know/s/ing/n/knew
assumption/s	detect/s/ing/ed	knowledge
bas/e/es/ing/ed	determin/e/es/ing/ed	not/e/es/ing/ed
belief/s	discover/s/ing/ed	notic/e/es/ing/ed
believ/e/es/ing/ed	discovery/discoveries	observ/e/es/ing/ed
calculat/e/es/ing/ed	disproof	opinion/s
calculation/s	disprov/e/es/ing/ed	premise/s
claim/s/ing/ed	distinctly	present/s/ing/ed
clear/ly	evidence	proof/s
cognition	evident/ly	proposal/s
confirm/s/ing/ed	explain/s/ing/ed	propos/e/es/ing/ed

Table 3.1   Continued

proposition/s	representation/s	support/s/ing/ed
prove/e/en/es/ing/ed	representative	suppos/e/es/ing/ed
rationalis/e/es/ing/ed	respond/s/ing/ed	tell/s/ing/told
rationaliz/e/es/ing/ed	response/s	theor/y/ies
realisation	retort/s/ing/ed	think/s/ing/thought
realization	say/s/ing/said	thoughts
realis/e/es/ing/ed	see/s/ing/saw	unambiguous/ly
realiz/e/es/ing/ed	show/s/ing/n/ed	uncertain
reason/s/ing/ed	stat/e/es/ing/ed	unclear
recognis/e/es/ing/ed	statement/s	understand/s/ing/~stood
recognition	stipulat/e/es/ing/ed	undoubtedly
recogniz/e/es/ing/ed	stipulation/s	unlikely
repl/y/ies/ying/ied	suggest/s/ing/ed	verification
represent/s/ing/ed	suggestion/s	verif/y/ies/ying/ied

Using WordSmith Tools Version 4.0 (Scott, 2004), we then proceeded to identify from this list those words which were 'key' in our corpus of student writing.

The findings and discussion are presented below in two parts, first focusing on the pure disciplines of history and physics, and then contrasting these with the findings for the applied disciplines.

6   Findings and Discussion

6.1   Pure disciplines: History and Physics

Table 3.2 gives the KWs that occur with significantly different frequencies in the history and physics corpora as opposed to the reference corpus ($p<0.00001$).

Some words at lower keyness values not listed in Table 3.2 displayed similar frequency patterns across both study corpora. DISCOVERY, for example, was positively key in history (10.57) and physics (5.35), while TELL, CONSIDERATION and CONSIDER were negatively key in history (–9.61, –15.87, and –16.41) and physics (–5.91, –8.19 and –6.13). The most striking point about the lists in Table 3.2, however, is that so many of the positive KWs in history are negatively key in physics, and vice versa. ARGUE, BELIEF, SUPPORT and CLAIM were significantly more common in the history assignments but significantly less common in the physics assignments than in the BAWE corpus as a whole. DETERMINE, KNOW, CALCULATE, FIND and SHOW behaved in the opposite way.

It should be noted that our methods of lemmatizing and of retrieving KWs concealed the distinction between homonyms. For example the method did not distinguish between nouns and verbs such as CLAIM/n and CLAIM/v, or, more importantly, the derivationally unrelated STATE/n and STATE/v. Moreover,

Table 3.2 Verbal and mental process keywords in History and Physics

History		Physics	
Keyword	Keyness	Keyword	Keyness
Positive keywords			
		CALCULATE	350.5819
		THEORY	318.5182
ARGUE	257.5518		
		DETECT	193.5404
		FIND	185.4318
		KNOW	179.1548
		OBSERVE	178.9922
		SHOW	127.7502
		DETERMINE	89.02161
BELIEF	75.51513		
ASSERT	58.30611		
SUPPORT	56.45473		
		DISCOVER	55.50309
STATE	46.53033		
BELIEVE	41.65763		
		ABSTRACT	30.03328
CLAIM	27.87113		
		EXPLAIN	24.58973
PROPOSAL	23.85093		
CRITICISM	22.27682		
Negative keywords			
NOTICE	−19.8002		
		OPINION	−20.3714
		ARGUMENT	−20.8476
IDENTIFY	−29.2359		
		SUGGEST	−32.9894
DETERMINE	−34.0833		
ADD	−34.1799		
		RESPONSE	−37.5827
		BELIEF	−39.5892
BASE	−40.8858		
KNOWLEDGE	−43.1839		
CALCULATION	−50.2794		
KNOW	−52.8485		
		SUPPORT	−58.5883
CALCULATE	−99.0087		
FIND	−99.7343	CLAIM	−99.7004
		ARGUE	−135.263
SHOW	−177.149		
INFORMATION	−214.336		

the method did not distinguish between Process types such as SUGGEST/Verbal and SUGGEST/Relational; or SHOW/Verbal and SHOW/Relational. As Halliday points out (1994: 142) words like SUGGEST and SHOW may be judged to realize Verbal Processes if the subject (in the active voice) is a conscious being, and/or if the clause it is in is projecting (e.g. 'I suggested that there might be a risk'), or Relational Processes when the nominal elements are abstract (e.g. 'the data suggested a potential risk').

Therefore in order to explore the contrasts between the KWs more fully we also examined them in their wider context, using discourse analysis to make up for the fact that our corpus linguistics techniques did not allow for delicate analysis. For this purpose 20 examples of each of the positive KWs in each discipline were extracted and analysed.

The positive KWs from the history study corpus were found either to describe the interplay of claim and counter claim that constitute the practice of history, or to relate to the field of the history texts, describing past events where individuals and groups made claims and proposals, acted on beliefs and offered (or denied) support to each other. Agents were almost always human, and rarely implicit. Subjects, when present, were proper nouns or personal pronouns. ARGUE, ASSERT, BELIEVE, CLAIM/v and CRITICISM referred exclusively or chiefly to Processes whose agents were historians, including the student writers, while BELIEF, CLAIM/n, PROPOSAL and SUPPORT/n referred to Processes whose agents were historical figures or groups (SUPPORT was used chiefly to refer to political support offered by one group or state to another). In almost every case the forms *state* and *states* functioned as nouns, with the sense of 'nation', and therefore did not express the Process of 'saying'. The verb forms *stated* and *stating*, however, had interpersonal and epistemological or ideational roles in roughly equal numbers. Some examples of students' use of KWs expressing 'saying' and 'internal cognition' in history assignments are listed below. In these and subsequent examples the codes in square brackets are the unique identifiers of the BAWE corpus files.

1. . . . in his expressive Cold War jeremiad 'Gentleman: You are Mad,' social commentator Lewis Mumford **asserted** that 'madmen govern our affairs in the name of security'. [0005c]
2. Gareth Steedman Jones provides the seminal work. He **argued** that social movements, such as Chartism, could be constituted on ideological and political platforms . . . [0005a]
3. In this sense, it can be **argued** that for Marx and Engels, a primitive idea of democratic, or majority, rights served to justify a complex social theory of inevitable revolutionary struggle. [0003i]
4. I would **argue** that the 1917 revolution would not have occurred without it; [0010a]

5. Recent **criticism** has argued that the Cold War split has caused intellectuals to make an over-simplified distinction between 'individualistic liberalism and state collectivism' . . . [0003j]

The claims and beliefs expressed in these examples were only rarely supported by evidence (three times in the 20 occurrences of ARGUE: one time in the 20 occurrences of ASSERT: and in none of the 20 occurrences of BELIEVE). Nominalization, however, enabled arguments and beliefs to be explicitly evaluated. In eight of the 20 occurrences of ARGUMENT, for instance, there was some form of evaluation, as in Example 6:

6. . . . in order to show the validity of the basic realist **argument** that there are two distinct realms of reality . . . [0004d]

In physics, all but three of the positive KWs referred to the establishment of facts from direct observation, measurement or calculation. The exceptions turned out not to express Verbal or Mental Processes. These were ABSTRACT (all examples referred to the abstract section of a research report), THEORY (in its usual role as the title of a section heading) and DETERMINE (when it was occasionally used to refer to causal relationships between states or events, as in Example 7).

7. The colour of the particular area of phosphor that the electron is fired at **determines** the colour of the resultant light on the screen. [0051a]

Agents were for the most part human in both disciplines, but in physics human agency was much more likely to be implicit, as in Examples 8 and 9:

8. Gamma ray photons are uncharged and create no ionisation or excitation of any material they pass through and hence the methods of **determining** their energies are somewhat limited. [0051c]
9. . . . and this data was used to **calculate** a value for Planck's constant. [0074a]

Non-human agents representing physical phenomena, theories, models or textual elements such as tables or calculations were also present in the physics assignments. The positive keyword SHOW, for example, was used with text-internal agents in 14 out of the 20 cases, although there was still always a human observer and a projecting clause, as in Example 10, indicating Verbal rather than Relational Process.

10. Table 1 **shows** that as the intensity was decreased, the stopping voltage measured increased. [0074a]

DETERMINE and EXPLAIN sometimes referred to causal relationships holding between real world phenomena, as in Examples 11 and 12 and were therefore not always part of Verbal or Mental Process clauses.

11. The energy carried by a wave is **determined** by its intensity. [0074a]
12. The rotational motion easily **explains** the Earth's diurnal motion. [6097b]

History and physics thus have distinct sets of KWs, used to signal the epistemological value of the propositions they introduce. This reflects a fundamental difference in disciplinary cultures. KWs in the physics assignments referred to causal, logical and evidential relationships between physical phenomena and between phenomena and propositions (in the form of models, theories and properties of physical objects or systems). Thus in physics the identities of agents are commonly suppressed, to emphasize the fact that knowledge is derived from replicable laboratory activities, observations and measurements rather than from interpretation or discussion. On the other hand the KWs in history were more likely to have explicit agents; the identities of the authorities and sources referred to were important in establishing their validity and relevance.

6.2 Applied disciplines: Medicine, Engineering and Hospitality, Leisure and Tourism Management

Table 3.3 presents the keywords for the three applied disciplines that were significant at the level p<0.00001.

The three lists are strikingly different from those in the pure disciplines (Table 3.2). Unlike the KWs in history and physics, the majority of positive KWs in the applied disciplines that also featured in our WordNet list (Table 3.1) indicated a degree of uncertainty regarding the proposition being expressed, as in Examples 13–15. Many also turned out not to express Verbal or Mental Processes.

13. There were no abnormalities in other systems, which **indicate** that this diagnosis is less likely. (0194h, Medicine)
14. In most cases it is **uncertain** whether the individual factors act as initiators or promoters, due to the complex interactions between them (5). (0047a, Medicine)
15. Although the shaft and thrust bearings are being designed to take 50 per cent body weight, it is **unlikely** that this will be thrust onto the drill in its lifetime. (0023e, Engineering).

Table 3.3 Keywords in HLTM, Medicine and Engineering assignments

HLTM		Medicine		Engineering	
Keyword	**Keyness**	**Keyword**	**Keyness**	**Keyword**	**Keyness**
Positive keywords					
		UNLIKELY	457.670		
				DEFINITELY	192.077
INFORMATION	179.730				
				APPARENT	171.312
		CONFIRM	164.213		
				CALCULATE	130.768
				UNLIKELY	119.399
		EVIDENCE	116.112		
		PRESENT	113.433		
				CALCULATION	100.183
		UNCLEAR	107.263		
		PROPOSED	107.263		
APPARENT	98.174				
		INDICATE	94.824		
		APPARENT	85.810		
DEFINITELY	79.766				
		UNCERTAIN	71.509		
		SUPPORT	61.418		
UNLIKELY	55.223				
		FINDINGS	48.883		
UNCERTAIN	42.951	APPARENTLY	42.905		
				UNDOUBTEDLY	41.530
UNDOUBTEDLY	36.815				
				UNCLEAR	31.147
APPARENTLY	30.679				
CRITICIZED	30.679				
UNCLEAR	30.679				
IDENTIFY	28.561	ARGUES	28.603	DETERMINE	28.737
Negative keywords					
				SUGGEST	−30.157
				CLEAR	−31.859
		KNOWLEDGE	−32.247	SAY	−32.387
		CLAIM	−34.498		
				ANSWER	−35.177
CLAIM	−37.772				
		FIND	−39.780	FINDINGS	−39.718
		INFORMATION	−43.675	CLAIM	−43.993
				ABSTRACT	−46.375
		ARGUMENT	−55.292	PRESENT	−55.266
				BELIEVE	−56.917
				CRITICISM	−60.525
				BELIEF	−65.731
				EXPLAIN	−67.214
		FACT	−70.340		

Table 3.3   Continued

HLTM		Medicine		Engineering	
Keyword	Keyness	Keyword	Keyness	Keyword	Keyness
EVIDENCE	−74.028	SEE	−74.304		
		STATE	−97.101	FACT	−78.302
THEORY	−100.355			ARGUE	−119.530
STATE	−125.666	THEORY	−145.713		
				ARGUMENT	−161.501
				THEORY	−198.460
				EVIDENCE	−295.722
				STATE	−333.788

In HLTM assignments, UNCERTAIN also concerned the operation of businesses in an insecure environment:

16. Comprising mainly small businesses that rely on fluctuating demand, profitability is **uncertain** and production, wages and skills are low. (3013b, HLTM)

Most uses referred to uncertain inferences drawn from observations. Even those KWs that appear to have a high degree of certainty (CONFIRM, SUPPORT, DEFINITELY, UNDOUBTEDLY), however, often implied that the truth of the proposition had been established in the face of some doubt, as in Examples 17 and 18.

17. However, results using a higher piston velocity still would be needed to **confirm** this assumption. (0329f, Engineering)
18. Clinical signs such as pyrexia, dyspnoea and lung crackles were revealed during the physical examination which **supports** the diagnosis of a LRTI. (0047c, Medicine)

DEFINITELY and UNDOUBTEDLY were used to introduce assertions by the writer and had something of an interpersonal appeal, implicitly addressing objections that might be raised by the reader.

19. . . . SYSTECH Intl. can **definitely** do well in the marketplace . . . (0090a, Engineering)
20. **Undoubtedly**, since more new entrants are going into the industry, the industry competition will be stiffer. (3085c, HLTM)

As before, our methods of lemmatizing and of retrieving KWs did not distinguish between nouns and verbs with the same form, and it was necessary to examine them in context to see whether they formed part of Verbal or Mental Processes. SUPPORT/n, for example, turned out to function as a Participant in Material or Relational Process clauses, as in Example 21.

21. Die-cast aluminium would be used for the housing and lubricated ball bearings would be used for shaft **supports**. (0018d)

Becher and Trowler (2001: 36) describe applied approaches to knowledge as 'functional' and 'pragmatic', applying 'heuristic approaches' to develop products or procedures. These descriptions are consistent with our observations concerning the KWs; most of those from the applied disciplines reflected professional practices that use partial evidence to support the formation of opinions or decision making. The student writers used these words in order to determine causal relations (Examples 14, 20), diagnose properties of objects or systems (Examples 13, 17, 19) and assess likely outcomes (Example 15), always on the basis of imperfect evidence.

The data from Table 3.3 reveal both similarities and differences between the three applied disciplines. Becher and Trowler (2001: 36) identify medicine as a hard applied discipline, while HLTM is a social science and exemplifies a soft applied discipline. Nevertheless the two share many of the same KWs in Table 3.3. In contrast, engineering, the other hard applied discipline, yielded fewer positive KWs and more negative KWs. Engineering negative KWs included such words as ARGUE, BELIEF and SUGGEST, which were also negatively key in the hard pure physics assignments. These words are more likely to be positively key in soft disciplines because of the emphasis they place on the social development of knowledge and on the identity (and thus trustworthiness) of the researcher. Engineering and physics are both disciplines where argument and interpretation are less important than measurement and observation.

Once again our study of the KWs indicated a fundamental difference in disciplinary cultures, particularly between the pure and applied fields, but also between HLTM and medicine on the one hand, and engineering on the other.

7   Conclusion

In this study we have been able to quantify a distinction between student writing in hard, soft, pure and applied disciplines by looking at the keyness of selected lexical items. This distinction reflects a difference in knowledge construction, which students who have achieved acceptable grades seem to have learnt to apply in their university coursework. One pedagogical implication of the findings is that there is a need to recognize discipline-specific ways of

thinking and saying at nominal group and clause level, where they can indicate the extent and explicitness of expressions of human agency, for example, and the degree of certainty with which propositions are put forward. For teaching and learning purposes it is important to present appropriate discipline-specific exemplars not only of entire texts, but also of lexical items and the grammar of the clause.

Notes

[1] The BAWE corpus was developed 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 called 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).

References

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

Babaii, E., & Ansary, H. (2005). On the effect of disciplinary variation on transitivity: The case of academic book reviews. *Asian EFL Journal, 7*, 113–126.

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines.* Milton Keynes: Society for Research into Higher Education and Open University Press.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd edn). Buckingham: Society for Research into Higher Education and Open University Press.

Berber Sardinha, T. (2000). Comparing corpora with WordSmith Tools: How large must the reference corpus be? *Proceedings of the Workshop on Comparing Corpora.* Annual Meeting of the ACL, Volume 9 (pp. 7–13). Morristown, NJ: Association for Computational Linguistics.

Bloor, T., & Bloor, M. (1995). *The functional analysis of English: A Hallidayan approach.* London: Arnold.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19*, 61–74.

Ebeling, S., & Heuboeck, A. (2007). Encoding document information in a corpus of student writing: The experience of the British Academic Written English (BAWE) corpus. *Corpora, 2*, 241–256.

Fellbaum, C. (1998). *WordNet: An electronic lexical database.* Cambridge, MA: MIT.

Halliday, M.A.K. (1994). *An introduction to functional grammar* (2nd edn). London: Arnold.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing.* Harlow: Longman.

Lea, M., & Street, B. (2000). Student writing and staff feedback in higher education: An academic literacies approach. In M. Lea & B. Stierer (Eds), *Student writing in higher education: New contexts* (pp. 1–14). Buckingham: Society for Research into Higher Education and Open University Press.

Love, A. (1993). Lexico-grammatical features of geology textbooks: Process and product revisited. *English for Specific Purposes, 12*, 197–218.

Martínez, I.A. (2001). Impersonality in the research article as revealed by analysis of the transitivity structure. *English for Specific Purposes, 20*, 227–247.

Nesi, H. (2008). BAWE: An introduction to a new resource. In A. Frankenberg-Garcia, T. Rkibi, M. Braga da Cruz, R. Carvalho, C. Direito & D. Santos-Rosa (Eds), *Proceedings of the 8th Teaching and Language Corpora Conference* (pp. 239–246). Lisbon, Portugal: ISLA.

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

Scott, M. (1997). PC Analysis of key words – and key key words. *System, 25*, 233–245.

—(2004). *WordSmith Tools 4.0* [Computer software]. Oxford: Oxford University Press.

Scott, M., & Tribble, C. (2006). *Textual patterns: Keyword and corpus analysis in language education*. Amsterdam: Benjamins.

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

Chapter 4

In the Wake of the Terror: Phraseological Tools of Time Setting in the Narrative of History

Marina Bondi

1   Introduction

The role played by time setting is pivotal in the discourse of history (Coffin, 2006: 95–115; Martin, 2003). Historical texts are prototypically characterized by features of narrativity and tend to develop chronologically. Time adverbials are perhaps the most obvious signals of shifts in time settings (e.g. Silver, 2006: 90–100). They designate the kind of movement realized along the timeline (forward or backward movements), as well as the limits of the time-span identified: see, for example, the forward movement of a sequence like *one day . . . later . . . eventually*, or the limits signalled by adverbials like *by the time, as yet, not yet*, etc.

The combination of textual signals like time adverbials with more specific verbal signals (basically tense and modal verbs) will provide for the possibility to establish different time settings, in the typical double temporal perspective that characterizes all narratives with a time of the story and a time of the discourse (Chatman, 1989; Weinrich, 1971). As shown in the following example, writers can not only establish a main narrative line by setting it firmly in a specific time span or point of the past (*in his childhood*; *at her death in March 1229*), but they can also temporarily shift the axis forward in the past (*he would . . . in 1234*) or even to the present of discourse (*is known*):

Example 1
In his childhood Thibaud IV lived under the tutelage of his mother Blanche of Navarre, whose claim to the inheritance of the crown of Navarre he inherited *at her death in March 1229*. (He *would* accede to the throne *in 1234*, and *is known* in the Navarrese lists as King Teobaldo I) (JMH, 1999).[1]

The notion that historians can only look at the past from a present point of view has long been debated by historians themselves (Carr, 1961; Collingwood,

1994; Evans, 1999). History research articles, in particular, do not simply reconstruct 'facts', but also offer a representation of the writer's interpretation of facts, and it is this position that becomes a key to the professional identity of the academic writer. It is therefore inevitable to see the writer's position in the relationship between past event and present gaze. What becomes prominent in academic history – following Coffin (1997) – is the position of the *interpreter*, expressing judgements on people and events, rather than the position of the *recorder*, presenting factual chronicles, devoid of attitudinal lexis. While the position of the recorder may be highlighted to substantiate the plausibility of interpretation, it is also true that the academic writer needs to argue a position within disciplinary debates. History is thus recount, interpretation and *dialogic argumentation* of the interpretation (Bondi, 2007; Bondi & Mazzi, in press).

Using recent models of time and value in historical discourse developed by J.R. Martin and the systemic functional school of linguistics (Coffin, 1996, 1997; Eggins et al., 1993; Martin, 1993, 2002; Veel & Coffin, 1996, and especially Martin, 2003 and Coffin, 2006), we shall look at the role of phraseological tools of time setting.

Martin (2003) distinguishes time *sequence* – often implicit and inferable from the simple succession of events – from time *setting*. In time setting, narration highlights agency and aims at a generalization about people and events narrated. This requires the *periodization* typically realized by initial position adverbials, whose function lies in identifying *phases* within the world of the text: 'We hop through the past instead of walking through each event one after another' (Martin, 2003: 24–25). It is this distinction between a mere sequencing of events and their articulation into phases identified as periods (and often nominalized as such) that marks the transition from recount to history. The dynamism of events is fixed into a macro-structure which turns the sequence of events into the construction of periods inside other periods (Martin, 2003: 28).

Adverbial tools of time setting have long attracted the attention of linguists and discourse analysts, but little attention has been paid to the variety of tools employed to designate time periods or phases and their evaluative implications. The notion of *chrononyms* is well known to historians themselves, often debating around the issue of periodization and the denomination of historical periods or units. Discourse analysts, on the other hand, have not dealt adequately with the language of periodization. And yet it looks like an ideal place for an exploration of the role of phraseology in historical discourse: how do historians construct their representation of historical periods? How do they identify phases in historical development? An empirical study of such nominalizations and of the time adverbials they constitute can help answer these questions.

The rest of this chapter intends to explore the variety of phraseological tools employed in a corpus of academic journal articles to identify or classify

temporal units. Section 2 briefly presents the corpus of articles used for the study and the tools for analysis. Section 3 focuses on the most frequent phraseological sequences found in the corpus, and looks at chrononyms – noun groups identifying periodization – with their evaluative implications and their textual patterns. Section 4 looks more closely at adverbial phraseology, studying in particular, the co-text of complex prepositional phrases expressing 'transitional' time setting (*in the wake of, in the aftermath of, on the eve of*). In the conclusions, frequencies and patterns are interpreted in the light of factors characterizing academic discourse and specific disciplinary values.

2   Materials and Methods

The analysis is based on a corpus of about 2.5 million words collected at the University of Modena and Reggio Emilia (HEM-History module). The corpus contains 306 historical journal articles across a wide range of disciplinary perspectives. The articles were taken from the 1999 and 2000 editions of the following specialized journals: *Labour History Review* (LHR), *Historical Research* (HR), *Gender and History* (GH), *Journal of European Ideas* (JEI), *Journal of Medieval History* (JMH), *Journal of Interdisciplinary History* (JIH), *Journal of Social History* (JSH), *Studies in History* (SH), *American Quarterly* (AQ) and *American Historical Review* (AHR). Even though journals were partly identified through exogenous criteria such as availability in electronic form, recourse was made to disciplinary experts who suggested a set of reliable publications to choose from. The corpus consists of full texts, from which only footnotes, tables and bibliography have been removed. No attempt has been made to identify the first language of the writers or their affiliation. The aim of the the whole project was not to describe a specific regional variety of English, but rather to describe international standards of academic English. The corpus is part of a modular network of corpora, which allows for comparison with other disciplinary corpora of similar design.

The present investigation has a double focus, involving both the nature of chrononyms, that is, noun phrases identifying the time period, and the adverbial patterns determining time frameworks in historical texts.

Chrononyms show different degrees of phraseological complexity: for example, *the twenties, the middle ages, the age of reason*, etc. They function both as proper nouns and as general denominations that can be more or less definite (e.g. *the inter-war period, the post-war period*). They can also be shown to suggest an evaluation of the period itself (*the roaring twenties, the age of anxiety*).

Adverbial patterns isolate segments in chronological time and determine a framework for textual time. They often involve chrononyms in the definition of time settings (e.g. *on the eve of the war, in the late 19th century*), thus contributing to the interpretation of narrative sequences.

Adapting Coffin's (2006: 4–6) semantic categories, we can see that chrononyms typically help with the function of *Segmenting time* (*the middle ages*, *the second world war*), whereas adverbial patterns can involve them by expressing *Setting* in time (*soon after the second world war, in the middle ages*) or *Duration* in time (*for nearly half of the second world war*), often combining with *Phasing* time (*the onset of the second world war, by the end of the middle ages*).

The methodology adopted for this study combines a corpus and a discourse perspective. Phraseological tools are identified on the basis of a combination of frequency-based information and semantics. This means looking at the many different trends of recent work on phraseology, ranging from linguistic studies exploring the continuum of word combinations from the most opaque to the most transparent (e.g. Cowie, 1998), to work based on bottom-up statistical analysis of co-occurrences (starting from Sinclair's 1991 idiom principle and moving on to Moon, 1998; Hunston & Francis, 1999; Sinclair, 2002; and Hunston, 2004) and to various interdisciplinary approaches (cf. Granger & Meunier, 2008; Meunier & Granger, 2008).

Drawing in particular on studies of an EAP-specific phraseology (Biber, 2004; Biber et al., 2004; Charles, 2006; Groom, 2005; Nesi & Basturkmen, 2006; Siepmann, 2005), we look at both quantitative and qualitative factors in order to identify our data. Quantitative analysis of repeated strings of words – whether referred to as n-grams, clusters or lexical bundles (Biber, 2004; Biber et al., 2004) – can be an excellent starting point, but this needs to be related to significant functions, such as discourse relations in work by Siepmann (2005) or Nesi and Basturkmen (2006). Mere lexical repetition, furthermore, does not account for repeated patterns involving a range of semantically related lexical items, as shown in Groom (2005) and Charles (2006). Starting from the frequencies of word forms and multi-word-units, this chapter looks at the extended lexical unit (Sinclair, 1996) – with its corollary of semantic preference and semantic prosody – trying to identify both the potential semantic associations between otherwise different forms and the association of the unit with further semantic or textual-pragmatic meanings, thus looking at semantic fields and semantic sequences (Hunston, 2008).

The first step in the analysis consists in looking at frequency data – wordlists and keywords, as calculated by means of *WordSmith Tools 4.0* (Scott, 2004) – and providing an overview of quantitative variation, with special attention being paid to four- and five-word clusters. These are studied both in frequency lists, highlighting the most frequent items in our corpus, and in 'keyword' lists, highlighting the items of unusual frequency when our corpus is compared to other reference corpora (Scott & Tribble, 2006). We looked in particular at the first 300 words of the frequency wordlist and at the first 100 five-word lists. Keywords were obtained by comparison with corpora of similar design and size (2.5 million words) containing articles in the field of economics and business.[2]

The study is also based on the analysis of concordances; the co-text of the nodes is analysed with a view to determining their collocational profiles and textual patterns, so as to bring out the evaluative and pragmatic implications of the expressions under investigation. Contextual analysis focuses on (a) potential chrononyms, that is, expressions that can be used to denote temporal units as time settings for the narrative; and (b) time adverbials. Special attention is paid to 'transitional patterns', that is, expressions like *on the eve of Charlotte's wedding*, *in the aftermath of 1945*, *in the wake of the Terror*, relating events and their interpretations to other landmark events (*before/after something*). These expressions are used to focus on transition phases: they highlight features of the landmark as 'divide', by pointing at patterns of matching and contrast in the time settings delimited by the landmark event. Analysis of the most frequent items is carried out in order to identify: (a) lexico-grammatical patterns in time setting: phraseology, collocation and semantic preference, that is, the 'entities' and 'relations' involved; (b) textual patterns: the pragmatic (and argumentative) moves involved.

3   Chrononyms

An analysis of the wordlists of the corpus shows that a number of generic nouns potentially referring to time units are actually very frequent in the corpus. Words like *century, year/s, period, age* are all found within the 300 most frequent words. And so are lexical units like *war, government, king*, all potentially referring to landmark events that can be used to identify time settings. Table 4.1 lists the most frequent potential chrononyms and provides frequency data (position in frequency list, absolute and relative frequency, distribution).

These generic nouns are often qualified by a set of more varied and less frequent expressions, such as numerals (*seventeenth century, twelfth century*),

Table 4.1   Potential chrononyms

Position	Word	Frequency	%	Texts
6	#[a]	50.258	2.04	306
61	Century	3.292	0.13	275
75	Time	2.757	0.11	298
106	Years	1.860	0.08	268
108	War	1.840	0.07	187
168	Government	1.212	0.05	162
173	Period	1.197	0.05	241
233	Year	897	0.04	202
258	King	845	0.03	112
292	Age	773	0.03	175

[a]This symbol represents any number, thus including those identifying dates.

generic quantifiers (*many years*) or identifiers (*Churchill years*). The co-text of the generic nouns listed above reveals that they are often part of wider phraseological patterns combining signals of Sequencing/Setting in time – meant to collocate events on the time axis – with summary representations of historical interpretations of the event – highlighting the Evaluating/Interpreting function of historical discourse. An expression like *the First World War* identifies a time setting by qualifying a potential landmark event (*war*) in terms of an interpretation of its dimensions (*world*) and of later awareness of subsequent events (*first* only substitutes *great* when a Second World War is categorized).

In a discourse perspective, the basic descriptive tools of the noun group are those of identification. Nominalization turns temporal reference into an element that can be analysed with the tools normally applied to the identification of participants in processes (cf., for example, Martin & Rose, 2003: 145–174). Distinctions such as the one between *classifying* participants and *identifying* participants can thus be used: in an example like '*the sixteenth century* was *a century of believers*'(SIH, 1999) the first noun group is used to identify a time setting (*the sixteenth century*), while the second is used to categorize it or classify it (*a century of believers*). A similar double function (Identifying and Classifying) is played by apposition: 'the eighteenth century, a century so prominently marked by the profusion of ideas on the subject' (HOEI, 2000). The tendency of the two discourse functions is to be associated with different determiners: the definite article for the vastly dominant function of identification (accounting for almost three quarters of the 3,622 occurrences of *century/centuries*) and the indefinite for classification.[3]

The most frequent phraseological patterns in which chrononyms are used are basically two, and can be identified in terms of patterns or semantic sequences (Hunston, 2008):

Table 4.2   THE + [Phase/Ordinal/Qualifier] + [Time unit/event]

The 19th	Century
The early nineteenth	Century
The Victorian	Age
The French	Revolution
The early	Twenties

Table 4.3   THE + [Phase] + OF + [Time unit/landmark]

The turn	Of	The century
The summer	Of	1943
The height	Of	The Terror
The end	Of	The war

An analysis of 400 occurrences of *century* shows, for example, that the premodification pattern (Table 4.2) is by far the most frequent, accounting, as it does, for 369 occurrences (92.25% of total). Phasing pre-modification is significant: 104 occurrences of *century* (26% of total) are qualified by a phased numeral (e.g. *early nineteenth*, *mid-twentieth*, *late eighteenth*). The second pattern (complex noun phrases, Table 4.3) accounts for another 65 occurrences, where phasing is expressed by *turn* (10 occurrences), *end* (11), *decades* (8), *half* (8), *quarter* (7), *years* (5), *middle* (5), *course* (5), *part* (2), *beginning* (2), *days* (1) and *80's* (1).

The patterns involving the most frequent potential chrononyms are often characterized by external objective reference, as when *centuries* or *years* are qualified by numerals, or by way of some kind of relative position, like in *later/ earlier, post-war years*. The same 400 occurrences of *century* show for example that, apart from 10 occurrences referring to duration (*century* as a time unit), 362 occurrences (90.5%) are involved in ordinal reference and 26 in deictic or anaphoric reference, whereas only 2 point at relative position.

We can notice, however, that the vagueness of *years* (as is also the case for *period*) allows for more varied associations than *centuries*. Although the vast majority of the occurrences are identified by numerals, quantifiers and sequencers, as well as endophoric and exophoric deictics (*those, these, the*), numerous expressions – perhaps the most interesting – relate the years to a major landmark event, like: *Jubilee years, the Weimar years, the Civil War years, the inter- and post-war years, the pre-Civil War years, the war years, the war years of 1198 to 1214*. In a random sample of 200 occurrences of *years*, this is the case with 29 occurrences (14.5%). This identifying strategy also allows for a combination of pre- and post-modification that often adds classifying elements to the identification. The functions isolated by the two nominals in a sentence like 'Jubilee years were "years of forgiveness and grace, of safety and peace, of exultation and pardon"' (IOMH, 1999) are somehow conflated in nominals like *The waning years of the Han dynasty, the worst years of the Depression, the wilderness years of the nineteen-thirties*, etc.

An expression like *age* – quite obviously less of a time unit than *century, year* or *period*, and more often referring to stages in human life than stages in the narrative (495/765 occurrences, 64.75%) – is also qualified by a whole range of expressions when used to identify time settings for historical research. The phraseology covers both largely accepted conventional chrononyms, like *the pre-columbian age, the Augustan age, the Augustean age, the Great Age, the Hanoverian Age, the Golden Age* or *the Gilded Age*, and more generic expressions – *the modern technological age* or *the industrial and technological age, the space age, the computer age, the communal age* – typically realized by adjectival premodification (*modern/millennial/military/meritocratic/mechanized/mechanical/inventive/industrial/classical/axial/ aristocratic/apocalyptic*). Nominal post-modification is also often associated with widely accepted chrononyms: *the Age of the Crusades, of the French Revolution, of the Enlightenment* etc. Scare quotes may be used for less recognized expressions (*'sullen*

age') or to signal disalignment with the ordinary meaning of the chrononym, often as a prelude to putting forward a new denomination, for example, when talking about the '*stone age*', before introducing the idea of *the new stone age*.

Nominals with *age* are also quite often accompanied by indefinite determiners (65/270, i.e. 23.33%), as in *a new age, a golden age, an early age* (the most frequent items), or singular occurrences like *a liberal, democratic age* or *an age of great and illustrious princes*, etc. These indefinite expressions – extremely varied in post-modification patterns – may not be immediately perceived as chrononyms, but textual patterns suggest that they are almost always used to classify an age and identify it by reference to some of its peculiarities. It is interesting to note that these classifying uses often coincide with an attempt to get at the best definition of the distinguishing features of an age: for example, 'envisaged the result of the downfall of patriarchy not as *a new mother-age* but as *an age of gender equality*' (AHR, 1999).

Concordance analysis also shows that adverbial phrases (*in an age of*) often signal emphasis of two kinds. On the one hand, they signal some kind of contrast with (or unexpectedness of) other features in the context: 'he seeks to preserve the essential *consistency* of her character, in an age characterised by *disjunction*'. (HOEI, 1999), 'What is the value of *local* specificity in an age of *global* capital?' (AQ1999), 'a chivalrous gentleman in an age of rapacious soldiers of fortune'(AHAR, 2000). On the other hand, they are also used to point at the very feature that stands at the centre and origin of the phenomena recorded: 'In an age of family income, taxation could be raised, not on the person, but on the household or hearth' (JOMH, 2000) or 'In an age of great and illustrious princes, James was inferior to none in "vigour of mind"' (HOEI, 2000).

Language analysis thus seems to confirm the importance often attributed to chrononyms by theoretical historical debate. Even a simple overview of the most frequent chrononyms and their phraseological patterns shows that they offer an important interpretative framework to the structuring of historical discourse and argument.

4   Clusters and Adverbial Phraseological Patterns

If we consider four- and five-word clusters generated automatically when indexing the corpus, we may be able to see that the vast majority have temporal reference: 62 of the first 100 five-word clusters potentially refer to temporal notions and most of them are part of adverbial patterns. Table 4.4 below reports the adverbial clusters with respective frequencies. All the phrases reported are found across a reasonable span of different texts in the historical corpus (ranging from 13 to 74 of the 306 comprised in the corpus).

Many of these clusters also turn out to be key-phrases, when the corpus is compared to other academic corpora. The modularity of our network of

Table 4.4   Five-word adverbial clusters

Cluster	Frequency
At the end of the	110
In the #s and #s	96
By the end of the	63
At the turn of the	60
In the course of the	44
In the late 19th century	44
Between # and #	39
In the early 19th century	39
In the summer of #	39
During the # and #	38
In the wake of the	38
Between # and # the	37
At the beginning of the	35
In the second half of	35
While at the same time	33
In the first half of	31
In the early 20th century	30
For the first time in	29
In the middle of the	28
At the same time as	27
On the eve of the	27
At a time when the	24
At the time of the	24
In the aftermath of the	24
In the spring of #	22
In the early years of	21
In the later middle ages	21
Over the course of the	21
In the late # and	19
At the same time that	18
From the # to the	17
In the history of the	18
At the heart of the	17
During the course of the	17

corpora allows for comparison with other social sciences like economics or business. The keyword function of *WordSmith Tools* can be used to compare lists of clusters. Comparison reveals statistical significance (based on log-likelihood) for many of the clusters above. Table 4.5 lists those that are found to be key-phrases of historical discourse (H) when compared to both economics (E) and business (B), with respective frequencies and keyness index.

From a lexico-semantic point of view, adverbial patterns show a few general trends in our corpus. Reference to duration in time is very limited.[4] Most of the phraseology lies in the field of setting in time (*in the #s and #s*, *in the course of the*, *in the summer of #*), often combined with phasing (*at the end of the*, *by the end*

Table 4.5   Key-clusters in history (reference corpora: economics and business)

CLUSTER	Frequency (H)	Frequency (E)	Frequency (B)	Keyness (E)	Keyness (B)
At a time when the	24	0	0	55.71	36.71
At the turn of the	60	0	0	139.27	91.77
By the end of the	63	30	8	51.83	56.39
During the # and #	38	20		28.51	
During the course of the	17	0	0	39.46	26.00
For the first time in	29	12		26.76	
In the # and #	96	88	28	34.25	49.43
In the course of the	44	10	0	57.90	67.30
In the early 19th century	39	0	0	90.52	59.65
In the early 20th century	30	0	0	69.63	45.88
In the first half of the	31	12	5	30.06	24.67
In the late 19th century	44	14	0	48.54	67.30
In the later middle ages	21	0	0	48.74	32.12
In the midst of the	17	0	0	39.46	26.00
In the spring of #	22	0	0	51.06	33.65
In the summer of #	39	12	0	43.89	59.65
In the wake of the	38	14	5	38.15	33.47
On the eve of the	27	0	0	62.27	41.30

Table 4.6   Prep. + THE + [Transition N] + OF +[Chrononym]

On	The	Eve	Of	Charlotte's wedding
In	The	Aftermath	Of	1945
In	The	Wake	Of	The Terror

of the, at the turn of the, in the late #th century, in the early #th century, at the beginning of the, in the first/second half of, in the middle of the, in the wake of the, on the eve of the).

Apart from distributional expressions of phasing – signalling the beginning, middle and end of a time segment – a small group of phasing devices throws an interesting light on the complexity of time expressions in narrative discourse. They also throw light on the need to integrate syntagmatic considerations with semantic perspctives. I have called these patterns 'transitional', in that they set the narrative around a landmark event, while blending temporal and causal meanings.

The transitional patterns in focus here create phasing by setting an event/state of the narrative around a landmark event (in a transition phase that follows or precedes the landmark event). The sequence can be represented as shown in Table 4.6.

Transitional markers relate events and their interpretations to other landmark events. The semantic implication is one of blending the notions of Time and Causality: temporal sequence and cause-and-effect implications are often interwoven, irrespective of the explicit criticisms of the *post hoc/propter hoc* fallacy

raised by historians and argumentation scholars alike. The textual and pragmatic focus of these expressions is on *transition* phases: these are isolated as segments in time, while features of the landmark as 'divide' are highlighted.

From the point of view of the development of the writer's argument, these adverbials carry a strong comparative implication: the gaze of the writer focuses on the relevant features of the contrast between times before and after a landmark event and attention to one side of the divide strongly predicts reference to the other. Example 2 below shows this pattern of matching and contrast at play:

Example 2

The next phase in the demise *came with the American Revolution*. Like the British after the Seven Years' War, American authorities picked up on Amherst's aborted designs. They, too, attempted to dictate the terms of intercourse. Furthermore, the national independence of the American republic removed the restraining influence that British policy had attempted to exert on the expansion of colonial settlement. *In the wake of the revolution*, swarms of westering settlers pursuing personal independence through private land ownership poured into the Ohio Valley. *As never before*, the lands of Great Lakes Indians became the targets for European occupation. *This was a decisive moment in the shift* from borderlands to bordered lands.

But the borderland era *was not over yet*. What gave it new life was the short-lived rivalry between the American republic and the holdover British domain in Canada. (AHR, 1999)

The relational meaning of transitional markers easily transfers to the narrative and suggests looking at textual sequences a bit more closely. The next subsections focus on *in the wake of, in the aftermath of* and *on the eve of,* respectively. From the point of view of the writer's (and the reader's) temporal perspective, the three examples illustrate different patterns: *in the wake of* and *in the aftermath of* illustrate a retrospective standpoint, presenting the landmark event as an antefact to the main event/state in focus, whereas *on the eve of* has a prospective dimension, pointing forward to an event that could somehow help the reader understand the time setting in focus.

4.1   In the wake of

The corpus contains 65 occurrences of this expression, mostly followed by reference to critical events, ranging from explicit conflict (*Terror, Great/Spanish-American/World War, Indian/1680 Pueblo Revolt*) to violent or brutal ending of something/someone (*assassination, liquidation, disintegration, crash, defeat, Owenites' own failure*), and times of unrest (*riots, brutal beating, 1775/French revolution, Nat Turner/1715/Jacobite/Great rebellion*). Only 13 of the 65 occurrences

(20%) do not refer to negative events. This may be as much a feature of the discourse of history as of the phraseological pattern. But a look for comparison at other corpora confirms that the expression is often associated with negative events and critical change.

The semantic preference of the whole adverbial *in the wake of* + [Chrononym] is for propositions of two kinds: they either refer to critical states/events (as signalled, for example, by collocates like *tension*, *serious consequences*, *threateningly . . .*) or to taking/changing a direction/position (as signalled by verbs like *introducing* or *re-evaluating*, adjectives like *new*, all pointing to change, often in a direction that is wrong or unsuccessful).

Analysis of larger textual sequences in our corpus shows that little more than half of the occurrences (34/65, 52.30%) are thematized. The quantitative role of thematized position varies widely across the three adverbials considered: basically dominant for *in the wake of* and *on the eve of* (cf. Section 4.3), as we will see in Section 4.2, it is dispreferred by *in the aftermath of*.

Initial position seems to emphasize the textual function of the phraseological unit, which could be described in terms of a double action. The expression introduces a statement – often attributed – that something is the case at a certain point in time, while at the same time predicting a subsequent claim on different conditions or different interpretations. The textual sequence is illustrated in Examples 3 and 4 below.

Example 3
In the wake of the Terror, it appeared as if the stain of bad social origin was unremovable and incurable. *It took the war to realize and institutionalize* Stalin's dictum in Soviet political life. (AHR, 1999)

Example 4
Some earlier scholars believed that appropriation had come to England *in the wake of the Conquest, but this should be seen as one of those myths* that often issue from great historic events. After sporadic occurrences of appropriation in the mid-twelfth century, t*he practice only gained* significant momentum in the eighties of the twelfth century. (JMH, 2000)

4.2   In the aftermath of

The occurrences of *in the aftermath of* are more limited. The expression occurs 42 times and is only thematized and characterized by a wider scope in 15 occurrences (35, 71%). The pattern is once again followed mostly by explicitly *negative* events (except for 13, i.e. ca. 30%), as explicitly suggested by collocates like *flood*, *assassination*, *war*, *revolution*, *reformation*, *oil crisis*, *incident*, *attack*, *defeat*, *collapse*. More implicitly, words like *Crusade* or *Reformation* typically refer to major critical events that are rich in historical consequences, much in the

same way as reference to a specific *law* or *report*, a *contest*, a *presidential vote* or a *(hotly contested) election* identifies typical nodes in historical argument narrative chains.

From a textual/rhetorical point of view, *in the aftermath of* also shows a clear tendency to occur in contexts where its merging of temporal and causal meanings can be used for the argumentative purposes of the writer. The expression is mostly part of textual patterns of two kinds: (a) presenting the main claim or (b) supporting local claims. The two textual patterns are exemplified below in Examples 5 and 6. Example 5 shows how the ground is prepared by evaluative metadiscourse for a major claim which points at 1067 as a major divide leading to the state of affairs in focus.

Example 5

The important point is that when Eustace rebelled in 1067, he did so by attacking Dover castle, in other words Odo's castle. *The significance of this may well have been underestimated. It is my contention that* a correct understanding of why and by whom the Tapestry was made depends on a true understanding of the relationship between Odo and Count Eustace *in the aftermath of the* incident at Dover in the autumn of 1067. (JMH, 1999)

Example 6 shows a similar function in developing a subclaim through an example (one of two 'discrete ways' in which the press and television were used by Labor modernisers). *In the aftermath of* is used to briefly sum up the antefact and focus on the specific case exemplifying the general trend.

Example 6

To be more precise, the Labor modernisers used the press and television to attain their ends *in two discrete ways. Firstly* they used already existent 'Labor Party crises' aired in the media, in order to argue for modernisation. *A good example of this process occurred* in the early part of 1964. Labor had stumbled to a serious electoral defeat in 1963, suffering from repeated press attacks concerning the practices of internal decisionmaking. *In the aftermath of that defeat,* the mainstream press began to report dissatisfaction with Arthur Calwell's leadership. The Sydney Morning Herald openly editorialised that 'many Labour supporters' believed that Calwell should be unseated. (LHR, 2000)

4.3   On the eve of

On the eve of reverses the temporal perspective of the other phraseological units examined and offers a prospective standpoint. The expression is frequently rankshifted to specify a nominal so as to identify an issue prospectively (*the political culture of the French elites on the eve of the Revolution*): 9 cases out of 34 have this rankshifted structure and basically tend to qualify an issue suggesting

the existence of an indefinite relation with future events. Of the 25 adverbials found – mostly topicalized (15/25, i.e. 60%) – only one is to be interpreted literally (*once a year, on the eve of the Assumption*). All other adverbials identify a fuzzy period of time represented as leading up to a future event.

From an argumentative point of view, the adverbial points to a future landmark event that highlights the significance of the process or state reported or determines a difference in interpretation. The argumentative effect is based on the prospective meaning of the expression and the double temporal perspective that is thus realized in the text: the writer's awareness of future events can be contrasted with the limited knowledge of historical characters or other interpreters.

The adverbial points to one of the most distinguishing features of historical discourse: its double temporal perspective. The historian looks at the past (*Then*) from a present standpoint (*Now*). This is clearly shown in examples like 7 below, explicitly contrasting present historiographic awareness (*today*) with past perspectives (*at that stage*), actually constructing present interest in past western approaches to ideologically biased interpretations of science (*on the eve of the cold war*) as having been subject to inquiry and deserving further exploration.

Example 7
Perhaps there is something comforting about the idea that Butterfield was in every sense the founder of history of science at Cambridge, in that this allows us to rest easy in the belief that whatever quarrels we may have *today* with the unwieldy concept of the Scientific Revolution, at least that approach constituted the pinnacle of historiographic sophistication attainable *at that stage*, in Cambridge and elsewhere. We know of some of the scorned alternatives, of course: what, *on the eve of the Cold War*, the academic protocols in the West dictated vis-à-vis so-called 'ideologically biased', 'externalist' interpretations of science and its past has been the subject of a number of historiographical inquiries and deserves further, detailed exploration. (SIH, 2000)

Most of the contexts in which the phraseological pattern is employed can be related to a contrast between different interpretations, typically understood as a contrast between the universe of knowledge of historical characters/experts and the awareness of the writer and his/her scientific community.

5   Discussion and Conclusions

This analysis of phraseological tools of time setting in historical discourse has focused on the most frequent features found in a corpus of academic articles,

with their evaluative implications and textual patterns. Frequency data have helped us identify the most frequent phraseological patterns used to designate time periods: phraseological chrononyms turned out to be particularly interesting when they were not built on numerical time units. The range of phraseological combinations offered for example by *years* or *age* revealed an interesting interplay between identifying time and classifying time, while pointing at the frequent need of historians to discuss the denominations they introduce and to express their evaluative stance in summary representations of a time setting.

The overview of chrononyms has suggested looking more closely at some of the adverbials in which chrononyms are found, studying in particular the co-text of complex prepositional phrases – or prepositional phrase bundles (Biber et al., 1999: 1012–1013) – expressing transitional time setting (*in the wake of, in the aftermath of, on the eve of*). Transitional patterns include both retrospective signals – allowing the writer to introduce antefact or to shift time forward while temporarily reminding the reader of antefacts – and prospective signals – allowing the writer to temporarily shift the narrative forward, while still focusing on a time segment that is set before the future landmark. In both cases there is a blurring of the notions of temporality and causality.

The study of transitional patterns has aimed at illustrating the value of combining corpus and discourse approaches. Retrospective signals like *in the aftermath of* or *in the wake of* were mostly found to be associated with negative events or the negative consequences of events. Analysis of discoursal functions and features also revealed these signals to be strongly evaluative and attitudinal. *In the wake of* was seen to introduce a claim while at the same time predicting a subsequent different interpretation. *In the aftermath of* was seen introducing a major divide as an explanation of the state of affairs in focus, and ultimately of the interpretation proposed by the writer.

Prospective temporal markers like *on the eve of* were seen to contribute to shifting the reader's attention to a contrast between differing interpretations in the double temporal perspective that all narrators can adopt and that historians often exploit in references to a 'future in the past'. The flash-forward movement inscribed in the expression contributes not only to the interpretation of the event, but also to the construction of the writer's persona, whose scholarly awareness of future events contrasts with the limited knowledge of historical characters or other interpreters.

Lexical choices and phraseological patterns can also be related to the epistemology of history, with its emphasis on causal sequences and narrative interpretation of factual data. Frequencies and patterns are thus interpreted in the light of factors characterizing academic discourse and specific disciplinary values, in line with recent interest in cross-disciplinary perspectives on academic discourse (Hyland, 2000; Hyland & Bondi, 2006; Hyland & Tse, 2007). Adverbials are shown to become resources by which the author negotiates his/her position with the reader.

We thus hope to have contributed to current debate in at least two directions. On the one hand the study has paid attention to phraseology as the ideal starting point for an analysis of the relationship between text and language system, while combining frequency data and discourse semantics. On the other hand, in the field of corpus and discourse analysis, we hope to have shown how attention to patterns and semantic sequences can lead to fruitful exploration of the writer's point of view and epistemology.

Notes

1 Here and in the rest of the chapter, examples are taken from a corpus of scholarly journals. Quotations are identified by the acronym of the journal (listed in the methods and materials section), followed by the year of publication. The emphasis is mine.

2 The economics corpus contains articles from the following journals: *European Economic Review, European Journal of Political Economy, International Journal of Industrial Organization, International Review of Economics and Finance, Journal of Corporate Finance, Journal of Development Economics, Journal of Socio-Economics, The North American Journal of Economics and Finance.* The business corpus contains articles from the following journals: *Academy of Management Journal, Administrative Science Quarterly, Business and Society Review, Business Strategy Review, Journal of Marketing Research, Journal of World Business, Marketing Science.* The years collected are the same for all three corpora.

3 Apart from the two constructions above, however, all other occurrences of *a century* (86/88) refer to duration and use *century* as a measure unit (*a century later, a century earlier, for almost a century, half a century of liberal condescension*). A closer look at our data suggests that the distinction may be blurred in a number of contexts.

4 Even adverbial clusters like *During the # and #* or *during the course of the*, when looked at in context, reveal that they refer to position in time, rather than duration. *During the course of the*, for example, refers to duration only in 4 occurrences of the 17 total, typically accompanied by adverbials like *gradually* and *increasingly* and verbs of state or continuous process.

References

Biber, D. (2004). Lexical bundles in academic speech and writing. In B. Lewandowska-Tomaszczyk (Ed.), *Practical applications in language and computers (PALC 2003)* (pp. 165–178). Frankfurt: Peter Lang.

Biber, D., Conrad, S., & Cortes, V. (2004). '*If you look at . . .*': lexical bundles in university teaching and textbooks. *Applied Linguistics, 25,* 371–405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Bondi, M. (2007). Authority and expert voices in the discourse of history. In K. Flottum (Ed.), *Language and discipline perspectives on academic discourse* (pp. 66–88). Newcastle: Cambridge Scholars.

Bondi, M., & Mazzi, D. (in press). Writing history: Argument, narrative and point of view. In M. Shiro, P. Bentivoglio & F. Erlich (Eds), *Analizando el discurso. Homenaje a Adriana Bolívar.* Caracas: Universidad Central de Venezuela.

Carr, E.H. (1961). *What is history?* New York: Alfred A. Knopf.

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*, 310–331.

Chatman, S. (1989). *Story and discourse: Narrative structure in fiction and film.* Ithaca: Cornell University Press.

Coffin, C. (1996). *Exploring literacy in school history.* Sydney: Metropolitan East Disadvantaged School Program, Erskineville, NSW.

—(1997). Constructing and giving value to the past: An investigation into secondary school history. In F. Christie & J.R. Martin (Eds), *Genres and institutions: Social processes in the workplace and school* (pp. 196–230). London: Pinter.

—(2006). *Historical discourse.* London: Continuum.

Collingwood, R.G. (1994). *The idea of history.* Oxford: Oxford University Press.

Cowie, A. (Ed.). (1998). *Phraseology: Theory, analysis and applications.* Oxford: Oxford University Press.

Eggins, S., Wignell, P., & Martin, J.R. (1993). The discourse of history: Distancing the recoverable past. In M. Ghadessy (Ed.), *Register analysis: Theory and practice* (pp. 75–109). London: Pinter.

Evans, R.J. (1999). *In defense of history.* New York: Norton.

Granger, S., & Meunier, F. (Eds). (2008). *Phraseology: An interdisciplinary approach.* Amsterdam: Benjamins.

Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes, 4*, 257–277.

Hunston, S. (2004). The corpus, language patterns, and lexicography. *Lexicographica, 20*, 100–113.

—(2008). Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics, 13*, 271–295.

Hunston, S., & Francis, G. (1999). *Pattern grammar: A corpus-driven approach to the lexical grammar of English.* Amsterdam and New York: Benjamins.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing.* Harlow: Longman.

Hyland, K., & Bondi, M. (Eds). (2006). *Academic discourse across disciplines.* Bern: Peter Lang.

Hyland, K., & Tse, P. (2007). Is there an 'Academic Vocabulary'? *TESOL Quarterly, 41*, 235–253.

Martin, J.R. (1993). Life as a noun. In M.A.K. Halliday & J.R. Martin (Eds), *Writing science: Literacy and discursive power* (pp. 221–267). London: Falmer.

—(2002). Writing history: Construing time and value in discourses of the past. In C. Colombi & M. Schleppegrell (Eds), *Developing advanced literacy in first and second languages* (pp. 87–118). Mahwah, NJ: Erlbaum.

—(2003). Making history: Grammar for interpretation. In J.R. Martin & R. Wodak (Eds), *Re/reading the past: Critical and functional perspectives on time and value* (pp. 19–57). Amsterdam: Benjamins.

Martin, J.R., & Rose, D. (2003). *Working with discourse: Meaning beyond the clause.* London: Continuum.

Meunier, F., & Granger, S. (Eds). (2008). *Phraseology in foreign language learning and teaching.* Amsterdam: Benjamins.

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach.* Oxford: Clarendon Press.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics, 11*, 283–304.

Scott, M. (2004). *WordSmith Tools 4.0 [Computer software].* Oxford: Oxford University Press.

Scott, M., & Tribble, C. (2006). *Textual patterns: Keywords and corpus analysis in language education.* Amsterdam: Benjamins.

Siepmann, D. (2005). *Discourse markers across languages: A contrastive study of second-level discourse markers in native and non-native text with implications for general pedagogy and lexicography.* London and New York: Routledge.

Silver, M. (2006). *Language across disciplines.* Boca Raton: Brown Walker Press.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

—(1996). The search for units of meaning. *Textus, 9*, 75–106.

—(2002). Phraseognomy. In S. Nuccorini (Ed.), *Phrases and phraseology: Data and description* (pp. 17–25). Bern: Peter Lang.

Veel, R., & Coffin, C. (1996). Learning to think like an historian: The language of secondary school history. In R. Hasan & G. Williams (Eds), *Literacy in society* (pp. 191–231). London: Longman.

Weinrich, H. (1971). *Tempus: Besprochene und erzählte Welt* (2nd edn). Stuttgart: Kohlhammer.

Chapter 5

Formulaic Language in Biology:
A Topic-specific Investigation

Diane Pecorari

1   Introduction

Few specialists within the field of English for Academic Purposes (EAP) have expert knowledge of their students' academic disciplines. Appropriate content, teaching approaches and materials for the EAP classroom depend, therefore, on accurate descriptions of how English is used in the disciplines, and a common objective of EAP research is, accordingly, pedagogical applications. However, for research findings to translate to the classroom, their descriptions of writing in the disciplines must cover not only what features occur, but also how and why they are used. The first of these questions – what linguistic features characterize a given genre within a given subject area – is most directly answered by corpus approaches. In order to understand the hows and whys, though, it is necessary to examine the rhetorical or discoursal functions of those features. A full account of writing in the disciplines requires, therefore, a synthesis of corpus linguistics and discourse analysis. The overarching purpose of this chapter is to describe an investigation which is situated at this intersection.

The linguistic feature under investigation in the research reported here is multi-word units, chunks which are represented orthographically as more than one word, but which resemble a single unit in various ways. Such units have been studied from a number of perspectives, and under a varied range of labels, including lexical phrases (DeCarrico & Nattinger, 1988), prefabs (Erman & Warren, 2000), idioms (Simpson & Mendis, 2003), formulaic sequences (Schmitt, 2004) and lexical bundles (Biber, Johansson, Leech, Conrad & Finegan, 1999). These units are of interest to applied linguists because they are a prevalent feature of language; Erman & Warren (2000) found that prefabs made up more than half of their corpus. A good command of them is thus an important part of fluent and idiomatic production. However, and, perhaps not surprisingly, this is an area in which the performance of learners sometimes diverges from that of more proficient language users. For

example, Cortes (2004) found that students in two disciplines, biology and history, make less frequent use of some lexical bundles that are common in published writing.[1]

Multi-word units are, therefore, one aspect of language use which deserves attention in the second- or foreign-language classroom. However, in order for teachers to address them, good descriptions of the lexicon of multi-word units are needed. This is complicated to some extent by the fact that these units are sensitive to context. In various studies they have been shown to occur more frequently in speech than in writing (Biber et al., 1999; Erman & Warren, 2000) and to be longer in writing than in speech (Erman & Warren, 2000). Differences among academic disciplines have been found both in the frequency of such units and in terms of which units are used (Cortes, 2004; Oakey, 2002). Frequency and type of bundle also vary in broader academic contexts, for example, between the language of the classroom and administrative support (Biber & Barbieri, 2007). Clearly, then, any effort to map multi-word units must take context into account. The research reported here is based, accordingly, on a homogeneous corpus made up of a single text type, research articles, within a single field, biology.

This selectivity with regard to research field is a common feature of EAP research. In recognition of the existence of variation among the 'academic tribes' (Becher, 1989) in terms of which textual features are considered conventional, studies of academic language are often based upon corpora consisting of texts from one or more specific disciplines (e.g. politics and materials science, Charles, 2006). Such studies have been very successful in identifying areas of disciplinary difference, and in showing that the variation reflects still more fundamental differences in the ways that knowledge is constructed and received across disciplines.

While a considerable body of research exists to document differences among academic disciplines, possible differences *within* broad disciplinary groupings have, to date, received little attention, although there are suggestions in the literature that such attention could be worthwhile. Samraj (2004) examined student writing from two courses on the same master's programme, Wildlife Behavior and Conservation Biology. She found differences between the two, including the fact that in Conservation Biology, citations to works from other fields were prized, while the texts in Wildlife Behavior did not cite earlier research outside of the immediate area, apparently a result of the fact that the former field is multidisciplinary in nature, while the latter is not. The working and writing practices of botanists as described by Swales (1998) is in sharp contrast with the more frenetic pace found in other natural science research areas. MacDonald did not investigate the relationship between specific topic areas and the larger disciplines they fall within, but her decision to study specific areas and her conscious methodological choice to include in her corpus 'writers who cited each other or were in other ways demonstrably

participating in the same subdisciplinary discourse' (1992: 557) reveals that she perceived the question of topic relationship among articles in a corpus as relevant.

The lack of attention given to more finely tuned subject areas is perhaps surprising, since from the perspective of the novice academic writer, the pressing task is to learn to write a competent research article (or thesis, research proposal, etc.) *on a given topic*. It is therefore of interest, from a pedagogical perspective, first to understand whether smaller subject groups may be useful units of analysis, and if so, to begin the process of documenting the ways in which smaller areas resemble and differ from the larger disciplinary divisions to which they belong. Another objective of this study is, therefore, to examine a corpus of texts which not only were produced in the same field, biology, but which are all related to the same topic, the yeast *Candida albicans*.

2   Methods

In order to do this, a corpus was built consisting of research articles on the topic of *Candida albicans*. This was done by searching the SpringerLink database,[2] which includes a large number of journals in the natural sciences, for articles with *Candida albicans* in the title. The resulting corpus consisted of 181 articles and approximately half a million words.

The type of multi-word units examined here were lexical bundles, following Biber et al. (1999). Unlike some other forms of multi-word units, lexical bundles are defined only with respect to the frequency with which they occur in a text, and are fixed units, with no open or variable slots. Other characteristics, such as transparency of meaning or idiomaticity, are not taken into account. Biber et al. (1999) define lexical bundles as groups of three or more words which occur with a frequency of at least ten occurrences per million words; in addition, to avoid the effects of a single author's idiosyncratic preferences, the minimum number of tokens must be distributed across at least five separate texts. In this study the frequency threshold adjusted for the size of the corpus was five occurrences, and the criterion for distribution was maintained. Although Biber et al. (1999) applied a lower frequency threshold for longer lexical bundles, in the present investigation a minimum of five occurrences was used for bundles of all lengths.

The corpus was processed with the n-gram function in the *AntConc* concordancing software (Anthony, 2008) in order to identify identical strings of a given length, starting at four and moving up until no hits were returned. In this way 1,483 different lexical bundles (and a total of 14,894 types) were identified. The distribution was then checked, and the resulting four-word bundles were examined to determine their function. During this part of the analysis,

three broad categories were identified: content, internal reference and external reference (these categories are described in greater detail below). In some cases the choice to assign a bundle to a given category was obvious; *adhesion of* C. albicans, for example, could easily be classed as having to do with the content of the article. In other cases, though, it was necessary to look at the bundles in context to determine their functions. When at least 50 per cent of the tokens of a given bundle were related to a given function, the bundle (i.e. the type) was assigned to that category. A small number of bundles were used for two or three of these functions with more or less equal frequency, and were not assigned to any category.

These categories are far from impermeable. Although a surprisingly large number of bundles serve the same discoursal function whenever they occur, some are used for more than one function. For example, *demonstrate the presence of* occurs six times in the corpus, and in five cases it is involved in a comment on the paper in which it appeared:

> *In the present paper, we **demonstrate the presence of** pleiotropic drug-resistance genes in* C. albicans.

It was therefore classed as internal reference. However, one occurrence

> *To **demonstrate the presence of** oxygen in the thin-layer cultures, a photogenic bacterium isolated from a marine fish was used*

relates most closely to the objectives of the research, and thus the category 'content'.

As these examples also illustrate, discourse markers, citations and other forms that would have been classified here as 'reference' all serve, ultimately, the purpose of putting forward the scientific content of the paper. The distinctions between these three groups are not, therefore, absolute, but rather a question of which function is served most directly and immediately.

3   Findings

The stock of lexical bundles identified in this corpus differed considerably from those found in the earlier studies which are most relevant for the present investigation, Biber et al.'s (1999) findings based on a more general corpus of academic texts, Cortes's (2004) study of lexical bundles in published writing in biology (*inter alia*) and Hyland's (2008) corpus of student and published articles from four disciplines, including biology. This section describes the lexical bundles in the corpus from three perspectives: number and length; how they

function; and the specific types and purposes within one category, external reference.

3.1   Quantitative characteristics of lexical bundles

Lexical bundles are typically rather short strings; Biber et al. report no figures for bundles in their corpus longer than four words, but say that longer ones are not common (1999: 994), while Cortes (2004) analyses only four-word bundles. In this corpus, however, one bundle was as long as 12 words:

C. albicans *strains used in this study are listed in table 1*

Longer lexical bundles contain shorter ones, of course. Thus the 12-word bundle above contains two 11-word bundles:

albicans *strains used in this study are listed in table 1*
C. albicans *strains used in this study are listed in table*

three ten-word bundles:

C. albicans *strains used in this study are listed in*
albicans *strains used in this study are listed in table*
strains used in this study are listed in table 1

and so on. The presence of long bundles, therefore, contributes in part to the frequency of shorter ones, a significant point since, as Table 5.1 shows, a surprisingly high number of shorter bundles, both tokens and types, were found.

Table 5.1   Length and frequency of lexical bundles

Length	Types	Tokens
12 word	1	6
11	2	12
10	3	18
9	4	31
8	6	43
7	30	201
6	74	540
5	272	2,325
4	1,091	11,718

The corpus contained over 1,000 four-word units which occurred at least five times and in five different texts, thus qualifying as lexical bundles. Four-word bundles occurred just under 12,000 times, for a frequency of 23,436 four-word bundles per million words of text. This is considerably higher than Biber et al.'s finding of 'over 5,000 times per million words' (1999: 994) in their academic corpus, and higher still than the results for two studies of biology writing. In Cortes's (2004) published biology corpus, four-word bundles occurred with a frequency of 3,546 per million words, while in Hyland's (2008) biology corpus, which included both student and published writing, the figure was 3,663 per million words, comparable to Cortes.

It should be noted that these two studies used the more stringent frequency criterion of 20 occurrences per million words, instead of Biber et al.'s (1999) ten per million, which was adopted here. However, that alone is unlikely to account for the differences between their results and those found here. Other possible explanations are taken up in the final section of this chapter.

3.2   Types of lexical bundles

An analysis of the four-word bundles showed that they fell into three broad categories according to their discoursal functions: those used in presenting the scientific content of the article, those used for metatextual or internal reference and those used to refer outside of the text. This last category consists of external reference, that is, all references to phenomena (e.g. researchers, organizations, institutions) lying outside the paper, the research it describes and the individuals who performed it. That category is described in detail in Section 3.3.

The category 'content' included bundles with one or more words from the specialist register within which the articles were written, such as *cultures were grown to*, *fungal pathogen in humans* and *in the fungal cell.* This category also included bundles used to describe research procedures, such as *in order to investigate* and *harvested by centrifugation at.* In addition, a number of bundles were placed in this group which were not particularly technical in and of themselves, but which were involved in delivering the technical content of the paper. So, for example, *however, there was no* and *in an attempt to* do not have any obvious connection to research on *Candida albicans.* However, that connection could be seen when they were examined in context:

however, there was no *difference in plasmid copy number*

in an attempt to *identify the conditions and factors favourable for the pathogenicity of fungi*

and thus they were placed in the 'content' category. This category accounted for approximately[3] 68 per cent of the four-word bundle types in the corpus.

The second category, internal reference, was smaller, consisting of just under 10 per cent of types of the four-word bundles. It included bundles containing section headings (e.g. *Materials and methods Preparation* and *Introduction* Candida albicans *is*) and references to figures and tables (*As shown in Table, are shown in Figure*). In addition, bundles which made reference to the article they appeared in were grouped under this heading (*in the present study, in this report we*). Here too it was not always obvious whether some bundles performed this metatextual function. *These results suggest that* could in principle mean 'the present results', making it an internal reference bundle, or 'the results of the earlier study which was first described', which would be external reference. However, when examined in context it was clear that in the large majority of cases the first sense was intended:

> *. . . did not significantly compromise the synergistic killing by fluconazole in MDM cultures (Table 5). These results suggest that in iron overload . . .*

3.3   Bundles for external reference

Approximately 20 per cent of the lexical bundles in the corpus were used exclusively or primarily for external reference. In addition, another 38 (3%) bundles were used both for external reference and other functions. The most obvious sort of external reference in research articles is perhaps citation, or references to the existing research literature. That was indeed one of the functions served by lexical bundles in this category. However, the bundles classified as 'external reference' also served two other, sometimes overlapping but often quite distinct, functions: acknowledgement and describing the materials and methods used.

The category 'acknowledgement' (*authors wish to thank*) consisted of approximately 50 separate lexical bundles, used for several different purposes. The most common were acknowledging financial assistance (*by a grant from, was accomplished with the*); discussions about the article content or help with writing it (*for critical reading of*); or the provision of experimental samples (*was a gift from*). Some bundles were used to acknowledge more than one type of help:

> *We **would like to thank** J. Ernst for providing the genomic* C. albicans *library.*

> *We **would like to thank** Wayne Murray and William Nuttley for critical reading of the manuscript and helpful discussion.*

> *We **would like to thank** the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior (CAPES) from Brazil for financial support.*

Interestingly, some organizations were so frequently acknowledged that their names formed parts of lexical bundles:

> *Sequencing of* Candida albicans *was accomplished with the support of the NIDR and **the Burroughs Wellcome Fund**.*

The bundles used to acknowledge individuals or institutions for a gift of experimental material overlapped with the next category, Materials and Methods.

This next category consists of approximately 60 bundles which were used either to identify the provenance of experimental materials (*was obtained from the*); to describe the types of materials used (*on Sabouraud dextrose agar*); or to describe the experimental procedures (*was prepared as described*); or the equipment used (*was performed using the*). Here too some proper names occurred frequently enough, and in sufficiently similar contexts, to be part of a lexical bundle. This example:

> *Laboratory standard strain SC5314 (ATCC MYA-2876, used in the* Candida *Genome Sequencing Project) **was obtained from the American Type Culture Collection**.*

contains several four-word bundles embedded in the eight-word bundle which appears in bold type.

The bundles used to recount research methods make the corpus texts more concise, in that they come not in the context of a detailed description of methodological procedures (which would have caused them to be categorized as 'content' bundles under the criteria applied in this study), but rather as an assurance that the procedures were carried out using established methods, which are described elsewhere.

> *The experiment was **performed according to the** instruction from the manufacturer.*
>
> *The antifungal agents and yeast inocula were prepared **in accordance with the** M27-A recommendations of the National Committee for Clinical Laboratory Standards (NCCLS) [19].*
>
> *When required, proteins were deglycosylated using endoglycosidase H (New England Biolabs, Beverly, Mass.) **in accordance with the** manufacturer's instructions.*

Often these methodological references were, as in the examples above, to professional bodies and standards, or the manufacturer of laboratory equipment.

Others, however, made reference to a method described in a scholarly publication, thus overlapping with the final subcategory of 'external reference':

> $[Cu(phendio)_3](ClO_4)_2 \bullet 4H_2O$ and $[Ag(phendio)_2] \ ClO_4$ were synthesized **in accordance with the** procedures outlined in McCann et al. (2003).

This final subcategory was labelled 'literature'. This category included some bundles used for generalization about the topic (*is well known that, are known to be*), and others used for describing a gap in the literature (*little is known about, best of our knowledge*). The largest group consisted of reporting verbs: *been suggested to be, has been found to, been reported to be.* In this group, too, some proper nouns occurred frequently enough to form part of lexical bundles. In this example, three four-word bundles reside in the six-word bundle in bold type.

> *Standard molecular biological techniques for plasmid isolation, restriction enzyme analysis, PCR amplification and* E. coli *transformation were used* **as described by Sambrook et al.** *(1989)*

This group of reporting forms, obviously, lent itself to the mention of specific published works, as in these examples:

> Candida albicans *virulence and pathogenicity is complex and it* **is believed to be** *correlated to different virulence factors [6].*
>
> *If flanking nucleotides are also included, this sequence almost perfectly matches the consensus TAAATAAA(G/A), which* **is believed to be** *important for transcription termination in yeast genes [1].*

Interestingly, though, writers sometimes took advantage of the opportunity created by the use of the passive to leave agency unspecified. In cases such as these, the external reference is an implied one, to the others in the field who believe the proposition that is asserted to 'be believed'.

> *Because of these important roles, the enzyme* **is believed to be** *encoded by all species.*
>
> *Conversion of* Candida albicans *from yeast to mycelial growth* **is believed to be** *associated with the organism's virulence.*

Another interesting feature of the reporting verbs is that there are in fact rather few of them, as indicated in Table 5.2, which shows all the reporting verbs present in lexical bundles used for external reference. This list rather generously includes some research process verbs that most frequently co-occur with a citation, but which could arguably be classed as content bundles, as well as some which are used only occasionally for reporting and more frequently for other purposes. Nevertheless, only 19 different verbs are present.

Table 5.2 Reporting verbs in lexical bundles

Reporting verbs	Examples of bundles
Be believed to be	is believed to be
Report	has also been reported
	was reported to be
Be in agreement with	is in agreement with
	results are in agreement
Be known to be	are known to be
	is known to be
Be similar to	are similar to those
	is similar to that
Be in contrast to	is in contrast to
Find	have been found to
	has been found to
Show	been shown that in
	has recently been shown
Suggest	been suggested that the
	been suggested to be
Use	been used as a
	been widely used to
Demonstrate	demonstrated the presence of
	it has been demonstrated
Identify	has been identified as
	have been identified as
Be associated with	have been associated with
Describe	have been described in
	described by Sambrook et
Consider	is considered to be
Be consistent with	which is consistent with
Isolate	have been isolated from
Be thought to be	is thought to be
Define	is defined as the

4   Discussion and Conclusion

This investigation of a topic-specific corpus produced results which would have been difficult to predict based on the findings of earlier studies of lexical bundles in academic writing. While lexical bundles are common in all kinds of discourse, they occurred in this corpus at a frequency much higher than that found in earlier studies. In addition, quite long bundles were identified at lower frequencies.

As was noted above, the presence of longer bundles goes some way to explaining the frequency of shorter ones, since longer bundles contain shorter ones. In and of itself, though, that is unlikely to account for the high frequency of shorter bundles. However, the *reason* for the existence of longer bundles may help explain the frequency question as well. This can be illustrated by looking

at the longest bundle in the corpus.

   C. albicans *strains used in this study are listed in table 1*

It is clear that this particular string is unlikely to occur, or at least unlikely to occur sufficiently frequently to qualify as a lexical bundle, in more general corpora of academic writing, or indeed of biology writing; it is likely to occur with that frequency only in a corpus with a focus on *Candida albicans.* The same point is true for many of the other bundles, most notably those that fell into the category 'content'. While not all name *Candida albicans*, many are unmistakably related to some element – organisms, experimental procedures, implications – with a direct connection to the nature of this research topic, as the following examples illustrate:

washed three times in
to buccal epithelial cells
role in the pathogenesis
of the fungal cell
Materials and methods Organisms

The frequency of lexical bundles appears to be due, at least in part, to the composition of the corpus, and specifically to the fact that the texts of which it consisted were on the same, narrow topic.

   The fact that the topic led to some multi-word units being used frequently has two possible explanations. One is that these lexical bundles may reveal something about the conventional discoursal moves in this field (cf. Flowerdew & Forest, this volume), or that these moves are often realized with a small range of common forms, or both. That is, lexical bundles which are involved in acknowledging funding, like

This work was supported in part by
This work was partially supported by

may be common in this corpus because acknowledging external funding is a conventional move in research articles in experimental biology. (That, in turn, is due to the fact that biologists are more likely to obtain external financial support for their research than, say, linguists.) However, there are, in principle, many ways in which funding could be acknowledged. It is therefore noteworthy that a small range of forms were used frequently enough to qualify as a lexical bundle. It is possible that conventional forms of expression, as opposed to more original ones, are common among this group of writers.

   The other possible explanation is the relationship between topic and lexis. It is only to be expected that texts on the same topic share a higher

proportion of lexical items than unrelated texts; it is therefore not surprising if they also share a store of multi-word lexical units. The fact that such a high proportion of the bundles related to the scientific content of the articles tends to suggest that this is at least a contributing factor. While there are many ways of acknowledging assistance, it is presumably more difficult to paraphrase

> *The plates were incubated at X°C for Y days.*

A group of papers which have the same need to express that idea are bound to do so in rather similar language.

Regardless of how the phenomenon is explained, though, the fact that lexical bundles were so common in the corpus has implications both for researchers and for classroom practitioners. For the language teacher, the implications are primarily about what students need to learn. At the undergraduate level there is scope for teaching about the forms in broad use in academic discourse. However, those who come to the writing classroom as postgraduates, or as researchers in later stages of their careers, are primarily concerned with learning to present their research findings, which are on a specific topic. The community of readers who will judge their work is, therefore, rather narrowly constructed. For a student working within the subject area studied here, it would arguably be more useful to teach a small range of reporting forms as well as bundles used for functions such as acknowledging the gift of laboratory samples, than to teach a wider range of reporting forms. Whether resources are available to tailor EAP instruction and materials to the individual areas of course participants is an important question, but the evidence of this study is that it might well be beneficial.

From a corpus perspective, this study illustrates the importance of the question of how to slice the pie. Significant differences have emerged from this very specialized corpus with respect to earlier, broader corpora of academic writing. This suggests that the decision of how broadly or narrowly to define an area when constructing a corpus is an important one, and one that should be made consciously.

Finally, an additional point deserves consideration. As has already been suggested, one of the reasons why lexical bundles are of practical interest is that they are part of fluent, native-like expression, and are therefore an important aspect of language learning. Not all language learning happens in the classroom, though; an important source of learning for novice writers is their reading of the texts in their fields. From this perspective it is worth wondering whether all frequently occurring forms are equally appropriate, or equally appropriate in all contexts. Although it fell outside the scope of this project to investigate this question, on casual observation some bundles appeared to be used in a questionable way. An example is *performed according to the*, which

occurred 12 times, usually in contexts like these:

> *Restriction-enzyme digestions and DNA ligations were **performed according to the** recommendations of the manufacturers.*
> *E test was **performed according to the** manufacturer's instructions [23], using RPMI 1640 agar 1.5% (Sigma) with 2% glucose.*

However, in one case, the use of 'instruction' in the singular gives the sentence an unidiomatic feel:

> *The experiment was **performed according to the** instruction from the manufacturer.*

Although *according to the instructions* is a bundle in its own right, there is only one other occurrence of '*according to the instruction*', singular, and that comes in the same article, and gives an equally unidiomatic impression:

> *The experiment was conducted according to the instruction supplied by the manufacturer.*

It is possible that the authors of this paper (who, based on their names and institutional affiliations, do not have English as their first language), were aware of *performed according to the* as a familiar unit, but were not able to complete it entirely successfully. If so, and if this is an issue for other English L2 and/or novice academic writers, then the question of what constitutes successful use of documented lexical bundles also deserves attention. This is a question for future study, and one which must be investigated at the intersection between corpus and discourse.

Notes

[1] It should be noted that Cortes's (2004) student corpus included assessment writing, and thus the different frequency of lexical bundles in the two corpora may be partially explainable by generic differences.

[2] I gratefully acknowledge Springer Verlag for kind permission to include these articles in my research corpus.

[3] 'Approximately', here and elsewhere, because of the point made above: some bundles could arguably be classed as belonging to more than one group.

References

Anthony, L. (2008). *AntConc 3.2.2* [Computer software]. Retrieved from www.antlab.sci.waseda.ac.jp/antconc_index.html

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines.* Milton Keynes: Society for Research into Higher Education and Open University Press.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 263–286.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English.* Harlow: Longman.

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*, 310–331.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 397–423.

DeCarrico. J., & Nattinger, J.R. (1988). Lexical phrases for the comprehension of academic lectures. *English for Specific Purposes, 7*, 91–102.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*, 29–62.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4–21.

MacDonald, S.P. (1992). A method for analyzing sentence-level differences in disciplinary knowledge making. *Written Communication, 9*, 533–569.

Oakey, D. (2002). Formulaic language in English academic writing: A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines. In R. Reppen (Ed.), *Using corpora to explore linguistic variation* (pp. 111–129). Philadelphia: Benjamins.

Samraj, B. (2004). Discourse features of the student-produced academic research paper: Variations across disciplinary courses. *Journal of English for Academic Purposes, 3*, 5–22.

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing and use.* Amsterdam: Benjamins.

Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly, 37*, 419–440.

Swales, J.M. (1998). *Other floors, other voices: A textography of a small university building.* Mahwah, NJ: Erlbaum.

Part II

Focus on Interpersonal Discourses

This page intentionally left blank

Introduction to Part II

Susan Hunston

The four chapters in this section share a concern with discourse as an inter-action between writer and reader. In addition, each of them takes the writing of novice academics – undergraduate or postgraduate students – as its focus, and thereby engages with the question of how students negotiate their academic status and persona. The chapters also demonstrate the variety of approaches that are used to interpret the discourse–corpus interface. As a result the question of methodology can be regarded as one to which a definitive answer is not (yet) available and as one which continues to generate fruitful debate and experimentation. Each chapter finds its own balance between treating each instance of a word or phrase in context as unique in meaning / function and finding commonalities within a substantial body of data.

Ken Hyland's chapter 'Corpus informed discourse analysis: the case of academic engagement' asks how writers construct their readers, respond to their expectations as members of the discourse community, and attempt to guide their reactions to the work being presented. As a way of researching this very discourse-oriented question, Hyland identifies and tests through examining examples in context, a large set of language features which realize the processes of engagement and which are open to quantification. He is then able to ask a further question: using the evidence of student writers, what is the role of engagement in different disciplines? Hyland widens his investigation in another dimension, by including data from interviews with student writers. He therefore not only knows what the students wrote but also has information about their assumptions and attitudes, especially those relating to what is permissible in a document which a student writes for assessment. This chapter reflects a common theme in this section, and indeed this book, in that the research moves from a reading of texts (to identify functions and wording) to a quantitative investigation of texts in corpora (to establish quantities) and back to the effectiveness and character of individual texts.

The next chapter, by Ann Hewings, Caroline Coffin and Sarah North, interprets 'interaction' somewhat differently, in that the texts it investigates comprise an e-conferencing discussion among students studying Health and Social Care. Writers are therefore interacting with a known, and virtually present, set of readers. Hewings et al. carry out two disparate but interleaving forms of analysis. One is a move analysis that treats each constituent text as a

whole within a sequence. The second is a corpus analysis that disregards the individuality of the text, though it does respect move boundaries. The corpus analysis is based on the identification of keywords and their collocates and the inspection of concordance lines. The keywords turn out to be parts of personal phrases such as *I think*, and these in turn are found to be most frequent in those moves that constitute the discussion in the e-conference. The role of the personal in negotiating knowledge is thereby demonstrated.

Like Hyland's chapter, the chapter by Maggie Charles uses corpus investigative techniques applied to work by students (postgraduates in this case) in contrasting disciplines. The approach is somewhat different, however, in that she begins with just four adverbs: *only, just, simply* and *merely* and investigates each in considerable detail. She argues that these adverbs construct the writer's stance and play an important role in construing consensus between writer and reader. This argument is very much in line with the view that evaluative meaning in academic discourse is often expressed in an implicit and subtle manner, making its identification difficult. Charles goes further than this, however, and widens the context beyond the word and its immediate collocates. She establishes typical rhetorical functions for the clauses in which the target words appear, namely, in adversative clauses, and in expressions of cause and effect. Investigating the latter, Charles links the adverbs with extended examples of discourse patterning. Charles's approach to the data, is very much 'bottom-up', but this does not prevent her from making statements about the ways that discourse is organized or, indeed, about the preferred discourse strategies in different disciplinary areas.

In the final chapter in this section, Ramona Tang takes a 'discourse', rather than a 'corpus' approach to her texts. That is, she treats each text as a differentiated whole and takes individual segments of text as her starting point rather than words or phrases in context. In this way her approach is a striking contrast to that used by Charles. Her analysis is based on the model of Engagement developed by White (2003; see also Martin & White, 2005). This model avoids linking conceptual categories, such as 'postulate' or 'acknowledge' with any defined set of language features, thereby making corpus searches less applicable. Tang's topic is the degree of authority demonstrated by various student writers in their essays, and she argues persuasively that the combination of particular categories in the Engagement model leads to the positive perception of some writers as 'more authoritative' than others.

Each of the writers in this part proposes their own response to the potential disparity between corpus and discourse approaches. While 'top-down' discourse methods might be argued to pay too little attention to the significance of individual phraseologies, and to generalize from too few examples, and 'bottom-up' corpus methods might be accused of according too much significance to that which can be counted, the writers in this part all find their own way of combining the best of both worlds.

It is interesting that all these writers, with their focus on the interactive, take student writing as their data. (Or, that writers who are concerned with student writing arrive at a focus on the interactive.) In different ways, all the papers draw attention to the student's dilemma: how to find a voice that establishes the writer as a member of an academic community while negotiating the multiple relationships – student and teacher, fellow-student, student and examiner – involved. That this dilemma leads to a particularly rich and subtle use of the interpersonal, one that can be investigated in ways that marry the general and the specific, is not surprising.

References

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

White, P. (2003). Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text, 23*, 259–284.

Chapter 6

Corpus Informed Discourse Analysis: The Case of Academic Engagement

Ken Hyland

1   Discourse Analysis, Corpora and Academic Writing

Corpora and discourse approaches are perfect bedfellows. While they are two aspects of applied linguistics which have not always had a lot to do with one another, they are increasingly seen as complementary approaches which can inform and enrich each other, thereby leading to more insightful analyses of language use.

Essentially, corpora bring evidence of typical patterning to discourse studies, providing language data which represent a speaker's experience of language in a restricted domain. In other words, corpus analysis is a method which moves away from individual preferences to focus on community practices, dematerializing texts and approaching them as a package of specific linguistic features employed by a group of users. As a result, it is particularly valuable in research into academic discourse (e.g. Biber, 2006; Hyland, 2004; Swales, 2004), providing insights into language variation across disciplines, genres and languages. Corpus studies also require a focus on 'action' to balance the focus on 'language', however, and this necessitates rematerializing these features to understand how and why writers make the choices they do when they write. This kind of qualitative interpretation is greatly assisted by interviews with text users, grounding patterns of text meanings in the conscious choices of writers and readers.

This chapter offers one illustration of how corpus analysis can contribute to our understanding of academic discourse, exploring the ways that writers seek to explicitly establish the presence of their readers in their writing. Drawing on a corpus of reports written by Hong Kong undergraduates together with transcripts of student interviews, I show how corpus analysis can reveal aspects of audience engagement that are not possible with other methods.

2   Interaction and Engagement

While academic writing is often distinguished by its apparent absence of explicit appraisal and attitude, it is nevertheless clearly structured to evoke affinity

and engagement (Hyland, 2004; Swales, 2004). In presenting their work, writers must also adopt interactional and evaluative positions, anticipating readers' expectations and responses to participate in what amounts to a virtual dialogue with them. To view writing as interactive, then, means examining discourse in terms of the writer's projection of the perceptions, interests and needs of a potential audience. Discourse analysts have sought to elaborate the ways by which such interpersonal meanings are expressed, describing these linguistic resources as *evaluation* (Hunston & Thompson, 2000), *appraisal* (Martin, 2001; Martin & White, 2005), *stance* (Biber & Finegan, 1989; Hyland, 1999) and *metadiscourse* (Ädel, 2006; Hyland, 2005).

These concepts, however, are largely concerned with *writer-oriented* features of language, what has been called *stance,* or the ways that writers explicitly intrude into the discourse to stamp their personal authority onto their arguments. Stance is a writer's community-recognized persona as expressed through his or her rhetorical choices, conveying epistemic and affective judgements, opinions and degrees of commitment to what they say. It therefore suggests something of how authors construct a credible academic identity. In this chapter I am interested in elaborating the flipside of the interactional coin: *engagement,* or the ways writers pull readers along with the unfolding discourse: recognizing their uncertainties, including them as discourse participants and guiding them to interpretations (Hyland, 2001, 2005).

This use of 'engagement' has been developed independently of that proposed by Martin (2000) and Martin and White (2005) who use the term to refer to locutions writers use to position themselves to other voices. For them it refers to the resources for conceding, averring, attributing, hedging, boosting and otherwise modalizing the status of an utterance. While there is a great deal of overlap between this view and my own, Martin and White are largely concerned with representing the writer's attitude or opinion towards the propositions he or she is setting out and so it is closer to what I have here called 'stance'. In contrast, I am concerned with exploring the ways that language is used to anticipate possible reader objections, acknowledge their interpersonal concerns and bring readers into a text. Engagement in this chapter therefore refers to the ways writers explicitly mark the presence of what Thompson (2001) calls the 'reader-in-the-text', and the most obvious indication of this dialogic awareness occurs where the writer overtly refers to readers by asking questions, making suggestions and addressing them directly (Hyland, 2001).

In contrast to Martin and White then, I am not concerned with providing a comprehensive description of the resources available in English to express appraisal, but with exploring some of the ways that argument differs across disciplinary contexts. The analysis therefore focuses on mapping and, to some extent attempting to explain, the use of language in particular circumstances of use, depicting what is usual in those contexts rather than just what is grammatically possible. By identifying these features I hope to show something of

how writers construe and position their readers, suggesting one way in which language choices are related to the ways disciplines conduct research and negotiate the construction of knowledge.

Since engagement is a writer's choice to introduce readers as real players in the discourse, rather than merely as implied observers of the discussion, the points at which this occurs must be explicitly marked. Devices signalling reader engagement have been proposed in the literature (e.g. Ädel, 2006; Biber, 2006; Bondi, 1999; Hyland, 2004, 2005) and I illustrate these below with examples from a corpus of research articles. Some of these features are very easily identified through a corpus word-search; others are much less easily recognized and involve careful reading of individual texts. They are as follows: (1) interrogatives; (2) inclusive first person pronouns, second person pronouns and items referring to readers; (3) directives, including imperatives, obligation modals referring to actions of the reader (*must, ought, should, have to, need to*), and adjectival predicate controlling a *to-* clause directing readers to a particular action; (4) references to shared knowledge; and (5) asides addressed to the reader.

Example 1

But what is the physical nature of the Schwarzschild horizon? What is going on there? Is it a real singularitity of spacetime or merely an artifact of the chosen coordinate system? (Philosophy)

Example 2

In this extract <u>we</u> can note how the lecturer stresses how he is trying to make things simple. (Applied Linguistics)

Cluster analysis makes no prior assumptions about important differences within a population and it also allows <u>you</u> to define groups by all the variables involved. (Business Studies)

<u>Some readers</u> will want to argue that this is a comparative analysis of neighborhood associations more than social movements. (Sociology)

Example 3

<u>See</u> Smith and Smith (1996) for a critical discussion of this issue. (Biology)

The temperature of the transistor <u>must</u> be accurately determined and maintained during the duration of the measurement. (Physics)

However, <u>it is important to note that</u> our discussion is not intended to reflect how strongly these feelings are held. (Business Studies)

Example 4

<u>Obviously</u>, in all these cases, a gyrotropic medium, instead of the chiral medium would render a vanishing total rotation, like in Figure 13. (Physics)

Example 5
And – <u>as I believe many TESOL professionals will readily acknowledge</u> – critical thinking has now begun to make its mark, particularly in the area of L2 composition. (Applied Linguistics)

In these examples, engagement works both to meet readers' expectations of inclusion and to rhetorically position them by capturing their attention and focusing them on key issues. Basically, engagement features are important as they offer insights into writers' perceptions of audience and the ways that the current text is aligned with other texts (Bakhtin, 1986). In other words, writers address an audience by drawing on their knowledge of earlier texts to shape their writing so that readers will recognize intertextuality between texts.

3   Methods and Corpora

This study is based on an analysis of these dialogic features in a corpus of 64 project reports written by final year (Year three) Hong Kong undergraduates and interviews with students in eight fields. The final year report is a major assessment genre in many universities around the world and typically represents a year's supervised research work and a sustained piece of writing of between 8,000 and 13,000 words. Reports follow guidelines based on the research paper formats of the discipline and are assessed by two examiners in terms of how well students meet the objectives of the project and on the quality of the written work. This, then, is a high stakes genre for students and is by far the most substantial and sustained piece of writing that they will do in their undergraduate careers.

Reports were collected from a cross section of disciplines: biology (Bio), mechanical engineering (ME), information systems (IS), business studies (BS), TESL, economics (Econ), public administration (PA) and social sciences (SS). These reports produced a corpus of 630,000 words. The corpus was searched for over 100 items from the five categories listed above which potentially initiate writer–reader dialogues using *WordPilot 2000,* a commercially available concordancer. Each example was then examined in its individual concordance line to ensure it functioned as an engagement device and eliminated from the analysis if it did not. The text data were supplemented with interviews conducted with supervisors and focus group data from student writers. The results are discussed below.

4   General Patterns of Engagement in Reports

Corpus studies are based on both qualitative and quantitative methods, using evidence of frequency and association as starting points for interpretation.

Frequency is a key idea for if a word, string or grammatical pattern occurs regularly or behaves differently in a genre, then it can be taken to be significant in how that genre is routinely constructed. The frequency of dialogic features in this corpus reveals the extent to which students feel able to use these features to engage their readers. The target features occurred about 24 times in each report, about one every two pages or so, with inclusive first person pronouns and directives being the most common devices overall and amounting to about two thirds of all devices. The results also show some interesting cross-discipline comparisons. Directives were most heavily used by students in IS and ME, for example, while questions and inclusive pronouns were most frequent in the more discursive soft fields. Table 6.1 shows the use of these features per 10,000 words across the disciplines of the corpus.

These frequencies indicate that the students are aware that academic writing is not altogether impersonal, but they are only half as frequent as in academic research articles (Hyland, 2001). This reveals how writers' choices reflect wider social and discursive practices which carry assumptions about participant interactions and how these should be structured and negotiated. Writer–reader relationships in research papers, for instance, are ostensibly egalitarian and effective interaction involves addressing readers as if social distinctions of power, status and standing do not exist, or at least are irrelevant to how the paper will be received. The final year report, in contrast, carries a heavy assessment burden and involves writers demonstrating an appropriate degree of intellectual autonomy while recognizing readers' greater knowledge of the field.

The tendency to acknowledge the reader's authority is exacerbated in this context, where the writers are second language students from a culture which tends to place a certain emphasis on respect for authority and the importance of face (Scollon & Scollon, 1995). Culture intrudes into our communicative practices in significant ways, and undergraduates familiar with a different writing tradition and conception of teacher status have little incentive to challenge the authority of reader/examiners, particularly as the judgements of these readers have material

Table 6.1   Frequency of engagement features in student reports (per 10,000 words)

Discipline	Questions	Reader pronouns	Directives	Shared knowledge	Asides	Totals
Info Systems	2.2	5.7	24.5	3.5	0.0	35.9
Mechanical Eng	3.0	3.6	23.7	4.7	0.0	35.0
Social Sciences	8.8	6.3	7.7	0.3	0.2	23.5
Public Admin	6.0	10.9	3.3	2.0	0.7	23.0
TESL	6.7	3.3	9.2	2.8	0.0	22.0
Biology	1.0	5.3	11.9	1.7	0.3	20.1
Economics	1.5	3.1	8.9	3.8	1.0	18.3
Marketing	1.1	6.0	3.7	2.2	0.2	13.3
Overall	4.3	6.1	10.6	2.5	0.3	23.9

consequences. So, while writers can always resist the relationships implied in a genre, awareness of audience is typically revealed in rhetorical choices that recognize the reader's authority. But while this institutional context helps structure writer–reader relations in this genre, a number of supervisors still expressed frustration at students' reluctance to use more engaging language:

> *I sometimes feel they could do more to argue with me and get me involved instead of mechanically trotting out the results. It might be something cultural I suppose, but I get no sense of real commitment to what they are doing or in getting me onside with it all.* (IS supervisor)

> *It is not just a question of content, getting the ideas or the findings down, but how they manage this. I get an impression of the writer when I read these reports, and often my impression is that they are trying to hide themselves and talk like a book instead of getting me excited and curious about what they've done.* (BS supervisor)

In the following sections I elaborate some of the ways that the most frequent engagement features are used and how they are understood by participants.

4.1   Reader pronouns: soliciting solidarity

Direct reference to the reader with personal pronouns or other devices can both acknowledge and claim disciplinary affinity with an active audience, but students significantly underuse these compared with expert academic writers. Second person pronouns are the clearest textual acknowledgement of the reader but occur only rarely in academic writing. This is, perhaps, because *you* and *your* are mainly found in more informal and interpersonal registers, being 25 times more common in casual conversation than in academic writing, for instance (Biber et al., 1999: 334). Their relative scarcity in research discourse means that students are often taught to avoid them and my interviewees had little difficulty in giving reasons for this:

> *Science writing is neutral. I know my supervisor will read my project but I cannot talk to him like in the tutorial. I must just put down the facts without personal idea, just show that I understand the books and that I follow the method.* (Bio student)

Where this pattern is used, then, it has a less personal reference closer to the indefinite pronoun *one* and referring to people in general rather than specific discourse participants:

Example 6
The CATV game distribution business allows <u>you</u> to enjoy game software while staying at home. (BS)

Whenever <u>you</u> run Windows or any Windows application, <u>you</u> are seeing the API in action. (IS)

Here, *you* seeks to engage readers through shared experience rather than direct personal interaction, and this also helps account for the use of the inclusive pronouns *we, us, our* and *ours*, which occurred over ten times more frequently than second person in the corpus. The fact that direct references to the reader were five times more frequent in the published articles, however (Hyland, 2001), suggests something of the institutional and intertextual constraints which underpin relationships in this genre. The undergraduate project report involves a novice communicating with an expert for an assessment purpose and there is little room here for the illusion of equality cultivated in research articles. The reports challenge students to demonstrate an appropriate degree of rhetorical sophistication while recognizing readers' greater status and disciplinary knowledge. This is clear in the comments of these students:

> *My report is a long assignment and I am writing to show what I know. My writing has to be accurate and not the same as talking in a tutorial. I can't use 'we' or 'you' as my supervisor might not agree what I think is true. I might be wrong.* (SS student)

> *I must be careful when I write. I don't want to make myself important. Of course it is my project and my result, but I am just ordinary student. Not an academic scholar with lots of knowledge and confident for myself.* (TESOL student)

As a result, students not only used inclusive pronouns far less often than expert writers, but also used them in different ways.

Only rarely, for example, did students dare to use them, as experts often do, to position the reader by guiding them to a preferred interpretation, as in these examples:

Example 7
If <u>we</u> agree that reproductive rights can promote happiness to the human well being, then <u>we</u> should determine the standard of reproductive rights in order to minimize the harm that bring up from the use of HRT. (PA)

<u>We</u> can conclude that the cadmium chloride solution will greatly affect the eye size, embryos length, heart beat, activity and the number of somite of the embryos when the embryo was malformed. (Bio)

This is a highly sophisticated use of reader pronouns and a tricky one to get right as it involves assuming the perspective of the reader to head off objections and secure agreement with the writer's position. It is a use which attempts to

weave the potential point of view of the audience into the argument and suggest what any reasonable, educated person might think. It therefore addresses the reader from a position of confidence as it allows the writer to take responsibility for leading the reader's thinking. Many students are understandably reluctant to assert this kind of control.

So, rather than claim an uncertain equality, students largely rejected an authorial role which directs their reader-assessors, for one which simply shortened the distance between them. More often, writers chose to draw on everyday knowledge and principles of reasonableness rather than appeal to specialist understandings and ideas. These examples are typical:

Example 8

In Hong Kong, because of the late development of representative democracy, we only have a very short history of election. (PA)

Such advancement is gradually changing our lifestyles, work habits, social structure and ultimately the whole world. (Econ)

In other words, these Hong Kong undergraduates shunned uses which might presuppose a risky egalitarianism with more knowledgeable readers to appeal to a solidarity based on less specialized forms of knowledge.

4.2   Questions: constructing involvement

Academic research is done with questions in mind or problems to solve and these occasionally emerge as explicit interrogatives. Questions are explicit engagement features as they invite collusion with readers: addressing them as someone with an interest in the problem posed by the question, the ability to recognize the value of asking it and the good sense to follow the writer's response to it. Largely confined to the soft disciplines in research articles, questions serve up an invitation for readers to respond, to orientate themselves in a certain way to the argument presented and to enter a discourse arena where they can be led to the writer's viewpoint (Hyland, 2002a; Webber, 1994).

So while questions seek to involve readers in the argument, they also construct unequal social relationships. Questions convey authority along with familiarity, implying an imbalance of knowledge and suggesting that the writer is in full control of both his or her material and audience. This explains their popularity with textbook authors, but not all genres confer such rights on writers. This is clear in the comparisons shown in Table 6.2 between the student reports, research articles and chapters from 56 textbooks in eight disciplines (Hyland, 2002a).

Clearly, an overt display of authority may have advantages for textbook authors, whose efforts to directly engage readers can also mark expertise and

Table 6.2   Genre variation in use and forms of questions

	Totals (per 10,000 words)	Wh forms (%)	Yes/No (%)	Alternative (%)
Textbooks	13.3	73.9	24.2	1.9
Research articles	5.5	77.8	21.0	1.2
Student reports	4.3	57.6	42.1	0.3
Overall	7.2	69.8	29.1	1.1

Source: Hyland, K. (2002a). What do they mean? Questions in academic writing. *Text, 22*, 529–557.

distinguish knower from novice, but it might be less effective when addressing one's supervisor. A biology student commented on this:

> *I never think to ask a question in the report. How can I ask a question in my report? Teachers ask questions and I am answering the questions. I think my supervisor doesn't want me to ask him questions but to answer questions. I leave it out.* (Bio student)

This sophisticated rhetorical awareness is reflected in the overall figures, where students tended to reject the strategies of the writers they encountered in their textbooks to produce frequencies similar to those in the research articles they were given to read. But despite the similarities in overall frequencies, they tended to use different forms of questions and employed them in different ways. Questions in the articles functioned to establish a research niche, convey a claim forcefully, express an evaluation and suggest further research. The students, in contrast, were more cautious, using far more yes/no questions ('Do students like using dictionaries?') and avoiding the more complex alternative forms ('Is government control of the internet adequate or do young people need more protection?').

The preponderance of polar forms reflects students' apparent preference for using questions to organize their discourse, with almost 70 per cent of all questions in the report corpus framing the text and guiding the reader through it. The students typically used the introduction to set out their research questions, with over 80 per cent of the texts containing questions to focus the paper, and while some writers sought to engage the reader's interest (Example 9), many simply produced an agenda for the discussion (Example 10):

Example 9
Is this pattern more obvious in Chinese dating couples? Do men and women behave differently in handling conflicts? How does Chinese culture influence this sex differences in handling the conflicts? Will the high-context

society, like Chinese, with well-defined gender roles in dating and marriage, affect the conflict-management skills of couples? Would the women still use accommodation in handling the conflicts? (SS)

Example 10
The following research questions are addressed in this study:

1. Is there any correlation among the triarchic abilities?
2. Is there any gender difference in the triarchic abilities?
3. Does any particular type of intelligence predict academic performance?
4. Do students with extra-curricular activities have higher scores in any particular type of ability than those without responsibilities? (TEFL)

To some extent all questions in academic prose represent the writer's awareness of audience and we can see here that the students were clearly taking the trouble to direct the reader in one direction rather than another. The students, however, used a much narrower range of functions with their questions than the experts. These textual choices suggest that their priority was to manage the structure of the argument and the flow of information rather than grab the reader's interest and engage them personally in the unfolding text. This is clear from the fact that many students did little more than recycle their research questions as section headings:

Example 11
Chapter 2: How does metal fatigue? (ME)
1.3 Why choose cadmium to study? (Bio)
Why use Windows environment? (IS)
Introduction: I. What are Wholesale Travel Agents? (BS)

So while punctuating a text with questioning sub-heads in this way is a recognized readability strategy, identifying what is to follow, text organization is far less adventurous than the expert uses.

Where students did employ in-text questions to engage readers with their argument, they employed expert-like practices in following them up with answers. That is, while questions appeared to invite the reader into the discourse, they actually anticipated the writer's own response, often posing a question which receives an immediate reply:

Example 12
Why only focus on tour operators? The reason is that only the tour operators can perform all the mentioned behaviours. (BS)

If one does think that pornography implies, woman as a class, enjoys being degraded, is it appropriate for us to say that this thought is derived from the

misinterpretation of the readers? The answer is YES! The readers exagger-
ate the notion of degradation existing in pornography. (PS)

What are mechanical properties? Fundamental mechanical properties are
strength, stiffness, elasticity, and plasticity. (ME)

Questions can therefore be an important resource as they bring readers
into the argument as participants, introducing an interactive dimension which
acknowledges readers' concerns, helps guide them though a text, and works
to position them in relation to the writer's claims. But while a question always
adds an interpersonal dimension, these students generally chose to ignore the
full range of engagement functions which questions offer. This suggests a clear
generic difference compared with their use in research articles and doctoral
theses (e.g. Hyland, 2002a).

4.3   Directives: managing readers

Directives were the most frequent devices used to initiate reader participation
in the student texts, comprising 45 per cent of all features. They represent an
explicit recognition of the dialogic dimension of argument as writers inter-
vene to direct the reader to some action or understanding, as in these exam-
ples from the student reports:

Example 13
It is important to note that the process of getting meaning is not so simple.
(TESOL)

Please refer to the result table below. (IS)

It is necessary to understand the initial vibration analysis and pick out the
points at which maximum vibration were obtained. (ME)

A directive utterance is therefore one which expresses an obligation on the
reader to do or not to do something and as such represents a risky strategy for
student writers. They convey a very definite attitude to the reader and so have
the potential to seriously affect the writer–reader relationship, claiming an
authority which many of these L2 students did not wish to display:

These words are too strong. It is like a demand and I cannot demand my supervisor to
agree with me. (Econ student)

I've seen them in the Readings but I don't use them. It is a command to the professor.
Is that OK to do that? I don't think so. (IS student)

Since reports are written primarily to gain credit for a research project, the
decision to use a directive might seem a perilous strategy, and perhaps this is

the main reason why the student reports contained only about half the number of directives found in the articles (per 10,000 words).

But while directives are often seen as constructing status or power differences in interaction, it is simplistic to see them as simply bossing the reader. In fact, we can classify directives according to three main forms of activity which they direct readers to engage in: *textual acts*, referring them to another part of the text or to another text; *physical acts*, instructing them to perform a real world action; or *cognitive acts*, steering them to certain interpretations (Hyland, 2002b). These different functions are reflected in the extent to which directives may impose on the reader and suggest that their distribution will be different across genres (Table 6.3).

The most imposing use of directives involves positioning readers, directing them to some cognitive action by requiring them to note, concede or consider some aspect of an argument. Typically these directives lead readers towards the writer's conclusions by emphasizing what they should attend to in the argument. But while about half the directives in the research papers guided readers to see things from the writer's perspective in this way, cognitive directives comprised only 20 per cent of those in the student reports. This marked reluctance to avoid taking control of the reader's thinking is also underlined by the preference of students to adopt the less threatening forms in this category and to simply lead readers through an exposition:

Example 14

Let's first review some theories of conflict management. (BS)

Suppose that an array containing 49 data is pushed into the sorting function, this function will then shuffle and sort the order of the data from small to large. (ME)

Let the stochastic variable Ut represent deviations of PBt from its long run path P. (Econ)

Related to this expository use, the students employed a similar proportion of textual directives to steer their readers to tables, examples, appendices and other sections of the report to support their argument. Only rarely did they point intertextually to other sources.

Example 15

Refer to the fatigue specimen of fatigue test. (ME)

Table 6.3 Functions of directives by genre (%)

Genre	Textual	Physical	Cognitive	Total
Student reports	21.1	58.3	20.4	100
Research articles	36.3	15.4	48.3	100

This should be compared to the answers to question 3 in the last section. (BS)

. . . See Friedman, M. (1982), pp. 2. Page 21 of 49. (PA)

Using directives in this way is very guarded, avoiding the more imposing cognitive function of telling readers how they should understand an argument. It suggests a sensitivity to the interpersonal impact of rhetorical options.

This interpretation of students' language choices is supported by their overwhelming preference for directives which address real world actions, typically to guide readers through research procedures. These accounted for over half of the forms in the report corpus and were particularly common in the hard science disciplines, perhaps influenced by the traditions of precision, tight space constraints and highly formalized argument structures in these fields:

> **Example 16**
> Then <u>rinse the plastic seaweeds</u> by tap water and <u>pour out the embryos</u> through the fish net. (Bio)
>
> The interfaces of the proposed system <u>should be designed</u> in user-friendly manner with clear instructions provided. (IS)
>
> <u>It is important to use</u> only a Ti-B-N film surface which is smooth otherwise pin-hole-free growth is impossible. (ME)

Once again, this is an extremely cautious use. Telling someone how he or she should navigate a text or carry out an experimental procedure is far less likely to impede their freedom of action and decision-making than directing the way they should follow a line of argument or the significance they should give to a claim. Several student respondents saw this research use as a conventional means of describing procedures with no potentially face-threatening implications:

> *In engineering we must be clear in describing our method so it can be easily followed. If we are direct then it can be done by another person without problems. I am only reporting what I did and how the method needs to be. It is a general procedure.* (ME student)
>
> *Yes, I use 'should' here to show how I tested the programme. It is like this in the textbook, I think. This is how we have to describe our work in the report. It is just normal, saying how anyone can do this not just us.* (IS student)

Consequently, the reports contained a much smaller proportion of the relatively more imposing *uses* of directives, but they also contained fewer of the more imposing *forms*. The modal *must*, for example, is generally regarded as carrying the strongest sense of obligation as it suggests the writer's clear

authority (Leech & Svartvik, 1994: 165). For this reason it is almost always replaced by the less confrontational *should* in both face-to-face contexts and in these reports. Where *must* did occur in student writing, it was relatively infrequent, always expressed in a passive form, and almost entirely restricted to procedural explanations in the science and engineering reports:

Example 17
After the regeneration, the column <u>must</u> be rinsed for excess regenerant. (Bio)

The probe <u>must</u> also be calibrated for the specific material. (ME)

The system <u>must</u> show which stage the user is working with. (IS)

Interviews with students showed that they recognized the personal authority of *must* and several expressed their discomfort at using it:

I try to not use it because it is too strong. It's like telling my supervisor what he must think. Of course he knows more than us so how can we tell him. I never use it. (Econ student)

Did I see this in the textbook? Maybe, but I think it is a very hard word: 'the interface must be like this'; 'the design must do that'. The grammar is OK but it is a hard word. (IS student)

Should is a weaker imperative, conveying something closer to an advisable course of action rather than inescapable obligation (Downing & Locke, 1992; Perkins, 1983), and for this reason was a more frequent directive:

Example 18
The data taking, measuring and balancing <u>should</u> be done simultaneously. (ME)

. . . and debugging <u>should</u> be carried out in the final stages of this process. (IS)

a thin layer of agarose <u>should</u> be added at the bottom of the petri dish. (Bio)

There seems, then, to be an awareness among these writers of the interpersonal implications of directives and of the authority they can lend to an argument and an author. This is not always an authority that students writing for their supervisor-examiners are willing to claim, as we can see from some of the very mitigated directives in some of the examples given above, using 'please', 'it is necessary to', and so on.

4.4   Knowledge appeals: claiming membership

Less imposing than either questions or directives and less directly personal than reader pronouns, is the use of appeals to shared knowledge. These are

fairly common in professional research writing where academics seek to position readers within the apparently naturalized and unproblematic boundaries of disciplinary understandings (Hyland, 2001). Readers can only be brought to agreement with the writer through building on what is already implicitly agreed, and by explicitly referring to this agreement writers construct themselves and their reader as members of the same discipline or academic community.

The students, however, were once again reluctant to employ such direct and explicit calls for the reader to recognize some community-specific perception, with less than half the frequencies of the research article corpus. Some students were bold enough to use the strategy, however, and did so effectively:

Example 19

[I]t is well-known that these shares differ in at least one important way . . . (Econ)

We know that some of the toxicant of zebra fish will exert a similar effect to the similar types of target organisms. (Bio)

A good deal is known about the acute and chronic effect of industrial exposure to chromium and its compounds. (ME)

More usually, however, they avoided claiming disciplinary membership with their examiners and instead drew on wider community understandings. Most examples therefore appealed to understandings which might be expected to be more generally held. Here for instance, reference is made to the 'common knowledge' of educated Hong Kong society:

Example 20

It is commonly known that in Hong Kong secondary schools a 'core' textbook would be chosen for students' use throughout the whole academic year. (TESL)

As we all know, the labour relation is an important concern of Japanese-style management, in relation to lifetime employment, the wage structure, on-the-job training and welfare system. (BS)

This is clearly a less risky option as it allows the writer to attempt to share understandings with the reader and recruit him or her into agreement without claiming equal status as a knower of specialized knowledge.

The most common ways of signaling the reader's presupposed understandings involved using *of course* and *obviously*. Both forms are often regarded as markers of epistemic stance, indicating the writer's certainty of a proposition. This is not always the case, however, as they can also realize engagement

meanings by moving the focus of the discourse away from the writer to shape the understandings of the reader:

Example 21

<u>Obviously</u>, the mean score of the items in the intrinsic orientation group are significantly lower than those in the social/practical group. (PA)

<u>It is obvious that</u> most of the consumers for electronic games are male, as most of the games contain violence and sex. (BS)

<u>Of course</u>, better CPU processor will be able to guarantee a shorter response time and better performance. (IS)

Here, then, we see students using a sophisticated rhetorical strategy to imply that the audience already knows, or will readily accept, the accompanying statement, recruiting them as partners in the argument. In using these forms writers are stepping personally into their texts to address the reader as someone with similar background knowledge and interests, presenting themselves as someone who has something in common with the reader and is able to make the same connections.

4.5   Asides: intimating sharedness

A final engagement strategy I want to briefly mention here is the ways writers address readers directly through asides and interruptions to the ongoing discussion, briefly breaking off the argument to offer a meta-comment on an aspect of what has been said. Like explicit references to shared knowledge, these features seek to forge a momentary alliance with the reader, but they do so by stepping outside the ongoing argument to offer a more personal comment which is often peripheral to rhetorical development, as we can see from these examples:

Example 22

Subjects were interviewed in an informal setting, <u>many of them in the student canteen with a cup of coffee</u>, using a semi-structured approach. (SS)

Xerox Corporation is an multi-national enterprise with its headquarters in Stamford, U.S.A. (<u>its offices in Hong Kong are in central, by the way</u>). It ranked 51st on Fortune 500 in 1996 with revenues for year USD19,521 million. (Mkt)

. . . this is usually (<u>but incorrectly, from my point of view</u>) referred to as . . . (Bio)

These insertions allow the writer to intrude into the text to break off from the argument and offer a comment that contributes more to a writer–reader relationship than to propositional development.

While asides express something of the writer's persona and willingness to take up a commentary role, they can also be seen as an essentially reader-oriented strategy. By turning to the reader in mid-flow, the writer acknowledges and responds to an active audience, often to initiate a brief dialogue that is largely interpersonal. This kind of direct engagement builds a relationship between participants which is not dependent on an assessment of what needs to be made explicit to elaborate a position, anticipate an objection or ease processing constraints. The writer introduces the audience into the text because he or she wants to reinforce the dialogic relationship at that point. It is an intervention simply to connect, to show that they are all, writer and readers alike, engaged in the same game and are in a position to draw on shared understandings, if not of actual content, then at least of what might be considered a relevant aside.

It has to be admitted, however, that this kind of engagement is rare in undergraduate reports, and is relatively uncommon in research articles (Hyland, 2001). It is also mainly restricted to a few soft knowledge disciplines, principally economics and public administration in this genre, as writers seek to solicit reader collusion in negotiating agreement on interpretations. As we have noted in the use of other engagement forms, this reflects the greater visibility of discourse participants and the creation of a dialogue in an explicitly interpretative framework. The relative infrequency in these reports, however, is perhaps due to the evaluative nature of the genre and the risks of getting too familiar with a reader whose judgements can have serious consequences for the writer.

5   Conclusion

Engagement is a crucial element of most types of argument as we need to encourage our audience to at least continue reading, if not accept what we have to say. Taken together the different features I have discussed here are important ways of situating academic arguments in the social interactions which occur between writers and readers. Through analysis of directives, personal pronouns, interjections, questions and so on, we can recover something of how writers construct their readers by drawing them into both a dialogue and a relationship. Of importance in this paper is the role that corpus analyses can play in uncovering these discoursal features and contributing to our understanding of them. My main argument has been that by abstracting away from any specific writer to examine recurring features in a large number of texts, we can infer more subtle relations between writers and readers and between linguistic choices and contexts than is possible through the intensive study of a few texts or of writers in the act of writing.

While a corpus approach may lose something in human richness and omit the local, contingent factors which can influence particular cases of writing, it nevertheless offers a powerful description of community practices. Essentially it brings a distributional perspective to language study to suggest something of the extent of variation and similarity in texts and of the complex interactions that occur in writing. Understanding writing involves both looking at what people write and how they comprehend what they write in order to see it as a socially situated practice. Corpus analyses are a key tool in uncovering such practices.

References

Ädel, A. (2006). *Metadiscourse in L1 and L2 English.* Amsterdam: Benjamins.

Bakhtin, M. (1986). *Speech genres and other late essays.* Austin: University of Texas Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* Amsterdam: Benjamins.

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text, 9,* 93–124.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow: Longman.

Bondi, M. (1999). *English across genres: Language variation in the discourse of economics.* Modena: Edizioni Il Fiorino.

Downing, A., & Locke, P. (1992). *A university course in English grammar.* Hemel Hempstead: Prentice Hall.

Hunston, S., & Thompson, G. (Eds). (2000). *Evaluation in text.* Oxford: Oxford University Press.

Hyland, K. (1999). Disciplinary discourses: Writer stance in research articles. In C.N. Candlin & K. Hyland (Eds), *Writing: Texts, processes and practices* (pp. 99–121). London: Longman.

—(2001). Bringing in the reader: Addressee features in academic writing. *Written Communication, 18,* 549–574.

—(2002a). What do they mean? Questions in academic writing. *Text, 22,* 529–557.

—(2002b). Directives: Power and engagement in academic writing. *Applied Linguistics, 23,* 215–239.

—(2004). *Disciplinary discourses: Social interactions in academic writing.* Ann Arbor: University of Michigan Press.

—(2005). *Metadiscourse: Exploring interaction in writing.* London: Continuum.

Leech, G., & Svartvik, J. (1994). *A communicative grammar of English* (2nd edn). London: Longman.

Martin, J. (2000). Beyond exchange: APPRAISAL systems in English. In S. Hunston & G. Thompson (Eds), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 142–175). Oxford: Oxford University Press.

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

Perkins, M. (1983). *Modal expressions in English.* London: Frances Pinter.

Scollon, R., & Scollon, S. (1995). *Intercultural communication.* London: Blackwell.

Swales, J. (2004). *Research genres.* Cambridge: Cambridge University Press.

Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics, 22*, 58–78.

Webber, P. (1994). The function of questions in different medical English genres. *English for Specific Purposes, 13*, 257–268.

Chapter 7

E-Conferencing: Corpus and Discourse Insights

Ann Hewings, Caroline Coffin and Sarah North

1   Introduction

Over the past decade a new type of academic writing space has emerged online which features, not the polished prose of a journal article or essay, but a discourse still under negotiation, where goals and conventions are evolving and unstable (perhaps inherently so). Within this online space, the use of asynchronous electronic conferences (e-conferences) is growing in distance education and blended learning environments. Claims are made that this environment is suited to the building of learning communities and the exchange and negotiation of ideas (Andriessen, 2006; Andriessen et al., 2003; Cousin & Deepwell, 2005; Marttunen, 1997; Ravenscroft & Pilkington, 2000). These claims, however, are not uncontentious (Joiner & Jones, 2003).

  A major focus in approaches to analysing academic writing has been on how writers negotiate their disciplinary knowledge claims (Bazerman, 1988; Chang & Swales, 1999; Hunston & Thompson 2000; Hyland, 1998, 2000; MacDonald, 1994; Myers, 2001). Citations, informal language, grammatical subjects, stance markers, pronouns, reporting verbs, genre patterns and politeness markers are among the features analysed with regard to how writers convey themselves in relation to their disciplines and how, in turn, their claims are received by their readers. A growing number of studies have looked at such features in student writing, much of it to inform pedagogy for international students (L2) studying in the medium of English (Coffin & Hewings, 2004; Hood, 2004; Swales, 1990; Tang & John, 1999), though increasingly also looking at the practices of first language English students (L1) (Berkenkotter et al., 1991; Charles, 2006; Drury, 2001; Harwood, 2005; Hewings, 2004; Hewings & Hewings, 2002; North, 2005). The language of e-conferencing has received relatively little detailed language-focused analysis. The collection of papers introduced by Androutsopoulos (2006) on sociolinguistic research into e-conferencing is possibly indicative of growing interest and some studies from a computer supported collaborative learning perspective have examined linguistic features such as qualifiers and intensifiers (Fahy, 2002; Jeong, 2006).

This chapter is a contribution to the exploration of e-conferencing using a linguistic lens. We report findings from a study[1] which used both a discourse and a corpus analysis framework to focus on interaction in e-conferencing within a group of Health and Social Care undergraduate students. The research questions addressed in this chapter are:

- What types of discourse moves are common in the e-conferences examined?
- Can corpus analysis support or extend these findings?
- What insights do the two approaches provide into student and tutor interaction in e-conferences?

Within many distance education programs e-conferencing has become a common means of creating a virtual learning community – bringing together students, otherwise separated by time and geography, to engage with course content at a time of their choosing. Interaction within an asynchronous environment consists of messages sent to all members of the e-conference which can be read, re-read and responded to by conference members at any time during a designated period. Discussions are often focused on articles, video or audio material or other course texts and are usually relevant to subsequent assessment. Much e-conferencing in higher education is based on groups working together in order to reach shared understandings or solutions or to create a product (Littleton et al., 2000; Wasson et al., 2003). In the social sciences and humanities e-conferencing is often used as a forum for students to exchange their views and perspectives on contentious issues and ideas, typically in response to a task set by their tutor. The discussions that may take place in e-conferences can be a particularly important form of collaboration, stimulating belief revision and conceptual change (Ravenscroft, 2000; Ravenscroft & Pilkington, 2000).

Experience has shown, however, that students are not necessarily eager participants in these exchanges, with concerns over how they present themselves and their opinions to people they may never have seen in a medium which preserves their contributions for all to see and reflect on (Hewings & Coffin, 2006). Students are developing ways of understanding and articulating disciplinary knowledge and knowledge-making practices. However, in communicating on an e-conference their thoughts, ideas and often personal beliefs are in a relatively public and lasting form which is available for scrutiny. Participants often take more time to plan and compose their contributions than would be the case in face-to-face encounters, resulting in more expansive turns than in casual conversation. However, there is no obligation to respond at all to any particular message – keeping silent is an option in e-conferencing that would be highly unusual in a face-to-face context and is open to negative interpretations. There is, therefore, considerable interpersonal risk involved; it is not

only the content discussions of e-conferencing that are worthy of investigation but also the personal investment in them by students.

Analysis of e-conferencing is located in different disciplines and draws on a variety of methodologies, many of which were originally developed to study other contexts. One such is discourse analysis, which has been adapted within psychology and applied to analyses of both e-learning and e-conferencing. Schrire (2006), for example, in addition to analysing cognition, also investigates interaction using a model of discourse analysis based on Wells' (1999) adaptation of Sinclair and Coulthard's (1975) approach to classroom discourse. This model involves a hierarchy of five levels – lesson, transaction, exchange, move and act – in which a typical exchange consists of initiating, responding and follow-up (or evaluating) moves. The focus in such studies is on the function of interaction rather than the identity and persona of the interactants.

Corpus analysis facilitates the investigation of groups of texts and as such is a promising approach for looking at the large amount of data generated by e-conferencing. Relatively little research so far has applied corpus analysis tools to investigating e-conferencing. Exceptions have been in the area of second language learning (Fitze, 2006; Montero et al., 2007). Corpus analysis, however, has been used in a number of studies to focus on the significance of pronominal reference and address the issue of how writers present themselves and their opinions. Such studies are particularly useful in highlighting aspects of authorial voice, an aspect of writing which has particular significance within the interactive setting of e-conferences. Most studies, however, have focused on published academic writing (Harwood, 2005; Hyland, 2000; Kuo, 1999) and traditional forms of student writing (Harwood, 2003; Hyland, 2002; Tang & John, 1999). Tang and John, for example, looked at essays written by Singaporean undergraduate students and constructed a typology of possible identities indicated through choice of pronouns. The most powerful authorial presence was described as '*I* as originator'. However, they found that students were more likely to assume less powerful authorial positions as they felt 'insecure about the validity of their claims, seeing themselves to be at one of the lowest rungs of the academic ladder' (Tang & John, 1999: S34). Hyland (2002), in a study of Hong Kong undergraduate students, had similar findings but attributed the unwillingness of students to take an authoritative stance to their cultural background, which discourages the promotion of an individual self. In a study of L1 postgraduate student master's dissertations in two discipline areas, Harwood (2003) found that the discipline itself was an important factor in how visible the writer's presence was, with relatively high uses of *I* and *we* in Computing Science dissertations as compared to Business and Management. In a study of pronouns in essays and e-conferencing by predominantly L1 masters students of Applied Linguistics, Hewings and Coffin (2007) found that a powerful authorial voice was often associated with the collective *we* which built upon shared professional knowledge.

The study described in this chapter combines corpus analysis with a specially developed discourse analysis framework for e-conferencing. The discourse analysis identifies the broad categories of interaction taking place as well as the moves within it. Corpus analysis techniques (Scott, 2004) are used to identify 'keywords', those words from the discourse which occur outstandingly frequently in comparison to some norm. Certain pronouns are highlighted by this means and are examined in the light of the research into authorial voice reported above; the findings are then compared to those from an analysis of pronouns in essays and e-conferencing in the field of Applied Linguistics. The corpus analysis findings are cross-referred to the results of the discourse analysis to facilitate greater understanding of the interaction taking place.

Below we outline the background to the development of our discourse analysis framework and the major move categories. For more detailed discussion see Hewings et al. (2006) and North et al. (2008).

2   The Discourse Analysis Framework

The discourse analysis framework developed for this research grew out of genre analysis (Martin, 1989) and developments in the analysis of casual conversation (Eggins & Slade, 1997), influenced by Sinclair and Coulthard's (1975) model of classroom discourse. The Eggins and Slade model of discourse structure analysis is similar to the exchange structure analysis of Sinclair and Coulthard in that it involves identifying the function of the various moves used by participants in a discussion. While Sinclair and Coulthard identified initiating, responding and follow-up moves in classrooms, Eggins and Slade distinguish opening and sustaining moves in casual conversation. Both models are relevant to the interactive exchange of views within a discussion, but require adaptation to deal with the different nature of asynchronous e-conferencing.

The discourse analysis framework was designed to contend with the particular characteristics of e-conferencing. Turns are often long and need to be segmented in order to identify different functions. This creates a problem in deciding on the unit of analysis. We use the t-unit, which consists of an independent clause together with clauses dependent on it. Once the text was segmented in this way, each t-unit was coded according to the functional move that it realized; where a move comprised more than one t-unit, coding was simply continued over all the relevant units. Messages in e-conferences are often not directly related to each other. A response message may be sent some time after the message to which it is responding and intervening messages may well have been sent. We have recorded the messages in the order that they were sent, in the knowledge that this is not necessarily the order in which participants viewed or responded to them. A numbering system (not discussed here) enables relationships between moves in the discussion to be tracked.

The type of interaction which is going on in e-conferencing has been subject to a variety of categorizations, distinguishing for example between task-related and non-task-related material (Schellens & Valcke, 2004), interpersonality and impersonality (Beuchot & Bullen, 2005), or between social, organizational and intellectual moves (Burnett, 2003). Since our initial focus in developing the framework was argumentation, we began by classifying argumentative moves as distinct from social, procedural and other instructional moves. The key criterion for identifying a move as argumentative was that it formed part of the negotiation of claims, either by proposing, supporting or challenging a position. However, making decisions about whether moves are claims or claim-related is not straightforward (Erduran et al., 2004) particularly as much reasoning is implicit. In the example below, the first two t-units constitute a claim and it is followed by reference to personal experience which is taken to be in support of the claim. While the reasoning in this case is implicit, the lexical signalling (*alternative, CAM, complementary orthodox medicine, medical model*) ties 1 and 2 with 3.

1. I have used the word alternative
2. but I also know for some people CAM can be used as complementary to orthodox medicine.
3. Many of my friends choose some form of CAM before visiting the doctor but still seem to need the reassurance of the medical model. (Julie 4/05)[2]

Such moves would be classified as Discussion, the major category within the discourse analysis framework, which also categorizes non-argument focused discourse under the headings: Social, Procedural and Other field-related. Since, as noted above, our original focus was the way that students argued in the e-conferences, we aimed to analyse moves in the 'Discussion' category exhaustively. Within the categories Social, Procedural and Other field-related we indicated only particularly salient types of move. The move categories within the framework as a whole can be seen in the Appendix. (See North et al., 2008, for detailed discussion of the argumentation framework).

Fourteen moves were classified as Discussion, relating to the topic under discussion in the e-conference and forming part of (or potentially contributing to) the on-topic argument. **Claim** moves, that is, contestable propositions, are central to the argumentation. For example:

Yes, I think GPs will have more confidence if there is either statutory or voluntary regulation. (Lucinda 4/05)

These can be challenged with moves such as **counterclaim** or **refute**. Alternatively, claims may be supported through **agreement** moves or **informing** moves. Informing, as a very large move category, is subdivided into

different types of information or reasoning that may be used to support a claim. Common informing moves used in support of claims were personal assertions, personal experience and professional experience. Also in the discussion category are moves such as **concessions**, which recognize the validity of alternative viewpoints, and **argument prompts**, such as the tasks given by the tutors.

Within Social are those moves which relate primarily to constructing or negotiating solidarity/community. Many e-conference participants used **salutations** such as '*Hi*' and **signing off** moves and **encouragement** moves were also common. For example:

> I'm glad you checked in with our conference, and thanks to Mary and Laura for your thoughtful support. (Julie 1/06)

Moves relating not to the discussion of the topic, but to establishing and maintaining the conditions which allow the discussion to take place are categorized as Procedural. These include both technical and organizational issues. **Problem** moves describe and/or ask for assistance usually relating to technical computing issues. These are responded to with **help** moves. For example:

> Instead of going to your tutor group you will see a column for OU Community Under that, click on Open University, then OU Students Association, then OUSA Signpost. (Lucinda 1/05)

Directives are moves in which participants are given instructions on how to carry out the e-conference task, usually by the tutor.

The final category is Other field-related, which covers moves that can be roughly classified as 'classroom talk', and cannot be classified under any of the other three categories as defined above. **Elicitation** moves include factual queries and responses not related to the intended topic of discussion, and **informing** moves provide background related to the wider educational contexts, but also not directly to the topic of discussion. For example:

> Did anyone else watch the programmes on BBC 2 about CAM? I thought they were excellent and provide some valuable insights and information. (Naomi 1/06)

3   Data Collection and Methods

3.1   Research context

The pedagogic setting was two cohorts of students following an undergraduate course *Perspectives on complementary and alternative medicine* (hereafter CAM)

in the Faculty of Heath and Social Care at the Open University, UK. The over-all aim of the course was to provide 'an accessible but rigorous introduction to complementary and alternative approaches to health' and 'to stimulate lively debates about this controversial and topical subject and to equip [students] with information and analytical frameworks with which to enter the debates' (www.open2.net/alternativemedicine/courses.html, accessed 14 March 2008).

We investigated 16 tutor groups in which participants were expected to participate in several e-conferences throughout the academic year. Our data shows that although there were between 15 and 20 students in each group the average number of active participants ranged from 2 to 13. The tutors attached to each tutor group had been given technical training in the use of the e-conferencing software (the commercially available *FirstClass* asynchronous system) but the extent of each tutor's experience in managing e-conference discussions was diverse. It was not possible to ascertain the extent of students' previous e-conferencing experience but based on interview and questionnaire data it appeared quite varied. Each conference lasted for approximately three weeks and was organized around different discussion tasks.

3.2 Data collection and preparation

Four tutors from the CAM course and two cohorts of their students were selected for analysis and their agreement obtained. Table 7.1 shows the data collected for two e-conferences for each tutor (one held at the beginning of the year and one held just after the half-way point of the course) and the assignments that related to the e-conferences. Questionnaire and interview data were also collected from the 2006 cohort. The main focus of the e-conference discussions concerned the assignment topics.

All the text data from the students and tutors was anonymized. Assignment question wording, end references and tutor comments were removed manually from the essays, so that only the students' own words would be analysed. In the e-conferencing data the duplicate text associated with copying messages or parts of messages that were replied to was removed. For the purposes of corpus analysis, these cleaned-up texts were converted into plain text. Initial

Table 7.1 E-conference and assignment data collected

	2005	2006
Number of tutor groups analysed	4	4
Number of e-conferences analysed	8 (2 per tutor) (31,507 words)	8 (2 per tutor) (17,541 words)
Number of assignments analysed	118 (224,779 words)	139 (246,483 words)

Note: The number of words was calculated using *WordSmith Tools 4.0.* (Scott, 2004) corpus software after the data was prepared for analysis.

corpus analysis was carried out using *MonoConc Pro* (Barlow, 2002) and subsequently using *WordSmith Tools 4.0* (Scott, 2004). Analysis using the discourse framework was carried out by the project team on data from the first two tutorial conferences, and the coding categories were gradually agreed on through discussion. All the text data was then coded by a single researcher, to maximize consistency.

4   Findings and Discussion

4.1   Discourse analysis

The discourse analysis framework was used for looking at a variety of elements in the e-conferences but below we only report on the frequency of particular categories of moves and aspects of interactivity. Figure 7.1 shows that the highest number of moves occurred in the Discussion category. This indicates that participants were willing to express views on the topic under discussion and was in line with what students said they valued and tutors were aiming for in the e-conferences. Social moves, such as greetings and encouragement, were consistently the next most frequent, which shows a concern for interpersonal aspects of e-conferencing.

A high number of Discussion moves is not necessarily indicative of interactivity; it is possible that claims are made but not picked up and discussed further. To capture the extent of the dialogic nature of the discussion we looked at whether and how claims were responded to. Figure 7.2 shows the percentage of claim moves which were responded to by the tutor or by other students, or which elicited no response at all, for all the e-conferences and for each tutor. It is noticeable that students make the most responses and that their rate of
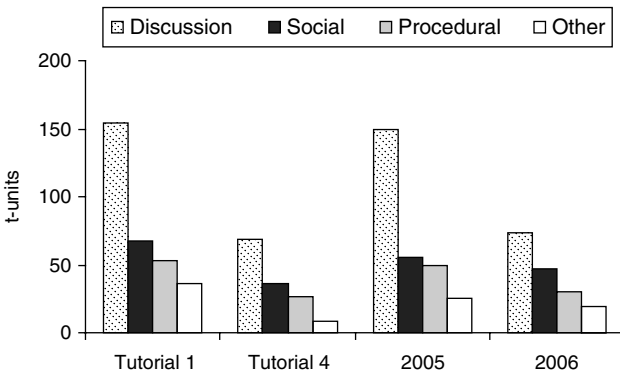
FIGURE 7.1   Frequency of moves in each category by tutorial conference and by year
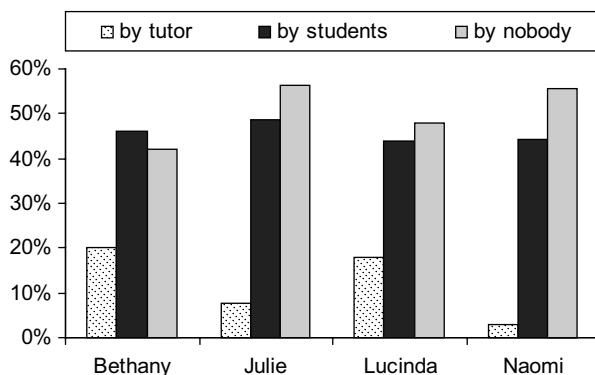
FIGURE 7.2   Responses to claims

Table 7.2   Agreements and challenges by tutorial and year

	Tutorial 1	**Tutorial 4**	**2005**	**2006**
Agreeing	64	41	92	13
Challenging	54	50	73	31

response is almost the same across the four different tutors (range 44.0–48.7%). Tutors, on the other hand, show greater variation. In groups where the tutor is more responsive, fewer claims go unresponded to. The student interviews suggest that receiving some form of response is important as some students may otherwise feel marginalized. Having summoned up the courage to put forward an opinion, one student clearly felt disconcerted by a lack of response:

> I think what upset me was most of the others would respond if you wrote anything in and I got no response, so at least at a face-to-face tutorial you would get a response, you know if what you were saying was right or wrong, I mean because nobody wrote back. Then I read what other people had written, but I sort of lost my confidence and I thought I haven't got anything valuable to say so I didn't write anything.

Figure 7.2 shows that this student's perception that only her postings failed to elicit a response is not borne out by the evidence. Unacknowledged posts clearly have the potential to undermine students' confidence in a medium which has no other feedback mechanisms.

The ways in which claims are responded to is also of significance in examining the ethos of the e-conferences. Two types of response are agreements and challenges. Table 7.2 shows that in 2005 there was slightly more agreeing than challenging, but that in 2006 there was a marked decrease in agreeing.

This decrease is set against an overall decline in activity in 2006. The higher number of challenging moves relative to agreeing moves accords with the 2006 cohort's positive views on the importance of challenging arguments, as reported in the questionnaire survey.

If success in e-conferencing can be judged by participation, particularly in on-topic discussion, then agreeing moves appear to have significance. Overall participation in 2005 was higher than in 2006 and agreeing moves were more frequent than challenges. This situation was reversed in 2006 with more challenges but less overall participation. Generally, agreeing moves appear to help create a collaborative ethos. They are frequently personalized, using *I* and building in the names of individuals (e.g. '**I** think **Abigail** has got a point in that regulation . . .'; '**I** also like the comment by **Chloe** about . . .'). Challenges also often include names, perhaps to make the interaction more personal but less threatening (e.g. '**Lucinda**, I'm not too sure about your predictions re. the NHS being in meltdown . . .'; 'What do you think they are saying, **Robert**?'). The frequency of naming may also be a by-product of the disruption of the turn taking sequence within an asynchronous environment. Naming may be a device that helps to identify not only *who*, but *what* is being agreed with or challenged.

The results presented here from the discourse analysis framework illustrate that most of the interaction in the computer conferences was on-topic discussion of the tasks set by the tutors, with social moves also proving significant. Students were prepared to put forward views and to support and challenge those of others. The framework allowed identification of claims that were not responded to. This was a common occurrence and interviews suggest that it was a factor in inhibiting participation.

4.2   Corpus analysis

Corpus tools were used to identify words and phrases within the e-conference discourse that were particularly salient, and to prompt further qualitative investigation and comparison with the discourse analysis framework. *WordSmith Tools 4.0* (Scott, 2004) was used to find out which words were 'key', that is occurred statistically more often in one wordlist when compared to another reference wordlist. We generated wordlists for each cohort for both the e-conferences and the assignments and compared them. This had the advantage of screening out those frequent words that were associated with the topics under discussion and instead focused on words which were significant only in the e-conferences. There were only 13 keywords in the 2005 e-conference corpus, of which only seven are of interest (Table 7.3). We have omitted *Subject, Re, tutorial, March* and *From*, which are words found in the message headers, along with *xquotex* which was used to replace text quoted from earlier messages. The pattern was similar for 2006 though only seven words in total were found to be key.

The keyness of these words across the different files within the e-conference corpora was checked using *WordSmith Tools'* key keyword function. This is a

Table 7.3 Top keywords measured by log likelihood (p=0.0000001) for e-conferences

	2005		2006	
	Keyness	**Raw frequency**	**Keyness**	**Raw frequency**
I	1,333	560	1,064	372
you	842	274	411	136
think	388	127		
my	318	124	336	96
your	272	106		
me	261	83		
Hi	214	51	365	74

Table 7.4 Keywords and their associates in the e-conferences

I	You	Your	Think	My
think	re	re	you	subject
your	think	think	re	tutorial
I'm	your	you	your	you
hi	I'm	message	tutorial	me
message	hi	I	I	I'm
my	message	subject	subject	I
from	my	tutorial	hi	your
re	from	from	I'm	hi
#	it's	thanks	#	re
agree	thanks	hi	from	am
am	xquotex	my	xquotex	get
get	#	I'm	message	think
its	agree	get	my	#
me	get	me	do	from
thanks	me	xquotex	agree	xquotex
xquotex	tutorial	don't	get	but
bit	do	#	it's	conference
do	don't	agree	me	first
don't	am	am	thanks	just
click	click	click	conference	do

'statistical measure of the "spread" of keywords across a corpus' which consists of a number of files (Baker, 2006: 142–143). It avoids the danger of considering as key a word that occurs multiple times but only in one file.

An analysis of the keywords and their associates (Table 7.4) highlights the verbs *think*, as well as forms of *be*, *do(not)*, *agree*, *get* and *click*. *Think* and *agree* are the most frequent lexical verbs and both convey mental processes (Halliday, 1994). Common clusters with *think* are *do you think*, *I think the*, *what do you*, *I think I*, and *I think that*. The question form occurs in both tutor and student messages. In the tutor messages it is a common strategy for prompting students to think more deeply. The example below from a tutor came in response

to a message on regulation of complementary and alternative medicine.

> Yes, I think GPs will have more confidence if there is either statutory or vol-
> untary regulation. But **do you think** it will make any difference to the pub-
> lic? (Lucinda 4/05)

In the student messages it is frequently used at the end of messages to make
claims less potentially face-threatening and to open up the discussion to oth-
ers, as in the following example.

> We are going back to a pre-modernity plural health market which sounds
> good to me.
> What do you think?
> In a hurry.
> Bye for now. (Lucinda 4/06)

I think is typically followed by *that*, a noun or pronoun or *the* followed mostly by
an abstract noun. It is often integral to the discussion, associated with giving
an opinion or supporting points made by someone else.

> **I think that professionalism** is changing (and should do) in response to
> public expectations (Naomi 4/06)
> One of the main reasons **I think hospital doctors and nurses** would be good
> is . . . (Bethany 1/05)
> **I think I** would have to agree with the comments on the subject of informed
> consent. As long as it is done well . . . (Bethany 4/05)
> **I think the same principle** applies with CAM. You try and find out about it
> but it is very hard to get accurate information as everyone is trying to sell you
> something. (Lucinda 1/05)

Agree is also used to show support either for points made or towards the
individuals making the points. It typically occurs in the clusters *I agree with*, *I
agree that* and *I do agree*. It is also frequently strengthened with boosters such
as *totally*, *strongly*, *have to*. The example below illustrates how both the person
being agreed with and the point they are making may be combined. It is then
elaborated by a new claim relating to religion.

> Hi Connor,
> **I do agree** that people may be using CAM for the spiritual side as well. My
> main theory is that this could do with the decrease in religious following.
> I also believe that Medicine has a down falling . . . (Bethany 1/05)

As *I* and *you* are the most strongly key items in the e-conferences, and pro-
nouns have been associated with authorial visibility, we investigated their use
and relationship to the wider discourse. Table 7.5 shows all discourse analysis

Table 7.5 Occurrences of 'you' and 'I' in discourse moves

	I (%)		You (%)	
	Tutors	**Students**	**Tutors**	**Students**
DISCUSSION				
Thesis				
Claim		5	2	1
Claim/Support				
Subclaim				
Recommendation				
Counterclaim				1
Informing				
recount		1		
professional recount				
personal recount				
procedure		2		
description		4		
counterfactual explanation				
other explanation		3		4
personal assertion		10		1
professional experience		1		2
personal experience		8		
other exemplification			1	1
other information	2	4		2
Agreement		12		5
Refute		3		
Concession		3		
Argument Prompt		1	4	2
Information Prompt			4	1
Issue				
Preview		3		1
Summary				
Sub-totals	**2**	**60**	**11**	**21**
SOCIAL				
Encouragement		2	5	
Teasing				1
Deferring		1		1
Salutation				
Signing off				
Other	2	7		3
Sub-totals	**2**	**10**	**5**	**5**
PROCEDURAL				
Problem	1	3		
Help	2		3	5
Directive			25	
Other	6	5	6	3
Sub-totals	**9**	**8**	**34**	**8**

Table 7.5   Continued

	I (%)		You (%)	
	Tutors	**Students**	**Tutors**	**Students**
OTHER FIELD-RELATED				
Elicitation				2
Informing	1	2	3	7
Other	2	2	4	
Sub-totals	3	4	6	9
UNCLASSIFIED	1	1	1	
Totals	17	83	57	43

Table 7.6   Focus of address in the use of 'you' in e-conferences

	Tutors (%)	Students (%)	Total (%)
Generic	4	8	12
Individual	2	10	12
Group	51	25	76
Total	57	43	100

categories and the uses of *I* and *you* by tutors and students. *I* is used predominately by students with 60 per cent of the uses coming in Discussion moves. The majority form part of informing moves which are information or reasoning put forward as part of the on-topic discussion. Personal experience (e.g. *Just after I had my daughter 6 years ago I was diagnosed with hypertension*) and personal assertion (e.g. *I do not want to be associated with this practice*) figure highly. The pattern *I think* is almost exclusively found in the Discussion categories used by students: claim, informing (mostly personal experience or assertion), agree, refute, and concession. Tutor uses of *I* were predominantly Procedural, often relating to technical or organizational considerations (e.g. *I should like to divide this tutorial into 3 streams*). 10 per cent of student uses of *I* were related to predominantly Social moves but only 2 per cent of tutor uses.

In contrast to the findings for *I*, the keyword *you* is used more frequently by tutors (57%) than students (43%). Table 7.6 indicates that the tutors' focus of address is largely the student group as a whole (51%).

The high figure for addressing the group by the tutors corresponds to the most frequently used move, which is Directive, in which the tutor typically instructs students how to carry out the task, often using imperative verb forms (e.g. *Think about the choices **you** have made in relation to your own health or well-being and the interaction **you** have had with health practitioners.*

Then look at the case study presented for TMA01 in the assignment booklet.) Group address is also used in the tutor Discussion moves, which are mainly prompts and often questions (e.g. *I like the pluralist approaches to treatment suggested so far! would* **you** *be wanting to know what the conventional treatment is, how effective it is and what the side effects are or would* **you** *trust your doctor's judgement?*). Most of the prompts using *you* in Table 7.6 are from one tutor. Examination of the discourse analysis results shows that she used information prompt moves more frequently than any of the other tutors. Students also used *you* to address the whole group using a variety of move types (e.g. *If* **you** *take the example of Louise in the course book; I know I'm WAY behind before* **you** *all shout!!*), but they also made frequent addresses to individuals, often in response to and particularly agreeing with earlier messages (e.g. **Elizabeth**, *I think* **you** *have made an interesting comment here*).

The corpus findings on key pronouns show both similarities and differences to those of an earlier study of e-conferencing (Hewings and Coffin, 2007). The students in that study were using the same e-conferencing system, but were studying a masters module in TESOL. *I* and *you* were keywords in the e-conference, but so too was *we*, which did not figure as differentiating the interaction on the e-conferences from the individually written assignments in the CAM data. *We* as a proportion of the TESOL e-conferences occurred twice as frequently as it did in the CAM e-conferences. The categorization from that research identified *we* as most commonly associated with an inclusive address. Sixty-six percent of occurrences referred to the writer and the others taking part in the conference. In contrast, only 32 per cent were inclusive in the CAM e-conference. The other major difference was in the number of uses of generic *we*. Only 4 per cent of occurrences in the TESOL e-conferences were of this type, whereas they accounted for 44 per cent in the CAM data. This seems to indicate that there is a lack of a group identity to call upon. Within the TESOL e-conferences the interactants were all teachers and *we* frequently invoked a sense of solidarity around 'we as teacher', which allowed students to link to the wider group thereby making their arguments less easy to challenge. There was also a shared professional background with the tutors which might indicate a less hierarchical context. Although the CAM course was aimed at CAM practitioners and interested others, there was much less evidence of a practice-based understanding or identity either among students or between students and tutors. Tutors were able to invoke an exclusive group identity linking them to other CAM practitioners (e.g. *in the society I am registered with* **we** *have excelling training . . .*). In contrast, the only group invoked by any of the students was other orthodox healthcare practitioners, usually nurses (e.g. *as nurses the registration* **we** *have gives us some protection . . .*). These differences suggest that an analysis of pronoun use can indicate how aspects of interaction are differently constructed

in e-conferences with greater or lesser degrees of field or disciplinary homogeneity among the participants.

The corpus analysis has identified the significance of personal pronouns and mental processes (Halliday, 1994) such as *think* and *agree*. Comparison with the assignments corpus found only a small number of words to be key and these were words associated with the interactive, interpersonal nature of discussions in e-conferencing. A comparative analysis with e-conferencing in a TESOL course where the pronoun *we* was also key, indicated a possible disciplinary aspect to the construction of claims based on whether or not students and tutors shared disciplinary or professional backgrounds.

5   Conclusion

Academic writing in e-conferences is still relatively new and under-researched. This chapter has outlined two language-based approaches to examining the interaction taking place. The discourse analysis framework was designed to account for the characteristics of asynchronous discussions and particularly focused on how students and tutors engaged in argumentation. In answer to our initial research question, the discourse moves found to be most common in the e-conferencing were those concerned with on-topic discussion and secondarily moves designed to construct social solidarity and community. Results suggested that a significant factor in encouraging debate was responses; qualitative findings that a large number of claims went unresponded to were supported by analysis of discourse moves. Agreeing moves appeared to be interpersonally significant in building up a collaborative and supportive ethos and were also associated with greater on-topic discussion.

Our second research question focused on whether corpus analysis could support or extend these findings. Keywords were identified in the e-conferences of the two student cohorts, predominantly personal pronouns with *I* and *you* being most key. *Think* was the only verb to be key for both cohorts. Analysis of key keywords revealed that the verb *agree* was strongly associated with *I, you, your* and *think*. Cluster analysis of the two verbs showed their use in Discussion moves in putting forward modalized claims and support for the claims of others. Extension beyond the findings of the discourse analysis framework came through a comparison of the keyword findings with an earlier study of e-conferencing. This suggested that concordance analysis of key pronouns is indicative of areas of difference in the way knowledge claims are made. In particular the use of *we* indicated a shared practice- or disciplinary-based background and could be exploited to make knowledge claims more inclusive of the group and less open to challenge. In the CAM data analysed here, similar strategies were followed by tutors, invoking the CAM profession, and by orthodox health professionals, invoking their peers and professional bodies.

However, the result was exclusive rather than inclusive as there was a much weaker shared identity within the e-conference groups.

Analysis of concordance lines allowed occurrences of *I* and *you* to be mapped against the discourse analysis framework. The combined analysis indicated that the uses made by tutors and students differ. *I* was common in student moves, particularly those relating to the on-topic discussion. Tutors made greater use of *you* particularly in directives, telling students what to do. These were associated with imperative verb forms. Of the Social moves examined, students used *I* and *you* to encourage, tease, and defer, whereas tutors mostly encouraged. The tutor moves highlighted by this analysis show tutors as rarely personalizing their Discussion moves. They do not tend to preface claims or other Discussion moves with '*I think*', an interpersonal marker often associated with hedging. The absence of tutors from the analysis of Discussion moves relating to *I* and *you* reflects the lower overall number of claims made by tutors, though this did vary depending on the tutor and their views on the purposes of e-conferencing. The tutor's pedagogic role was apparent in the use of directives. The focus of address when using *you* was predominantly on the group, though students also made use of individual address.

The combined analyses presented here have illustrated ways in which interaction and particularly the discussion of ideas can be observed to occur in e-conferencing. The discourse analysis framework enabled quantities and patterns of moves to be categorized and tracked through the unfolding of the e-conferences. The corpus analysis highlighted the significance of personal reference within these moves. Concordance analysis combined with the discourse categories indicated the personalized nature of many of the moves and also the differences between the moves of the tutors and students. *I* was much more commonly used by students, particularly in combination with *think*. This enabled claims to be made more tentatively and the discourse analysis framework revealed that such claims were more likely to be taken up and discussed. The tutors made relatively little use of *I* except in Procedural moves relating to technical or organizational matters, though *you* was much more common and used mostly to instruct or direct students. The picture emerging is of a relatively hierarchical pattern of interaction, with tutors directing and being less personal in their Discussion moves. An alternative interpretation would be that tutors were trying to set up the tasks and then either taking a back seat in order for students to interact together and/or couching their Discussion moves in less personal and more abstract or theoretical terms as models of academic discussion for students. Further analysis of tutor Discussion moves and their linguistic realizations is necessary to explore these interpretations. Corpus analysis of keywords in the assignments corpus could also be the foundation for examining what is more salient in a traditional written genre and which discourse moves this was associated with. The implication of the analysis presented here is that claims

in assignments are less personalized, so the question remaining is what strategies are employed and have these built upon any strategies rehearsed in the e-conference discussions.

Appendix

Discussion

(The first five all involve contestable propositions that may be challenged/ supported)

Claim
A contestable proposition relating to how things are (analytic)

Thesis
An overall position on an issue (at a higher level of generality than a claim) is put forward (i.e. a thesis statement)

Recommendation
A contestable proposition relating to how things should be (hortatory)

Counterclaim
A claim which takes an alternative position to a previous claim

Claim/Support
A claim which includes supporting evidence or reasoning in the same move

Informing
Information or reasoning which is put forward as part of the on-topic discussion; these moves may be either integrated (used to support a claim) or unintegrated (not linked to any particular claim, but available as potential support for a claim)

- *recount*: A recount of a series of actions or events
- *procedure*: Information about how a procedure is being/has been/will be carried out
- *description*: Information about the nature or condition of a person, place, object or concept
- *counterfactual explanation*: Reasoning that speculates on what might have happened
- *other explanation*: Other logical reasoning, involving explicit causal relationships
- *personal assertion*: A comment related to the on-topic discussion which describes the writer's affective response and is therefore not open to challenge
- *professional experience*: Reference is made to professional experience provided by the writer
- *personal experience*: Reference is made to personal experience provided by the writer
- *other exemplification*: One or more specific examples of a general point

- *other information:* Any other material which is part of the specified on-topic discussion, but does not fall into one of the above categories

Agreement
A previous claim is confirmed by a participant agreeing with it

Refute
A questioning or criticism of an argument or claim made in a previous turn, (or in a forum outside the conference such as a text book, academic article etc.) No new claim is made, unlike Counterclaim

Concession
Recognizes the validity of an alternative viewpoint expressed in a previous turn. This move is subsidiary to a claim being put forward by the writer

Argument Prompt
A question designed to stimulate and prompt participants' views on an issue

Information Prompt
A question designed to stimulate participants to provide information as part of the on-topic discussion

Issue
The overall issue to be debated is identified (without indication of the stance or approach to be taken by the writer)

Preview
The direction of the forthcoming discussion or section of discussion is explicitly introduced

Summary
Preceding discussion points are explicitly summarized or completed

Social

Encouragement
Participants motivate and encourage each other

Teasing
Participants denigrate each other or each others' contributions, playfully or otherwise (opposite of Encourage)

Deferring
Participant minimizes own contribution and/or seeks reassurance from others

Salutation
Participants open contributions with a greeting

Signing off
Participants close contributions

Other

Procedural

Problem
Describes and/or asks for assistance with a procedural problem (relating to technical issues or other conditions that affect the ability to carry out the task)

Help
Provides information intended to help with procedural matters

Directive
Moves in which a participant (normally the tutor) instructs participants how to carry out the task

Other

Other field-related

Elicitation
Any move intended to elicit factual information which is related to the wider educational field but not part of the specified on-topic discussion itself

Informing
Any move providing factual information which is related to the wider educational field but not part of the specified on-topic discussion itself

Other
Includes explicit teacher evaluation of student contributions, or student evaluations in same style

Notes

[1] This study was funded by the Higher Education Academy, UK and the project report is available at www.heacademy.ac.uk/ourwork/research (accessed 12 May 2009).

[2] E-conference examples are referenced by the tutor's pseudonym, the conference number and the year.

References

Andriessen, J. (2006). Collaboration in computer conferencing. In A.M. O'Donnell, C.E. Hmelo-Silver & G. Erkens (Eds), *Collaborative learning, reasoning and technology* (pp. 197–231). Mahwah, NJ: Erlbaum.

Andriessen, J., Baker, M., & Suthers, D. (2003). Argumentation, computer support, and the educational context of confronting cognitions. In J. Andriessen, M.J. Baker & D. Suthers (Eds), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 1–25). Dordrecht: Kluwer.

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics, 10,* 419–438.

Baker, P. (2006). *Using corpora in discourse analysis.* London: Continuum.

Barlow, M. (2002). *MonoConc Pro 2.0* [Computer software]. Houston: Athelstan.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science.* Madison, WI: University of Wisconsin Press.

Berkenkotter, C., Huckin, T.N., & Ackerman, J. (1991). Social contexts and socially constructed texts: The initiation of a graduate student into a writing research community. In C. Bazerman & J. Paradis (Eds), *Textual dynamics of the professions: Historical and contemporary studies of writing in academic and other professional communities* (pp. 195–215). Madison, WI: University of Wisconsin Press.

Beuchot, A., & Bullen, M. (2005). Interaction and interpersonality in online discussion forums. *Distance Education, 26*(1), 67–87.

Burnett, C. (2003). Learning to chat: Tutor participation in synchronous online chat. *Teaching in Higher Education, 8,* 247–261.

Chang, Y-Y., & Swales, J. (1999). Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers? In C.N. Candlin & K. Hyland (Eds), *Writing: Texts, process and practices* (pp.145–167). London: Addison Wesley Longman.

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25,* 310–331.

Coffin, C., & Hewings, A. (2004). IELTS as preparation for tertiary writing: Distinctive interpersonal and textual strategies. In L. Ravelli & R. Ellis (Eds), *Analysing academic writing: Contextualized frameworks* (pp. 153–171). London: Continuum.

Cousin, G., & Deepwell, F. (2005). Designs for network learning: A communities of practice perspective. *Studies in Higher Education, 30,* 57–66.

Drury, H. (2001). Short answers in first-year undergraduate science writing: What kind of genres are they? In M. Hewings (Ed.), *Academic writing in context* (pp. 104–121). London: Continuum.

Eggins, S., & Slade, D. (1997). *Analysing casual conversation.* London: Cassell.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education, 88,* 915–933.

Fahy, P.J. (2002). Use of linguistic qualifiers and intensifiers in a computer conference. *American Journal of Distance Education, 16*(1), 5–22.

Fitze, M. (2006). Discourse and participation in ESL face-to-face and written electronic conferences. *Language Learning and Technology, 10,* 67–86.

Halliday, M.A.K. (1994). *An introduction to functional grammar* (2nd edn). London: Arnold.

Harwood, N. (2003). Person markers and interpersonal metadiscourse in academic writing: A multidisciplinary corpus-based study of student and expert texts. Unpublished PhD thesis, University of Kent at Canterbury.

—(2005). 'We do not seem to have a theory . . . The theory I present here attempts to fill this gap': Inclusive and exclusive pronouns in academic writing. *Applied Linguistics, 26,* 343–375.

Hewings, A. (2004). Developing discipline-specific writing: An analysis of under-graduate geography essays. In L. Ravelli & R. Ellis (Eds), *Analysing academic writing: Contextualized frameworks* (pp. 131–152). London: Continuum.

Hewings, A., & Coffin, C. (2006). Formative interaction in on-line writing: Making disciplinary expectations explicit. In K. Hyland & F. Hyland (Eds), *Feedback on ESL writing: Contexts and issues* (pp. 225–245). Cambridge: Cambridge University Press.

—(2007). Writing in multi-party computer conferences and single authored assignments: Exploring the role of writer as thinker. *Journal of English for Academic Purposes, 6*, 126–142.

Hewings, A., Coffin, C., & North, S. (2006). *Supporting undergraduate students' acquisition of academic argumentation strategies through computer conferencing*. Research Project Report for Higher Education Academy. Retrieved from www.heacademy.ac.uk/ourwork/research (accessed 12 May 2009).

Hewings, M., & Hewings, A. (2002). 'It is interesting to note that . . .': A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes, 21*, 367–383.

Hood, S. (2004). Managing attitude in undergraduate academic writing: A focus on the introductions to research reports. In L. Ravelli & R. Ellis (Eds), *Analysing academic writing: Contextualized frameworks* (pp. 24–44). London: Continuum.

Hunston, S., & Thompson, G. (Eds). (2000). *Evaluation in text*. Oxford: Oxford University Press.

Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam: Benjamins.

—(2000). *Disciplinary discourses: Social interactions in academic writing*. Harlow: Longman.

—(2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics, 34*, 1091–1112.

Jeong, A.C. (2006). The effects of linguistic qualifiers and intensifiers on group interaction and performance in computer-supported collaborative argumentation. *International Review of Research in Open and Distance Learning, 6(3)*. Retrieved from www.irrodl.org/index.php/irrodl/article/view/258/843 (accessed 12 May 2009).

Joiner, R., & Jones, S. (2003). The effects of communication medium on argumentation and the development of critical thinking. *International Journal of Educational Research, 39*, 861–871.

Kuo, C.H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes, 18*, 121–138.

Littleton, K., Faulkner, D., Miell, D., Joiner, R., & Hakkinen, P. (Eds). (2000). Special issue: Learning to collaborate, collaborating to learn. *European Journal of Psychology of Education, 15*(4).

MacDonald, S.P. (1994). *Professional academic writing in the humanities and social sciences*. Carbondale: Southern Illinois University Press.

Martin, J.R. (1989). *Factual writing: Exploring and challenging social reality* (2nd edn). Oxford: Oxford University Press.

Marttunen, M. (1997). Teaching argumentation skills in an electronic mail environment. *Innovations in Education and Training International, 34*, 208–218.

Montero, B., Watts, F., & Garcia-Carbonell, A. (2007). Discussion forum interactions: Text and context. *System, 35*, 566–582.

Myers, G. (2001). 'In my opinion': The place of personal views in undergraduate essays. In M. Hewings (Ed.), *Academic writing in context* (pp. 63–78). London: Continuum.

North, S. (2005). Different values, different skills? A comparison of essay writing by students from arts and science backgrounds. *Studies in Higher Education, 30*, 517–533.

North, S., Coffin, S., & Hewings, A. (2008). Using exchange structure analysis to explore argument in text-based computer conferences. *International Journal of Research and Method in Education, 31*(3), 257–276.

Ravenscroft, A. (2000). Designing argumentation for conceptual development. *Computers & Education, 34*, 241–255.

Ravenscroft, A., & Pilkington, R.M. (2000). Investigation by design: Developing dialogue models to support reasoning and conceptual change. *International Journal of Artificial Intelligence in Education, 11*, 273–298.

Schellens, T., & Valcke, M. (2004). Fostering knowledge construction in university students through asynchronous discussion groups. *Computers and Education, 46*, 349–370.

Schrire, S. (2006). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers and Education, 46*, 49–70.

Scott, M. (2004). *WordSmith Tools 4.0* [Computer software]. Oxford University Press, Oxford.

Sinclair, J.M., & Coulthard, R.M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils.* Oxford: Oxford University Press.

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

Tang, R., & John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes, 18*, S23–S39.

Wasson, B., Ludvigsen, S., & Hoppe, U. (Eds). (2003). *Designing for change in networked learning environments.* London: Kluwer.

Wells, G. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education.* Cambridge: Cambridge University Press.

Chapter 8

Stance, Interaction and the Rhetorical Patterns of Restrictive Adverbs: Discourse Roles of *Only, Just, Simply* and *Merely*

Maggie Charles

1   Introduction

In recent years there has been increasing recognition that the interpersonal dimension is of key importance in the construction of successful academic texts. Accordingly, much research on academic discourse has been concerned with the identification and analysis of interpersonal features. For example, work on evaluation or stance has been carried out by, among others, Hunston (1993, 1994), Hyland (1999, 2005), Charles (2003, 2007) and Biber (2006). Other researchers have examined the rhetorical strategies employed by writers to persuade or influence their readers (e.g. Coffin & Hewings, 2004; Dressen, 2003; Hoey, 2000; Koutsantoni, 2006; Shaw, 2003). Such studies show how these techniques can be distinguished and manifest themselves differently in different disciplines and genres. This concern with the effect of the text on its audience is taken further in the work of Thompson (2001), who reveals how written academic discourse can be seen more generally in terms of the interaction between writer and reader. This chapter addresses these three interpersonal aspects of academic discourse – stance, interaction and rhetoric – showing how they form a cluster of interlocking concerns that together shape the expression of the writer's persuasive purposes. Corpus techniques have an important role to play in the analysis, as they provide quantitative data for identifying items which are potential markers of the interpersonal discourse features studied.

One significant way of expressing interpersonal meanings is through the use of adverbs (Biber & Finegan, 1988; Hoye, 1997) and specific groups of relevant adverbs have been identified, including 'attitudinal disjuncts' (Quirk et al., 1985), 'modal adjuncts' (Halliday, 1994) and 'stance adverbials' (Biber et al., 1999). Following this line of research, stance adverbials have been the focus specifically of work on written academic discourse. Conrad and Biber (2000) compare their occurrence across registers and

find that academic prose has about half as many stance adverbials as conversation, but almost twice as many as news reportage. They conclude that there are substantial differences between registers in the use of stance adverbials, due to differences in the communicative purposes and production circumstances of each register. In a contrastive study of genres in academic discourse, Bondi (2002) examines a number of epistemic stance markers (e.g. *typically*, *generally*) in abstracts, research articles and textbook chapters in the discipline of economics. She shows not only that the frequency of the markers differs between the genres, but that they perform different functions in the argument.

This research has added greatly to our understanding of the category of stance adverbials; however much less attention has been devoted to the study of other types of adverb. As defined by Biber et al. (1999: 556), restrictive adverbs are a sub-category of circumstance adverbs; they 'emphasize the importance of one part of the proposition, by restricting the truth value of the proposition either primarily or exclusively to that part'. In the following example, taken from the materials science corpus described in the next section, the nature of the *discussion* is restricted and emphasized:

> *There will only be a limited discussion of the observations, because this will form part of the subject of Chapter V . . .* (Materials)

This chapter uses corpus methods to identify frequent restrictive adverbs and their associated patterns, while discourse analytic techniques are employed to shed light on the stance and rhetorical effect of the adverbs. I argue that restrictive adverbs perform an important interactive role, both in indicating the writer's stance and in structuring the discourse through the rhetorical patterns of which they form part.

2   Corpora, Method and Data

The data are drawn from two corpora of theses written by native-speakers: roughly 190,000 words (eight MPhil theses) in the discipline of politics/international relations and 300,000 words (eight doctoral theses) in materials science. Further details of the corpora may be found in Charles (2006). Using *WordSmith Tools* software (Scott, 1999), word lists were compiled for each corpus and examined in order to identify restrictive adverbs. The investigation covered only the most frequent adverbs, those which occurred with a combined frequency of over 20 instances per 100,000 words. This procedure led to the identification of four restrictive adverbs: *only*, *just*, *simply* and *merely*. Concordances were then made for each adverb and the lines were individually examined to eliminate instances of adverbs used

as modifiers as well as non-adverbial and non-restrictive uses of the search items. It should be noted, however, that there can be considerable overlap in the meanings constructed by an individual adverb. For example, there are two possible readings of the sentence below: (1) that nothing else happened to the sample apart from cleaving; and (2) that the sample was cleaved in a simple way.

> *The sample wafer was **simply** cleaved to obtain a specimen with smooth cross-sectional (110) surface prior to examination.* (Materials)

It is not possible to decide exactly which meaning the writer intended and indeed it may be that the aim was to convey both simultaneously. Thus indeterminate examples such as this were included in the study. Frequencies for the four adverbs were normed to 100,000 words and are presented in Table 8.1.

The higher figures in the Politics corpus may be explained by differences between the two disciplines in the construction of knowledge. Becher and Trowler (2001) note that knowledge in the social sciences tends to be concerned with particulars, while that in the natural sciences deals primarily with universals. It is likely, then, that the narrower focus required for the study of particular events and entities contributes to the higher use of restrictive adverbs in the Politics corpus. Further, as I will show in Section 4, restrictive adverbs play a considerable role in the construction of the writer's stance. Thus the higher frequencies in the Politics corpus also lend further support to earlier findings that the occurrence of stance features tends to be greater in social science than in natural science disciplines (Hyland, 1999).

It should be noted, however, that the frequencies of these adverbs are still substantial in the Materials corpus. In particular the figure for *only*, the most frequent restrictive adverb, is only about 20 per cent lower than in the Politics corpus. Indeed in both corpora, this adverb is among the most frequent of all the adverbs that appear. Given their frequency of occurrence, then, the functions and patterns associated with these adverbs would seem to merit further investigation; this forms the subject matter of the following sections.

Table 8.1   Frequency of restrictive adverbs per 100,000 words

Restrictive adverb	Politics	Materials
Only	143.7	118.0
Simply	23.7	12.3
Just	18.4	6.7
Merely	21.6	2.7
Total	207.4	139.7

3   The Restrictive Function of *Only, Just, Simply* and *Merely*

The prime function of these adverbs is to indicate that the truth value of a proposition is limited and in so doing, the adverb focuses attention on one element of the message (Biber et al., 1999). The restriction may apply at phrase or clause level, as shown in the following two examples. In the first, it is the noun phrase that is perceived as focused and restricted, while in the second, the force of the adverb is seen to apply to the whole clause:

> *For powder in tube superconductor fabrication this process performs more than **just** the function of shape alteration.* (Materials)

> *Moreover, clause 9 specified **only** that UNPROFOR was to take all necessary measures in reply to bombardments against the safe areas . . .* (Politics)

The use of a restrictive adverb also evokes a contrast with a wider proposition or entity, however. This can most clearly be seen when the two contrasting elements appear in the text, a phenomenon primarily signalled by the presence of negation. In the following example *only* indicates the restricted entity, while the wider element is marked by *not*:

> *It is of concern that **only** one set of <110> type dislocations are present, and **not** the usual two sets.* (Materials)

However, even when the less restricted element does not appear in the text itself, I would argue that the potential for this contrast is nonetheless evoked by the restrictive adverb. For example, the use of *simply* below implies that there is a wider set of other ways in which the properties of metals can be controlled. If *simply* were omitted, all sense of the existence of other possible methods would be lost:

> *The business of controlling the properties of metals by **simply** heating them up has been going on for a long time . . .* (Materials)

Similarly, *only* in the following example implies that there are other circumstances where the use of force could potentially occur, but would not be permitted:

> *As Article 51 of the Charter stipulates, the employment of force is **only** permissible in self-defence.* (Politics)

Thus the use of a restrictive adverb can be seen to evoke a contrast with a wider proposition or entity, which may or may not be explicitly referred to in the text itself.

4　The Stance Function of Restrictive Adverbs

4.1　Restrictive adverbs and the construction of stance

Given the contrast suggested above, the designation of a certain element as restricted comes about through a comparison of its scope with that of some other possible element. The outcome of this comparison is an evaluation, since the writer expresses an opinion or attitude concerning the proposition or entity he/she is discussing. Thus I would argue that restrictive adverbs are essentially evaluative and have an important role to play in the writer's construction of stance.

In the example below, the use of *merely* indicates the writer's attitude towards a *problem* in the research. The nature of this problem has changed but has not been solved, and since the purpose of scientific research is to solve problems, the implication is that this is an unsatisfactory outcome. Thus the use of the restrictive adverb stresses the writer's opinion that the shift is of limited value, as the problem remains unresolved:

> *As for micelles, it appears that the problem has **merely** been shifted to finding a method of producing monodisperse aerosol droplets.* (Materials)

The following example from Politics reveals the writer's negative stance towards one way of dealing with *the varied behaviour of the middle powers*. The use of the restrictive adverb *simply* indicates that *to avoid the question* will not provide an adequate response to the issue, an assessment which is further supported by the use of the verb *limit* in the following clause:

> *There have been several attempts to account for the varied behaviour of the middle powers. One way is **simply** to avoid the question: limit the middle power category to those states whose actions fit a particular 'middle power behaviour'.* (Politics)

Thus, although the inclusion of the restrictive adverb may be regarded as optional in grammatical terms, its function is to alert the reader to the fact that an evaluative judgement is being made and omitting it would remove a key indicator of the stance of the writer towards the information given.

In taking a stance in a text, however, writers also position themselves in relation to others. They engage in an interaction with their readers and indeed such interaction can be seen as a fundamental property of all texts (Hoey, 2001). Just as the intended readership of a text influences the stance taken by writers, so writers use stance to influence readers in their reaction to the text. For thesis writers, readers include not only supervisors and examiners but also other specialists in the field. I would argue that in order to satisfy such readers, the writer must construct the stance of an insider in the discipline, one who is

familiar with its norms, conventions and on-going debates. In order to do so, writers build consensus with the disciplinary community by evoking shared knowledge and expectations.

One of the ways in which this interaction is carried out is through the use of restrictive adverbs, which enable the writer to create a consensus over shared disciplinary knowledge, but to indicate that some limitation exists which would not be expected by the reader. In this way, writers take an insider stance, but show that their knowledge actually exceeds reader expectations. This, in turn, leads to the defeat or modification of those expectations and the substitution of the writer's own view. In Thompson's (2001) terms, these adverbs constitute an 'interactional resource'; they bring the reader into the text, constructing a dialogue which involves the reader in the development of the argument.

In the first example below, the implication is that the reader would expect to understand *safe area policy* in other ways as well, while in the second, *these equations* would be expected to apply elsewhere too. The use of the adverb *only* serves to counteract these expectations, with the result that the reader's view is aligned with that of the writer:

> . . . *the emergence of a safe area policy in Bosnia can **only** be grasped in the context of the continuing ground war* . . . (Politics)
>
> . . . *we must note that these equations **only** apply to the restraining forces.* (Materials)

With the use of restrictive adverbs, then, the writer not only constructs shared disciplinary expectations, but also reveals their limitations. In this way writers show themselves as particularly insightful, and therefore valuable, members of the discipline.

4.2 Restrictive adverbs, negation and stance

The construction of stance may be seen particularly clearly when the restrictive adverb is associated with negation. There are two main ways in which this occurs: either the restrictive adverb itself is negated or the negation and restriction occur as separate elements. First I will examine instances in which the restricted adverb is negated. With a normed figure of 34.2 per 100,000 words, the Politics corpus contains six times as many instances of this feature as the Materials corpus, which has a normed frequency of only 5.7. Although the scale of the difference here is large, these findings again reflect the expected higher incidence of stance features in a social science.

In almost all cases, the negated, restricted element is followed by an element of wider scope. Thus the negated adverb signals the start of a two-part pattern which structures the discourse and allows the reader to predict the

way in which the text will continue. The wider statement is mostly signalled by the correlative co-ordinator *but*, either with or without the adverb *also*. In both corpora *only* is the most frequent adverb used in this pattern, although in Politics, there are also appreciable numbers of *just* and *merely*. In Hoey's (2005) terms, these patterns form one of the 'textual primings' of the negated form of these adverbs:

> *This relation is **not only** important for operational considerations, **but** is **also** highly relevant to degradation behaviour.* (Materials)

The negation of the restrictive adverb indicates that the writer accepts the statement, as far as it goes, but that there is something more that has been left out. In the wider element, the writer then gives their own, broader and more complete view. In the following example, the writer accepts the determination of *Cpn*, but adds to it that of *Cni* and *Cpi*:

> *For silicon treated with dilute HF solution, **not only** can Cpn be determined, **but** so can Cni and Cpi . . .* (Materials)

This two-part pattern contributes directly to constructing interaction between the reader and the writer. The negated statement contains information that is shared by both reader and writer and hence conforms to reader expectations, while the wider element holds new or unexpected information, known only to the writer. Thus the reader's expectation is accepted, but then modified and the writer gives additional information which takes the reader further than anticipated. This is illustrated in the next example, where the reader is portrayed as expecting that *Mazzini* reversed *French values*. This expectation is accepted by the writer, but modified by the statement that Mazzini *incorporated* and *downgraded* the values, which is presented as information known to the writer, but surprising to the reader:

> *. . . Mazzini did **not simply reverse** what he saw as French values: he **incorporated** them in his system but **downgraded** their status.* (Politics)

The implication of the example below is that *gas-phase collisions* are, indeed, prevented, but that there is an additional, more surprising and hence more newsworthy effect: the limitation of *impurity inclusion*:

> ***Not only** does this prevent gas-phase collisions occurring elsewhere in the chamber, **but also** serves to limit impurity inclusion.* (Materials)

In all these examples, the writer first uses the negated restricted adverb to establish the consensus view, which is shared by others in the discipline.

He/she then moves on to show originality in the field by adding a new or unexpected element to disciplinary knowledge. However it is important to note that, although Biber et al. (1999) describe the two elements of this pattern as forming a 'contrast', they are not mutually exclusive; both the negated restricted statement and the wider statement remain true. In clause-relational terms, they form a matching relation of compatibility (Winter, 1994). In fact the contrast is between the reader's expectation and the writer's wider reality, rather than between the truth value of the two statements themselves.

I now turn to consider the pattern in which negation occurs separately from the restricted adverb. In such cases, there are still two closely connected elements, one restricted and one wider in scope, but the adverb itself is not negated:

> *Under such conditions there is **no** real co-operation in society. Society can **only** be held together by force.* (Politics)

This pattern is much less frequent than that with the negated adverb, but is worth mentioning, since its use constructs a rather different stance. The number of instances is again much higher in the Politics corpus at 10.0 per 100,000 words in comparison with a normed figure of 3.7 for Materials. *Only* is the most frequent adverb used. In this pattern, the negated element incorporates the reader's expectation, which is defeated through the negation and replaced by the writer's restricted assertion. Thus although the two statements are both true, the reader's expectation is rejected rather than modified and it is the restricted statement that represents the writer's view. In the following example, the reader is portrayed as expecting that *jus cogens rules* operate in the same way as *regular rules*. This expectation is defeated by the writer, who substitutes their own opinion in the final sentence:

> *Jus cogens rules **cannot** give way to new legal regimes in the same way that regular rules do: through violation, acquiescence and recognition. They **can only** be replaced by a new rule of the same stature.* (Politics)

In several examples in both corpora, this pattern is used metadiscoursally, specifically to draw attention to the writer's aims and intentions. Such instances show particularly clearly the role that these restrictive adverbs play both in constructing interaction with the reader and in managing the development of the discourse.

> *I take **no** view on whether the world role should have been maintained. I seek **only** to show that it could have been maintained . . .* (Politics)

In the example above, the reader's expectation is that the writer would have a *view on whether the world role should have been maintained.* This expectation is

defeated through the negation and the writer substitutes their own restricted aim, which is *to show that it could have been maintained.*

The use of restrictive adverbs, then, contributes to the construction of writer–reader interaction and the text can be seen as dialogic in the sense used by Bakhtin (1979/1986). However, it is important to note that the expectations implied or stated in the text are those that any member of the field would share. Thus the writer creates a consensus with the readers over disciplinary expectations, and shows that they can evaluate knowledge according to the norms of the discipline, determining which are the most newsworthy items and highlighting them by means of the two-part pattern.

5   Rhetorical Patterns of Restrictive Adverbs

The rhetorical patterns associated with clauses which contain a restrictive adverb can be seen by investigating the clause relations in which they participate. Winter defines clause relations as 'the shared cognitive process whereby we interpret the meaning of a clause or group of clauses in the light of their adjoining clauses' (1994: 49). Examination of the concordance lines shows that restrictive adverbs often co-occur with a marker of contrast or of reason/consequence in the same clause or in the immediately preceding or following clause. These markers provide evidence of the clause relations of matching contrast and logical sequence respectively and show how restrictive adverbs play a role in the development of extended arguments. I deal first with the relation of contrast.

5.1   **Restrictive adverbs with markers of contrast**

Markers of contrast found in conjunction with restrictive adverbs include: the co-ordinator *but*; linking adverbs (e.g. *however* and *yet*); and subordinators (e.g. *although, whereas* and *while*). Such contrast markers occur with normed frequencies of 30.0 per 100,000 words in the Politics corpus and 32.3 in Materials. In both corpora, the two most frequent patterns are *however* or *but* in the same clause as the restrictive adverb *only*:

> *Macmillan was liked and supported by his Party and the Foreign Office, **but** in the end was in the post for **only** a few months . . .* (Politics)

In the Politics corpus, many of the examples refer to political events and actions. Here the writer constructs a position that reveals their own view and which may also reflect the stance of those involved at the time. Thus in the example above, the writer indicates that the brief extent of Macmillan's term as Foreign Secretary was surprising.

In the Materials corpus, however, these two patterns are often used to indicate the writer's position towards their own work or that of other researchers. In particular, the majority of instances of *however* in the same clause as *only* are used with reference to the writer's own research and most of these indicate a shortcoming or ongoing problem:

> *The use of low accelerating voltage and beam current is necessary to maximise contrast levels sufficiently for the analysis of these materials to be performed.* ***However, only*** *the difference in p-dopant concentration could be determined in this investigation . . .* (Materials)

In such examples *however* generally occurs sentence initially, thereby linking the restriction signalled by *only* to a statement in the previous sentence. This two sentence pattern again enables the writer to interact with the reader by showing that they are aware of the disciplinary expectations before indicating that they are unable to fulfil them to complete satisfaction. Thus the first sentence of the example above shows that the writer knows what is required for the analysis, while the second sentence using *However, only* indicates their awareness that their work has certain limitations. A similar example is seen in the Politics corpus:

> *These results do indicate somewhat more scalability . . .* ***However,*** *there is still* ***only*** *one scale (privacy cases, VALUE=5) which would be accepted on Guttman's criteria . . .* (Politics)

Using this pattern, then, the writer constructs the stance of a competent researcher who is capable of evaluating their own work according to the standards of the discipline. A similar stance is also expressed using *but* with *only* in the same clause, although the limitation signalled by the adverb may not necessarily be a flaw:

> *The magnitude of this effect depends upon surface roughness and the orientation of the surface plane with respect to the crystallographic axes,* ***but*** *in this treatment* ***only*** *the dependence on dopant type is of concern.* (Materials)

This combination of *only* with *but* is also used to comment on the work of other researchers:

> *Kapustin has undertaken a thorough experimental survey of the effects of temperature on the voltage signal over the course of a field pulse* ***but*** *the author is aware of* ***only*** *one systematic assessment of the effect of applied transport current on these systems in pulsed magnetic fields (Rayroux et al. 1967).* (Materials)

In such examples, the initial clause puts forward information concerning others' research which is generally positive, while the contrasting clause signalled by *but* places a restriction upon this information and thereby diminishes its value or importance. Hence the restrictive adverb is used here in order to construct a gap or problem in the field and the contrast is between what has been achieved in the discipline so far and what still remains to be done. The following example from the Politics corpus, using *yet* together with the restrictive adverb *simply* operates in a similar way. The effect is of a rather grudging acceptance of the work of others, before the writer points out its limitations:

> *The most important further mechanism which must be clarified systematically is the precise role played by domestic dynamics in producing state socialization. Of the writers examined in this thesis, Armstrong goes the furthest in this direction; **yet** he **simply** notes that socialization often requires the presence of powerful domestic advocates of conformist policies.* (Politics)

However, by using this two-part pattern the writer shows that they are aware of and acknowledge other research and this makes it more likely that, when they subsequently draw attention to the existence of a gap in the field, their opinion will be considered acceptable by other members of the discipline.

There is one further pattern of note, which occurs only in the Materials corpus. Here the restriction signalled by *only* occurs in a clause which immediately precedes the *however* clause:

> *Contrast mechanisms failed to reveal the grain structure, and electro-etching was successful **only** with pre-treated specimens. **However** the pre-treatment did not alter the important features exhibited by the alloy.* (Materials)

In such examples, the restricted clause operates as a concession: it anticipates a criticism that the reader might make and serves to acknowledge its validity. In the subsequent contrast clause the writer moves to counter this possible criticism by stressing a positive outcome and thus affirming the value of the work. The occurrence of this pattern in the Materials corpus may be because it is used to refer to experimental limitations, which must be mentioned, but can be acknowledged without undermining the worth of the study:

> *The electron-beam current density was, however, **only** ~0.25 A cm-2 with the gas in. **However,** the features compare well with figure 3.1(1) . . .* (Materials)

By mentioning and then dealing with potential problems in their research in this way, writers construct an interaction with their readers and show that they are able to evaluate their own work according to the norms of the discipline. At the same time, this two-part rhetorical pattern offers writers the opportunity to respond to possible criticism by stressing the achievements of their research.

So far, then, we have seen how restrictive adverbs in conjunction with markers of contrast create rhetorical patterns that enable writers to position themselves in relation to their own work and that of other researchers in the discipline. In so doing writers construct the stance of a competent member of the field, whose judgements and opinions merit attention and acceptance from the disciplinary community.

5.2 Restrictive adverbs with markers of reason or consequence

In both corpora, mention of a restriction often occurs in conjunction with the provision of a reason for or consequence of this restriction, thus forming the clause relation of logical sequence (Winter, 1994). This is illustrated in the following example, where the information in the restricted clause marked by *just* leads to the consequence in the final clause signalled by *thus*:

> On this argument, Justices' votes **just** are less predictable in significant cases . . . and **thus** the composition of coalitions is, also, less predictable. (Politics)

Reasons or consequences can be signalled by a wide variety of means, including linking adverbs (e.g. *thus*, *therefore*) and subordinators (e.g. *because*, *as*, *since*), but also by complex prepositions (e.g. *due to, because of*) and lexical choices (e.g. *this results in*, *a possible reason*). As the concern here is to examine the relationship between the restriction and the reason/consequence, rather than to give a precise quantification of the phenomenon, the approach adopted was to examine only the subordinators and linking adverbs that co-occur with restrictive adverbs. It should be stressed, however, that this procedure leads to some underestimation of the extent to which restriction is associated with reason/consequence.

Although the occurrence of these reason/consequence markers in combination with restrictive adverbs is slightly lower than that of contrast markers, it is still substantial. Counting markers that occur in the same clause as the restrictive adverb or in the immediately preceding or following clause gives a normed frequency of 18.9 per 100,000 words in the Politics corpus, while in the Materials corpus the figure is higher at 28.7. In both corpora by far the most frequent restrictive adverb occurring in these combinations is *only*, with normed figures of 14.2 for Politics and 25.3 for Materials. It is probable that the higher figures recorded for the Materials corpus are due to the fact that the construction of knowledge in Materials proceeds by experimentation. Thus it is likely that there will be practical restrictions upon the circumstances or outcomes of the research that must be explained:

> The metal screening of the wires may **simply** serve to increase this thermal mass and **thus** damp out the temperature fluctuations due to boiling. (Materials)

In general terms, reason and consequence statements construct the logical relation of cause and effect and within this relation, the restriction marked by the adverb may function in one of two ways: as the cause of another phenomenon or as its effect. However, each of these possibilities is associated with several different grammatical patterns. In both corpora, the restriction occurs most often as the cause of another phenomenon and is most frequently realized by the use of the subordinator *as* with the restrictive adverb *only* in the subordinate clause:

> **As** the high temperature activation energy has been **only** crudely calculated from two data points . . . there is no justification in performing any minor correction. (Materials)

Such examples, in which the restriction and reason occur within the same clause, are often used in both corpora to refer to the writer's own work. In such cases, the restriction is given as the reason for a potentially questionable research decision. Thus in the example above, the writer anticipates that the reader may object to the fact that no *minor correction* has been performed and gives the crude calculation as the reason for this potential shortcoming. A similar pattern is seen with the subordinator *since*:

> However, **since** Cook's measure is limited to the Burger Court **only**, the focus of this thesis will be on Burger Court cases. (Politics)

By putting the restriction into the subordinate clause, the writer presents it as given information, thereby downgrading its importance in comparison with the statement of the research procedure, which appears in the main clause (Winter, 1982). Thus the distribution of information achieved through the use of this grammatical sequence serves to highlight the research procedure and make it appear well-justified. In this way, the writer again interacts with the reader, showing that they can anticipate and deal satisfactorily with potential criticisms of their work. This stance is particularly noticeable when, as in the examples above, the subordinate clause occurs before the main clause.

However the restriction may also function as a cause when it occurs in a main clause followed by linking adverbs such as *thus* and *therefore*. In the Materials corpus these are the next most frequent patterns that occur and again they tend to refer to the writer's own work:

> The magnitude of the screened field in the high temperature superconductors used in this work is **only** of the order of 5mT and **thus** an almost immediate change in area after the field reaches its maximum value is expected. (Materials)

Here, the restriction and its consequence appear to carry equal weight as they both occur in main clauses. These patterns are generally used just to explain

research phenomena, rather than to justify decisions that could attract reader criticism. Similarly, in the politics corpus such examples tend to refer to political events and actions:

> . . . *the realisation that Britain could no longer play such a role came* **only** *after Suez. The Suez crisis was* **thus** *a definitive moment in British involvement in the region . . .* (Politics)

There are fewer examples of restrictions functioning as effects. However, the Materials corpus makes use of the linking adverbs *therefore* and *thus* with *only* in the same clause:

> . . . *the first two experiments determined that the fibreboard was completely eroded in 5 minutes.* **Therefore** *all subsequent experiments were* **only** *performed for a 1 min duration.* (Materials)

Again most examples refer to the writer's own work, but this time it is the research decision that is limited in some way and the initial clause gives a justification for the restricted research action. The presence of the restriction is seen to demand some sort of explanation, which is simultaneously evoked and provided by the use of this two-part pattern.

However, it is noticeable that in both corpora, restrictions also occur as effects when the writer refers to the work of other researchers. In this pattern, the initial restricted clause points out a limitation in others' work, while the subsequent clause gives a reason for this limitation. Here the two-part pattern enables the writer to give an explanation for a gap or problem in the field and thereby to mitigate the criticism of other researchers that could be implied:

> *Even 'high' energy studies tend* **only** *to consider electrons with tens of keV of energy, such as 40 keV*[31]. *This is* **because** *at higher electron-energies, the various interaction cross-sections become much smaller.* (Materials)

One further rhetorical pattern may be distinguished, in which the restrictive adverb occurs in association with both a marker of contrast and of reason/consequence. In the most frequent pattern of this type the restriction provides the reason for the contrasting statement:

> *A redesign was considered* **but as** *failure* **only** *occurs once every six months it was not considered worth it.* (Materials)

> . . . *it is hard to see why . . . Justices vote with such ideological rigidity.* **However, since** *the results are based on a sample of* **only** *22 votes . . . they must be taken as preliminary.* (Politics)

These examples reveal a three-part rhetorical pattern consisting of an initial clause or clauses that give a research action or opinion, followed by a clause of contrast and a restricted clause of reason. However the reason, given in a subordinate clause, interrupts the contrast clause immediately after the marker and the pattern concludes with the contrasting information, which occurs in a main clause. Both the use of a subordinate clause and its front-position serve to downgrade the importance of the reason by presenting it as given information, already known to the reader.

This pattern again shows evidence of the interaction between writer and reader. The initial clause acknowledges the reader's expectations, but the subsequent contrast marker gives a clear signal that they will not be fulfilled. The restricted clause presents the reason as information the reader already shares, while the main clause gives the writer's view. The effect of the pattern is to let the reader know that the writer will counter their expectations and to construct a consensus as to the reason for this, before asserting the writer's own position. Thus the use of this pattern makes it more likely that the writer's point will be accepted. The restrictive adverb signals both an explanation and a justification for an aspect of the research which is open to criticism.

The use of restrictive adverbs in conjunction with markers of reason/consequence constructs a similar stance to that achieved by the combination of restrictive adverbs and contrast markers. In both cases writers take account of reader concerns and show themselves to be competent members of the discipline, capable of appropriate judgements and evaluations with regard to their own and others' research.

6   Conclusions and Pedagogical Implications

This study of restrictive adverbs has shown their particular importance as a key element in rhetorical patterns that consist of two, or even three, parts. It would seem, then, that mention of a restriction often leads to the need for a statement that can supplement or make up for that limitation. This tendency is evident both in patterns using the negated adverb as well as in those which include markers of contrast or reason/consequence. With the negated adverb, writers construct a consensus with the disciplinary community and then show how their own research takes the work of the discipline forward, while the use of markers of contrast or reason/consequence allows the writer to indicate gaps or weaknesses in their own or others' work and to provide an explanation or point out mitigating circumstances.

Essential to the way these patterns function is the interaction between writer and reader. A dialogue is set up in which one part represents the reader's view, incorporating their expectations as perceived by the writer, while the other modifies or even rejects this position and, in so doing, establishes the writer's

opinion as correct or appropriate. The ultimate purpose, then, of these patterns is to persuade the reader to accept the writer's position. However acceptance is made more likely by the acknowledgement of others' views. This recognition of others' opinions goes to the heart of what it means to be a successful academic writer. In order to achieve success it is necessary for writers to persuade others to accept their ideas. However, persuasion is not just about making a strong case for one's own view, but rather about taking account of others' views in the furtherance of one's own. Thus these patterns, in which the restrictive adverb plays a pivotal role, are seen to be fundamental to the construction and acceptance of knowledge and constitute an important element in the disciplinary and genre awareness that expert academic writers bring to bear upon the task of writing.

There is some evidence to suggest that student writers who are non-native speakers of English have difficulty with the use of adverbial expressions in academic writing (Hinkel, 2003). The study of restrictive adverb patterns can therefore be of value in several different respects. For example, the development of rhetorical awareness and competence can be fostered by working on patterns with contrast and reason/consequence markers in order to show how writers deal satisfactorily with weaknesses in their own work and how they draw attention to gaps in the field appropriately. In terms of discourse structure, the study of extended rhetorical patterns can also help students to see how texts are constructed through strategic choices that determine the way in which relatively long stretches of text are organized. Finally, focusing on the interactive aspect of the patterns would allow students to enhance their understanding of the writer's relationship with the reader, in particular the notion that the reader has certain expectations that require the writer's attention and response.

In terms of approach, this chapter has shown how the techniques of corpus linguistics and discourse analysis complement each other and can thus provide a particularly rich analysis of academic writing. First, the use of corpus methods enables the frequency of a linguistic item to be established, which gives an indication of its potential interest and significance. More importantly, the analysis of the regularities apparent in concordance lines allows the researcher to see how items occur within their environments and how they combine with each other. This draws attention to possible sites which can then be investigated in greater detail using discourse analysis methods. Further, by enabling access to expanded concordance lines or even the original source material, corpus software also facilitates work on longer stretches of text. Discourse analytic techniques can then be used to understand and explain the ways in which lexical, grammatical and rhetorical choices combine to form appropriate and persuasive disciplinary writing. Thus the combination of corpus linguistic and discourse analytic approaches offers an extremely powerful tool for gaining insight into the way in which texts are constructed and the means by which their persuasive purposes are achieved.

References

Bakhtin, M.M. (1986). *Speech genres and other late essays* (McGee, V.W., Trans.). Austin: University of Texas Press.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd edn). Buckingham: The Society for Research into Higher Education and Open University Press.

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*, 97–116.

Biber, D., & Finegan, E. (1988). Adverbial stance types in English. *Discourse Processes, 11*, 1–34.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Bondi, M. (2002). Attitude and episteme in academic discourse: Adverbials of stance across genres and moves. *Textus, 15*, 249–264.

Charles, M. (2003). 'This mystery . . .': A corpus–based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes, 2*, 313–326.

—(2006). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics, 27*, 492–518.

—(2007). *Argument* or *Evidence*? Disciplinary variation in the use of the noun *that* pattern in stance construction. *English for Specific Purposes, 26*, 203–218.

Coffin, C., & Hewings, A. (2004). IELTS as preparation for tertiary writing: Distinctive interpersonal and textual strategies. In L. Ravelli & R. Ellis (Eds), *Analysing academic writing: Contextualized frameworks* (pp. 153–171). London: Continuum.

Conrad, S., & Biber, D. (2000). Adverbial marking of stance in speech and writing. In S. Hunston & G. Thompson (Eds), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 56–73). Oxford: Oxford University Press.

Dressen, D. (2003). Geologists' implicit persuasive strategies and the construction of evaluative evidence. *Journal of English for Academic Purposes, 2*, 273–290.

Halliday, M.A.K. (1994). *An introduction to functional grammar.* (2nd edn). London: Arnold.

Hinkel, E. (2003). Adverbial markers and tone in L1 and L2 students' writing. *Journal of Pragmatics, 35*, 1049–1068.

Hoey, M. (2000). Persuasive rhetoric in linguistics: A stylistic study of some features of the language of Noam Chomsky. In S. Hunston & G. Thompson (Eds), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 28–37). Oxford: Oxford University Press.

—(2001). *Textual interaction*. London: Routledge.

—(2005). *Lexical priming*. London: Routledge.

Hoye, L. (1997). *Adverbs and modality in English*. London: Longman.

Hunston, S. (1993). Evaluation and ideology in scientific writing. In M. Ghadessy (Ed.), *Register analysis: Theory and practice* (pp. 57–73). London: Pinter.

—(1994). Evaluation and organization in a sample of written academic discourse. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 191–218). London: Routledge.

Hyland, K. (1999). Disciplinary discourses: Writer stance in research articles. In C.N. Candlin & K. Hyland (Eds), *Writing: Texts, processes and practices* (pp. 99–121). London: Longman.

—(2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*, 173–192.

Koutsantoni, D. (2006). Rhetorical strategies in engineering research articles and research theses: Advanced academic literacy and relations of power. *Journal of English for Academic Purposes, 5*, 19–36.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* London: Longman.

Scott, M. (1999). *WordSmith Tools* (Version 3.0) [computer software]. Oxford: Oxford University Press.

Shaw, P. (2003). Evaluation and promotion across languages. *Journal of English for Academic Purposes, 2*, 343–357.

Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics, 22*, 58–78.

Winter, E.O. (1982). *Towards a contextual grammar of English.* London: Allen & Unwin.

—(1994). Clause relations as information structure. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 46–68). London: Routledge.

Chapter 9

A Dialogic Account of Authority in Academic Writing

Ramona Tang

1   Introduction

The issue of how and whether authority is projected in student academic writing has received significant attention over the years from researchers and teacher-practitioners alike (e.g. Bartholomae, 1985; Brannon & Knoblauch, 1982; Greene, 1995; Hyland, 2002; Ivanič, 1995, 1998; Koutsantoni, 2006; Peritz, 1993; Starfield, 2002; Thompson, 2005). The continued efforts made to understand this particular quality of academic writing is, I suggest, an indication of our recognition within this field that 'authority' is a crucial element of good academic writing. Indeed, as Starfield (2002: 121) has argued, the ability to 'construct a powerful, authoritative textual and discoursal identity' is central to the success of student writers.

This chapter, then, seeks to complement and extend existing ways of exploring authority in student academic writing by proposing another way of framing our discussion of this phenomenon. Drawing on the Bakhtinian notion that all language is dialogic in its orientation, and borrowing terminology from the branch of APPRAISAL[1] research referred to as ENGAGEMENT (Martin & White, 2005; White, 2003), I show how a unified and varied exploration of 'authority' can be facilitated through the adoption of a dialogistic perspective of student academic writing. In particular, this chapter demonstrates how student writing may be viewed as sites of dialogue not only between the student writer and a specific tutor-reader, but also between the writer and the wider disciplinary community, and one key idea that will be developed here is that 'authority' in student writing is associated with those writers who succeed in maintaining their voice as 'dominant' within the heteroglossic diversity typical of academic discourse.

2   'Authority' in Academic Writing: A Brief Review of the Field

The term 'authority' as it relates to academic writing has been used in various ways in the literature. One major strand of research sees 'authority' as

a linguistic construct, created and negotiated by writers in their writing. Specifically, an academic writer's ability to demonstrate his/her familiarity with the conventions and practices privileged within his/her disciplinary discourse community is seen as central to success in the academy and the conveyance of authority in writing. (See, for instance, Harris, 1987; Hendricks & Quinn, 2000; Ivanič & Simpson, 1992; Krause, 2001; Lillis, 1997, 2001; Read, Francis & Robson, 2001; and Thompson, 2005, who focus on novice undergraduate writers; Kirsch, 1993, and Matsuda & Tardy, 2007, who deal with faculty, advanced 'enculturated' students, and those seeking to publish in academic journals; and Delpit, 1986 and Sunderland & Barton, 2001, who are concerned with researchers crossing from one academic culture into another.)

A second, related, and major strand of research sees 'authority' as associated with the extent to which a writer presents himself/herself as being an 'author', a 'maker of meaning' (Ivanič, 1994: 12), a social actor who 'owns' his/her writing and takes responsibility for the ideas expressed within. The discoursal construction of this second type of authority has also been addressed in numerous studies. Greene (1995: 187–188), for instance, appears to have this type of authority in mind when he writes that '[t]he source of an author's authority derives from an ability to create and support his or her vision'. Tang and John (1999) argue that some uses of the first person pronoun in student writing construct more 'authority' than others because they front writer roles associated with the origination of ideas. Also focusing on first person pronouns, Hyland (2002) has explored how and whether L2 student writers and published academics in a range of hard and soft disciplines use such pronouns 'to establish a stance towards their propositions, to get behind their words and stake out a position' (1094). Based on the premise that first person pronouns are 'a powerful means by which writers express an identity by asserting their claim to speak as an authority' (1093–1094), Hyland's study, which reveals that the writers in his published corpus are more likely than those in his student corpus to use the first person while making interpretations and claims, offers support for the idea that textual authority is associated with the presentation of oneself as an active 'maker of meaning'.

Finally, a third major strand of research into 'authority' in academic writing views it not as a construct created by writers through their writing, but as a kind of autonomy and entitlement bestowed upon writers by their readers. Thus, Greene (1995), for instance, has addressed the question of who it is who 'authorizes' a student text. Authority in student writing, according to Greene, is 'always provisional, depending not only on the authors' ability to develop intellectual projects of their own, but also upon the authorizing principles that exist in the social structures of schooling and the conventions of academic inquiry' (188). It is argued, in other words, that teacher-readers need to *allow* students to have authority, and this may involve teachers responding to student texts not only in ways which position students as the rightful authors of their own texts (Tang, 2000), but also in ways which empower students by

making known to them the mechanisms by which the academic discourse community works (Hyatt, 2005). (For more on this third sense of 'authority', see Berkenkotter, 1984; Lutz & Fuller, 2007; and Raymond, 1993.)

This, then, has been a brief overview of some of the ways in which the notion of 'authority' in academic writing has been understood over the years. It is against this backdrop that the following framework for conceptualizing authority in academic writing is offered.

3.   Laying the Groundwork for a New Account of 'Authority': Dialogism and ENGAGEMENT

While exploring academic writing through the lens of dialogism is not without precedent (see, for instance, Lillis, 2003), the explicit pairing of the notion of authority in student writing with the notion of dialogism is a much newer idea. In this section, I introduce in brief the Bakhtinian concept of dialogism, including how it features within the analytical framework of ENGAGEMENT (Martin & White, 2005; White, 2003). This brief theoretical account serves to underpin the dialogic account of authority which will be developed in the following sections.

The Bakhtinian view of language as being inherently dialogic is, by now, familiar to most of us. For Bakhtin, all utterances contain what he calls 'dialogic overtones'. Even utterances which are ostensibly monologic (e.g. written sentences in an academic essay) are 'dialogic' in that they exist 'against the background of other concrete utterances on the same theme, a background made up of contradictory opinions, points of view and value judgements . . . pregnant with responses and objections' (Bakhtin, 1981: 281). All language users, in other words, whether speakers or writers, are constantly engaged in a dynamic process of responding to or anticipating the real or imagined utterances of others. This notion of dialogism is one of the basic tenets underlying the account of 'authority' to be set out in this chapter. In essence, we will see that 'authority' in student writing can be made the subject of critical discussion by focusing on the ways in which student writers manage this dynamic process of engaging with the real and projected utterances of others.

The Bakhtinian notion of dialogism also underlies the system of ENGAGE-MENT (Martin & White, 2005; White, 2003), an analytical framework which has been developed in recent years within the field of Systemic Functional Linguistics, and some of whose metalanguage, I suggest, can very usefully be employed to facilitate a discussion of authority from a dialogistic perspective. Central to ENGAGEMENT is the Bakhtinian notion of heteroglossic diversity, the notion that there always exists a variety of alternative positions one might take up with respect to a single phenomenon. All utterances, therefore, are made and understood in the context of such alternatives. The system of ENGAGEMENT,
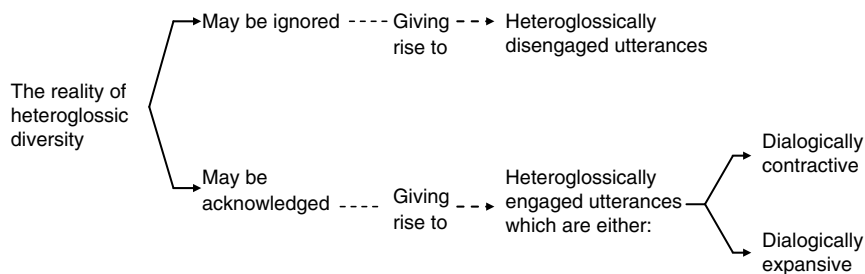
FIGURE 9.1   Basic underlying premise of ENGAGEMENT

then, is a systematic semantic map of the different ways in which 'heteroglossic diversity' might be handled in discourse. For instance, are the possible views of others *acknowledged* or *ignored*? If acknowledged, are they *embraced* (through dialogically *expansive* wordings) or *dismissed* (through dialogically *contractive* wordings)? If embraced, how are the alternative views embraced? And, if dismissed, how is the dismissal done? Figure 9.1 represents the basic underlying premise of ENGAGEMENT.

From among the many categories of heteroglossic engagement identified by White (2003), the following, I suggest, are particularly useful to have in our metalinguistic toolkit as we develop a dialogic account of authority.

3.1   Dialogically expansive strategies

With these strategies, writers acknowledge the existence of viewpoints/ positions which are different from those which they are forwarding in their text, and also embrace the viability of these diverse viewpoints, hence 'expanding' the potential for dialogue.

- POSTULATE – Writers present their viewpoints not as foregone conclusions, but as possibilities. (e.g. *Jane Austen may be one of the most well-loved writers of all time.*)
- EVIDENTIALIZE – Writers convey that it is only the evidence currently at hand, or the surface appearance of things which has led them to take up the positions they have. (e.g. *It appears from our survey that Jane Austen is an extremely well-loved writer.*)
- HEARSAY – Writers present a proposition as something voiced by unnamed others. (e.g. *Some say that Jane Austen is one of the most well-loved writers of all time.*)
- ACKNOWLEDGE – Writers attribute the ideas being put forward to a named person or persons in a neutral manner. (e.g. *Smith (2005) states that Jane Austen was a very savvy writer.*)

- DISTANCE – Writers attribute the ideas being put forward to a named person or persons in a distancing manner. As with the preceding strategy 'ACKNOWLEDGE', because the introduced proposition is ascribed to a specific Other, its inherent subjectivity is foregrounded, and it follows that there is room for other viewpoints to enter into the dialogic mix. (e.g. *Smith (2005) claims that Jane Austen was a very savvy writer.*)

3.2   Dialogically contractive strategies

With these strategies, a writer acknowledges viewpoints contrary to the one being forwarded, but does so in a way that dismisses them as less viable (White, 2003). Because one particular position is being privileged over its alternatives, the dialogic potential is contracted.

- PRONOUNCE – Writers emphasize their personal backing of a proposition. The heightened personal investment attached to the proposition raises the 'interpersonal cost' (White, 1998, chapter 3) to a reader of challenging it, as any disagreement could be deemed a personal attack. (e.g. *I believe Jane Austen was a very skilful writer.)*
- Signal CONCURRENCE – Writers present a proposition as uncontentious. This makes it difficult for readers to disagree since any deviation on a reader's part from the forwarded proposition would mark that reader as being in some way not 'normal'. (e.g. *Of course, Jane Austen was a very skilful writer.*)
- ENDORSE – Writers align themselves with and endorse a named, possibly more authoritative, source. (e.g. As Smith (2005) argues, Jane Austen was a skilful writer.)

A fuller account of these and other ENGAGEMENT resources can be found in Martin and White (2005) and White (2003). In the interest of ensuring the accessibility and pedagogic functionality of the current approach to exploring 'authority' in student writing, I suggest that this slightly simplified set of ENGAGEMENT strategies serves well enough to offer us the additional metalanguage with which to talk about the ways in which students enact the dialogic negotiations in their writing. This, taken together with the general Bakhtinian understanding that all language is inherently dialogic in its orientation, provides us with the basis from which to embark on an exploration of 'authority' from a dialogistic perspective.

4   Exploring 'Authority' from a Dialogistic Perspective

The following sub-sections set out a conceptually integrated account of authority in student writing which is centred around the notion of 'dialogue'

or 'dialogism'. Focusing specifically on sites of heteroglossic engagement in student writing, I show how such textual locations can be read as sites of 'dialogue' on two levels: (i) between the writer and a specific tutor-reader, and (ii) between the writer and the wider disciplinary community. The text extracts used for exemplification here are from a corpus of 30 academic essays of approximately 1,000 words each, written by English Language majors for a first year undergraduate linguistics module offered by a university in Singapore. Collectively, the extracts exhibit a variety of different possible modes of dialogic negotiation by students, and I will demonstrate how the notion of dialogism and the metalanguage of ENGAGEMENT can facilitate a critical discussion of 'authority' in them.

4.1   The unique dialogicality of undergraduate essays

The undergraduate essay is inherently dialogic. Essay prompts are often presented as questions eliciting an opinion (e.g. *Do you agree with this view expressed about accents of English?)* or commands requiring a linguistic response (e.g. *Argue for or against the view that standard languages are* '*weapons of social class*'), and students are often described as 'answering essay questions'. The dialogism displayed by students in their essays, then, always operates on two levels.

On one level, the dialogism displayed by students must be interpreted as operating with respect to the demands of a specific reader, the lecturer/tutor who formulated the 'question' and who will be reading its 'answer'. On another level, however, the dialogism displayed by students can also be interpreted as operating with respect to an abstract readership. On this second level, a student writer's heteroglossic engagement can be seen as a signal of his/her participation in the 'dialogue' of a wider discourse community.

As mentioned earlier, an important theoretical premise underlying the idea of dialogism is that all utterances exist 'against a background of other concrete utterances on the same theme, a background made up of contradictory opinions, points of view and value judgements . . . pregnant with responses and objections' (Bakhtin, 1981: 281). Utterances which acknowledge the existence of such a multiplicity of viewpoints are regarded as heteroglossically engaged or dialogic utterances, while utterances which do not acknowledge the existence of viewpoints other than the one being forwarded are regarded as heteroglossically disengaged or monologic.

Given that heteroglossically engaged utterances, by definition, acknowledge the existence of viewpoints other than the one being forwarded, we might say that heteroglossically engaged utterances serve to situate a text within a wider 'dialogue' comprising related alternative positions. We could argue, therefore, that writers who use heteroglossically engaged utterances, whether they choose to embrace the multiplicity of viewpoints implied in them or discount it, can be seen as recognizing that their writing does not

exist in a vacuum, but in fact contributes to a wider on-going 'dialogue'. Such writers are, in other words, positioned as participating members of a wider discourse community. The exact nature of their participation within this community, we could then argue, will depend on the kind of dialogism employed in their texts.

That these two levels of dialogicality operate at the same time in undergraduate essays can be illustrated by considering the following extract from the start of an essay written by a student, Yee Hwee, in response to the prompt *Argue for or against the view that standard languages are 'weapons of social class'*:

[1] [1.1] *Ferguson defines a standard language as 'the process of one variety of a language becoming widely accepted throughout the speech community as a supradialectal norm – the 'best' form of the language – rated above regional and social dialects'.* [1.2] *Yet David Crystal perceives it as 'a standard variety that cuts across regional differences, providing a unified means of communication'.* [1.3] *Here, we are presented with two rather contrasting views on standard languages.* [1.4] *A standard that is considered as being the 'best' form as proposed by Ferguson need not necessarily mean that it is the commonest, most widely used form.* [1.5] *It may be used to separate, or even distance, classes of people in the quest for exclusivity and status.* [1.6] *Yet, in Crystal's definition of a standard language, he chooses to highlight the unifying function of standard languages.*

[1.7] *Indeed, a standard language appears to close up the communication gap across the classes.* [1.8] *However, I strongly believe that the evidence for the separatist function of a standard language far exceeds its unifying function.* (Yee Hwee)

Reading this as a response to a specific question prompt, the dialogism Yee Hwee displays can be interpreted as operating with respect to the demands of her tutor-reader. Reading this as a (simulated) contribution to the field, however, the dialogism displayed can be interpreted as operating with respect to an abstract discourse community. For instance, when she (in the terms of ENGAGEMENT) ACKNOWLEDGES Ferguson's and Crystal's views on standard languages (in lines 1.1 and 1.2), she is not merely situating herself in the midst of disciplinary differences in opinion; she is at the same time responding to the question prompt by providing one argument for and one argument against the proposition that standard languages are 'weapons of social class'. When she writes that *a standard language <u>appears to</u>* (EVIDENTIALIZE) *close up the communication gap across the classes*, she is not only allowing for a disciplinary community whose members may have contradictory opinions; she is in effect setting up her argument in response to the requirement that she argue either for or against a particular proposition. And when she PRONOUNCES *I strongly believe that the evidence for the separatist function of a standard language far exceeds its unifying function*, this is dialogic both in the sense that she is making a particular contribution to the

'disciplinary dialogue', and in the sense that she is explicitly responding to the demands of the question prompt, 'answering' the question posed by her tutor.

Heteroglossically engaged utterances in undergraduate essays, then, can be interpreted as being dialogic on two levels, and we shall see that an exploration of both types of dialogicality can offer us insights into how 'authority' is negotiated in student essays.

4.2   The undergraduate essay prompt: writers in dialogue with their teachers

Question prompts position students as having to 'respond' to an external Other. The undergraduate essay, then, is generically geared towards *teachers'* agendas, not the agendas of the students writing them. This is a largely inescapable aspect of this type of writing, and does not in itself contribute to or subtract from the possible authority in it. However, what does have implications for textual authority is how students deal with this contextual reality in their writing, how (and whether) they assert themselves in staging this unique form of dialogue. Consider the following opening sentences from two student essays:

[2] QUESTION PROMPT

'I find the whole fuss about accents unnecessary. We should focus on what people say instead of how they say it' (Dr Tony Hung in the *Sunday Times*, 11/8/96). Do you agree with Hung's view expressed about accents of English? Why then are some English speakers (whether in or outside Singapore) so conscious about accents?

OPENING SENTENCES FROM TWO ESSAYS

It is certainly strange that something as frivolous and insignificant should receive so much attention. (Pei Gee)

To a certain extent, I would agree with Hung that the fuss about accents is unnecessary and that what is more important is the content of people's speech. (Pow Hong)

In these examples, the reader is positioned to read the essays not as discussions initiated by these writers, but as student responses to a question posed by someone else. This is evidenced by Pei Gee's omission of the object of her evaluation (accents) from her introductory sentence, and the abrupt way in which Pow Hong proclaims her agreement with a person she has not yet written about. Borrowing from the terminology of Exchange Structure analysis (e.g. Sinclair & Coulthard, 1975), we might say that these writers have begun with a *Respond* move rather than an *Initiate* move. This behaviour, I would

argue, has the potential to undermine textual authority, because it reinforces the notion that undergraduate essays are concerned with engaging with the *teacher's* pre-set agenda.

Some might suggest that it is not unacceptable, given the genre and the wording of the question prompt, for students to start their essays with a *Respond* move. While this may be true, it does not change the fact that student writers have the choice of linguistically positioning themselves in their texts as either responders or initiators. Positioning themselves as responders rather than initiators of the 'dialogue' they are engaged in, Pei Gee and Pow Hong in effect write themselves into a teacher-centred setting in which they have little say over what gets discussed. This way of managing the inherent dialogicality of undergraduate essays is in sharp contrast to that modelled in the following example, where the writer, Tad, clearly attempts to foreground the centrality of his own agenda:

[3] QUESTION PROMPT

Argue for or against the view that standard languages are 'weapons of social class', the preserve of the elite, with particular reference to the development of Standard English.

OPENING PARAGRAPHS FROM TAD'S ESSAY

[3.1] *If we ever take note of the languages we know, there always seems to exist a notion of a standard or 'center' which is generally associated as the 'proper' way to speak or write.* [3.2] *It is inevitable that English, one of the most popular and varied language in the world share this common trait.* [3.3] *However, we should take note that given the long history of the English language, this norm changes century-to-century and region-to-region.* [3.4] *In this essay, we shall attempt to survey the main trends of development of the Standard English in history and see if there is any basis for the claim that it is a 'weapon of social class' at any time of any place.* [3.5] *For completeness, we shall look briefly into cases of other languages for comparison with the main focus on English.*

[3.6] *Before we start, it remains for us to draw the lines of analysis – namely along the aspects of grammar, lexis, phonology, reasons for language change and the distinction of language use.* (Tad)

Although Tad makes occasional references to the question prompt, there are clear linguistic indicators pointing to the fact that he is trying to carve out for himself an area for investigation. We note that he does not start his essay *in medias res* as Pei Gee and Pow Hong have done, but instead carefully introduces the premise of his writing (3.1–3.3), his aim (3.4), and the direction he will take (3.5–3.6). In the second paragraph as well, he positions himself as an initiator, someone who dictates the issues to be considered in his essay.

We see, then, that textual authority can be either undermined or enhanced by the way a student chooses to 'answer' an essay question. Certainly, the contextual reality surrounding undergraduate essays may dictate that the teacher's agenda be addressed, and the way in which some question prompts are formulated seems to position students as 'responders' instead of 'initiators'. (Pei Gee and Pow Hong's essay prompt, for instance, seems to do this in a way that Tad's does not.) However, despite these external pressures, I would argue that it is possible for students to choose *through discoursal means* to position themselves within their writing as either prioritizing the teacher's agenda, or pursuing their own. As Koutsantoni (2006: 21) notes, 'even individuals thought to possess lower status and power can contest, dispute and challenge the roles assigned to them by others'. The different modes of response to question prompts, then, impact on the authority projected in student texts, and exploring the choices made by students in staging this written 'dialogue' with their tutors is, I suggest, one entry point for discussing the construction of textual authority.

4.3   Writers in dialogue with the wider disciplinary community: identifying two interpersonal moves central to authority in student writing

In this section, I shift from looking at heteroglossically engaged utterances as signals of a direct response to a question prompt to looking at them as signals of a writer's participation in the dialogue of a wider discourse community. I consider two extracts here, one each from Sue's and William's essays, with the aim of demonstrating yet another way of applying the dialogistic perspective to the study of 'authority' in student essays. Both extracts address the topic of accent-consciousness, and touch on the different levels of prestige attached to different accents. This is from Sue's essay:

[4] [4.1] *There are many English speakers who are very conscious about the kind of accents they are using.* [4.2] *There is a general misconception that a particular accent will put a person in a position seen as more 'prestigious' than the rest.* [4.3] *'Prestigious' as in widely accepted by the famous people in our community and or commonly used by the intellectuals.* [4.4] *It may also be due to the different values that each community possesses which cause some accents to be admired, approved and some, despised.*

[4.5] *We cannot deny the fact that an accent may enable you to be 'looked up' to because our society seemed to have proven this theory.* [4.6] *However, what is the worth of an accent if the speaker delivers a speech with no substance or content in it?* [4.7] *Furthermore, if the listeners are not comfortable with his accent, or are unable to comprehend what he is saying, he has already defeated the primary purpose of delivering such an oration.* (Sue)

We note that Sue's opinions are very evident in this extract, and that she is fairly assertive (she uses many dialogically contractive wordings). Note, for instance, her PRONOUNCEMENT in 4.5 (*We cannot deny the fact . . .*) and her assumption of CONCURRENCE in 4.2 (*There is a general misconception . . .*), 4.5 (*We cannot deny . . .*), and 4.6 (*what is the worth of an accent . . .?*).[2] These, together with her dialogically expansive wordings – *seemed to have proven* (EVIDENTIAL), *it may also be due to . . .* (POSTULATION), *an accent may enable you . . .* (POSTULATION) – make this extract heavily dialogized. If we accept that the use of heteroglossically engaged utterances positions one as participating in a wider 'dialogue' (as suggested earlier), then it seems clear that Sue is participating actively in a wider discourse community.

However, if we look closely at Sue's dialogic utterances, we might ask just who she is constructing as her dialogic partners. We see, for instance, that her argument is made to rest on EVIDENCE from 'society' (*our society seemed to have proven this theory*), and that the proposition she expresses as CONCURRENCE (*what is the worth of an accent if the speaker delivers a speech with no substance or content in it?*) appears to be grounded in nothing more than common sense. Further, that which she initially expresses as HEARSAY in 4.2 ('*prestigious*') is later revealed to have roots in the views of 'famous people' (4.3).

While setting up a reasonable argumentative rhythm, then, Sue appears to be positioning herself and her readers as lay people, rather than as members of the linguistics community. Thus, although this extract is heavily dialogized, and although this positions Sue as participating in a wider discourse community, we could say that the kind of dialogism she displays positions her as dialoguing within the wrong discourse community, and this undermines her discoursal authority. Contrast now the following extract from William's essay:

[5] *An accent portrays a certain educational background or the social class of a person. In England the 'prestige' accent is known as Received Pronunciation, or RP' (Crystal, 1990). Hence, a person who wanted to be associated with 'prestige' in England would most probably use RP. . . .*

'*In New York City, the non-prevocalic /r/ is associated with high social status and prestige' (Graddol, Leith, Swann, 1996). The use of the non-prevocalic /r/, also known as rhoticity, is a feature of certain accents. Labov carried out a study in New York to assess whether people's speech was rhotic or non-rhotic, according to their social class. He used a range of different styles of speaking, from very casual to very careful, and found out that 'lower middle-class informants use a higher proportion of non-prevocalic /r/ in the formal reading situations than the speakers in the social groups above them: the upper middle class' (Graddol et al., 1996). Labov went on to conclude that this was because the lower middle class speakers were conscious of the prestige value of rhoticity and wanted to 'be accepted and recognised as members of the upper middle class' (Graddol et al., 1996). This experiment clearly showed that some people were conscious of their accents because it reflected on their social background.* (William)

Here, it is clear that William, unlike Sue, has situated himself within the 'conversations' of his discipline. The terminology he employs (*non-prevocalic /r/, rhoticity*) is specialist terminology, and although by no means flawlessly executed, his argument that different accents may enjoy different levels of prestige is suitably heteroglossic, being supported by external voices from the field of linguistics (Crystal, and Graddol et al.), and by his reference to Labov's linguistic experiment.

Crucially, in the midst of this dialogicality, William's voice is not lost. Most notably, he puts forward a proposition at the end (*some people were conscious of their accents because it reflected on their social background*) which relies for its evidence on the linguistic study he has just cited (*this experiment clearly showed . . .*). This move effectively reasserts control over the discourse, preventing the paragraph from ending on an external voice.

Looking at the discoursal behaviours exhibited in Sue's and William's essays, and drawing on the senses of 'authority' identified in the research literature, I want to argue that two kinds of interpersonal moves work together to construct 'authority' in student essays:

a. the assertion of a writer's voice (encompassing his/her opinions, stance, and agenda); and
b. the locating of that voice within the ongoing or past 'conversations' of his/her discipline.

In other words, authority in student writing might be interpreted for the purposes of critical discussion as being the product of a particular manner of engagement with disciplinary ideas, a manner of engagement which positions a writer as being a *contributing participant* in the intellectual exchanges of his/her disciplinary community. Such a view of authority is supported by Hyland, for instance, who has observed that '[w]riting in the humanities stresses the individual creative thinker, but always within the context of a canon of disciplinary knowledge' (2000: 37).

One way for discourse analysts and teachers of academic literacy to have an analysis-grounded discussion of 'authority' in student essays, then, is to draw on the notion of dialogism and the metalanguage of ENGAGEMENT to explore texts for (a) how writers assert their own voices and (b) how they display, through heteroglossically engaged utterances, participation in a wider discourse community. Either of these interpersonal moves on its own, I suggest, offers only a partial picture of what it means to construct 'authority' in an academic text. We have seen that it is possible for writers to assert their voices while dialoguing in the wrong discourse community, and I have suggested that this undermines discoursal authority. Similarly, it is not difficult to imagine writers positioning themselves within the right discourse community (by providing a detailed review of relevant literature, for instance), but not making

clear what their own contributions to the disciplinary conversation are. This, clearly, would also undermine discoursal authority.

In the following sub-sections, I demonstrate in greater detail how a critical discussion of authority can be centred around the exploration of these two interpersonal moves by offering some further examples of different kinds of dialogic negotiations in student essays.

4.3.1   Yee Hwee: evaluating prevailing positions within the disciplinary debate

Consider again Example [1] from Yee Hwee's essay. At the start of the extract, Yee Hwee ACKNOWLEDGES the views of Ferguson and Crystal on standard languages (1.1–1.2). We note however that she not only sets up a site of dialogue in her text where Ferguson's voice is pitted against Crystal's, but also demonstrates her readiness to participate in the dialogue herself, confronting both the external voices (in 1.3) with her assessment of how they relate to each other. Sentences 1.4 and 1.5, expressed as POSTULATIONS (*need not necessarily mean, may*), represent a further engagement in the disciplinary dialogue as Yee Hwee offers her own take on a position already prevailing within the community. Finally, when Yee Hwee in 1.8 explicitly PRONOUNCES her position, we can interpret this as being her overt statement as to what her contribution to this disciplinary debate is to be.

The progression from the dialogic expansiveness of Yee Hwee's initial ATTRIBUTIONS and POSTULATIONS through to the dialogic contraction of her final PRONOUNCEMENT is in fact a move common in the staging of academic arguments, and I would suggest that such an analysis of the dialogic moves in a text can form the basis of a critique of whether or not a student text conveys authority.

4.3.2   Cowan: representing one side in a disciplinary debate

The following is an extract from Cowan's essay, which addresses the issue of whether New Varieties of English (NVEs) can be likened to battlegrounds upon which one set of norms is challenged by another set of counter norms:

> [6] [6.1] *I therefore propose that this playing field of norms versus counter norms is not a lamentable scenario.* [6.2] *Detractors are inclined to term NVEs as 'lesser forms of the language, the results of deviations from the original, proper/correct language, which are unsystematic, random collections of errors, aberrations, etc., made by those who do not know the rules and lack proficiency.' (T. Kandiah, Encounters with the English Language)* [6.3] *Firstly, it is a fallacy to label NVEs as rule-less, because there ARE rules to follow, the consequences of erring would be the same ostracization accorded*

by fellow users just the same if the language in contention was standard English. [6.4]
It is vital to recognize that NVEs do have rules of the linguistic nature just like any
other language forms. [6.5] *In the areas of syntax and grammar there are guidelines*
for usage which are very much entrenched in everyday speech . . . (Cowan)

Here, again, a dialogistic perspective can facilitate a discussion of textual
authority. Like Yee Hwee in the previous example, Cowan constructs the dis-
ciplinary community as being in dialogue. However, unlike Yee Hwee, who
stages her dialogue by introducing two external voices, Cowan stages his dia-
logue by *himself* taking up a position contrary to that held by some members of
his community. This, we could say, is a highly authoritative move as it positions
Cowan as being an integral part of the disciplinary debate. He is, in effect,
what makes the debate possible.

Cowan's staging of his contribution to the disciplinary conversation can be
further analysed. His PRONOUNCEMENT in 6.1 makes clear what his position is.
That he intends to stage a 'debate' to refute what he sees as prevailing miscon-
ceptions about NVEs is signalled with the first word of sentence 6.2 – *detractors*.
This evaluative term positions Cowan as DISTANCED from the external voice he
is about to introduce. With 6.1 and 6.2, then, the scene is set for a dialogue
between Cowan and the 'detractors'. In 6.3, Cowan begins his systematic refuta-
tion of the *detractors'* viewpoint. He states that it is *a fallacy to label NVEs as rule-less*
and proceeds to justify his position. 6.4 is again a PRONOUNCEMENT, underlining
Cowan's conviction in the position being forwarded. Additionally, I suggest that
the linguistic realization of this PRONOUNCEMENT, involving an emphatic expres-
sion of obligation (*it is vital to recognize . . .*), adds another layer of dialogicality to
the essay; there is a sense in which Cowan seems to be trying to persuade some
(perhaps sceptical) Other. That Cowan has himself taken on one side of the
disciplinary debate suggests that he constructs himself as the dialogic equal of
the people whose views he is critiquing. This adds to his textual authority.

4.3.3   Exploring authority negotiation in texts where voices are in agreement

The 'academic dialogue' does not always involve disagreement. It could also
involve being part of a community of people whose voices validate each others'.
The dialogistic perspective introduced here is also useful for considering text-
ual locations where this is the case:

[7] *This is the beauty of any NVE: its versatility and variety, the fact that they are*
'indicative of the acculturation of English in new socio-cultural or linguistic contexts
and reflect . . . its acceptance as a vehicle of (indigenous) social norms and ecological
needs' (Kachru, 1983: 28). (Cowan)

How might we critique the negotiation of authority in this extract? I suggest that the same two interpersonal moves – the assertion of a writer's voice, and the locating of that voice within the relevant disciplinary conversation – are as vital for the construction of textual authority in cases of disciplinary agreement as in cases of disciplinary dissent.

In extract [7], we note that the writer, Cowan, first proffers his view in his own words (*This is the beauty of any NVE: its versatility and variety . . .*). He then invokes Kachru's voice as a way of emphasizing the validity of his assertion. This, in my terms, constitutes a site of 'dialogue' as Cowan's voice is brought into contact with Kachru's. Additionally, Cowan's choice of this particularly apt quote from an authority in the field points to his familiarity with the topic, and positions him firmly within the disciplinary 'dialogue'.

However, merely being aware of the conversations taking place within a discipline is not sufficient to construct textual authority. A writer has to demonstrate an ability to assert his/her own voice within that conversation. With extract [7], we note that Cowan's voice almost 'bleeds into' Kachru's; it is Kachru's words which complete Cowan's sentiments. Because of this, some might argue that Cowan allows Kachru's voice to take over his discourse, and that this shows that Cowan has abdicated 'propositional responsibility' (Groom, 2000: 15) to Kachru.

While this might be true if Cowan simply makes a neutral ACKNOWLEDGEMENT and gives no indication of his stance towards Kachru, it is patently not the case here. Although it is, in one sense, Kachru's words which lend legitimacy to Cowan's, it is Cowan's voice which is foregrounded throughout.

Cowan begins his sentence with a strong positive evaluation: *This is the beauty of any NVE: its versatility and variety . . . .* In the absence of any linguistic indicators to the contrary, we assume this to be Cowan's own 'averral' (Hunston, 2000). When he goes on with the words *the fact that they are . . .,* the entire noun phrase with *fact* as its head is clearly meant to be a continuation of his assertion about the beauty of NVEs, and thus can be read as having his personal backing. We would expect, then, that what would in grammatical terms be the Complement of the clause which begins with *they are . . .* would also have his personal backing, since this clause is embedded within the aforementioned noun phrase. We could argue, then, that Kachru's words, which serve as the said grammatical Complement, have in fact been introduced under a higher level blanket of ENDORSEMENT by Cowan. In this reading of the extract, Cowan's voice is very much in evidence throughout, and the way he subsumes Kachru's voice under his own constructs a moment of mutual ENDORSEMENT in his text. This positions Cowan as someone who is able to engage with the 'experts' on a level footing, and this, I argue, adds to his textual authority.

Groom (2000) writes that 'a successful argumentative text is one which always positions the writer as its *dominant* voice: other voices must be allowed to speak, but they must ultimately be subordinated by, and thus subordinate to,

the textual subjectivity of the writer herself or himself' (19). Cowan, we have seen, manages to subordinate the external voice to his own. Not all students, however, are able to retain/assert control of their own voices in the midst of heteroglossic dialogicality. Consider this next example:

[8] *Thus, it may seem so far that French's impact upon English was sharp and imme-diate, but, 'It should be remembered too, that, at first, French was confined to the ruling classes and that it took more time for its impact upon English to be felt.' (Orr 1948: 20).* (Fiona)

This extract starts out in Fiona's voice. The dialogically expansive EVIDEN-TIAL *it may seem so far that French's impact upon English was sharp and immediate* is presented as Fiona's own take on the discussion at hand. However, a shift soon takes place, resulting in Orr's voice completely overshadowing Fiona's. I would argue that Fiona does not come across as authoritative here because she does not 'control' both halves of what is essentially the same discourse move. Her dialogically expansive EVIDENTIAL *seem*, which projects open-endedness, prospects a subsequent contractive move. However, although it is Fiona who expands the negotiatory space of the text, it is in fact Orr whom Fiona relies upon to perform the subsequent, expected, dialogic contraction.

Fiona's abdication of writer responsibility (cf. Sinclair, 1986) and her failure to see the dialogic negotiation she initiates through to its conclusion, position her as unable to fully participate in the academic dialogue on her own behalf, and undercut her textual authority.

5   A Dialogic Account of Authority:
Concluding Remarks

In this chapter, I have shown how a unified and varied exploration of dis-coursal authority is possible when we adopt a dialogistic perspective of student writing. Centring my discussion here largely around extracts from student essays, I have illustrated how the Bakhtinian notions of dialogism and hetero-glossic diversity, and the insights and metalanguage offered by White's (2003) ENGAGEMENT system, can help to facilitate a critical discussion of the ways in which 'authority' is projected in such texts. Specifically, I have proposed and demonstrated here that one fruitful way of exploring authority in student essays is to see it as being the product of two kinds of interpersonal moves working together: (a) the assertion of a writer's voice (encompassing his/her opinions, stance, and agenda), and (b) the locating of that voice within the ongoing or past 'conversations' of his/her discipline.

I would emphasize here that this dialogic account of authority is intended to extend, not supplant, existing approaches to the study of this phenomenon.

It will thus be noticed that the ideas raised here are compatible with the work already done in the areas of attribution and intertextuality (e.g. Fairclough, 1992; Groom, 2000; Hunston, 1995; G. Thompson & Ye, 1991). Issues such as who is projected as bearing responsibility for what is being asserted (cf. the notion of attribution versus averral (Hunston, 2000; Sinclair, 1986; Tadros, 1993), and Groom's (2000) notion of 'propositional responsibility'), how evaluation is performed during citation (see G. Thompson & Ye, 1991), and how different voices in a text are assimilated or kept separate (see, for instance, Fairclough's (1992: 119) notion of 'boundary maintenance') are clearly all related to my discussion here. But, as Casanave (2002) has pointed out, the way in which a phenomenon is framed for discussion impacts on how we understand it. My exploration of authority through the lens of dialogism, then, is an attempt to frame these issues in a new way, and to offer researchers and teachers of academic literacy a conceptually integrated and accessible approach to the study of 'authority' in academic writing.

Notes

[1] Following the norm in much of the early APPRAISAL literature, small caps will be used when ENGAGEMENT sub-systems and categories are being referred to in this article.
[2] Rhetorical questions can be read as encoding CONCURRENCE because they assume that one and only one answer to the question is at all possible.

References

Bakhtin, M. (1981). *The dialogic imagination: Four essays.* (trans. C. Emerson & M. Holquist). Austin: University of Texas Press.

Bartholomae, D. (1985). Inventing the university. In M. Rose (Ed.), *When a writer can't write: Studies in writer's block and other composing-process problems* (pp. 134–165). New York: Guilford Press.

Berkenkotter, C. (1984). Student writers and their sense of authority over texts. *College Composition and Communication, 35*, 312–319.

Brannon, L., & Knoblauch, C.H. (1982). On students' rights to their own texts: A model of teacher response. *College Composition and Communication, 33*, 157–166.

Casanave, C.P. (2002). *Writing games: Multicultural case studies of academic literacy practices in higher education.* Mahwah, NJ: Erlbaum.

Delpit, L.D. (1986). Skills and other dilemmas of a progressive black educator. *Harvard Educational Review, 56*, 379–385.

Fairclough, N. (1992). *Discourse and social change.* Cambridge: Polity Press.

Greene, S. (1995). Making sense of my own ideas: The problems of authorship in a beginning writing classroom. *Written Communication, 12*, 186–218.

Groom, N. (2000). Attribution and averral revisited: Three perspectives on manifest intertextuality in academic writing. In P. Thompson (Ed.), *Patterns and perspectives: Insights into EAP writing practice* (pp. 14–25). University of Reading: Centre for Applied Language Studies.

Harris, J. (1987). The plural text/the plural self: Roland Barthes and William Coles. *College English, 49*, 158–170.

Hendricks, M., & Quinn, L. (2000). Teaching referencing as an introduction to epistemological empowerment. *Teaching in Higher Education, 5*, 447–457.

Hunston, S. (1995). A corpus study of some English verbs of attribution. *Functions of Language, 2*, 133–158.

—(2000). Evaluation and the planes of discourse: Status and value in persuasive texts. In S. Hunston & G. Thompson (Eds), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 176–207). Oxford: Oxford University Press.

Hyatt, D.F. (2005). 'Yes, a very good point!': A critical genre analysis of a corpus of feedback commentaries on Master of Education assignments. *Teaching in Higher Education, 10*, 339–353.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing.* Harlow: Longman.

—(2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics, 34*, 1091–1112.

Ivanič, R. (1994). I is for interpersonal: Discoursal construction of writer identities and the teaching of writing. *Linguistics and Education, 6*, 3–15.

—(1995). Writer identity. *Prospect, 10*(1), 8–31.

—(1998). *Writing and identity: The discoursal construction of identity in academic writing.* Amsterdam: Benjamins.

Ivanič, R., & Simpson, J. (1992). Who's who in academic writing? In N. Fairclough (Ed.), *Critical language awareness* (pp. 141–173). London: Longman.

Kirsch, G.E. (1993). *Women writing the academy: Audience, authority, and transformation.* Carbondale and Edwardsville: Southern Illinois University Press.

Koutsantoni, D. (2006). Rhetorical strategies in engineering research articles and research theses: Advanced academic literacy and relations of power. *Journal of English for Academic Purposes, 5*, 19–36.

Krause, K-L. (2001). The university essay writing experience: A pathway for academic integration during transition. *Higher Education Research and Development, 20*, 147–168.

Lillis, T. (1997). New voices in academia? The regulative nature of academic writing conventions. *Language and Education, 11*, 182–199.

—(2001). *Student writing: Access, regulation, desire.* London: Routledge.

—(2003). Student writing as 'academic literacies': Drawing on Bakhtin to move from critique to design. *Language and Education, 17*, 192–207.

Lutz, J., & Fuller, M. (2007). Exploring authority: A case study of a composition and a professional writing classroom. *Technical Communication Quarterly, 16*, 201–232.

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

Matsuda, P.K., & Tardy, C. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes, 26,* 235–249.

Peritz, J.H. (1993). Making a place for the poetic in academic writing. *College Composition and Communication, 44*, 380–385.

Raymond, J.C. (1993). I-dropping and androgyny: The authorial *I* in scholarly writing. *College Composition and Communication, 44*, 478–483.

Read, B., Francis, B., & Robson, J. (2001). 'Playing safe': Undergraduate essay writing and the presentation of the student 'voice'. *British Journal of Sociology of Education, 22*, 387–399.

Sinclair, J. (1986). Fictional worlds. In M. Coulthard (Ed.), *Talking about text: Studies presented to David Brazil on his retirement* (pp. 43–60). Discourse Analysis Monographs Number 13. Birmingham: English Language Research.

Sinclair, J., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils.* Oxford: Oxford University Press.

Starfield, S. (2002). 'I'm a Second-language English speaker': Negotiating writer identity and authority in Sociology One. *Journal of Language, Identity, and Education, 1*, 121–140.

Sunderland, J., & Barton, D. (2001, September). 'Coming across to a new academic discourse community'. Paper presented at the 34th Annual Meeting of the British Association for Applied Linguistics (BAAL): *Unity and Diversity in Language Use,* University of Reading, UK.

Tadros, A. (1993). The pragmatics of text averral and attribution in academic texts. In M. Hoey (Ed.), *Data, description, discourse: Papers on the English language in honour of John McH. Sinclair* (pp. 98–114). London: HarperCollins.

Tang, R. (2000). Do we allow what we encourage? How students are positioned by teacher feedback. *The Australian Journal of Language and Literacy, 23*, 157–168.

Tang, R., & John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes, 18*, S23–S39.

Thompson, C. (2005). 'Authority is everything': A study of the politics of textual ownership and knowledge in the formation of student writer identities. *International Journal for Educational Integrity, 1*(1). Retrieved 25 March 2008, from www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/18/8

Thompson, G., & Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics, 12*, 365–382.

White, P. (1998). *Telling media tales: The news story as rhetoric.* Unpublished PhD dissertation, University of Sydney.

—(2003). Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text, 23*, 259–284.

Part III

Focus on Learner Discourses

This page intentionally left blank

Introduction to Part III

Diane Pecorari

The five chapters in this section investigate various aspects of the textual production of student writers, and have, as a consequence, a common interest in the pedagogical applications of EAP corpus research. In Chapter 10, Sylviane Granger and Magali Paquot report the findings of an investigation into the lexical verbs used by students and professional writers. The quantitative portion of their study shows that these two groups differ strikingly in their use of lexical verbs. In addition, they demonstrate that the differences are found not only in which verbs (i.e. lemmas) are overused or underused by the novice writers, but that distinctive patterns of use exist for specific verb forms. They then examine a selection of these verbs in context, and are able to conclude that 'a dual approach – combining both lemmas and word forms – gives us a more precise picture of the diversity of form-meaning mappings that characterizes the use of EAP verbs' (p. 210).

The next chapter, by Philip Shaw, similarly combines quantitative findings about the frequency of a feature with an analysis of its function in context. Shaw's focus is on linking adverbials, adverbials which provide metadiscoursal intersentential ties, in writing in literary studies. Like Granger and Paquot, Shaw compares the occurrence of these adverbials in a student corpus with similar findings for professional writers (both within literary studies and from another discipline). The comparison of the student and the professional corpora shows a higher frequency of linking adverbials among the students. The subsequent discourse analysis of these adverbials reveals differences both in the frequency with which individual adverbials occur, and in their placement, and suggests that these differences may be explained by the fact that the two groups of writers produce different types of text for different purposes.

Intersentential links are also the subject of Chapter 12, by Amina Gardezi and Hilary Nesi. They examine the feature they term 'conjunctive ties' in the writing of two groups of undergraduates, all of whom have English as a first language, but who come from two different national and cultural backgrounds. Here, too, differences were found between the two groups, in terms of the frequency of conjunctive ties, which ones were used, and the position in which they occurred. Gardezi and Nesi conclude that these differences are an indication of local preferences in rhetorical style, and note an important pedagogical implication of their findings: while both sets of writers succeeded

in producing texts which were judged successful in the contexts in which they were writing, should they venture into another discourse community they would need to adapt to a new set of rhetorical preferences.

Student writing is again the focus in the next chapter, by Sheena Gardner and Jasper Holmes. Their study examines the use of section headings in a corpus of assessed writing done by British university students. Their fine-grained analysis shows that the question of which subject headings are commonly used, or indeed whether they are commonly used, depends on year of study, discipline and type of assignment. From a pedagogical perspective these findings provide another useful implement in the writing teacher's tool kit, since, as Gardner and Holmes note in their introduction, the ability of the writing teacher to help students across the disciplines depends on an awareness of what is conventional in those disciplines.

In the final chapter in this section, and in the volume, Suganthi John shifts the emphasis from specific textual features to one of their effects, namely the visibility of the academic writer. In this pedagogically motivated study, John first examines how the specific linguistic choices in her corpus (which consists of the methods sections of MA theses) contribute to making the writer more or less visible. She then adopts a case-study approach to show how students and supervisors can work on the issue of identity during the revision process to help novice writers position themselves effectively.

These five chapters, then, like the others in this volume, aim to contribute to what we know about academic texts, and therefore to what we know about helping student writers to meet the expectations of their discourse communities. Significantly, they succeed in this objective by going beyond simply documenting the frequency with which certain features occur, to showing how they work in their discoursal contexts to achieve the rhetorical objectives of the text.

Chapter 10

Lexical Verbs in Academic Discourse: A Corpus-driven Study of Learner Use

Sylviane Granger and Magali Paquot

1 Introduction

In spite of their relative infrequency in English for Academic Purposes (EAP) as compared to other genres, notably conversation and fiction (Biber et al., 1999: 358), lexical verbs contribute significantly to some major EAP functions such as expressing personal stance, reviewing the literature, quoting, expressing cause and effect, summarizing and contrasting. They enable writers to modulate their ideas and position their work in relation to other members of the discipline. Hinkel (2004) classifies them into the following five categories: activity verbs (*make, use, give*), reporting verbs (*suggest, discuss, argue, propose*), mental/emotive verbs *(know, think, see)*, linking verbs (*appear, become, keep, prove*) and logico-semantic relationship verbs (*contrast, follow, cause, illustrate*). Among those it is undeniably the category of 'reporting verbs' that has received the most attention (Charles, 2006a, 2006b; Hyland, 1999; Shaw, 1992; Thomas & Hawes, 1994; Thompson & Ye, 1991). Reporting verbs are important in academic discourse, as 'they allow the writer to clearly convey the kind of activity reported and to precisely distinguish an attitude to that information, signaling whether the claims are to be taken as accepted or not' (Hyland, 1999: 344). Other categories, such as that of 'coming-to-know verbs' (Hiltunen, 2006; Meyer, 1997) have also been the subject of detailed investigation. In general, EAP studies have tended to focus on one specific category of verbs rather than give a general overview of the use of lexical verbs in academic discourse. Williams (1996) is an exception in this respect as he investigates all lexical verbs of a particular frequency used in medical reports.

Although verbs figure prominently in EAP word lists, it is difficult to draw up a list of EAP lexical verbs from existing lists as they often fail to give any indication of word category membership. The most popular EAP list, Coxhead's (2000) Academic Word List (AWL), contains many words like *conduct, focus, approach, survey* or *function*, which can be nouns and verbs. Another

characteristic of the AWL is that it excludes the top 2,000 words in the language, that is, those that figure in the General Service List (GSL). This is justified to some extent as many high frequency verbs are rarely used in EAP. For example, Biber (1988: 105) demonstrates that 'private verbs' like *love, want, like, feel* or *hope*, which 'are used for the overt expression of private attitudes, thoughts, and emotions', are typical of involved discourse, notably conversation, and rarely used in academic texts. However, there are in fact several high frequency verbs which turn out to play a major role in EAP and are therefore worth including in EAP syllabuses. For example, Meyer's (1997) study of the acquisition of knowledge in the process of academic investigation includes high frequency verbs like *find* or *show* which show 'all the vaguenesses, polysemies, and ambiguities of everyday language', but 'are used to discuss matters lying at the very heart of the scholarly process' (368). In order to give these verbs the coverage they deserve in EAP, Paquot (2007) has included in her Academic Keyword List (AKL) verbs like *aim, argue, cause, claim, effect* or *suggest,* which are absent from Coxhead's list.

Insufficient knowledge of verbs that are typically used in academic written discourse is a serious handicap for learners as it prevents them from expressing their thoughts in all their nuances and couching them in the expected style. As pointed out by Swales (2004: 17), 'a formal research report written in informal English may be considered too simplistic even if the actual ideas and/or data are complex'. Presenting learners with lists of EAP verbs and the exact meanings they convey is therefore undoubtedly an important first step but unless it is complemented with a detailed description of their use, results are bound to be highly disappointing. One of the strengths of EAP verbs, their ability to help modulate the message via tense, aspect, mood and voice, creates a minefield of difficulties for learners (Hinkel, 2002; Swales & Feak, 2004). Research has tended to focus largely on these areas of difficulty, in particular on the issue of tense and aspect and the question of the transferability of General English rules to EAP (Swales, 1990: 151). However, this is not the only problem that learners are faced with. They also have to deal with the fact that each EAP verb has its own preferred lexico-grammatical company, namely, subjects (*this study shows that*; *the evidence suggests that; these results suggest that*), objects (SUPPORT *the view / hypothesis that . . .;* PROVIDE *evidence/information*) and adverbs (DIFFER *significantly;* VARY *considerably/widely;* APPLY *equally; closely related; widely used; generally accepted*) and tend to appear in routinized structures (*as discussed in; there is [no, some, little] evidence that; it should be noted that*) (cf. Gilquin et al., 2007b; Paquot, 2007). Generalities such as 'the passive is very frequent in academic discourse' are not very helpful as some EAP verbs are hardly ever used in the passive while others are typically (if not exclusively) used in the passive (cf. Swales, 2004: 12).

Lexico-grammatical restrictions of EAP verbs are often disregarded in EAP textbooks, which tend to present verbs separately from nouns and adverbs when in fact, as demonstrated by several recent learner corpus-based studies, it is their interaction that causes difficulty for learners. This is confirmed by Nesselhauf's (2005) investigation of German-speaking English as a Foreign Language (EFL) learners' misuse of collocations in verb-noun combinations. Similarly, Hyland's (2008: 50) analysis of word clusters in Cantonese-speaking students' academic writing shows that 'many of the clusters most frequently used in published academic writing were never, or only rarely, found in the student texts' (see also Ädel, 2006; Altenberg & Granger, 2001).

All these studies show that it is phraseology in the wide sense, namely, including both highly fixed and much looser routinized sequences, that EFL learners find most difficult. Some of these phraseological difficulties, in particular those related to pragmatic appropriacy and discourse patterns, are shared by novice native writers. Hyland and Milton (1997: 192) show that both Cantonese learners and novice native writers mix 'informal spoken and formal written forms and transfer conversational uses of academic genres'. Similarly, Neff et al. (2004: 152) compare the expression of writer stance in various corpora of argumentative texts written by EFL learners, novice and professional native writers and show that 'all of the student writers (native and non-native) have the novice-writer characteristic of excessive visibility'. However, it would be wrong to conclude that native student writers and EFL/English as a Second Language (ESL) learners face exactly the same difficulties in academic writing and can therefore be considered as belonging to one and the same category of novice writers. As pointed out by Gilquin et al. (2007a) and further argued below, a wide range of lexico-grammatical difficulties are exclusive to L2 learners and therefore deserve specific attention.

The main objective of this chapter is to give a detailed description of the use of lexical verbs in L2 learners' academic writing compared to both expert and novice native writing. The investigation is based on expert and learner corpora of academic writing and the method is corpus-driven rather than corpus-based, that is, it 'relies heavily on data and (largely) automatic procedures' (De Cock, 2003: 197; cf. also Tognini-Bonelli, 2001). The investigation attempts to tackle the following questions: which (categories of) verbs do learners use in their EAP writing? Is the set of EAP verbs used by L2 learners different from that of both expert and novice native users? Do L2 writers use EAP verbs in their typical lexico-grammatical patterning?

In Section 2 we describe the corpora and the methodology used to extract EAP verbs. Section 3 discusses the advantages and disadvantages of taking word forms or lemmas as units of analysis. Section 4 gives the results of the analysis of lexical verbs in EFL and professional academic writing. Section 5

addresses the issue of text type and domain comparability by revisiting the findings of Section 4 in the light of a comparison between EFL and native novice writing. Section 6 contains concluding remarks.

2   Data and Methodology

This study makes use of two large collections of academic discourse to describe the use of EAP verbs by expert and learner writers. The learner data comes from the second edition of the *International Corpus of Learner English* (henceforth ICLE) (Granger et al., 2009), which contains over 3 million words of argumentative essay writing by high-intermediate to advanced EFL university students of 16 different mother tongue backgrounds: Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana and Turkish. The focus of our study is on EFL learners rather than ESL students. The two populations are rarely distinguished in the literature and yet they are quite different. For example, the use of phrasal verbs instead of the more EAP-appropriate single word equivalents is often presented as a major problem for EAP students (cf. e.g. Swales & Feak, 2004). It may well be a problem for ESL learners exposed to informal English on a daily basis or for novice native writers who may transfer their everyday English to their academic texts. However, it is not a major source of difficulty for EFL learners, who make scant use of phrasal verbs (cf. Liao & Fukuya, 2004; Sjöholm, 1998).

A large collection of expert writing, which will be referred to as ACAD, is used as a reference corpus. It is composed of the academic sub-parts of the *MicroConcord* corpus collection (Johns & Scott, 1993) and the Baby British National Corpus (cf. Burnard, 2003), which together contain 2 million words. Both corpora consist of published academic prose (book samples and articles) and are divided into five sub-corpora of c. 200,000 words, each of which corresponds to a broad academic discipline (i.e. humanities, social science, applied science, technology and engineering).

The main advantage of these two corpora is that they are large collections of academic texts (cf. Table 10.1) and thus highly valuable in providing a general overview of the use of lexical verbs in academic writing. An important

Table 10.1   Description of the corpora

Corpora	Number of words	Professional status	L1 or L2	Text type
ACAD	2,027,880	Professional	L1/Proficient L2 writer	Expository
ICLE	3,233,214	Non-professional	L2 learner	Argumentative

caveat, however, is that the two corpora are not fully comparable. Expert texts are expository in nature, that is, they are topic-oriented (cf. Britton, 1994) and rely on the comprehension of general concepts (cf. Werlich, 1976), while argumentative essays start 'from the assumption that the receiver's belief must be changed' (Gramley & Pätzold, 1992: 193). In addition, expert texts are discipline-specific while learners' essays discuss a range of general topics such as feminism, the impact of television, drugs, etc. Special care therefore needs to be taken to interpret results in the light of genre analysis as some differences between learner essays and expert texts may simply reflect differences in their communicative goals and settings (cf. Neff et al., 2004). Another issue concerns the use of professional native writing as a standard of comparison in learner corpus research. This has been criticized by several authors, among others Lorenz (1999: 14), who considers this practice to be 'both unfair and descriptively inadequate' and Hyland and Milton (1997: 184), who take a stand against the 'unrealistic standard of "expert writer" models' and argue that native student writing is a better type of comparable data to EFL learner writing if the objective is to describe and evaluate interlanguage(s) as fairly as possible.[1]

The two corpora were lemmatized and part-of-speech tagged with the *Constituent Likelihood Automatic Word-tagging System* (CLAWS) C7 (cf. Garside & Smith, 1997).[2] The tagset includes six different tags for lexical verbs: VV0 (base form, e.g. *drink, work*), VVD (past tense, e.g. *drank, worked*), VVG (-*ing* participle, e.g. *drinking, working*), VVI (infinitive, e.g. *drink, work*), VVN (past participle, e.g. *drunk, worked*), VVZ (-s form, e.g. *drinks, works*). We applied a Perl program[3] to CLAWS output to create corpora consisting of lemmas + simplified POS-tags. POS-tags were automatically simplified to match the level of specificity of lemmas, i.e. the six tags available for lexical verbs (VV0, VVD, VVG, VVI, VVN, and VVZ) were replaced by a single VV tag.

We made use of *WordSmith Tools 4* (Scott, 2004) to create lists of word forms + POS-tags and lemmas + POS-tags for each corpus. In this study, we analyse all lemmas and word forms that were assigned a VV or VV* tag.

3   Verb Forms versus Verb Lemmas

Any corpus-driven investigation of lexical verbs needs to consider the advantages and disadvantages of using verb lemmas or verb forms as units of analysis. If lemmas are used, the different inflectional forms, for example, *claim, claims, claimed, claiming*, are merged. This is a useful option if the aim of the analysis is to give a general overview of learners' lexical repertoires and/or detect patterns of use that cut across verb forms (e.g. the use of a *that*-clause with the lemma CLAIM). However, as rightly pointed out by Sinclair (1991),

lemmas are an abstraction and only using lemmas amounts to losing important information as each word form has its own individual patterning. Sinclair sees a future for a new branch of study that focuses on the interrelationships of a lemma and its forms as 'it is not yet understood how meanings are distributed among forms of a lemma' (1991: 41). He even goes as far as to suggest that lexicographers change the traditional practice of using the 'base' or uninflected form as headword and use 'the most frequently encountered form' instead (42), a pioneering view that has so far gone unheeded. In a previous study (Granger & Paquot, 2005), we carried out an automatic comparison of a 1-million-word corpus of academic writing and a similar-sized fiction corpus. Using the criteria of keyness, frequency, range and evenness of distribution, we identified 233 verb lemmas that figured more prominently in the academic corpus than in the fiction corpus. One of the interesting results of the study is that verbs regularly function as EAP keywords in only one or two inflectional forms. As shown in Figure 10.1, nearly half of the verbs (47%) appear as distinctive EAP items in only one word form and almost a quarter of them (23%) in two word forms. A minority appear in three (19%) or four (or five) word forms (11%).

Table 10.2 lists some of the verbs in each category. It shows that the verb lemma ASSOCIATE,[4] just like several others such as BASE, CONFINE and LINK, appears as a distinctive EAP item in only one word form, that is, the –ed form. For LACK or COMPRISE, it is the –ing form that is distinctive and for ENTAIL and REVEAL, the –s form. This shows that, as suggested by Hyland and Tse (2007: 243), we need to 'be cautious about claiming generality for families whose meanings and collocational environments may differ across each inflected and derived word form' (cf. also Oakey, 2005). This word of caution has been at the forefront of our analysis of lexical verbs in learner and native writing, the results of which are presented in the following section.
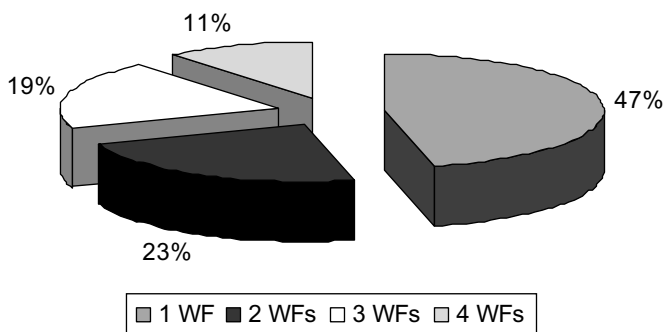
FIGURE 10.1   Number of key word forms (WF) per key lemma

Table 10.2   EAP word forms versus EAP lemmas

	EAP word forms	EAP lemmas
1 word form	*associated*	ASSOCIATE
	based	BASE
	confined	CONFINE
	linked	LINK
	observed	OBSERVE
	summarized	SUMMARIZE
	undertaken	UNDERTAKE
	lacking	LACK
	comprising	COMPRISE
	inducing	INDUCE
	entails	ENTAIL
	predicts	PREDICT
	reveals	REVEAL
	seeks	SEEK
	assert	ASSERT
	benefit	BENEFIT
	coincide	COINCIDE
	participate	PARTICIPATE
2 word forms	*indicate/indicates*	INDICATE
	amount/amounts	AMOUNT
	conclude/concludes	CONCLUDE
	explain/explains	EXPLAIN
	emerge/emerges	EMERGE
	assume/assumes	ASSUME
	achieve/achieved	ACHIEVE
	adopt/adopted	ADOPT
	specify/specified	SPECIFY
	assess/assessing	ASSESS
	characterizes/characterized	CHARACTERIZE
	contrasts/contrasting	CONTRAST
	designed/designing	DESIGN
3 word forms	*argue/argues/argued*	ARGUE
	suggest/suggests/suggesting	SUGGEST
	show/shown/shows	SHOW
	discuss/discussed/discussing	DISCUSS
	illustrate/illustrates/illustrated	ILLUSTRATE
4 word forms	*include/included/including/includes*	INCLUDE
	exist/existed/existing/exists	EXIST
	develop/develops/developed/developing	DEVELOP

4   Lexical Verbs in Learner Academic Discourse

In this section we draw up lists of the lexical verbs used in ICLE and compare the results with those used in ACAD. We first focus on verb lemmas for the insights they provide into learners' lexical stock of EAP verbs and then on verb

forms to uncover new perspectives on learners' preferred and dispreferred EAP patterns.

4.1 EAP verb lemmas

The lists of the top 100 verb lemmas in ICLE and ACAD were first compared. Table 10.3 shows the degree of overlap in the top 100 verbs in each corpus. Of the 148 different verbs, about 35% (N = 52) are shared by the two corpora and around a third (32.4%; N = 48) are only found in either the ICLE or the ACAD corpus. Among the shared verbs quite a number display marked differences in ranking: WANT (rank 8 in ICLE vs 46 in ACAD), TRY (rank 19 vs 49), HELP (21 vs 66), SHOW (28 vs 9), PROVIDE (40 vs 16).

In addition, many of the top 100 verbs (84.5%, N = 125) display marked differences in frequency: 55.4% (N = 82) are significantly overused in ICLE and 29% (N = 43) are underused. Only 15.5% (N = 23) are used with similar frequencies. The top 50 underused verb lemmas in ICLE are presented in Table 10.4 in decreasing order of keyness.

Approximately half (23/50, i.e. 45.1 %) of the 50 most underused verb lemmas in ICLE are EAP words according to Coxhead's (2000) AWL. These appear in bold in Table 10.4. All the other words except one (COLLIDE) are words from the General Service List (GSL).[5] If Paquot's (2007) Academic Keyword List (AKL) is used instead, the proportion of underused EAP words rises sharply to reach a staggering 88% (44/50). The AKL proves to be highly useful in uncovering all the words highlighted by the comparison with the AWL plus a large

Table 10.3 Top 100 verbs: ICLE versus ACAD

ICLE only	ICLE and ACAD	ACAD only
AFFECT, AGREE, BAN, BUY, CLAIM, COMMIT, DECIDE, DIE, DREAM, EARN, EAT, ENJOY, FACE, FIGHT, FORGET, GROW, HAPPEN, HEAR, IMAGINE, IMPROVE, KILL, LEARN, LET, LIKE, LOSE, MEET, MENTION, PAY, PLAY, PREPARE, PREVENT, PROTECT, PROVE, READ, REALIZE, SMOKE, SOLVE, SPEND, START, STATE, STAY, STOP, STUDY, SUFFER, SUPPORT, TALK, TEACH, WATCH	ACCEPT, ALLOW, ASK, BECOME, BEGIN, BELIEVE, BRING, CALL, CAUSE, CHANGE, CHOOSE, COME, CONSIDER, CREATE, DEVELOP, DISCUSS, EXIST, FEEL, FIND, FOLLOW, GET, GIVE, GO, HELP, INCREASE, KEEP, KNOW, LEAD, LEAVE, LIVE, LOOK, MAKE, MEAN, NEED, PROVIDE, PUT, REDUCE, SAY, SEE, SEEM, SHOW, SPEAK, TAKE, TELL, THINK, TRY, TURN, UNDERSTAND, USE, WANT, WORK, WRITE	ACHIEVE, ACT, ADD, APPEAR, APPLY, ARGUE, ARISE, ASSUME, BASE, CARRY, COMPARE, CONTAIN, CONTINUE, DEAL, DEFINE, DEPEND, DESCRIBE, DETERMINE, DRAW, ESTABLISH, EXPECT, EXPLAIN, EXPRESS, FORM, HOLD, IDENTIFY, IMPROVE, INDICATE, INVOLVE, MOVE, NOTE, OBTAIN, OCCUR, OFFER, POINT, PRESENT, PRODUCE, RECEIVE, REFER, REGARD, RELATE, REMAIN, REPRESENT, REQUIRE, SET, SUGGEST, TEND, TREAT

Table 10.4 Top 50 underused verb lemmas in ICLE in decreasing order of keyness

Lemma	Frequency in ICLE	Frequency in ACAD	Log-likelihood
DESCRIBE	273	1080	947.8
OCCUR	324	947	664.5
NOTE	73	527	622.8
SUGGEST	500	1079	558.6
REQUIRE	589	1072	444.5
CONTAIN	233	655	444.3
OBTAIN	310	728	414.4
IDENTIFY	120	471	411.2
INVOLVE	497	939	410.4
ASSUME	186	565	409.4
DERIVE	73	372	377.1
FOLLOW	767	1127	327.1
INCLUDE	468	805	306.8
RECORD	37	252	291.1
DETERMINE	236	531	288.4
REMAIN	555	869	283.5
APPEAR	593	901	278.5
ATTEMPT	69	294	270.1
DEMONSTRATE	98	337	268.7
MEASURE	72	296	266.1
RESPOND	35	224	252.3
ASSESS	34	211	234.6
HOLD	627	881	233.9
PRODUCE	737	979	230.2
ASSOCIATE	144	367	227.2
INTERPRET	71	267	226.6
REPORT	175	403	224.6
GENERATE	81	276	218.6
DEFINE	271	498	209.3
REFER	283	507	205.4
ESTABLISH	391	618	205
RETAIN	52	220	201.2
CONSTITUTE	114	305	197.8
YIELD	17	152	193.4
RELATE	342	554	191.6
COLLIDE	7	126	190
ILLUSTRATE	120	304	187
INDICATE	286	488	183.6
VARY	116	288	173.7
SPECIFY	22	149	171.8
CALCULATE	46	189	169.8
EMERGE	72	226	168.1
ARISE	297	481	166.3
RECOGNIZE	197	373	163.5
EXTEND	141	306	159.4
CONSENT	4	98	155.2
ADD	305	474	152.6
REPRESENT	331	498	151.1
OUTLINE	7	104	151
REMOVE	130	285	150.3

number of other words, such as DESCRIBE, SUGGEST, NOTE or INCLUDE, which fill important roles in EAP and therefore deserve to be brought to students' attention (AKL words are underlined in Table 10.4). As most of the verbs are polysemous, a fine-grained semantic classification would require manual scanning of each verb use in context, which clearly falls beyond the scope of this article. However, even without an examination of the verbs in context, it appears clearly from Table 10.4 that the majority of the underused verbs fall into three categories: communication verbs (DESCRIBE, SUGGEST, NOTE, DEFINE, RESPOND, REPORT, ADD, SPECIFY); cognition verbs (ASSUME, DERIVE, INTERPRET, ASSESS) and relational verbs (APPEAR, REQUIRE, REMAIN, INCLUDE, INVOLVE).

By contrast, the large majority (45, viz. 90%) of the top 50 overused words (see Table 10.5) belong to the General Service List (in bold). Besides topic-dependent verbs like DREAM, BAN or SMOKE,[6] the list contains several verbs that are marked by Biber et al. (1999) as typical of conversation (e.g. THINK, GET, GO, KNOW, LIKE, WANT) and/or highlighted by Hinkel (2004) as not appearing in EAP texts (e.g. FEEL, LIKE, TRY, WANT). Most are activity verbs (HELP, PUNISH, WORK, TEACH, PLAY) and mental verbs of cognition, perception and affection (THINK, LOVE, FEEL, REALIZE). The list also contains the overused verb of communication SAY. One overused word that is not in the GSL (CREATE) belongs to the AWL but the other four (BAN, IMPORT, RECYCLE, REHABILITATE) are neither in the AWL nor in the AKL. Five overused verbs (STUDY, USE, SOLVE, BECOME, CREATE) appear in the AKL list (underlined in Table 10.5).

4.2   EAP verb forms

With a view to assessing the relative merits of a lemma versus word form approach, we replicated the analysis described in the preceding section with verb forms instead of lemmas. While the analysis revealed a wide area of overlap between the two analyses, it also demonstrated that an exclusive focus on lemmas is liable to distort the picture and hide some major differences between expert and learner use. This distortion can take two different forms: (1) similar frequencies at the lemma level hide over- and/or underuse at the verb form level (cf. Table 10.6); (2) overuse or underuse at the lemma level affects only some of the verb forms (cf. Tables 10.7 and 10.8). A good example of the first type of distortion is the verb CONCLUDE (Table 10.6), which displays no difference in frequency at the lemma level, but in fact turns out to display an overuse of the infinitive form (*conclude*_VVI) coupled with a significant underuse of the 3rd person singular of the simple present tense (*concludes*_VVZ) and the simple past form (*concluded*_VVD). The second type can be illustrated by the verb ARGUE (Table 10.7) whose overall underuse at the lemma level conceals an overuse of the simple present form (except for the 3rd person singular) and the verb CAUSE (Table 10.8) whose overall overuse conceals an underuse of the *–ing* form.

Table 10.5 Top 50 overused verb lemmas in ICLE in decreasing order of keyness

Lemma	Frequency in ICLE	Frequency in ACAD	Log-likelihood
THINK	8711	1331	3245.8
GET	7531	1113	2887.9
DREAM	2453	19	2231
WANT	5169	677	2182.6
WATCH	2331	97	1666
LIVE	4110	578	1641.4
BAN	1358	20	1167.2
LEARN	2768	426	1023.7
PAY	2385	335	953.12
LIKE	2039	266	863.05
GO	5268	1524	837.72
BUY	1464	119	822.81
NEED	3928	1027	753.39
SMOKE	921	24	729.3
SPEND	1732	230	723.28
HELP	2632	555	694.64
TRY	2907	661	693.06
FORGET	1057	83	603.71
KILL	1450	206	574.03
STUDY	1612	268	554.73
PLAY	1963	392	554.25
IMPORT	653	12	546.05
BECOME	5066	1763	521.22
START	1884	391	507.32
EARN	751	40	498.95
KNOW	4941	1742	490.09
FEEL	2530	663	483.2
BELIEVE	2303	582	467.46
TEACH	1243	202	437.15
WORK	2996	917	423.03
SAY	5567	2159	408.81
PUNISH	645	41	402.54
CHANGE	2128	569	392.31
USE	6785	2808	389.66
MAKE	8863	3897	388.57
IMAGINE	970	140	379
FIGHT	944	135	371.66
CREATE	1891	498	358.03
RECYCLE	388	3	352.83
SOLVE	1072	191	343.97
HAPPEN	1621	421	314.32
AFFORD	564	55	288.41
REHABILITATE	311	3	278.27
REALIZE	794	130	277.24
LET	1349	339	276.28
KEEP	1870	571	265.35
LOVE	542	58	262.41
MASTER	331	9	260.05
SAVE	664	96	259.05
EDUCATE	450	36	254.78

Table 10.6 Lemmas versus verb forms: lemmas with similar frequencies in ICLE and ACAD

Lemmas with similar frequencies	Overused verb forms	Underused verb forms
ACCESS	*access* (VVI)	*accesses* (VVZ)
ALLOW	*allowed* (VVN)	*allowing* (VVG)
CONCLUDE	*conclude* (VVI)	*concludes* (VVZ), *concluded* (VVD)
DISCUSS	*discuss* (VVI)	*discusses* (VVZ), *discussed* (VVN)
LEAD	*lead* (VV0), *lead* (VVI)	*led* (VVD), *led* (VVN)
PROVIDE	*provide* (VV0)	*provided* (VVD), *provided* (VVN)

Table 10.7 Lemmas versus verb forms: underused lemmas

Underused lemmas	Underused verb forms	Overused verb forms
SEE	*see* (VV0), *saw* (VVD), *seen* (VVN)	*see* (VVI)
SHOW	*showed* (VVD), *shown* (VVN)	*showed* (VVN)
ARGUE	*argued* (VVN), *argued* (VVD), *argues* (VVZ), *arguing* (VVG)	*argue* (VV0)

Table 10.8 Lemmas versus verb forms: overused lemmas

Overused lemmas	Overused verb forms	Underused verb forms
BELIEVE	*believe* (VV0), *believe* (VVI)	*believed* (VVD)
CAUSE	*cause* (VV0), *cause* (VVI), *causes* (VVZ), *caused* (VVN)	*causing* (VVG)
FIND	*find* (VVI)	*found* (VVN), *found* (VVD)
GIVE	*give* (VV0), *give* (VVI), *gives* (VVZ), *giving* (VVG)	*given* (VVN), *gave* (VVD)
KNOW	*know* (VV0), *know* (VVI), *knows* (VVZ)	*known* (VVN)
MAKE	*make* (VV0), *make* (VVI), *makes* (VVZ)	*made* (VVN)
SEEM	*seem* (VV0), *seems* (VVZ)	*seemed* (VVD)
SPEAK	*speak* (VV0), *speak* (VVI), *speaking* (VVG)	*spoke* (VVD)
TAKE	*take* (VV0), *take* (VVI), *taking* (VVG)	*took* (VVD), *taken* (VVN)
UNDERSTAND	*understand* (VV0), *understand* (VVI)	*understood* (VVN)
USE	*use* (VV0), *use* (VVI), *using* (VVG)	*used* (VVD), *used* (VVN)
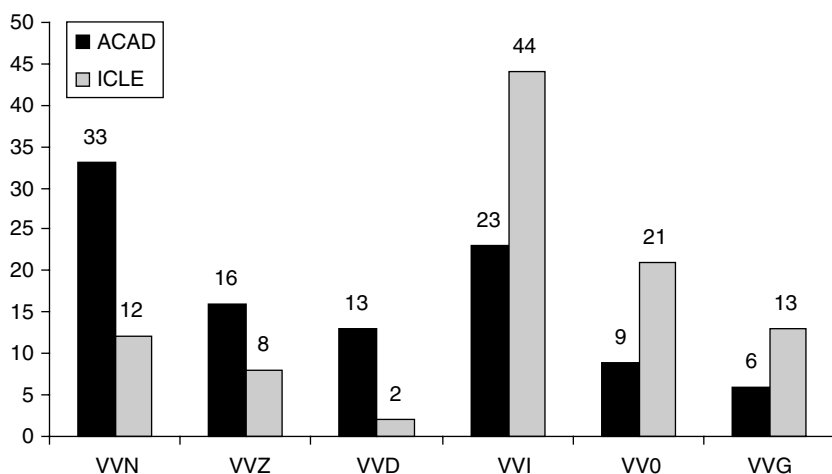
Figure 10.2 Top 100 verb forms in ACAD and ICLE: breakdown per word category

It is possible to form a more general picture of the use of EAP verb forms by investigating the breakdown of the different VV tags displayed by the top 100 verb forms in each corpus. As shown in Figure 10.2, the analysis shows striking differences, notably learners' predilection for infinitive forms (X^2 = 9.9, $p < 0.01$) coupled with a seeming avoidance of past participle forms (X^2 = 12.6, $p < 0.01$).

4.3 Lexico-grammatical patterns of EAP verbs in ICLE

The quantitative differences at the verb form level are indicative of marked differences in phraseological patterning. In order to illustrate this link between verb forms and lexico-grammatical patterns, two representative verbs were selected – CONCLUDE and ARGUE – and submitted to close scrutiny. Concordances proved invaluable in highlighting the patterns of use typical of each corpus.

As shown in Table 10.6, the lemma CONCLUDE is used with similar frequencies in ACAD and ICLE. However, the word form analysis shows that EFL learners significantly overuse the infinitive form. This is due to a significant overuse of the connector 'to conclude' in sentence-initial position (130 out of 419 occurrences of the lemma CONCLUDE; 31%), a use that is very infrequent in ACAD (7 out of 208; 3.4%). The contrast between the repetitive use of 'to conclude' in ICLE and the wider range of patterns used in ACAD appears clearly from Examples (1) to (7).

Learner writing

Example 1
To conclude we can say that the social position is 'gradually' improving. (ICLE-DUTCH)

Example 2
To conclude I would like to say once more that we have to have faith and hope within ourselves. (ICLE-SWEDISH)

Example 3
To conclude, I think that the government should ban smoking in restaurants as the health of the public can be improved. (ICLE-CHINESE)

Expert writing

Example 4
Finally, *the chapter concludes* by providing some reflections about the prospects.

Example 5
He concludes that the effectiveness of a given system should be based on its ability.

Example 6
It is reasonable to conclude from this that, although there are colliding plane wave space-times . . .

Example 7
We may conclude that, in all cases, the opposing waves mutually focus each other.

By contrast, the lemma ARGUE is significantly underused by learners: it is almost twice as frequent in ACAD as in ICLE (401 vs 222 per 1 million words). However, as shown in Table 10.7, this underuse does not affect the base form (VV0), which is overused, due to a recurrent use of the verb ARGUE with *people* and *I* as subject. Here too the contrast between the wide range of patterns displayed by ACAD and the limited range displayed by ICLE is striking (see Examples 8 to 19).

Learner writing

Example 8
Some people argue that television is the greatest invention of the 20th century (ICLE-POLISH)

Example 9
Many people argue that criminals should be punished physically again. (ICLE-GERMAN)

Example 10
I argue that all humans have the same rights to live compared to other humans; whether rich, poor, majority or minority this being deserves the same chance we were all given. (ICLE-TURKISH)

Example 11
That is why *I argue* in favour of rehabilitation, and against the prison system. (ICLE-NORWGIAN)

Expert writing

Example 12
It could be argued that this gives the work a sense of coherence.

Example 13
Integration*, it is argued*, will only work in areas . . .

Example 14
Moreover, *as argued above*, a major reason for having rules . . .

Example 15
In previous chapters *I have argued that* the decline of this investigatory response . . .

Example 16
Gergen (1979) also *argued that* social events are openly competitive.

Example 17
In the theatre*, he argues*, there is an internal dramatist.

Example 18
Spinoza shows this *by arguing that* God is the creator . . .

Example 19
He laid great emphasis on the unity of the Trinity, *arguing that* root, stem and bark together . . .

These two examples effectively illustrate the strength of the verb form approach, which functions as a quick way into learners' phraseology. It also shows that over- and underuse need not be taken as negative terms. They can – and indeed should – be taken as prompts for lexical expansion and used with learner populations who wish to attain a native-like mastery of EAP and would benefit from increasing their repertoire of EAP patterns.

5   Issues of Text Type and Domain Comparability

Some of the differences between learner and academic writing highlighted in Section 4.3 may be due to differences in text type. As explained in Section 2, ACAD consists of book samples and articles which are expository in nature while the learner texts are short argumentative essays. A large proportion of the verbs that are significantly underused in ICLE (cf. Table 10.4) perform two essential rhetorical functions in professional academic writing: they are used to quote and report what other scholars have written (e.g. ARGUE, COMMENT, EXPLAIN, NOTE, OBSERVE, PROPOSE, REMARK, REPORT, SUGGEST and WRITE) and to refer to tables, figures and other parts of the text (e.g. DESCRIBE, ILLUSTRATE, SHOW). Learners' underuse of these verbs may thus be (at least) partly explained by a difference in text type as there is no need in argumentative writing to situate one's opinion against what has been written in the literature and typically, argumentative essays do not contain tables and graphs and are too short to include internal reference to chapters and sections.

This provides justification for those who argue that learner writing should not be compared with professional academic writing (cf. Hyland & Milton, 1997; Lorenz, 1999). Indeed, if the French L1 sub-corpus of ICLE is compared with a subpart of the Louvain Corpus of Native Speaker Essays (LOCNESS), namely, a 150,000-word corpus of argumentative essays written by American university students (cf. Granger, 1996), EFL learners' underuse of these verbs is less marked and at times nonexistent. Similarly, French learners and American students seem to share a preference for active structures with first person subject pronouns, which may also partly reflect a difference in text type. In argumentative essays such as those contained in ICLE and LOCNESS, 'personal references and subjective attitudes are certainly hard to avoid' (Recski, 2004), since essay prompts explicitly encourage learners and native students to give their opinions (e.g. 'In the 19th century, Victor Hugo said: "How sad it is to think that nature is calling out but humanity refuses to pay heed." Do you think this is still true nowadays?').

Both EFL learners and novice writers draw more extensively on high-frequency verbs and employ a more restricted number of lexico-grammatical patterns when compared to professional academic writers. Syntactic structures that are frequently used in academic written discourse such as the *–ing* supplementive clause and the passive construction are less frequent in LOCNESS and, to a larger extent in ICLE, than in ACAD. These differences are indicative of novice writers' – both French learners and American students – imperfect grasp of academic writing conventions (see also Gilquin et al., 2007a; Hyland & Milton, 1997; Neff et al., 2004).

However, a close analysis of LOCNESS reveals that novice L1 writing often appears to occupy an intermediate position between academic writing and EFL learner writing (cf. Gilquin & Paquot, 2008; Neff et al., 2004). Not only

do novice L1 writers make far fewer semantic and syntactic errors overall than EFL learners (cf. Gilquin et al., 2007a), they also underuse fewer verbs when compared to academic expert writing: *argue, assess, attempt, determine, illustrate, indicate, include, involve, occur, outline, respond, vary* and *yield* are examples of verbs that are underused by learners but *not* by novice writers. Novice L1 writers also better approximate expert writers' use of EAP verbs. For example, they do not make repeated use of *to conclude* in sentence-initial position or the verb *argue* in active structures with *people* or *I* as subject. They use a wider range of patterns similar to those found in ACAD.

As put forward by Howarth (1998: 186), '[a] much greater diversity in non-standard phraseology is found in non-native writing, reflecting learners' general lack of awareness of the phenomenon'. EFL learners use lexical verbs in phraseological patterns that are not found in native writing. Example 20 shows that, in ICLE-FR, sentence-initial 'to conclude', is very often followed by a hedging device introduced by a first person pronoun (in italics).

Example 20

(20) *To conclude, I will say that* we have to be careful . . .

To conclude, I would rather consider Europe as a nation . . .

To conclude, my opinion is that television can be considered as the opium of the masses . . .

To conclude, we should acknowledge that although television is the new opium . . .

To conclude, I shall once more insist on the fact that a world in which dreams . . .

In addition, a significant proportion of learner specificities are transfer-related (cf. Nesselhauf, 2005; Paquot, 2008). For example, French learners use the erroneous colligation *discuss *about* as a translation of Fr. *discuter de* (Example 21).

Example 21

You can *discuss *about* several points of view and compare the different opinions. (ICLE-FR) (Correction: *discuss several points*)

These results clearly show that learners and novice L1 writers cannot be included in one undifferentiated category of novice writers. The two types of writing have some shared characteristics as both writer populations are students who are learning academic writing conventions. However, there is only a partial overlap between the difficulties of EFL learners and novice L1 writers. Learner writing contains many examples of difficulties that arise from the fact that the learners are writing in a foreign language (difficulties, for example,

with selecting the appropriate preposition after a verb, or the right delexical verb with nouns such as *claim, decision* and *argument*) and are strongly influenced by their L1.

6   Conclusion

The field of EAP vocabulary has so far been largely dominated by a lemma-based approach. Our study shows that a dual approach – combining both lemmas and word forms – gives us a more precise picture of the diversity of form-meaning mappings that characterizes the use of EAP verbs. Automatic retrieval of verbs from academic texts produced by EFL learners and expert writers is a powerful first step towards our goal of understanding EFL learner difficulties and enhancing EAP teaching tools. Three main findings emerge from our study. The first is that EFL learners significantly underuse the majority of 'academic verbs', that is, verbs like *include, report* or *relate*, that express rhetorical functions at the heart of academic writing, and instead tend to resort to 'conversational verbs', that is, verbs like *think* or *like*, that are characteristic of informal speech. The second is that when learners use academic verbs, they tend to restrict themselves to a very limited range of patterns, which contrasts sharply with the rich patterning that characterizes expert writing. Our study therefore adds support to Ellis et al.'s (2008: 391) observation that, even at an advanced proficiency level, learners still 'need help to recognize the distinctive formulas that are special to EAP', a major prompt for including language awareness exercises targeting these formulas in EAP classes. Thirdly, a comparison between ICLE and LOCNESS data has demonstrated that, while novice native writers share a number of problems with EFL learners, the latter are faced with a much wider range of difficulties, many of which are exclusive to the learner population.

This corpus-based study has demonstrated the tremendous potential of corpus approaches to EAP but also the many challenges they pose to EAP researchers. Besides the issue of comparability which, as shown in our study, has a major impact on the results and conclusions that can be drawn from them, a series of other issues require further investigation. Prime among these is the very notion of academic prose as a single register, 'an overly blunt instrument' according to Hunston (2002: 103), and one whose very existence has been called into question in a number of recent EAP and ESP studies (cf. Hyland & Tse, 2007). A natural next step in our work is to investigate to what extent the verb patterns displayed in our large academic corpus hold across individual disciplines (cf. Granger & Paquot, 2009). This investigation is just one small step on the long journey toward mapping out the features of native and learner EAP corpora – still a largely underexplored territory.

Notes

1   See Ädel (2006: 205–208) for an excellent discussion of corpus comparability in learner corpus research.
2   See http://ucrel.lancs.ac.uk/claws/
3   Perl is a programming language that is most helpful to corpus linguists as it provides powerful text processing facilities (cf. Danielsson, 2004).
4   From here on, we use small caps to refer to lemmas and italics for word forms.
5   The version of the GSL used is the one that was created in 1995 by John Bauman and Brent Culligan. This list includes all 2,000 capitalized headwords from the original General Service List of West (1953), plus 284 more, ranked and presented in frequency order based on the Brown Corpus. It is available from www.auburn.edu/~nunnath/engl6240/wlistgen.html
6   This overuse is clearly topic related: for example, the verb BAN is used in essays which answer the prompts '*Citizens in the USA should not be allowed to own guns'* and '*The role of censorship in Western society'.*

References

Ädel, A. (2006). *Metadiscourse in L1 and L2 English.* Amsterdam: Benjamins.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics, 22*, 173–195.

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

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow: Longman.

Britton, B.K. (1994). Understanding expository text: Building mental structure to induce insights. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641–674). New York: Academic Press.

Burnard, L. (2003). Reference guide for BNC-Baby. Retrieved from www.natcorp.ox.ac.uk/corpus/baby/index.html (accessed 8 May 2009).

Charles, M. (2006a). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*, 310–331.

—(2006b). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics, 27*, 492–518.

Coxhead, A. (2000). A new Academic Word List. *TESOL Quarterly, 34*, 213–238.

Danielsson, P. (2004). Simple Perl programming for corpus work. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 225–246). Amsterdam: Benjamins.

De Cock, S. (2003). *Recurrent sequences of words in native speaker and advanced learner spoken and written English: A corpus-driven approach.* Unpublished PhD dissertation. Louvain-la-Neuve: Université catholique de Louvain.

Ellis, N.C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 41,* 375–396.

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* (pp. 102–121). New York: Addison Wesley Longman.

Gilquin, G., Granger, S., & Paquot, M. (2007a). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes, 6,* 319–335.

—(2007b). Writing sections. In M. Rundell (Ed.), *Macmillan English dictionary for advanced learners* (2nd edn) (pp. IW1–IW29). Oxford: Macmillan Education.

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1,* 41–61.

Gramley, S., & Pätzold, M. (1992). *A survey of modern English.* London: Routledge.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds), *Languages in contrast: Text-based cross-linguistic studies. Lund Studies in English 88* (pp. 37–51). Lund: Lund University Press.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *The International Corpus of Learner English. Version 2. Handbook and CD-ROM* (2nd edn). Louvain-la-Neuve: Presses Universitaires de Louvain. Available from www.i6doc.com

Granger, S., & Paquot, M. (2005, May). *The phraseology of EFL academic writing: Methodological issues and research findings.* Paper presented at AAACL6 & ICAME26, Ann Arbor, Michigan.

Hiltunen, T. (2006). Coming-to-know verbs in research articles in three academic disciplines. In *Proceedings of the 5th International AELFE Conference* (Asociación Europea de Lenguas para Fines Específicos) (pp. 246–251). Retrieved from www.unizar.es/aelfe2006/

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features.* Mahwah, NJ & London: Erlbaum.

—(2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar.* Mahwah, NJ and London: Erlbaum.

Howarth, P. (1998). The Phraseology of learners' academic writing. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161–186). Oxford: Oxford University Press.

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics, 20,* 341–367.

—(2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18,* 41–62.

Hyland, K., & Milton, J. (1997). Qualifications and certainty in L1 and L2 students' writing. *Journal of Second Language Writing, 6,* 183–205.

Hyland, K., & Tse, P. (2007). Is there an 'Academic Vocabulary'? *TESOL Quarterly, 41,* 235–253.

Johns, T., & Scott, M. (1993). *MicroConcord corpus (Collection B).* Oxford: Oxford University Press.

Liao, Y., & Fukuya, Y.J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning, 54,* 193–226.

Lorenz, G. (1999). Adjective intensification: Learners versus native speakers. A corpus study of argumentative writing. *Language and Computers: Studies in Practical Linguistics 27.* Amsterdam and Atlanta: Rodopi.

Meyer, P.G. (1997). *Coming to know: Studies in the lexical semantics and pragmatics of academic English.* Tübingen: Gunter Narr.

Neff, J., Ballesteros, F., Dafouz, E., Martínez, F., & Rica, J.P. (2004). The expression of writer stance in native and non–native argumentative texts. In R. Facchinetti & F. Palmer (Eds), *English modality in perspective* (pp. 141–161). Frankfurt: Peter Lang.

Nesselhauf, N. (2005). *Collocations in a learner corpus.* Amsterdam: Benjamins.

Oakey, D.J. (2005). Academic vocabulary in academic discourse: The phraseological behaviour of EVALUATION in economics research articles. In E. Tognini–Bonelli & G. Del Lungo Camiciotti (Eds), *Strategies in academic discourse* (pp. 169–183). Amsterdam: Benjamins.

Paquot, M. (2007). *EAP vocabulary in EFL learner writing: From extraction to analysis: A phraseology-oriented approach.* Unpublished PhD thesis. Université catholique de Louvain, Centre for English Corpus Linguistics.

—(2008). Exemplification in learner writing: A cross-linguistic perspective. In S. Granger & F. Meunier (Eds), *Phraseology in foreign language learning and teaching* (pp. 101–119). Amsterdam: Benjamins.

Recski, L.J. (2004). Expressing standpoints in EFL written discourse. *Revista Virtual de Estudos da Linguagem, 2,* (3). Retrieved from www.revel.inf.br/site2007/ed_anterior_list.php?id=3 (last accessed on 20 June 2008).

Scott, M. (2004). *WordSmith Tools 4.0* [Computer software]. Oxford: Oxford University Press.

Shaw, P. (1992). Reasons for the correlation of voice, tense, and sentence function in reporting verbs. *Applied Linguistics, 13*, 302–319.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Sjöholm, K. (1998). A reappraisal of the role of cross-linguistic and environmental factors in lexical L2 acquisition. In K. Haastrup & A. Viberg (Eds), *Perspectives on lexical acquisition in a second language* (pp. 209–236). Lund: Lund University Press.

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

—(2004). Then and now: A reconsideration of the first corpus of scientific English. *Ibérica, 8,* 5–21.

Swales J., & Feak, C. (2004). *Academic writing for graduate students: Essential tasks and skills.* Ann Arbor, MI: University of Michigan Press.

Thomas, S., & Hawes, T.P. (1994). Reporting verbs in medical journal articles. *English for Specific Purposes, 13*, 129–148.

Thompson, G., & Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics, 12*, 365–382.

Tognini–Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam & Philadelphia: Benjamins.

Werlich, E. (1976). *A text grammar of English*. Heidelberg: Ouelle and Meyer.

West, M. (1953). *A general service list of English words*. London: Longman.

Williams, I.A. (1996). A contextual study of lexical verbs in two types of medical research reports: clinical and experimental. *English for Specific Purposes 15*, 175–197.

Chapter 11

Linking Adverbials in Student and Professional Writing in Literary Studies: What Makes Writing Mature

Philip Shaw

1   Introduction

Much research on academic discourse has aimed at identifying linguistic differences between more and less proficient writing, to provide a basis for targeted teaching (Hewings, 2004). Such work compares the incidence and function of various elements in two corpora either by concordancing relatively large quantities of text or by closely analysing discoursal functions. A frequent parameter for such comparisons has been the use of what Biber et al. (1999) call 'linking adverbials' and Gardezi and Nesi (this volume) 'conjunctive adjuncts'. These are sentence adverbials like *however* and *for example* which perform metadiscoursal functions. Functionally they fall into Hyland and Tse's interactive category of metadiscourse (2004), mainly as transitions, frame markers and code glosses. Transitions (*however, therefore*) are much more common in academic writing than frame markers (*firstly, finally*) or code glosses (*that is*). However, 'linking adverbial' is a formal category from Biber et al. (1999) and many text items that fall into Hyland and Tse's functional categories are, for example, coordinating or subordinating conjunctions (*but, because*), sentence stems (*my purpose here is to . . .*) or (complex) prepositions (*such as*) and not linking adverbials.

   Biber et al. (1999: 887) divide linking adverbials into several semantic categories. Table 11.1 shows those that are most frequent in academic writing, classified by semantic category (including two stance adverbials that have linking functions, *in fact* and *indeed*, 858, 562). The most frequent categories in Biber et al.'s large sample of mixed-discipline academic writing are apposition (corresponding approximately to Hyland and Tse's code glosses), contrast-concession (or adversative) and result-inference (both corresponding to Hyland and Tse's transitions).

   Different text types have different profiles of linking adverbial use. In academic prose the most frequent types are *however, thus, therefore, for example*, and

Table 11.1   Frequencies of some linking adverbials

Frequency band (per million words)	Enumeration and addition	Apposition	Result-inference	Contrast/concession	Stance
1,100				*however*	
700			*thus*		
600		*for example*	*therefore*		
400			*then*		
200		*e.g., i.e.*	*so*		*indeed*
100	*first, finally, furthermore, in addition*	*that is*	*hence*	*rather, yet, on the other hand*	*in fact*

Source: Adapted from Biber D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

then (Table 11.1), while in conversation *so, then, though,* and *anyway* predominate (Biber et al., 1999: 887). Apart from their profile, researchers have considered the density (i.e. overall frequency) of linkers, and their position in the clause (i.e. initial, medial or final). Thus Biber et al. (1999) find for academic writing a relatively high density of linkers – some 720 per million words – with roughly equal proportions in initial and non-initial position, with medial predominating in the non-initial category.

Eighty per cent of the linking adverbials found by Biber et al. were single-word adverbs, and most of the rest are short stereotyped prepositional phrases. As Gardezi and Nesi (this volume) observe, linking adverbials are simple concrete items, easily taught and useful in drawing learners' attention to the importance of logical coherence. As a result, writers of learner genres, and particularly non-natives, are often encouraged to use linking adverbials to articulate the structure of their argument. Some justification for this pedagogical practice comes from evidence that non-native writers use fewer linkers and are more likely to place them initially (Aarts & Granger, 1998; Altenberg & Tapper, 1998). Other studies show overall frequencies for L1 and L2 users that are similar to one another, but under-use of adversative connectors and over-use of additive and appositive ones by L2 writers (Granger & Tyson, 1996). Most of the investigations cited by Wolfe-Quintero et al. (1998) found no correlation between essay quality and linker use, although Grant and Ginther (2002) found that higher scoring essays included more conjuncts (roughly = linking adverbials). Nevertheless longitudinal studies of L2 learners show increases in linking adverbial density (Shaw, 2004; Shaw & Liu 1998). These results indicate that learners of English as an L2 need to be taught, are taught and can learn this class of words, thus justifying the practice of teaching linking adverbials. However, learner genres are not ends in themselves, but means to ends including acquisition of transferable skills applicable in professional

or academic genres, and it is therefore relevant to ask whether the way linking items are used in learner genres is appropriate in real-life genres.

Some research has suggested a negative answer. A number of studies have shown a higher density of linking adverbials for L2 students compared to L1 students and/or professional L1 writers (Bolton et al., 2002; Green et al., 2000; Milton & Tsang, 1993). Differences in the usage profile have also been found, with L2 students using more non-adversative conjuncts (Green et al., 2000; Milton & Tsang, 1993) and a disproportionately high use of *however* by British students and of *so* and *and* by L2 students (Bolton et al., 2002). Several possible explanations exist for these differences. The high frequency of non-adversative conjuncts could signal either a discourse that includes much mere listing and addition, or a writer mistakenly marking relations that are not additive. Alternatively, the profile may reflect stylistically inappropriate conjuncts such as *besides*, which is simply not used as much in mature writing as it is by East Asian learners (Bolton et al., 2002).

There is a sort of paradox here, in that within one learner genre, like the test essay, higher-rated products may have more linkers, and development is associated with using more linkers, but comparing learner genres with professional ones, fewer linkers appear in the work of the presumably more skilled group. One possible explanation is simply that the texts being compared belong to different genres. Much of the material used to set norms in the studies just described was journalistic and generically quite different from student essays. Even within academic writing, register differences may be expected among educational, learner and research genres.

In fact, Biber et al. (1999) show that linking adverbials are much more frequent in academic writing than in journalism (about 7,000 versus 1,750 instances per million words). Within academic prose, Hyland (1999) shows a higher density of logical linkers in textbooks than in research articles. Textbooks address an audience supposed to be less knowledgeable than their writers, and thus feature 'heavy use' of logical connectives and 'extensive use' of frame markers and code glosses for pedagogical purposes (Hyland, 2000: 117), while research writers can assume that readers will fill in the links. It is possible therefore that a high density of linkers in learner genres is similarly a consequence of text purpose: the writers have to show the readers/graders explicitly that they understand relationships correctly. Furthermore, patterns of linker usage are discipline-specific, presumably reflecting different types of content. Tse and Hyland (2006: 187) show that within academic book reviews, both the overall density and the frequency of various types of metadiscoursal devices varied across disciplines.

Since both genre and discipline influence density and profile, one might question aspects of the studies which claim inappropriate over-use by students. The corpora use mixed or unspecified disciplines and compare student essays with research articles or even journalistic prose. But Bolton et al. (2002) found

higher density in L2 essays relative to closely comparable L1 student essays, and there is other evidence that L2 users differ from L1 users in this way. Field and Yip (1992) showed that Hong Kong schoolchildren writing L2 English had differences in position and profile from British schoolchildren: they 'over-used' sentence-initial linkers, especially (of course) *besides.* Narita et al. (2004) found that Japanese EFL learners, like other immature writers, used selected linking adverbials in sentence-initial position more often than native peers and that they had a characteristic profile: they overused *for example, first, moreover* and *in addition* while underusing *then, yet* and *instead.* The problem in these cases is probably not simply linguistic, but generic or rhetorical. Corpus data do not of course show whether the L2 writers used a discourse which enumerated more and argued less, or simply used linguistic marking for enumeration more and argument less, and there are arguments for both interpretations. One may speculate that the processing demands of writing in a foreign language, often under time pressure, lead to more enumeration and less argument across a wide range of cultural backgrounds. If this is so, differences in linker profiles are indices of different discourses, and this is consonant with the different profiles of disciplines mentioned above. However Gardezi and Nesi (this volume) found that writers of different cultural backgrounds but similar language proficiency and hence facing similar processing demands marked enumeration to different extents, so the difference may also be due to different cultural styles.

It is thus genuinely not uncommon in skill-display argumentative essays for L2 users to over-use (some) linkers. But skill-display essays are actually of a rather specialized type, described by Coffin and Hewings (2004: 169) as requiring 'a distinctive argumentative style which cannot be closely modelled on professional academic genres'. Does the same apply in more purposive and discipline-specific writing? At least one study has tried to investigate this question. Chen (2006) compared a generically mixed selection of writing by ten Taiwanese MA TESOL students with a corpus of 16 relevant journal articles, thus comparing the (advanced) student register with the researcher register within a discipline. Linker density in the student texts was somewhat higher, though her students wrote shorter sentences than her researchers, so that although they had rather more linkers per 1,000 words, they actually had somewhat fewer per thousand sentences. The researchers had a higher proportion of adversative linkers than the students, due mainly to their using *however* much more. Nevertheless, in other respects the profiles were similar. Both groups had *however* as the most frequent linking adverbial (as in Biber et al., 1999, cited above and in Gardezi and Nesi, this volume), and both had *therefore, for instance/example,* and *thus* in some order as the next three, again as in Biber et al. (1999). This suggests that, in this fairly soft applied area, advanced learner genres written by L2 users and researcher genres written by professionals shared a repertoire of linking devices and used them with similar frequency.

Chen seems to have done the only study of linkers in student and professional writing within a single discipline but her sample is small and confined to a discipline whose practitioners are likely to be highly conscious of linking adverbials. The implication of the other studies is that texts by students and researchers are different in density and profile of linker use, and it would therefore be interesting to find out what is the case in other disciplines and thus whether 'over-use' of linking adverbials is generic, developmental or teaching-induced.

The studies discussed above use the figures produced by the concordancer directly to comment on over-use and under-use of items by different groups of writers. They do not systematically refer back to the texts to see what functions the items have. Closer examination of this kind would enable the investigator to characterize the nature of the maturity or genre differences underlying the over-use or under-use. This study, like the volume in which it appears, aims to take this further step.

2   This Study

This investigation aims both to give information about student writing as such, and to investigate a disciplinary discourse which remains largely uncharted at the linguistic level. It aims, like Chen's (2006) study, to compare a learner genre and a research one to assess the extent to which texts with these two different statuses within the same discipline share common features in terms of the density, profile and position of linking adverbials. It aims to collect accurate frequency data by corpus-linguistic techniques and then use discourse-analytic techniques (move analysis as in Swales, 1990, 2004) to look beyond the numbers for an explanation of the differences.

The discipline chosen is literary studies. This is an area characterized by relatively weak generification, and published articles may seem very different from the conventional IMRD of other disciplines. Research articles in literary studies typically lack epistemically oriented introductions which follow something like the Swales 'creating a research space' (1990, 2004) model, and correspondingly have more phenomenally oriented ones which launch directly into discussion of the specific theme of the article, often via quotation or anecdote or *in medias res* with a topic sentence (MacDonald, 1994; Shaw, 2001). Nevertheless most of the texts include a great deal of straightforwardly argumentative material which presents grounds and makes claims just as natural-scientific writing does. A typical pattern is: Preview claim – quotation – interpretive recount – claim marker – claim (Shaw, 1998, 2001; cf. Tucker, 2004). Literary-critical articles thus belong to the broad genre of scientific research articles (cf. Afros, 2007) and we can thus compare across disciplines. Given the low level of generification and the effort to avoid hackneyed phrases (Elbow, 1991), it would, however,

be as well to find out how uniform the register of the discipline can be expected to be before making such a comparison.

An advantage of literary studies as an area for comparing student and professional writing is that both researchers and learners are expected to express an original response to the work of art in a form that is not openly stereotyped. The student has to display her skills by finding a new angle rather than by reproducing knowledge, so it is an area where research and learner genres might not be very different.

In this chapter I use concordancing software to examine the density, profile and position of linking adverbials in various corpora, and then analyse some discourse in more detail to provide explanations of the findings. First, two separate corpora of articles in literary studies are examined to assess how uniform the literary-critical register can be expected to be in terms of profile density and adverb position. These results are then compared to a hard-science corpus and to 'general academic' findings to assess what is characteristic of the literary-studies register. A corpus of student essays is then examined to assess the differences between professional and student writing within literary studies. Finally, uses of some linkers in their contexts are analysed to provide a thicker description of the differences.

3   Corpora

The student corpus (STULIT: 80,743 words) is a convenience sample consisting of 30 essays by Newcastle University first-year students (collected by Hsin-Hsin Yang as part of her PhD work: Yang, 1998) and nine by corresponding students at Sheffield University collected in 2002. The article corpus (PROFLIT: 254,263 words) consists of 15 articles from *English Literature in Transition* and 15 from *Nineteenth-Century Literature*. The journals were chosen to achieve a selection of published articles which was compatible generically with the student writing, in that it put forward theses about particular works of art supported by quotations and references to their content, and also compatible in content, in that many of the student essays were about nineteenth- or early twentieth-century novels and novelists. The articles were selected by taking one article, the first that seemed to meet the criterion above, from each issue over the period 1994–2005.

Two other corpora are referred to. One is OLDLIT (323,882 words), another corpus of published articles in literary criticism (used in Shaw, 2001) on specific works and writers, consisting of two articles each from 15 journals; this provides a check that features discussed are generalizable to other samples of literary-critical writing. Since we do not know that the register of literary-critical writing is actually a consistent one, it cannot be assumed without evidence that the characteristics of a sample are typical of the disciplinary register, but if two samples have features in common, the evidence is better.

The other corpus is MATH (81,392 words) a collection of five articles each from 2004–2005 issues of the *Mathematische Zeitschrift, Physics Letters A* and *Journal of Engineering Mechanics*, designed to represent text including a high proportion of mathematical expressions. This is used merely to highlight differences in usage between two different disciplinary discourses, as opposed to differences between a disciplinary discourse and the average of all academic writing, as in Biber et al.'s (1999) and Bolton et al.'s (2002) data.

4   Items Searched for and Method

Textual linking functions like those of *therefore* (Hyland's transition), *firstly* (frame marker) or *for example* (code gloss) can be performed by members of several grammatical classes, including among others coordinating conjunctions (*and, but, so, yet*) and subordinators (*because, since, although*). Writers adopting a functional viewpoint can have categories like Hyland's transitions whose members are quite disparate, but writers like those described above who look at a formal category of markers have to decide what counts as a linker (whatever term they use) and if possible do this explicitly enough for replication.

One problem is the range of linking functions counted. Linking adverbials link sentences, clauses and phrases to previous sentences, clauses or phrases (*He likes rain and therefore Ireland*). Prototypical linking adverbials like *therefore* can occur in all these roles and in a variety of positions in the clause: initial, medial of various types, and (occasionally) final. In this study all linking functions have been counted but only items functioning to link sentences have been classified as initial or non-initial, because placement in subordinate or coordinated clauses seemed to be governed by different factors.

Another issue concerns inclusion or exclusion of words like *and, but, so, or*, and *yet* used as linking adverbials, that is, starting a new orthographic sentence. Tolerance of such usage might well distinguish skill levels, genres or disciplinary styles, which argues for their inclusion. Furthermore, *but* competes directly with *however*, so information is lost if it is not considered. In fact Gardezi and Nesi (this volume) found that their two culturally different samples contained similar numbers of adversatives, but different proportions of *but* and *however*; ignoring *but* would have concealed the similarity in adversatives. Adverbial uses of these five words are therefore included in this study, but in the cases of *and, but* and *or* only if they are sentence-initial.

A third, related, difficulty is that, as Biber et al. (1999) point out, a given adverbial form may be an exponent of more than one functional category – *thus* may be a circumstance adverbial meaning 'in this way' as well as a linker – and many words like *then* and *so* have a variety of other functions than linking adverbial. In this study items have been hand-checked and non-metadiscoursal, non-adverbial uses excluded.

In the corpus-based phase of this project I used *AntConc* (Anthony, 2006) to search for the words listed in Appendix 1. The list aims to include all adverbials discussed in the previous literature which can be regarded as realizations of interactive metadiscourse, excluding subordinating conjunctions. I included two stance adverbials, *indeed* and *in fact*, because they are examined in many previous studies and because of their proposition-connecting function.

All instances which were not adverbial and not metadiscoursal were then laboriously excluded. The metadiscoursal linkers were then classified as marking intersentential links or intrasentential ones. Finally the intersentential category was subdivided into sentence-initial and non-initial. All cases of *and*, *but* and *so* counted were intersentential and initial, and all cases of *though* counted were intersentential, as a consequence of the selection criteria above.

The three parameters of density, profile and position were defined as follows. The raw numbers of linking adverbials in each position were summed to get totals for the corpus in question. Then all these numbers were reduced to frequency per 100,000 words for comparability among corpora of different sizes. This means that a figure of 5 refers to about 15 instances in 30 texts in PROFLIT and OLDLIT, and to about four instances in 38 or 15 texts in STULIT and MATH respectively. Clearly numbers at this low level are of little significance, but I did not aggregate numbers into larger categories like 'adversative' because it likely that each linker actually performs a different function (Charles, in press).

Overall density was the total frequency of all linking adverbials per 100,000 words in the corpus. The profile was a list of all the adverbials ranked in order of frequency and their individual frequencies, so that ranking and frequency could be compared with other such lists. (Tables given here omit frequencies under 5 per 100,000 words, for readability.) Position was measured by counting only linkers functioning intersententially, and dividing the frequency of initial placement by the frequency of non-initial placement, both for the sum of all items, and for each individual linker. If this ratio is 1, initial and non-initial placement are equally frequent; if it is greater than 1 initial placement is more frequent, and if it is less than 1 the majority of adverbials are placed non-initially. Thus density is a measure of overall frequency of all linkers or of a specific item, profile describes the relative proportions of individual linkers, and the position ratio shows the ratio of initial to non-initial uses.

Once these numerical measures had been ascertained by corpus-linguistic techniques, discourse analysis was used to obtain qualitative and explanatory insights. The figures suggested that interesting items, such as those which have different frequencies or position ratios in the corpora, be compared, and the concordance lines for these were scanned and where necessary expanded or referred back to the whole text. Thus the items were viewed in context to observe commonalities or contrasts in use. This enabled the investigator to explain the numerical differences found and to understand the discoursal differences that underlie them.

5 Results

5.1 The literary-critical register

The density of linking adverbials as defined above in PROFLIT was 540.1 and in OLDLIT it was 580.1 These numbers are sufficiently similar to suggest a range in which one can expect the density of such forms to lie in literary-critical discourse, well below the 720 or so Biber et al. (1999) found for mixed-discipline academic material.

The profile for individual expressions is given in Table 11.2. The adverbials *however, yet, thus, but, indeed, in fact,* and *then* occur above 20 times per 100,000 words for both corpora; the main differences seem to be that in PROFLIT

Table 11.2 Frequency of various adverbials in OLDLIT and PROFLIT

Frequency band (instances per 100,000 words)	PROFLIT	Frequency	OLDLIT	Frequency
60–69	*however*	69.2		
	yet	63.3		
50–59	*thus*	52.3	*thus*	58.0
			however	54.6
			yet	51.6
40–49			*and*	49.1
			then	46.3
			indeed	44.8
30–39	*but*	37.4	*but*	38.3
20–29	*indeed*	29.1	*in fact*	24.7
	in fact	27.9		
	for example	26.7		
	then	25.2		
10–19	*rather*	16.1	*rather*	19.5
	for instance	15.7	*therefore*	19.5
	instead	15.3	*for example*	17.3
	and	14.6	*that is*	15.7
	therefore	13.8	*moreover*	12.0
	hence	11.4	*finally*	11.4
	nevertheless	11.4		
5–9	*similarly*	9.0	*in other words*	9.9
	on the other hand	8.7	*similarly*	9.6
	in other words	7.1	*for instance*	8.3
	furthermore	6.7	*instead*	7.7
	nonetheless	5.1	*further*	6.5
	still	5.1	*nonetheless*	6.2
	in contrast	5.1	*on the other hand*	6.2

Table 11.3 Sentence position ratios (initial frequency/non-initial frequency) of selected adverbials functioning intersententially

	PROFLIT	OLDLIT	MATH
then	All non-initial	All non-initial	5.6
therefore	0.6	All non-initial	all initial
hence	0.8	All initial	7.5
for example	1.6	0.23	3.0
on the other hand	1.2	0.56	all initial
thus	1.9	0.86	28.5
however	0.5	0.28	1.8
in fact	2.1	1	0.7
indeed	6.1	4.5	all initial
moreover	9	3.9	18.0
yet	22.0	16	none
All	**2.1**	**1.6**	**5.5**

use is more concentrated on *however* and *yet* and that the writers or editors of OLDLIT have allowed more sentence-initial use of *and*. Even at lower frequencies the profiles of the two corpora seem very similar.

For PROFLIT the position ratio is 2.1 (around two-thirds initial) and for OLDLIT it is 1.6 (around three-fifths initial). Biber et al. (1999: 891) show a ratio of about 1, which is lower than I find in any sample. Some details are given in Table 11.3.

5.2   Comparison of literary-critical and other academic discourses

Chen (2006) used a similar definition of linking adverbials to that used here and found 720 per 100,000 in applied-linguistics articles, considerably higher than in OLDLIT and PROFLIT, but very similar to what Biber et al. (1999) found for general academic writing (with a somewhat different definition of linkers). The density in MATH of linkers as defined here was 893/100,000, much higher than in the literary-critical corpora. The figures in Gardezi and Nesi (this volume) suggest a range for student economics writing, for intrasentential connectors only, between 768 (British) and 996 (Pakistani). This suggests that literary-critical discourse is characterized by a relatively low density of markers.

Given that Biber et al. did not count *and* and *but*, and I did not count *e.g.* or *i.e* (on the grounds that they are rarely intersentential, Biber et al. 1999: 890), the profile of PROFLIT and OLDLIT differs from the general academic one in Table 11.1 mainly in *yet* being much more common. Chen's applied-linguistic profile is more different, with *in other words*, *that is*,

therefore and *furthermore* ranked higher and a corresponding lower ranking of *yet*, *but*, *indeed* and *rather*. This suggests an applied-linguistics discourse with more code glosses and a literary-critical one with more adversative transitions.

The profile of MATH is very different from those of the literary corpora, with much greater frequency above all of *then* (281/100,000), but also of *thus* (104), *so* (68), *hence* (36) and *therefore* (30) and lower frequencies of *yet* (no cases), and *but* (7.4/100,000), and to a lesser extent of *indeed* (15) and *in fact* (13). It is also very different from literary criticism in terms of adverbial placement. In the literary-critical corpora one-third or more of adverbials linking sentences were non-initial, as noted above. However, in MATH only a sixth were non-initial (position ratio 5.5). One reason for this is that the usage and function of the frequent *then* was completely different in the two disciplines. In MATH *then* is usually sentence-initial in logical-consequence constructions like *Let x = 2. Then y = 6*, while in PROFLIT and OLDLIT it is never sentence-initial and is used in summative constructions like *Such fears, then, help to form the bounded selfish love of Experience*. But the pattern of higher ratios in MATH, that is, predominantly initial placement, is similar for most other linkers, as Table 11.3 shows. (Note that in a sentence that begins *And yet, yet* is non-initial.)

It has thus been established that OLDLIT and PROFLIT are similar to one another and different in consistent ways both from mixed-discipline corpora and from corpora representing other disciplines. They would seem to represent a literary-critical discourse with a fairly low density of linkers, a profile marked by numerous adversatives (cf. Barton, 1995), and especially the word *yet*, and a proportion of sentence-initial adverbials that is very different from that in hard science.

5.3 Comparison of professional and student literary-critical writing

At 844.7, the overall density of linking adverbials in STULIT is much higher than in PROFLIT and OLDLIT and comparable to that in Gardezi and Nesi (this volume). As in several other investigations, students use more linkers than professionals, and the reasons for this must be established by discourse analysis.

Table 11.4 shows the forms in STULIT with a frequency of 5 or more. It also gives the average of their frequencies in OLDLIT and PROFLIT, which is nearly always lower than the figure for STULIT because STULIT uses more linkers overall (i.e. the overall density is higher), presumably for generic reasons. Table 11.4 therefore also gives the average multiplied by 844/560. This last figure, the adjusted average, is the frequency that the form would have if

Table 11.4 Frequency of selected linking adverbials in STULIT compared with the average of PROFLIT and OLDLIT

	1. STULIT	2. Average in OLDLIT and PROFLIT	3. Average in OLDLIT and PROFLIT adjusted for density
however	122.6	61.9	93.3
yet	104.0	57.5	86.6
thus	80.5	55.2	83.1
therefore	86.7	16.7	25.1
for example	50.8	22.0	33.2
indeed	40.9	37.0	55.7
in fact	34.7	29.7	44.8
then	37.2	35.8	53.9
again	31.0	1.2	1.8
but	27.2	37.9	57.0
though	22.3	1.8	2.8
similarly	21.1	9.3	14.0
hence	19.8	8.8	13.3
consequently	17.3	2.0	2.9
still	13.6	3.2	4.8
rather	12.4	17.8	26.8
nevertheless	12.4	8.2	12.3
for instance	8.7	12.0	18.1
instead	7.4	11.5	17.3
that is	7.4	7.9	11.8
furthermore	6.2	9.2	13.8
alternatively	6.2	0.8	1.2
and	6.2	31.9	48.0
at the same time	6.2	6.8	10.2
on the other hand	5	7.5	11.2
as a result	5	0.4	0.6
firstly	5	0.3	0.5
All linkers	**844.7**	**560**	–

the density were the same across samples, and thus gives a picture of the profile (i.e. the proportions of different linkers used) independent of the overall density. Column 1 of Table 11.4 shows that it is the density rather than the profile that is most different from other academic corpora and the literary corpora discussed here. STULIT has a typical academic-writing profile with high frequencies for *however, thus, therefore, for example* and *then*. In particular the profile of STULIT matches that of the professional literary writers (as against general academic and applied-linguistic profiles) in the high frequency of *yet* and *indeed*. There are some differences in the profile that are suggestive, however. Students seem to use *therefore, again*, and adverbial *though* considerably more than professionals, and to use sentence-initial *and* and *but* considerably less.

Table 11.5 Sentence position ratios (initial frequency/non-initial frequency) of selected adverbials functioning intersententially in OLDLIT, PROFLIT and STULIT

	OLDLIT	**PROFLIT**	**STULIT**
however	0.28	0.5	1.3
yet	16	22.0	34.0
thus	0.86	1.9	4.3
therefore	All non-initial	0.6	0.3
for example	0.23	1.6	1.7
indeed	4.5	6.1	6.7
in fact	1	2.1	2.5
then	All non-initial	All non-initial	All non-initial
again	none	0.5	1.7
but	All initial by definition		
though	All non-initial	All non-initial	All non-initial
similarly	2.1	4.5	all initial
hence	All initial	0.8	all initial
consequently	All initial	0.3	2.5
still	All initial	4.0	4.0
rather	3.2	5.5	1.0
nevertheless	2	3.3	5. 0
All	**1.6**	**2.1**	**1.9**

Finally, STULIT appears to be similar to PROFLIT and OLDLIT in terms of adverbial placement, with an overall ratio of 1.9. Table 11.5 shows that this is partly an artefact of the differences in profile, and in most cases adverbials in STULIT are slightly or considerably more often initial than in PROFLIT, which itself shows somewhat more initial use than OLDLIT. The students' writing is in the range of the professionals in terms of this parameter, but seems to tend to have more initial uses.

6 Explanatory Investigations

On the basis of these parameters it would be reasonable to say that the students were writing a high-density version of the literary-critical register. If a tendency to more initial placement is a marker of less skill, they are slightly less skilled (as one would expect). The profile differs from that of the published literary scholars in a lower frequency for *and* and *but* and a higher frequency for *again*, *though* and *therefore*.

Since *but* and *however* are close in meaning it is likely that the students' high score for *however* is in part compensation for their low *but* frequency. The reason for the low frequency of both *and* and *but* in the student corpus, as in MATH, is likely to be simple prescriptivism. Everyone (or possibly everyone

in Britain, cf. Gardezi and Nesi's figures in this volume on *But* in Pakistani student writing) is taught at some stage to be cautious in using co-ordinating conjunctions as adverbials (University of York, no date). Literary critics celebrate their freedom from convention by ignoring this advice.

The analysis that follows is based on reading concordance lines and selecting what appear to be typical examples and so is more suggestive than conclusive. I deal first with the linker *though*, which may be a marker of less mature or less 'written' style. Biber et al. (1999) found it to be almost confined to spoken language. Certainly Examples 1 and 2 using it from two student essays seem to have other stylistic features suggesting some lack of skill, like the nonfunctional repetition of *herself*, the somewhat unidiomatic use of *power*, and the use of *I believe* and *in my opinion*. Since there seems to be no generic reason for using *though*, and professionals do not seem to use the word as an adverbial, it is plausible that it is a feature of less mature writing. It should be noted, however, that the students use it in the medial position typical of written language, not the final one common in speech.

Example 1

Florinda attempts to save herself from rape by claiming that she is a lady; her power is short-lived though as she declares herself to be man's property. (sh08)

Example 2

Philosophy, though, cannot be limited to mere questioning – there must be some attempt to proffer an answer. I believe that to agree that Conrad is a philosophic writer, one must also agree that he does suggest some replies to his questions. In my opinion he does. (B14c1)

To understand the distribution of *again* and *therefore* it is necessary to refer to the preview – quotation – interpretive recount – claim marker – claim sequence often found in literary-critical writing. In student essays the writer is often required to state a position at the beginning, often in response to a prompt, and then to provide evidence for it. This means that the claim (or just the label) is often used to show that the interpretive recount confirms the position stated. The linker *again* seems to be used to show that the student is developing a consistent argument. One essay, responding to the prompt 'Conrad's narrative methods are complex but his themes are surprisingly simple: they include honour, loyalty and "solidarity". Discuss.' says, for example:

Example 3

(Para 14: (preview) *In all of the texts one will find recurrent themes such as honour, loyalty and 'solidarity', which can be challenging enough in their own right.* (interpretive recount) *For example, in Lord Jim . . . . The Nigger of the 'Narcissus', is another fine example of how . . . (14.7)* **Again,** (claim) *one can see here a consistency*

of ideas running through Conrad's work, yet it must also be recognised that if one looks further, a greater and more complex network of issues will arise.

(Para 15: preview) *Once **again** Heart of Darkness can provide us with some excellent material to validate this statement, for . . . .* (B18C1)

Another, responding to the prompt 'Give me the liberty to know, to utter and to argue truly, according to conscience, above all liberties.' (Milton) Does Milton condemn Eve for exercising liberty in Paradise Lost?' links two claims (marked by 'it is arguable' and 'can') with *again*, each time referring back to the relevant interpretive recount with a *such* phrase.

Example 4

It is arguable that by placing Eve upon such a pedestal Adam made his disappointment inevitable.

. . . (several paragraphs)

***Again,** a level of condemnation can be portioned to Adam for allowing Eve to develop such a dependant nature:* (sh0102)

It can be suggested that *again* is generically appropriate to an essay in which the grader's attention has to be drawn to the student's success in doing the task assigned. It is very rare in the professional corpora. The word functions to signal that the writer is aware that the same point is being made again. As a reviewer of this chapter has pointed out, repetition of this kind may be desirable in displaying one's knowledge of a number of books and ability to perceive repeated patterns, but less so if the aim is to enlighten the reader as in published articles.

The linker *therefore* occurs in all the literary corpora, but much more frequently in the student writing. In the professional writing it is often used, synonymously with *thus* (Biber et al., 1999: 889) to connect a claim to an interpretive recount, as in Example 5.

Example 5

Greifenhagen's imagery . . . [interpretive recount of picture content]. . . . (claim 1) *If Conrad's own description is unsettling, . . . Greifenhagen's image is doubly so, for it provides only one image of the coolies: an image that not only types them as unproductive, that not only effaces the specificity of their individual demeanor, but also one that crucially reinforces the unfathomable distance inscribed into the picture by the demarcating rope in the foreground – between Western spectator and Chinese coolie.*

(claim 2) *As Greifenhagen's illustration suggests, the reception of 'Dead Reckoning' and Typhoon **therefore** would have toed the line between the intricacies of authorial intent and more reductive readerly expectations informed by the pervasive cultural and literary climate encouraged by the imaginary Yellow Peril.* (Forman, 2004)

The example is quoted at some length to illustrate two typical features of the professional writing. One is its multiple rewriting: the picture is described interpretively, then a claim is made about the meaning of the image, then this is used as the basis for a much wider claim about the reception of the stories analysed. The other feature is the large inferential leaps made from data to successive levels of claim. Only the highest level of claim is marked by a linker in this example. Contrast this with a student extract (Example 6):

Example 6
[the narrator] *is however scrupulously fair in his handling of the material, honest and objective as this is his western code. (2.5) The bare facts of Razumov's actions and psychological state are **therefore** left to stand for themselves . . .* (B12C1).

The linker is used to indicate a more general claim about the novel based on the interpretive recount of the narrator's role, in the same way as in Example 5, but the inferential leap is shorter and the claims both shorter and simpler. Because the claims are shorter, there is less textual space between the ideas connected by *therefore.* It may be relevant that Bunton (1999) considered metatextual linking over longer stretches to be essential for coherence in dissertations. Moreover the student writing seems to use the marker for a first-level claim while the professional made two leaps and only marked the second.

If *therefore* is synonymous with *thus* (Biber et al., 1999), these differences actually have little to do with *therefore* as such. If, on the other hand, *thus* typically indicates a less direct connection and has a more summative function, it is easy to see why students, who make more immediate connections between shorter units, might have more instances of *therefore* in their texts.

Both groups of literary-studies writers, the professionals and the students, use *yet* more than the academic average, so it seems genuinely to be characteristic of a literary-critical or humanities register. It is twice as frequent in all three literary-critical corpora as in Biber et al.'s sample (and only appeared once in the ICE-UK material in Bolton et al., 2002). So it would be of interest to see how students and professional writers use this typical feature of their register.

Referring back to the texts from the concordance lines for *yet* shows that professional writers often use it in creating a research space. But this move is not part of the student essay genre (not if the topic has been assigned, anyway) and actually the most common use of *yet* in literary-critical writing at both levels is within interpretive recounts, just preparatory to the transition to the claim, as in student Examples 7 and 8 and professional Example 9.

Example 7
*In Lord Jim honour is explored most fully: the title character spends most of the story trying to recover the honour that he loses at the beginning of the novel. **Yet** because of*

Conrad's use of many narrators, as readers we are wary of any one perspective and so (claim) *are not entirely convinced that his dishonour is damning.* (B17C1)

Example 8

Jim forfeits his honour when he abandons the 'Patna' – that is, he betrays the code of conduct based on 'solidarity'. Yet his guilt and his attempts to redeem himself show, (claim) *more clearly than any number of good intentions or any display of rectitude, how loyal he is to the very code he has betrayed, and how much faith he has in its existence.* (B15C1)

Example 9

It not only represents Jane's frustration that she cannot control others' assumptions about her, but also describes the day-to-day pressures celebrity puts on authors hounded by followers demanding information about their lives. **Yet** *if we stand Jane side-by-side with other representations of celebrity from periodical articles and longer fiction, it is apparent* (claim) *that Sinclair's response to the dilemma of celebrity omits an alternative perspective in early twentieth-century debates.* (Troy, 2004)

These examples show that *yet* can function similarly in professional and student texts, confirming that they are parts of the same general discourse, but they also show what the difference is: the professional writers have longer, more complex propositions between the markers and make more daring and significant claims.

7   Conclusion

Discourse-analytic techniques complement frequency studies because they enable data from such studies to be explained and made more usable in educational applications. In particular, they enable 'over-use' and 'under-use' to be interpreted and the frequency differences which might point to a need for intervention to be distinguished from those which are necessary features of generic or other functional differences between text-types.

In literary studies, and perhaps in many disciplines, student and research writing share register features: not only lexical items (technical terms) but also function words like linking adverbials. Both sets of texts, for example, have a characteristic use of *yet* in terms of both frequency and function. Nevertheless it is indeed the case that students 'over-use' linkers in relation to professional academic writers. At least in literary studies this over-use is the result of complex differences in generic demands, stylistic maturity, and above all, disciplinary maturity. The demand that one adopts a position in relation to a prompt creates a need for a different profile of linking adverbials than the desire to establish a research space, and this is one genre difference

underlying the differences in profile between professionals and students. Students use some forms that are stylistically inappropriate (like *though*). If they are writing in their first language these are likely to come from conversational use; if they are writing in L2, they can come from a variety of sources including excessive emphasis by teachers (like *besides*). If they are not familiar with the rhetoric or discourse of the genre and discipline they are writing in their writing may seem immature because they construct texts with dispreferred types of relations between propositions and therefore dispreferred linkers. But above all students tend to have a higher density of linkers in their essays than researchers because the linked propositions are shorter and simpler ideas. The same sort of propositions are connected by the same sort of linking adverbials, but the propositions are better developed in the published essays, so there is more content between the linkers in professional writing than in student essays.

Appendix

Words searched for

above all	*firstly*	*in short*	*overall*
accordingly	*first of all*	*in sum*	*rather*
additionally	*for example*	*in the first instance*	*second*
again	*for instance*	*in the last instance*	*secondly*
also	*for that reason*	*in turn*	*similarly*
alternatively	*further*	*indeed*	*so*
and (sentence-initial only)		*instead*	*still*
as a result	*furthermore*	*likewise*	*that is*
at any rate	*hence*	*later*	*that is to say*
at last	*however*	*lastly*	
at least	*in addition*	*meanwhile*	*then* (metatextual)
at the same time	*in any case*	*moreover*	*thereby*
besides	*in conclusion*	*nevertheless*	*therefore*
but (sentence-initial only)	*in consequence*	*next*	*third*
by contrast	*in contrast*	*nonetheless*	*though* (only adverbial uses with intersentential function)
consequently	*in effect*	*on the contrary*	*thus*
conversely	*in fact*	*on the other hand*	*to sum up*
finally	*in other words*	*on the whole*	*yet*
first		*or* (sentence-initial only)	
		otherwise	

References

Aarts, J., & Granger, S. (1998). Tag sequences in learner corpora: A key to interlanguage grammar and discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 132–141). London: Longman.

Afros, E. (2007). Promotional (meta)discourse in research articles in language and literary studies. Unpublished MA thesis, University of Waterloo, Ontario, Canada.

Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on computer* (pp. 80–93). London: Longman.

Anthony, L. (2006). *AntConc* [Computer software]. Retrieved from www.antlab.sci.waseda.ac.jp

Barton, E.L. (1995). Contrastive and non-contrastive connectives: Metadiscourse functions in argumentation. *Written Communication, 12*, 219–239.

Biber D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Bolton, K., Nelson, G., & Hung, J.A. (2002). A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics, 7*, 165–182.

Bunton, D. (1999). The use of higher level metatext in Ph.D theses. *English for Specific Purposes, 18,* 41–56.

Charles, M. (in press). Corpus evidence for teaching adverbial connectors of contrast: *However, yet, rather, instead* and *in contrast*. In N. Kübler (Ed.). *Corpora, language, teaching, and resources: From theory to practice*. Bern: Peter Lang.

Chen, C.W.-Y. (2006). The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learners. *International Journal of Corpus Linguistics 11*, 330–345.

Coffin, C., & Hewings, A. (2004). IELTS as preparation for tertiary writing: Distinctive interpersonal and textual strategies. In L. Ravelli & R. Ellis (Eds), *Analysing academic writing: Contextualized frameworks* (pp. 153–171). London: Continuum.

Elbow, P. (1991). Reflections on academic discourse: How it relates to freshmen and colleagues. *College English*, 53, 135–155.

Field, Y., & Yip, L.M. (1992). A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers of English. *RELC Journal 23,* 15–28.

Forman, R. (2004). Coolie cargoes: Emigrant ships and the burden of representation in Joseph Conrad's *Typhoon* and James Dalziel's 'Dead Reckoning'. *English Literature in Transition 1880–1920, 47*, 398–429.

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes, 15*, 17–27.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing, 9*, 123–145.

Green, C.F., Christopher, E.R., & Mei, J.L.K. (2000). The incidence and effects on coherence of marked themes in interlanguage texts: A corpus-based enquiry. *English for Specific Purposes, 19*, 99–113.

Hewings, A. (2004). Developing discipline-specific writing: An analysis of under-graduate geography essays. In L. Ravelli & R. Ellis (Eds), *Analysing academic writing: Contextualized frameworks* (pp. 131–152). London: Continuum.

Hyland, K. (1999). Talking to students: Metadiscourse in introductory textbooks. *English for Specific Purposes, 18*, 3–26.

—(2000). *Disciplinary discourses: Social interactions in academic writing.* Harlow: Longman.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*, 156–177.

MacDonald, S.P. (1994). *Professional academic writing in the humanities and social sciences.* Carbondale: Southern Illinois University Press.

Milton, J., & Tsang, E.S. (1993). A corpus-based study of logical connectors in EFL students' writing: Directions for future research. In R. Pemberton & E.S. Tsang (Eds), *Studies in lexis* (pp. 215–246). Hong Kong: HKUST Language Centre.

Narita, M., Sato, C., & Sugiura, M. (2004). Connector usage in the English essay writing of Japanese EFL Learners. *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal,* 1171–1174.

Shaw, P. (1998). Patterns of argument in published articles in literary studies. In L. Lundquist, H. Picht & J. Qvistgaard (Eds), *LSP identity and interface: Research, knowledge and society. Proceedings of XI LSP Symposium Copenhagen, August 1997.* Copenhagen: Copenhagen Business School, pp. 155–164.

—(2001). Forms of argument in literary criticism at two levels. In F. Meyer (Ed.), *Languages for special purposes: Perspectives for the new millennium* (pp. 501–509). Tübingen: Gunter Narr.

—(2004). The development of Swedish university students' written English, appropriacy, scope, and coherence. In T. Caudery (Ed.), *Proceedings of the NAES Conference May 2004.* Retrieved from www.hum.au.dk/engelsk/naes2004/papers.html

Shaw, P., & Liu, E. (1998). What develops in the development of academic writing? *Applied Linguistics 19*, 225–254.

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

—(2004). *Research genres.* Cambridge: Cambridge University Press.

Troy, M. (2004). May Sinclair's *The Creators*: High-cultural celebrity and a failed comedy. *English Literature in Transition 1880–1920, 47*, 50–75.

Tse, P., & Hyland, K. (2006). Gender and discipline: Exploring metadiscourse variation in academic book reviews. In K. Hyland & M. Bondi (Eds), *Academic discourse across disciplines* (pp. 103–122). Bern: Peter Lang.

Tucker, P. (2004). Evaluation and interpretation in art-historical discourse. In G. Del Lungo Camiciotti & E. Tognini-Bonelli (Eds), *Academic discourse: New insights into evaluation* (pp. 161–179). Bern: Peter Lang.

University of York. (no date). *Arch-Help undergraduate handbook sec 3.1.* Retrieved from www.york.ac.uk/depts/arch/ugrad/bos/essays.htm (accessed 13 May 2009).

Wolfe-Quintero, K., Inagaki, S., & Kim, H-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity.* Honolulu: University of Hawaii.

Yang, H.-H. (1998). *Three worlds and two discourses: Social dimensions and discourse functions of grammatical subjects in students' English literary essay writing.* Unpublished PhD thesis, University of Newcastle, Newcastle upon Tyne, UK.

Chapter 12

Variation in the Writing of Economics Students in Britain and Pakistan: The Case of Conjunctive Ties

S. Amina Gardezi and Hilary Nesi

1   Introduction

Conjunctive ties are part of the metadiscoursal repertoire, and serve as an explicit means by which a writer can comment on a text and influence a reader's interpretation of it. A number of studies (e.g. Crewe, 1990; Green, Christopher & Mei, 2000; Milton & Tsang, 1993) have suggested that learners of English as a Foreign Language and English for Academic Purposes overuse these ties, particularly in sentence initial position. Various reasons have been proposed for this: it has been argued that some learners have only limited understanding of logical relationships, perhaps as a consequence of teaching that focuses on syntax rather than semantics (Bacha & Hanania, 1980: Zamel, 1983), that the overuse of conjunctive ties is an attempt to give a superficial appearance of academic style (Crewe, 1990), or that it is due to interference from the mother tongue, particularly with regard to fronting (Green et al., 2000). Such studies all assume a deficit on the part of the learner, and a need to approximate more closely to the academic writing conventions manifested by native speakers of English.

Contrastive rhetorical research suggests that metadiscoursal variations may sometimes be a matter of cultural choice, however, rather than level of language proficiency. Writers are influenced by local conventions, unconsciously acquired through intertextual processes, and divergence from British or American academic norms may thus be explained in terms of discourse community differences (e.g. Ahmad, 1997), socio-historic and socio-political factors (e.g. Salager-Meyer et al., 2003), and/or national intellectual styles (e.g. Clyne, 1987; Duszak, 1997). Rhetorical choices may also vary according to professional and disciplinary context (e.g. Dahl, 2004; Hyland, 2004, 2005; Mauranen, 1993; Shaw, this volume; Yakhontova, 2006).

The interaction between disciplinary and national factors has been examined by Mauranen (1993), Dahl (2004) and Yakhontova (2006). Mauranen studied

the work of a Finnish and an American economist writing in English, and found that the Finn used far less organizational metatext than the American. According to Mauranen, the Finnish school system 'consistently teaches that metadiscourse is not only superfluous, but the sign of a poor writer' (1993: 8), and she argues that the implicit Finnish style may be viewed by Finns as a negative politeness strategy, allowing readers the freedom to interpret meaning for themselves, without being patronized. Americans, on the other hand, seem to be aiming for positive politeness by explicitly providing readers with the information and guidance they need in order to interpret the text in the way the writer intends. As Dahl (2004: 1821) points out, 'within the Anglo-Saxon world . . . emphasis is put on communication with the reader, making this an explicit feature of the writing process'.

Dahl (2004) and Yakhontova (2006) examined the effect of two variables, language and discipline, on rhetorical choice. Yakhontova's comparative study of conference abstracts revealed a range of differences in structure, organization and pronominal use between applied mathematics and applied linguistics, and between English and two Slavic languages (Ukrainian and Russian). She concludes that such differences can best be explained in terms of inherited academic writing traditions within 'relatively closed national academic writing communities' (2006: 164). Dahl (2004) focused on metatextual variation, examining locational and rhetorical devices used in research articles across three disciplines and three languages (English, French and Norwegian). She found little national difference in the use of these devices in medical articles, but a significant difference in economics and linguistics texts (French writers used much less metatext than English and Norwegian writers). Medicine, chosen as representative of the natural sciences, had a low overall incidence of metatext; the smaller number of rhetorical markers indicated that writers tended to refer to their research rather than to the text (using 'research act' rather than 'text act' verbs, as classified by Bunton, 1999), while the smaller number of locational markers reflected the fact that medical research articles conform to a standard format, whereas economics and linguistics articles can be structured in a variety of ways and therefore require more navigational support. Dahl concludes that there is a more reader-oriented and writer-responsible academic writing culture in English and Norwegian, but she also notes that economics and linguistics research papers are more heterogeneous and require greater subjective interpretation than research papers in medicine, hence encouraging greater use of metatextual devices.

Hyland's studies of metatext in graduate student writing in Hong Kong (Hyland, 2004, 2005; Hyland & Tse, 2004) support Dahl's findings. When the distribution of metadiscourse markers was compared across six disciplines, totals were higher in applied linguistics, public administration and business studies than in computer science, electronic engineering and biology. This distribution was reflected across all of Hyland's ten metadiscourse categories,

including 'transitions', the commonest of his interactive markers, and the category which contains most conjunctive elements as defined by Halliday and Hasan (1976). According to Hyland (2005: 57–58) 'the figures reflect the greater role that explicit personal interpretation plays in the humanities and social sciences, where interpretations are typically more explicit and the conditions for establishing proof less reliable than in the hard fields'.

Such studies suggest that neither a more implicit nor a more explicit style is inherently superior, and that the appropriate amount of metatext depends on the readers' and writers' disciplinary, academic and cultural context. Mauranen (1993) and Duszak (1997) both advise writers seeking publication to take note of the dominant cultural style, however: 'texts with traces of alien patterns are dispreferred – sometimes returned for repair, or edited with a possible loss of the author's intentions' (Duszak, 1997: 21).

The problems surrounding learner writers' use of metatext are thus more complex than some writers in the field of English language teaching have implied. Learner writers may lack the language knowledge to realize metatextual functions, or may blindly apply textbook rules regardless of the message they want to convey, but prior studies designed to isolate national and disciplinary differences make it clear that they may also make apparently unconventional metatextual choices simply because they are conforming to an alien cultural norm, or to a disciplinary convention that their writing tutors are unaware of.

So far research has focused on rhetorical differences across disciplines and national languages, and there has been little exploration of the independent influence of local academic cultures. This study contributes to the debate by examining the essays of undergraduate students from Britain and Pakistan, two groups who share the same L1, and who study in the same broad field, but who belong to different local discourse communities. We presume that undergraduate students will have been influenced primarily by local rhetorical norms, whereas the expert writers and graduate students considered in most prior studies are likely to have had exposure to a wider community of international scholarship. This therefore makes undergraduates particularly appropriate writers for a study of this kind, which seeks to investigate the influence of the local discourse community in isolation from other factors.

Metatext is acknowledged to be a concept that is difficult to define and delimit (Dahl, 2004; Hyland, 2005; Mauranen, 1993) but conjunctive relations are probably the least problematic metatextual markers, and the easiest to identify. This, combined with the fact that their overuse is a noted feature in EAP and EFL student writing (Crewe, 1990; Green et al., 2000; Milton & Tsang, 1993), encouraged us to focus on conjunctive elements in this study. Our experience with both Pakistani and British student writers led us to hypothesize that the Pakistani group would use far more conjunctive ties in their writing than their UK counterparts. If this was the case, it would indicate

the need for awareness raising, especially on the part of writing tutors and Pakistani undergraduates planning to study at postgraduate level and perhaps publish outside their local setting.

2   Method

A small corpus of 56,142 words[1] was created, consisting of ten assignments written by British students (20,313 words), and ten assignments written by Pakistani students (35,829 words). The concordancing program *AntConc* (Anthony, 2006), was used to identify and compare the various types of conjunctive adjuncts occurring in the two subcorpora, and examine them individually, using the techniques of discourse analysis. The complementary use of the two types of analysis enabled us both to quantify the distribution of the formal lexico-grammatical feature across the 20 assignments, and consider the meaning and purpose of each type of conjunctive adjunct in context, within each text.

The British subcorpus, a subset of the BAWE corpus,[2] was composed of assignments written by ten different native speaker undergraduates studying social science subjects (economics, sociology and politics) at British universities. All the British writers had received all their secondary education in Britain, and all the assignments had received merit or distinction grades (equivalent to an upper second or first class honours degree).

The Pakistani subcorpus consisted of assignments written by 10 different undergraduates from the departments of Economics at Lahore University of Management Sciences (LUMS) and Kinnaird College (KC). Both these institutions have a good reputation in Pakistan and internationally, and all the Pakistani assignments had been awarded distinction or merit grades by LUMS or KC tutors (A+, A or B+). All the Pakistani writers had been educated in the medium of English throughout their school and college life and used English both inside and outside the classroom. They regarded English as a mother tongue (alongside Urdu, in most cases).

None of the assignments in either subcorpus had been written under time constraints. The students had prepared their work at home, over a period of days, if not weeks, with ample time to revise and correct any careless errors. The Pakistani and British contributors were well matched in terms of age (18–20 years), formal English-medium education (approximately 14 years) and disciplinary area. However Tables 12.1 and 12.2 indicate that although the topics were similar, some of the Pakistani essays were considerably longer. Most of the British essay questions were more detailed than the Pakistani ones, with more references to specific economic models, but this should not be taken to imply that the Pakistani students wrote on more general topics. Both groups prepared for the writing tasks in class, and had been given reading lists and specific oral guidance.

Table 12.1 Components of the British subcorpus

No	Discipline	Title	Words
1	Economics	The Heckscher-Ohlin model and its relevance to policymakers. Describe the Heckscher-Ohlin model and explain the Heckscher-Ohlin Theorem. Provide a critique of the assumptions of the model. Is the Heckscher-Ohlin Theorem robust to the underlying assumptions? Explain and illustrate important points by using diagrams. Based on this critique, analyse the relevance of the model for policymakers.	1921
2	Economics	The extent to which a monopoly induces economic inefficiency depends, among other things, upon factors such as: (i) vertical integration in the market, (ii) horizontal market contestability, (iii) technology, (iv) the role of advertising and (v) market demand elasticity. Discuss.	2194
3	Sociology	Taylorism was a new form of relationship between employer and employee, introduced because of the development of the large capitalist corporation. Discuss.	1487
4	Economics	What relationship between unemployment and the real wage is predicted by the Shapiro-Stiglitz shirking model? Explain how an increase in the unemployment rate could reduce firms' expenditure on monitoring workers' behaviour.	2158
5	Economics	To what extent should policy be used to stabilize the economy? Discuss with examples from the UK.	2192
6	Economics	To what extent should policy be used to stabilize the economy? Discuss with examples from the UK	2188
7	Economics	Was the collapse of the Bretton Wood System inevitable? And what were its long run consequences.	3311
8	Sociology	What is 'Racism' and how do you account for its persistence in modern societies?	1549
9	Politics	Evaluate Marx's conception of the capitalist social order as antagonistic and crisis prone.	1556
10	Politics	Assess the centrality of the concept of 'embeddedness' to Polanyi's 'The Great Transformation' and its significance for IPE scholarship	1757
Total			**20313**
Average word length			**2031**

Sentences in the British subcorpus were somewhat longer on average than those in the Pakistani subcorpus (about 25.5 words per sentence as compared to 20). Sentence length could only be calculated approximately, however, because of some vagaries in punctuation.

The study focused on conjunctive adjuncts in the four major conjunctive categories, as described in Halliday and Hasan (1976): additive, adversative, causal and temporal. Various other terms are also used in the literature to describe

Table 12.2 Components of the Pakistani subcorpus

No	Discipline	Title	Words
1	Economics	The underdevelopment of development thought. Development betrayed: paradigms and paradoxes. An equation gone wrong: difference between theory and practice	4389
2	Economics	The underdevelopment of development thought. The discourse of development, the practice of dominance. At the crossroads; alternatives?	4215
3	Economics	The free market is the panacea that will extricate the Third World from the vicious cycle of underdevelopment.	2588
4	Economics	The origins, performance and future of Islamic economics	2205
5	Economics	Dimensions of human resource development	2797
6	Economics	Economic development of Pakistan.	6145
7	Economics	The process of economic development	4839
8	Economics	Discuss the concepts of external debt, debt burden and debt sustainability in a developing country context	3407
9	Economics	Is inflation a fiscal phenomenon in Pakistan?	3035
10	Economics	What are some of the causes of recent banking crises and how do they connect with the setting of exchange rates and other aspects of macroeconomics performance?	2209
Total			**35829**
Average word length			**3582**

such items, including 'logical connectives' (Crewe, 1990), 'logical connectors' (Milton & Tsang, 1993) and 'linking adverbials' (Biber et al., 1999), but there is consensus that their role is to help the reader interpret links between ideas internal to the discourse, thus excluding from these categories any markers of addition, comparison or consequence relating to the outside world. Following Halliday and Hasan (1976), in this study conjunctions marking logical connections between clauses or phrases were not considered, and only intersentential links were counted.

3   Findings

All ten of the British assignments contained examples of at least three categories of conjunctive tie (see Table 12.3) and all the Pakistani assignments contained examples of all four categories of conjunctive tie (see Table 12.4). Marked differences are apparent in terms of the quantity of conjunctive elements.

Table 12.3   Findings for the British subcorpus

Essay no.	Additive	Adversative	Causal	Temporal	Total
1	12	5	3	3	23
2	2	10	5	1	18
3	–	2	2	2	6
4	3	6	5	1	15
5	5	10	1	2	18
6	7	11	11	2	31
7	7	10	3	1	21
8	–	5	1	1	7
9	3	2	7	–	12
10	–	1	3	1	5
Total	**39**	**62**	**41**	**14**	**156**

Table 12.4   Findings for the Pakistani subcorpus

Essay no.	Additive	Adversative	Causal	Temporal	Total
1	15	19	16	8	58
2	9	12	8	1	30
3	13	10	6	1	30
4	7	7	4	2	20
5	7	7	6	1	21
6	17	18	9	2	46
7	14	14	2	3	33
8	15	3	17	7	42
9	2	6	20	6	34
10	8	8	23	4	43
Total	**107**	**104**	**111**	**35**	**357**

Differences in distribution between the two subcorpora were calculated using Rayson's Log-likelihood Calculator (Rayson, undated). This compares the frequencies of a given linguistic feature against the total number of words in two different-sized corpora, using Dunning's G^2 ratio as the statistical measure (Rayson & Garside, 2000). The higher the G^2 value, the more significant is the difference between the two frequency scores. A G^2 of 3.8 or higher is significant at the level of $p < 0.05$, and a G^2 of 6.6 or higher is significant at $p < 0.01$. Results are listed in Table 12.5. In this and subsequent tables, the relative frequency figure indicates occurrences per 100 words.[3]

Table 12.5 shows that the frequencies for adversatives and temporals in the two subcorpora were similar, but that there were significant differences in the distribution of additives and causals ($p < 0.05$). The British writers used adversatives considerably more than other types of conjunctive tie, whereas the Pakistani writers used adversatives, causals and additives to a similar degree.

Table 12.5   Relative frequency and log-likelihood scores

Category	British (20313 words)	Relative frequency	Pakistani (35829 words)	Relative frequency	G^2 value
Adversative	62	0.31	103	0.29	0.14
Causal	41	0.20	111	0.31	5.84
Additive	39	0.19	107	0.30	5.94
Temporal	14	0.07	35	0.10	1.27
TOTAL	**156**	**0.77**	**357**	**1.00**	**7.60**

Table 12.6   The ten most frequently identified conjunctive ties

	Conjunctive ties	British	Relative frequency	Pakistani	Relative frequency	Total	G^2 value
1	however	53	0.26	44	0.12	97	−13.65
2	therefore	22	0.11	25	0.07	47	−2.22
3	hence	7	0.03	34	0.09	41	+7.30
4	and	0	0.00	32	0.09	32	+28.74
5	but	3	0.01	28	0.08	31	+11.54
6	thus	5	0.02	25	0.07	30	+5.59
7	for example	13	0.06	9	0.03	22	−4.75
8	furthermore	6	0.03	12	0.03	18	0.06
9	first(ly)	8	0.04	8	0.02	16	−1.27
10	consequently	2	0.01	12	0.03	14	+3.36

As expected, the Pakistani writers were found to use a significantly greater number of conjunctive ties overall ($p < 0.01$).

The most frequently used conjunctive ties and their comparative frequency in the two subcorpora are listed in Table 12.6. In the G^2 value column, + indicates over-use in the Pakistani subcorpus relative to the British subcorpus, and − indicates under-use in the Pakistani subcorpus relative to the British subcorpus.

The use of these conjunctive elements is illustrated and discussed below. British examples are marked 'B', and Pakistani examples 'P'. This letter code is followed by the essay number and, in the case of the British examples, the BAWE corpus assignment code.

4   The Use of *However* and *But*

However was the most frequent intersentential link in both subcorpora, as it was in Shaw's corpora of literary studies writing (this volume). In the BAWE corpus as a whole it has a relative frequency of 0.16 per 100 words, so there was

highly significant over-use in the British subcorpus with respect to BAWE ($G^2 =$ 10.17, $p < 0.01$) as well as with respect to the Pakistani subcorpus.

Sentence-initial *but* seems to serve a similar function to *however*, but marks the contrastive relation a little less strongly (see Example 1). The relative frequency of sentence-initial *but* was the same in the British subcorpus as in the BAWE corpus as a whole, but was significantly higher in the Pakistani subcorpus.

Example 1

It was just a plain biological theory designed to explain relation between mankind and its history. *But* ironically, development has become a justifiable victim of the evolutionary principle of 'the survival of the fittest'. *Source: P 2*

Bell (2007) claims that sentence initial *but* is frequent in published academic writing, and Shaw (this volume) found that professional literary critics made greater use of it than British undergraduates. Shaw suggests that the low use by British students is probably the result of prescriptivist schooling which discourages the use of coordinating conjunctions as adverbials. Perhaps the Pakistani students had not been given the same advice as the British students, or perhaps, like the professional literary critics represented in Shaw's corpora of published articles, they had chosen to ignore it.

Although Pakistani writers used *but* to connect sentences, none used *however* to connect clauses or phrases within sentences. There were three cases in the British subcorpus of *however* being used in this way, and Thompson (1997: 204) notes an increasing tendency to convert conjunctive adjuncts such as *however* and *therefore* into conjunctions marking intrasentential relations. Such cases were discounted from this study.

Of the instances of *however* as an intersentential link, 88 per cent in the British subcorpus and 93 per cent in the Pakistani subcorpus were sentence initial, whereas only 60 per cent are sentence-initial in the BAWE corpus as a whole. Shaw (this volume) suggests that 'a tendency towards more initial placement is a marker of less skill', and there do seem to be instances in the two subcorpora where it might have been more rhetorically effective to delay the placement of the adversative marker. Delayed placement allows for initial hedging, and enables other thematic elements to take up initial position, thus providing more scope for manipulation of the thematic structure of the text according to the perspective the writer wishes to take. This can be seen in Example 2, where the (British) writer places the first *however* in sentence-initial position, but draws attention to the conditional relation between the two clauses in the final sentence by placing the second *however* after *if*.

Example 2

This causes economic inefficiency in short term. *However*, in the longer run, when the patent has expired economic efficiency will resume as all firms

take advantage of the new technology, competing with the original firm. If, *however*, there were no patents, this long term gain in efficiency may not have taken place. *Source: B2, 0118a*

When instances of *however* and *but* are counted together, their frequency in the British and Pakistani subcorpora are not significantly different ($G^2 = 3.10$). This suggests that the two sets of essays have a similarly argumentative function, and the fact that adversative relations are signalled more often in these two subcorpora than in the BAWE corpus as a whole is likely to be due to the effect of genre, and possibly discipline.

5   The Use of *Consequently, Hence, Thus* and *Therefore*

Causal conjunctive ties tended to occur with greater frequency in the British and Pakistani subcorpora than they do in the BAWE corpus as a whole. Similarly Shaw (this volume) notes a much higher frequency of *thus, hence* and *therefore* in his maths, physics and engineering corpus than in his literary corpora; perhaps this reflects a particular requirement in these and related fields to express logical inferences and results, for example, in mathematical calculations, as in Example 3.

Example 3
If r* = n, both B* and y grow at the same rate. *Hence* b* (debt to output ratio) stays unchanged over time. *Source: P 8*

Although *consequently, hence, thus* and *therefore* seem to be interchangeable in many contexts, and are often defined similarly in dictionaries, their distributions vary. The intersentential markers *hence* and *consequently* are comparatively rare in the BAWE corpus as a whole (0.01 per 100 words) and *thus* is only slightly more frequent (0.02 instances per 100 words), whereas *therefore* is relatively common (0.10 instances per 100 words). The British subcorpus more or less mirrored this pattern of distribution, but in the Pakistani subcorpus *hence* and *consequently* were more common, and *therefore* and *thus* were used with equal frequency (0.07 per 100 words). In some contexts *thus* seemed to have a wider scope than *therefore*, signalling a conclusion gradually arrived at rather than simply a logical consequence of the preceding proposition. Shaw (this volume) suggests that *thus* may indicate 'a less direct connection and a more summative function'; Halliday and Hasan (1976) also acknowledge this difference by classifying *thus* as an additive and *therefore* as a causal marker of conjunction.

In Example 4 *thus* is used as an intersentential link and *therefore* is used as an intrasentential link. By using two different markers to express similar relations within the same sentence the writer avoids repetition, but *thus* also seems to

function as the marker of a conclusion to an extended argument.

Example 4
It is *thus* the specific set of socio-legal structures belonging to modern industrial capitalism which facilitate, legitimate, incentivize and *therefore* encourage an instrumental rationality geared towards material gain. *Source B 10 (0075e)*

In both subcorpora all instances of intersentential *consequently* and *hence* and almost all instances of *thus* were sentence initial. However, whereas the Pakistani writers also followed this pattern of use for *therefore*, 50 per cent of the placements of intersentential *therefore* in the British subcorpus were delayed. As with *however*, delayed placement made possible the thematization of other sentence elements.

6   The Use of *And, Furthermore, For Example* and *Firstly*

In the British subcorpus greater use was made of additives signalling an exemplifying role for subsequent information: *for instance* and particularly *for example*. The Pakistani writers, on the other hand, made significantly greater use of markers signalling a sequence of propositions of equal status, such as *furthermore* and particularly sentence-initial *and,* which was not used at all in the British subcorpus and is comparatively rare in the BAWE corpus as a whole (0.01 per 100 words). British writers probably avoid sentence-initial *and* for the same reasons that they avoid sentence-initial *but,* because teachers in British schools have traditionally advised against it.

In both subcorpora temporals were the least favoured conjunctive ties. They were sometimes used to enumerate the stages of an argument, as in Example 6, and additionally four of the introductions in the Pakistani subcorpus employed sequences of temporal markers to map out essay plans, as in Example 7.

Example 6
There are *three reasons* for this. *Firstly* the culture of colonialism and the belief that the 'Other' is inferior is still apparent in our society today. *Secondly* these beliefs are reinforced through the economy, as capitalism exploits ethnic minorities, and ensures that they are viewed as less skilled, and *finally* racism is embedded in our political system through the policies of political parties and the links they make between non-white immigrants and social problems. *Source B 1 (0111a)*

Example 7
First I will define external debt followed by a set of indicators to assess a country's debt burden. *Then* I will examine what it means for a country's debt to

be sustainable. *Lastly* I will present a simple framework depicting the dynamics between debt management, inflation and fiscal adjustment. *Source: P8*

Some of the British writers also provided introductory plans, but they employed fewer conjunctive ties, and signalled not only sequential but also causal clause relations (see Example 8). According to Martin (1992: 185–193) the causal ('consequential') as opposed to temporal signalling of clause relations enables connections between events to be 'modulated', so that 'one event is seen as *enabling* or *determining* the other rather than simply preceding it'.

Example 8

By firstly investigating the interlinked web of contributive causes of the Bretton Woods breakdown, before discussing the various ways in which events post 1971 have been shaped by the consequences of the preceding era, this essay aims to address both the inevitability of the breakdown and the long term consequences of such a development. *Source: B7 (0399c)*

7   Conclusion

The aim of this study was to ascertain whether use of metatext, and specifically conjunctive adjuncts, was affected by the local academic culture. We found a shared preference for causal and adversative as opposed to temporal conjunctive adjuncts in both subcorpora, reflecting both the relative sophistication of university-level writing, and the requirements of the argumentative essay which focuses on contributing factors and conflicting views (e.g. 'What are some of the causes of recent banking crises?' (*P 10*); 'Assess the centrality of the concept' (*B 10*)).

Pakistani writers tended to produce shorter sentences, with more frequent sentence-initial ties and greater use of ties which mark a sequence of equally important propositions. British writers were more likely to delay placement of conjunctive adjuncts, and to make more sparing use of those which simply announced the addition of new information. British sentences, being longer, also contained more clauses linked by coordinating conjunctions. Typical differences can be seen in Example 9, where all three ties are sentence initial, and Example 10, where the placement of *therefore* in sentence 4 is delayed and some propositions in sentences 3 and 4 are intrasententially linked by the conjunctive use of *hence* and *and*.

Example 9

(1). Statistics show that 18 per cent of the bank deposits were denominated in dollars. (2). *And* this made bank liabilities sensitive to changes in the exchange rate. (3). *Hence* when there was devaluation in the peso the debt

burden of the banks increased. (4). The private sector also found it difficult to service their debts. (5). *As a result* bad loans amplified liabilities and losses for banks. *Source: P10*

Example 10
(1). *However* we now turn to the situation in which trade can occur between the two countries. (2). *First*, recall that in autarky, p h is greater than p f that is (p x/p y) is greater in H. (3). If trade can occur, consumers in country H will observe that good X is relatively cheaper in country F, *and hence* make some of their purchases of X by importing from F. (4). Producers in F *therefore* respond by increasing supply, *and* this clearly causes p h to fall. *Source: B1 (0111a)*

Both sets of essays met departmental requirements, displayed similar levels of scholarship and demonstrated positively polite consideration of the reader by explicitly marking clause relations. Marked differences in style probably reflect local practice, and/or the prescriptions of local teachers. The Pakistani writers seem to have been encouraged to use sentence-initial conjunctive adjuncts to link each new proposition. The British writers may have had more exposure to expert argumentative discourse, introducing them to alternative initial thematic elements. At the same time they seemed to have been proscribed the use of *and* and *but* as conjunctive adjuncts linking sentences, although at least some of them did not seem to have been taught the complementary rule which proscribes the use of *however* as a coordinating conjunction.

Neither of these sets of writing is presented as a model to which learner writers should ultimately aspire. University students are still undergoing apprenticeship into the discourse of their field, and the argumentative essays that they are required to write belong in any case to a different genre from those they might produce in the future as professional writers. A comparison of the two subcorpora does have pedagogical implications, however. It illustrates the signalling choices available to writers of argumentative text, so that learners can increase their own repertoire, and it draws attention to differences in the choices made by writers in different local discourse communities, so that learners can adjust this repertoire to conform to a different set of expectations, if they wish. Conformity to the local style may be the safest choice during undergraduate studies in the home country, but as Duszak (1997) points out, students' success can be adversely affected if they persist with the local style when they move on to study or publish outside their local discourse community.

Acknowledgement

Notes

1 Word counts exclude titles, tables of contents, formulae, tables, figures, footnotes and bibliographies.
2 The British Academic Written English corpus, see acknowledgement above.
3 For comparison with Shaw's data (this volume) the occurrences of conjunctive adjuncts per 100,000 words were 768 for the British subcorpus, and 996 for the Pakistani subcorpus.

References

Ahmad, U.K. (1997). Research article introductions in Malay: Rhetoric in an emerging discourse community. In A. Duszak (Ed.), *Culture and styles of academic discourse* (pp. 273–303). Berlin: Mouton de Gruyter.

Anthony, L. (2006). *AntConc.* [Computer software]. Retrieved from *www.antlab.sci. waseda.ac.jp* (accessed 13 May 2009).

Bacha, N.S., & Hanania, E.A.S. (1980). Difficulty in learning and effectiveness of teaching transitional words: A study on Arabic speaking university students. *TESOL Quarterly, 14*, 251–254.

Bell, D. (2007). Sentence-initial *AND* and *BUT* in academic writing. *Pragmatics, 17*, 183–201.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow: Longman.

Bunton, D. (1999). The use of higher level metatext in PhD theses. *English for Specific Purposes, 18*, S41–S56.

Clyne, M. (1987). Cultural differences in the organisation of academic texts. *Journal of Pragmatics, 11*, 211–247.

Crewe, W.J. (1990). The illogic of logical connectives. *ELT Journal, 44*, 316–325.

Dahl, T. (2004). Textual metadiscourse in research articles: A marker of national culture or of academic discipline? *Journal of Pragmatics, 39*, 1807–1825.

Duszak, A. (1997). Cross-cultural academic communication: A discourse-community view. In A. Duszak, (Ed.), *Culture and styles of academic discourse* (pp. 11–39). Berlin: Mouton de Gruyter.

Green, C.F., Christopher, E.R., & Mei, J.L.K. (2000). The incidence and effects on coherence of marked themes in interlanguage texts: A corpus-based enquiry. *English for Specific Purposes, 19*, 99–113.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English.* Oxford: Oxford University Press.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

—(2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing, 13*, 133–151.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*, 156–177.

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

Mauranen, A. (1993). Contrastive rhetoric: Metatext in Finnish-English economics texts. *English for Specific Purposes, 12*, 3–22.

Milton, J., & Tsang, E.S. (1993). A corpus-based study of logical connectors in EFL students' writing: Directions for future research. In R. Pemberton & E.S. Tsang (Eds), *Studies in lexis* (pp. 215–246). Hong Kong: HKUST Language Centre.

Rayson, P. (undated). *Log-likelihood calculator*. Retrieved from *ucrel.lancs.ac.uk/ llwizard.html* (accessed 13 May 2009).

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. *Proceedings of the Workshop on Comparing Corpora*. Annual Meeting of the ACL, Volume 9 (pp. 1–6). Morristown, NJ: Association for Computational Linguistics.

Salager-Meyer, F., Ariza, A., & Zambrano, N. (2003). The scimitar, the dagger and the glove: Intercultural differences in the rhetoric of criticism in Spanish, French and English medical discourse (1930–1995). *English for Specific Purposes, 22*, 223–247.

Thompson, G. (1997). *Introducing functional grammar*. London: Arnold.

Yakhontova, T. (2006). Cultural and disciplinary variation in academic discourse: The issue of influencing factors. *Journal of English for Academic Purposes, 5*, 153–167.

Zamel, V. (1983). Teaching those missing links in writing. *ELT Journal, 37*, 22–29.

Chapter 13

Can I Use Headings in My Essay? Section Headings, Macrostructures and Genre Families in the BAWE Corpus of Student Writing

Sheena Gardner and Jasper Holmes

1   Introduction

Working in a university writing centre or a university EAP programme can be daunting when students appear for help in the hope that the tutor will have some idea about writing in their disciplines. Well-stocked centres will have local assignments on file from across disciplines, but many tutors find themselves relying on disciplinary norms they are familiar with, or contacting subject tutors for guidance. While some departments provide clear instructions in handbooks, in others there is greater variety, and asking three different subject tutors may yield three different answers. Descriptions of writing across many disciplines, based on actual student assignments, are virtually non-existent. Our investigation of genres of assessed university student writing aims to make a contribution to this area. This project (ESRC 000-23-0800, 2004–2007) includes the development of a corpus of 2,761 successful (i.e. awarded good marks) student assignments from across four disciplinary groups and years of study, which is now available to mine for descriptions of university student writing.

  While the task of detailed genre analysis is a long-term ambition, we have begun by classifying all the assignment texts into genre families and noting this information in the assignment file headers. As part of our investigation of student assignment disciplinary contexts, we used student reports, course documentation, student interviews (Gardner & Powell, 2006) and tutor interviews (Nesi & Gardner, 2006) to catalogue assignments across the disciplines. The various different assignment text types were grouped into 13 genre families, or groups of assignment text types with similar purpose and staging (Gardner & Nesi, 2008), to allow the comparison of assignment texts across disciplines. Genres such as **product evaluation, policy critique** and **book review** may be specific to different disciplines, but by grouping them together in genre

families that share the same functional and structural properties (these are all *critiques*) we are able to carry out cross-disciplinary comparisons.

Building on earlier studies of macrostructure in graduate theses (e.g. Paltridge, 2002; P. Thompson, 1999), we recognize that macrostructure can be identified through chapter or section headings, and have developed a classification of undergraduate assignment macrostructure (Gardner & Holmes, 2006, forthcoming). Our interest in section headings follows from their metadiscourse role as interactive resources (G. Thompson, 2001) and frame markers (Hyland & Tse, 2004). Understanding of metadiscourse has developed through earlier detailed studies of several texts, and corpus studies of lexico-grammatical items. With section headings XML tagged in our corpus, we are able to investigate their role in a large corpus of student writing (over 6.5 million words). By extracting headings for each assignment we can see their skeleton structure at a glance. This gives us a novel perspective on the assignments. These skeletons can be sorted by genre family and discipline to give us a good overview of the assignment macrostructures in the whole corpus and where headings are used in student writing.

In this chapter we focus on the section headings and assignment macrostructures they shape in the 13 genre families of the BAWE (British Academic Written English) corpus. Our aim is threefold. First, we aim to address the corpus–discourse interface theme in terms of the assignment macrostructure–genre interface, where macrostructures consist of section headings, and genres are social processes (Martin, 1992) identified through their educational purpose and generic stages. Second, we aim to explore the extent to which macrostructure can be used to expedite genre identification in our large corpus of student writing. Here we examine the match between specific assignment macrostructures and the 13 genre families identified across disciplines. Such information is potentially of value to those teaching and learning writing at university, which reflects our third, pedagogical aim of answering questions such as 'which good student essays use section headings?'

2   Genre and Macrostructure

Issues identified in the tension between discourse analysis and corpus linguistics are echoed in the different concerns of genre analysis and macrostructure analysis. Where genre analysis seeks to group texts with shared communicative purposes and to analyse them into stages which may overlap in their realization, the analysis into macrostructures tends to be 'somewhat formalistic' (Starfield & Ravelli, 2006), in that it tends to assume that chapter or section headings give accurate indications of the communicative purposes of the sections they govern, and that texts are thereby divided into parts which reflect their structure. We acknowledge that headings may sometimes mislead – a heading

such as *Conclusion* may not always be followed by a concluding section. Equally, the same headings may govern sections with variations in moves or even purpose across disciplines. Thus further more detailed, more functional analysis is required for a full description of each genre. Nevertheless, we find the analysis of macrostructure to be of value in itself as a means of describing university student writing. It not only shows us where section headings are included and where omitted and how texts are partitioned, but also provides a context for more detailed analysis of specific sections (e.g. conclusions) across years of study, genre families and disciplines. Its formal nature enables ready description of a large corpus of student assignments in a novel way.

3   Assignment Macrostructures

Earlier studies of macrostructure in student writing have examined graduate theses and dissertations (Dudley-Evans, 1999; Ridley, 2000; P. Thompson, 1999) and classified these primarily according to chapter headings. Building on such studies, Paltridge (2002) identifies three main classes of graduate thesis: the traditional IMRD type, topic-based, and compilations of research articles.

In our work on text partitioning, we developed a comparable classification of undergraduate assignment macrostructures based on the structural complexity of the assignment and the functional nature of the section headings, as shown in Table 13.1. This more elaborate classification is described in detail in Gardner and Holmes (forthcoming). Basically, however, assignments with a simple macrostructure have one main text part; assignments with a complex macrostructure have one main text part divided into sections; while assignments with a compound macrostructure have parts which themselves are texts.

Central to our approach is the distinction between assignments, which are submitted by students as one piece of work to be graded accordingly, and texts, which typically (i.e. for assignments with simple and complex macrostructures) correspond to assignments less front and back matter such as the name of the student, tutor, university and module, the date, plagiarism declarations, word counts and end notes. Thus because each of the 93 compound assignments in the corpus are realized through two or more texts, we have fewer assignments (2761) in the corpus than texts (2897).

In our classification of the functions of section headings we were influenced by the systemic functional metafunctions (Halliday & Matthiessen, 2004). We identified headings which foreground the textual organization of their section and the organization of the text as a whole (e.g. *Introduction*, *Methods*, *Conclusion*), headings which foreground the ideational-experiential content of their section (e.g. *Martyrs to the nation*, *Secular religiosity?*), and headings

Table 13.1   A classification of assignment macrostructures

Types	Structure	Examples
1 Simple	**FM ^ Text [1 section] ^ (BM)**	Philosophy essay, Sociology ethnography
2 Complex	**FM ^ Text [section a ^ section b (^ . . . section n)]^ (BM)**	
2A Genre based	Complex with generic or 'textual' headings	Lab report, SWOT analysis*
2B Topic based	Complex with specific or 'ideational' headings	Long history essay, Annotated bibliography
2C Context based	Complex with contextual or 'interpersonal' headings	Exercise, Seminar notes
2D Mixed	Complex with mixed headings	Biology essay, Engineering report
3 Compound	**(FM) ^ Text 1 ^Text 2 (^ . . . Text N) ^ (BM)**	
3A Colony	Parallel texts	Compilation of lab reports, or of essays
3B Portfolio	Complementary texts	Essay and review, Literature survey and proposal
3C Mixed	Parallel and complementary texts	Compilation of case notes with one reflection

Note: FM = Front Matter; BM = Back Matter; ^ = followed by; () = optional; [ ] = realized as
*SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis headings are found in agriculture, business, engineering, hospitality management and publishing in our corpus.

which foreground the interpersonal nature of their section as responses to tutor questions (e.g. *Question 1*). Section headings may also foreground logical connections between sections (e.g. *Part 1*, *3.2*). The scare quotes on 'textual', 'ideational' and 'interpersonal' indicate these terms are used as shorthand for 'foreground textual meaning' etc. All headings have all types of meaning, but in specific headings one or more will be salient or foregrounded.

Corpus–discourse interface issues arise not only in the relationship between macrostructures and genres, but also when functional categories such as inter-active resources are investigated through lexically-based corpus searches.

4   Manual and Automated Analysis of Frame Markers and Section Headings

Where both corpus linguistics and systemic functional linguistics are concerned with naturally occurring language and with language as text, they tend to differ in their focus on frequency in formal contexts *versus* meaning in social context (G. Thompson & Hunston, 2006: 4–5). Where corpus analysis tends to work well on formal lexico-grammatical features with large amounts

of text, functional analysis favours more meaningful, contextual or socially embedded features of individual texts. Often corpus studies are inspired by discourse analytical studies, and it is usual to find shunting back and forth (e.g. Miller, 2006) between automated and manual analyses, as understandings are mutually enhanced.

Hyland's work on metadiscourse is a case in point. He develops a category of frame markers, which are interactive (following G. Thompson, 2001) in that they 'help to guide the reader through the text' and framing in that they 'refer to discourse acts, sequences, or text stages' (Hyland, 2004: 139). This category is inspired by examination of individual texts, and then explored using automated corpus tools, where it relies heavily on formal cues for discourse functions. This makes his methodology explicit, and therefore replicable, in ways which traditional discourse analysis may not be. He examines frame markers by searching for 74 specific items such as the sequencing *finally*, the stage-labelling *in conclusion*, the goal-announcing *in this chapter* and the topic-shifting *digress* (Hyland, 2005: 219–220).

Such clearly circumscribed corpus analysis allows for descriptions of large amounts of data, with minimal markup, but issues of coverage can arise – have all (substantial) formal types been identified? One feature that seems to have slipped through the corpus-informed approach to metadiscourse in academic writing (Hyland, 2005) is section headings, though the description of frame markers as referring to 'text boundaries or elements of schematic text structure', which have four functions: 'to sequence, to label text stages, to announce discourse goals, and to indicate topic shifts' (Hyland & Tse, 2004: 186) could have been written about section headings. Because section headings can correspond in principle to any linguistic item, it is not possible to simply search for them all using basic concordancing tools in plain text. The automated identification of section headings in large amounts of text requires preparing the corpus prior to analysis so that such features can be extracted.

In our project the XML tagging language was used to identify the beginning and end of each section heading, as well as its level and any font modifications. Section headings were identifiable by a combination of font (e.g. bold), layout (e.g. indent), and numbering (e.g. *3.2*). This allows us to extract information about the prevalence of assignment texts with section headings across categories identified in the assignment files, such as year of study, discipline and genre family. It also allows us to extract the headings themselves and examine them in isolation for the meanings they convey and the different types of heading that occur in different disciplines and genres at different levels of embedding. In what follows we examine the nature of section headings in different years of study, genre family, discipline and level of embedding, with the aim of exploring the extent to which the formal properties of macrostructure can be used to expedite genre analysis. We conclude with an examination of the use of section headings in essays, which provides an answer to the question posed in our title.

5   Section Headings, Year of Study and Word Count

Our growing familiarity with the student assignments indicated that while some assignments were required to have section headings, in others there was more room for student choice. We hypothesized that the longer an assignment, the more likely it was to have section headings. Word length data is readily available for all assignment texts; as is level of study. Basically year 1 refers to first year coursework written by first year students, year 2 to second year, year 3 to third year, and year four to taught masters. Exceptions are detailed in Alsop and Nesi (2009). As expected, the average length of assignment texts (excluding front and back matter, formulae, tables, footnotes, references, appendices etc.) increases steadily with year of study from 1788 words for Year 1 to 2324 for Year 2, 2637 for Year 3, and 2903 for Year 4.

As Figure 13.1 shows, the proportion of assignments with section headings also increases from first to fourth year in three of the four disciplinary groups.

The greatest differences are between first year Arts and Humanities (AH), where just 6 per cent of assignments have section headings, and all year 4 and Physical Sciences (PS) assignments at more than 70 per cent. The four disciplinary groups are ranked from AH with relatively few section headings, through SS (Social Sciences) where the proportion with section headings increases most, to LS (Life Sciences) which is consistently in the sixties for
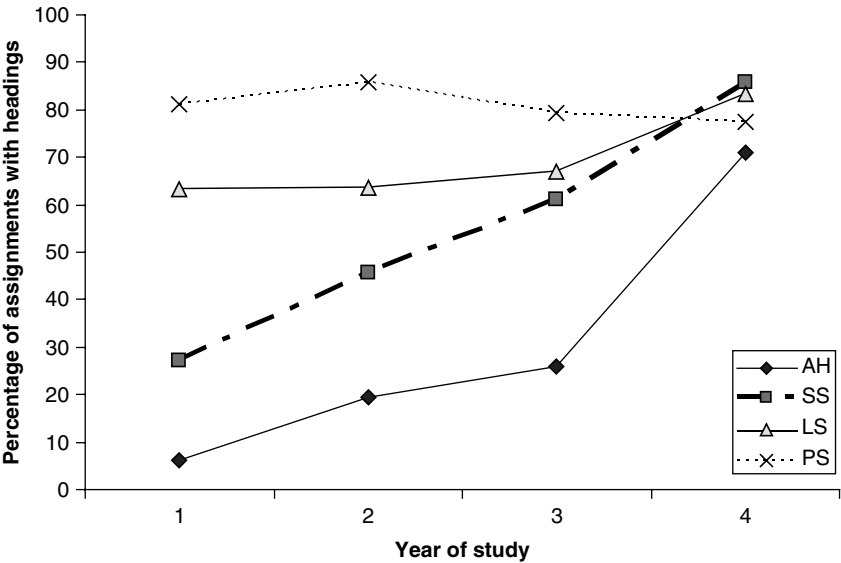
FIGURE 13.1   Prevalence of assignments with headings by year and group

Years 1–3, and PS where the proportion is highest overall at around 80 per cent In all disciplinary groups and years over 10 per cent of assignments have no section headings. We assume that an examination of different assignment macrostructures will help explain these differences.

6 Frequent Section Headings

Having established the spread of headings across our corpus, it is useful to find out more about the headings themselves. This was done by extracting the 18,493 section headings into *Wordsmith Tools 4.0* (Scott, 2004) and creating a list to search for frequent words and strings of words (n-grams). As Table 13.2 shows, *Introduction* and *Conclusion* are the most frequent content words in headings, occurring in around two thirds of the 1,579 assignments with headings. These are followed by *Results* and *Analysis* which occur in approximately one-third of assignments with headers and, with *Methods*, and *Discussion,* are associated with experimental report macrostructures found particularly in the science disciplines. This must largely explain the differences between disciplinary groups seen in Figure 13.1.

Table 13.2 Ten most frequent content words in all section headings

Word	Freq	Word	Freq	Word	Freq
Introduction	1,020	*Results*	569	*History*	472
Conclusion(s)	951	*Analysis/es*	527	*Problem*	328
		Method(/s/ology)	473	*Management*	319
		Discussion(s)	396	*Issues*	305

Perhaps surprising are the remaining four items, *History, Problem, Management* and *Issues.* Further investigation shows that these are all used in medical case studies.

Extract 1: Partial macrostructure of Medical case history (0065e)

Case Summary
Referral Information
History
    Presenting Complaint:
    History Presenting Complaint:
    Past Medical History:
    Drug History
    Social History:
Analysis of history and examination
    X-ray

Formulation of the patient's problem(s)
   Ideas
   Concerns
   Expectations
Management
3.1
3.2
. . .
Outcome
Evidence based care and issues for research
Commentary
   References
Impact on your learning

As shown in Extract 1, the word *history* occurs six times in headings in this one assignment. However, as *history, problem, management* and *issues* also occur in non-medical assignments, we cannot assume that medical portfolio case histories are the next most frequent assignment type after experimental reports. In fact these section headings have such salience in the corpus because they are written using a proforma, so that all 66 medical case histories from 12 different modules have (almost) all of the headings shown above.

Using word frequency counts we have investigated the most frequent words in section headings and thus far arrived at a sense of three typical macrostructures: one with *Introduction* and *Conclusion*; one with an **experimental report** (IMRD) macrostructure; and one that of **medical case histories**. This evidence strengthens the link between section headings, macrostructures and specific genres.

Table 13.3 shows the 50 most frequent 2- and 3-grams in section headings. The most frequent (*of the, in the,* etc.) represent common syntactic patterns that are also frequent outside section headings as evidenced with reference to the whole corpus.

More specific to section headings are those involving content words already identified as key in this context (*analysis, discussion, results*), often expanded (*materials and methods, observations and results*) or in combination with each other (*analysis and discussion of results*). There are more nominalized processes (e.g. *evaluation of, description of, determination of*) which provide useful indicators of the functions of the sections they govern. These also explain the overall frequency of *of* (second most frequent word [after *the*] occurring 3757 times in the headings), *of the* and *of a*, suggesting that '(the) x of (the/a) y' is a typical syntactic structure for section headings. This examination of n-grams strengthens the links between section headings, macrostructures and stages of genres in their indication of the functions of different sections.

Table 13.3   Fifty most frequent 2- and 3-grams in section headings

N-gram	N[1]	Frequency in h'ings[2]	Frequency in corpus[3]	N-gram	N	Frequency in h'ings
of the	630	0.79	0.96	a new	32	0.04
in the	181	0.23	0.46	and results	32	0.04
on the	106	0.13	0.17	to be	32	0.04
to the	97	0.12	0.33	discussion of results	31	0.04
of a	95	0.12	0.12	observations and	31	0.04
and the	89	0.11	0.21	the project	31	0.04
analysis of	83	0.10		the role	31	0.04
as a	78	0.10	0.12	apparatus and	30	0.04
for the	78	0.10	0.15	development of	29	0.04
what is	65	0.08		impact of	29	0.04
and discussion	63	0.08		part b	29	0.04
of results	58	0.07		the role of	29	0.04
and methods	50	0.06		analysis and discussion	28	0.04
results and	50	0.06		effect of	27	0.03
is the	48	0.06	0.09	part a	27	0.03
evaluation of	45	0.06		in a	26	0.03
role of	44	0.06		materials and methods	26	0.03
discussion of	43	0.05		observations and results	26	0.03
use of	43	0.05		structure of	26	0.03
analysis and	41	0.05		the UK	26	0.03
from the	40	0.05		literature review	25	0.03
description of	39	0.05		what are	25	0.03
materials and	35	0.04		control of	24	0.03
with the	35	0.04		discussion and	24	0.03
determination of	33	0.04		into the	24	0.03

Note: 1 = number (N); 2 = frequencies per 100 words in headings; 3 = frequencies per 100 words for the top 50 2- and 3-grams in the whole corpus.

7   Section Headings and Genre Families

Before we turn to essays, we shall consider the other genre families in turn to develop a sense of where macrostructure points clearly to genre and where it does not. The prevalence of section headings in the different genre families is presented in two tables. Table 13.4 shows categories where more than 75 per cent of the texts have headings. For example, 174 of the 191 case studies have headings; this amounts to 91 per cent of all case study texts.

Most of the **research reports** and **methodology recounts** have the experimental report format with first level section headings of the Introduction Method Results Discussion (IMRD) type. The difference between the two genre families resides in their purpose: whether the student has developed the design themselves in the context of the literature and presented it in a format similar to published research (**research paper**, one type of **research report**),

Table 13.4 Proportion of genre family texts with section headings (Part 1)

	Research report	Methodology recount	Design specification	Case study	Problem question	Proposal	Exercise
N with headings	60	314	88	174	31	66	90
Total N	61	359	93	191	40	76	114
Total (%)	**98**	**88**	**95**	**91**	**78**	**87**	**79**

or whether they have been told what to do and the assignment is basically a write-up of an experiment conducted (**lab report**, one type of **methodological recount**). The differences were made explicit in tutor interviews, assignment rubrics and other contextual information. In the texts they are evidenced through assignment titles and most obviously through the word length and nature of *Introduction, Theory, Discussion* and *Conclusion* sections. Thus in our data IMRD type section headings point to a number of different genres, rather than one specific genre.

There is also disciplinary variation in IMRD headings, with Engineering and Computer Science generally including a *Theory* section, while Food Sciences generally includes a *Calculation* section. Typical macrostructures for six disciplines are presented in Table 13.5.

Table 13.5 IMRD variations across disciplines

Biological Science	Computer Science	Engineering	Food Sciences	Physics	Psychology
(Abstract) 32/52	(Abstract) 16/64	(Abstract) 44/83		(Abstract) 15/18	(Abstract) 5/10
Introduction	1. Introduction 2. Theory	Introduction Theory	Objective Introduction	1. Introduction	Introduction
Materials and Method	3. Design	Apparatus and Methods	Method	2. Experimental Details	Method
Results	4. Implementation	Observations and Results	Results	3. Results	Results
Discussion	5. Results and Analysis	Analysis of Results	Calculation	4. Discussion	Discussion
(Conclusion) (Future Work)	6. Conclusion	Discussion Conclusion	Discussion		
(References) 22/52	(References) 29/64	(References) 63/83	(References) 53/69	(References) 15/18	(References) 8/10

Note: () = optional.

Further disciplinary variation is seen in the prevalence of optional sections. For instance, 15 of 18 Physics texts (83%) have *Abstracts*, compared with only 16 of 64 (25%) Computer Science texts.

Many assignments have a second level of textual headings under *Method*. For example, Psychology assignments have *Participants*, *Materials*, *Procedure* and *Data Analysis*. Some disciplines have second level ideational headings, such as *Banana*, *Ketchup* and *Peanuts* in Food Sciences. Third and fourth level headings occur in all the sciences in Table 13.5.

Design specifications, common in Engineering and Computing, are also easily recognizable from their section headings: they include a design brief and design details under headings such as *Objectives*, *User Requirements*, *System Specifications*, *Design Details*, *Performance and Cost Estimates*, *Implementation* and *Details*.

There are two main types of **case study** in our corpus, medical and business, both of which also occur outside their eponymous disciplines. They involve analysing aspects of a case and making recommendations, as reflected in Extract 1 above and Extract 2:

Extract 2: Business case study headings (0253h)

Executive Summary
Introduction
    1.1 Current situation of Glass Product Division
    2. Evaluation of existing systems and practices
        2.1. Control system
            2.1.1 Planning and budgeting procedures
                2.1.1.1 Sales Forecasting
                2.1.1.2 Plant Manufacturing Budget
                2.1.1.3 Comparison of Actual and Standard Performance
                2.1.1.4 Planning procedures in general
            2.1.2 Structure of accountability
                2.1.2.1 Sales Unit as a Revenue Centre
                2.1.2.2 Manufacturing Unit as a Profit Centre
            2.1.3 Reward Structures
        2.2 Customer focused approach
    3. Recommendations for change
        3.1 Planning and budgeting procedures
        3.2 Kaizen Costing
        3.3 Structure of Accountability
            3.3.1 Sales unit as a profit centre
            3.3.2 Plant Manufacturing unit as a pseudo-profit centre
        3.4 Balanced Scorecard Framework
    4. Conclusion
Appendix 1: Sales Budgeting Procedures
References

From their macrostructure, case studies may be similar to **problem questions**. The differences reside in whether the case or scenario is real or fabricated, and whether it is given to the students or not. Problem questions occur across Law, Business and Engineering, as in Extracts 3 and 4.

Extracts 3 and 4: Business and Law problem questions (0169b, 0196a)

3 Business	4 Law
PROBLEM INDENTIFICATION	Introduction
ANALYSIS	The law on bias
ALTERNATIVES	Application to the case
RECOMMENDATIONS	Conclusion
PLAN OF ACTION	

Proposal section headings can be less uniform than the genres outlined so far. Some resemble research reports with (Extract 5) or without (Extract 6) a heading to indicate they are proposals rather than completed studies:

Extracts 5 and 6: Computer Science and Hospitality, Leisure, and Tourism Management (HLTM) proposals (6169f, 3018a; Level 1 headings only)

5 Computer Science	6 HLTM
Abstract	Introduction
Aim	Service Style
Background	Menu
Critical review of relevant literature	Wine List
Objectives	Control Methods
Research Methods	Equipment
Project Plan	Conclusion
References	Bibliography

To differentiate these from research reports we have to look inside the sections for statements of intent.

Exercises tend to have interpersonal headings such as *Question 1*, *Question 2* and *Question 3*. They may also have ideational or textual headings as in Extract 7 from Archaeology.

Extract 7: Archaeology exercise (6157c)

Exercise 1: Descriptive statistics (L1)
         Hypotheses
Exercise 2: Normality of data (L1)
         Hypotheses

Exercise 3: ANOVA (L1, L2)
      Hypotheses
Exercise 4: Chi-squared test (L1, L2)
      Hypotheses
Exercise 5: Correlation analysis (L1, L2)
      Hypotheses
Exercise 6: Regression (L2)
      Hypotheses

Table 13.6 shows those genre families that have a smaller proportion of texts with section headings, but it is noteworthy that headings are found across all families.

Table 13.6   Proportion of genre family texts with section headings (Part 2)

	Literature survey	Explanation	Critique	Empathy writing	Narrative recount	Essay
N with headings	25	134	187	17	30	363
Total N	35	214	319	35	72	1183
Total (%)	**71**	**63**	**59**	**49**	**42**	**31**

Four of the **literature surveys** have headings that are bibliographic details of books and journals, but most literature surveys cannot be recognized as such from their section headings. The same is true of explanations.

Extracts 8 and 9: Explanations in Physics and Meteorology (6129b, 1629a)

8 Physics	**9 Meteorology**
Abstract	Solar radiation
How meniscus forms	Effects of the atmosphere
Molecular forces	Effect of cloud cover
Conclusion	Effect of latitude
References	Effect of Land and sea
	Appendix

Some, like Extract 8 from Physics, invoke a question (*how does meniscus form?*) which an explanation could answer, but most, like Extract 9, are lists of ideational headings which could be the headings of a literature survey, an explanation, an essay or a critique.

Indeed, headings in **critiques** seldom suggest their genre, but occasionally there are terms such as *strengths, weaknesses* or *critical review:*

Extracts 10 and 11: Critiques in Engineering and Business (0021c, 0169f)

10 Engineering	**11 Business**
Abstract	1. Critical review of SSM Mode 1
Introduction	2. Roles, norms and values in SSM Analysis 2
Strengths of the PRA technique	3. Politics, power and SSM Analysis 3
Weaknesses of the PRA technique	4. Conclusion
PRA in practice	Bibliography
Conclusions	
References	

Empathy writing headings are recognizable from their register as non-academic writing, *I've always been a bit on the plump side, but how can I tell if I'm really overweight or just a bit chubby?* (Food Science, 6023b) and *Dear Mr. Beswick,* (Publishing, 3089d), while many **narrative recounts** are accounts of group work on projects, and headings such as those in Extract 12 can indicate this:

Extracts 12 and 13: Narrative recounts in Health and Medicine (3034e, 0065g)

12 Health	**13 Medicine**
What happened in the 'Forming' phase, and how I felt about it?	Cairo
What happened in the 'Storming' phase, and how I felt about it?	The Kasr El Aini Teaching Hospital
What happened in the 'Norming' stage and how I felt about it?	The atrocity of the 7th of April, 2005
What happened in the 'performing' stage and how I felt about it?	Conclusion
If the situation arose again what would you do?	
Conclusion	
Reference List	

Other recounts may also be suggested by headings, though, as in Extract 13, these headings could apply to most other genres in Table 13.6.

Table 13.7 From section heading to genre

Headings		Genres
IMRD headings	→	research papers and lab reports
literature review	→	research report or proposal
executive summary	→	business case study
Multiple history headings	→	medical case study
ideational headings + bibliography	→	essay, explanation, critique
non-academic register in headings	→	empathy writing
first person references (*I, me*)	→	narrative recount
interpersonal headings (specific list)	→	exercise

This overview of section headings in genre families suggests that although identification of assignment macrostructure is not sufficient for genre identification, some section headings and macrostructure elements point to specific groups of genres. This is summed up in Table 13.7.

9   Section Headings in Essays

Finally we discuss **essays**. There are essays with headings across all disciplinary groups and most disciplines, though the numbers in Life and Physical Sciences are small; the smaller the essay pool, the less reliable are any inferences drawn from the data presented here. Overall, 31 per cent of the 1,183 essays in our corpus have section headings.

While essays account for only 65 (11%) PS texts, Table 13.8 shows 58 per cent of these have headings. In contrast, essays account for 554 (82%) AH texts, of which only 14 per cent have headings. This suggests that where students are used to writing assignments with headings, such as reports and case studies, this will carry over into their essay writing. Indeed multiple levels of headings are found in essays (e.g. *1.2.1 Homology Groups* [Mathematics] and *2.1.1 Negation Effects on Vegetation* [Biology]), though numbering to three levels is rare across all genres (occurring in only 42 texts) and numbering to four or five levels is extremely rare.

10   Functions of Section Headings in Essays

Essays tend to have headings which organize the ideational meaning or content (Field) of the essay, but tell us very little if anything about the rhetorical organization or the genre (discussion, exposition, challenge etc.). In addition to the ideational headings, many essays have an *Introduction* and *Conclusion*. Thus the typical macrostructure of essays with headings is (Introduction)^Ideational Heading 1-n^(Conclusion) (Bibliography/References) as illustrated in Extracts 13–16. Extract 13 has both *Introduction* and *Conclusion*; 14 has no *Introduction*; 15 has no *Conclusion*; and 16 has neither.

Table 13.8 Prevalence of essays with headings by discipline

Disciplines	Essays	Essays with headings	
	N	N	(%)
Archaeology	49	23	47
Linguistics/Applied Linguistics	75	33	44
Classics	78	12	15
Philosophy	98	15	15
English	89	6	7
Comparative American Studies	71	2	3
History	94	3	3
Total AH	**554**	**94**	**14**
Agriculture	27	18	67
Medicine	10	6	60
Biological Sciences	11	6	55
Health	15	7	47
Food Sciences	7	2	29
Psychology	57	2	4
Total LS	**127**	**41**	**32**
Cybernetics & Electronics	2	2	100
Physics	12	10	83
Architecture	4	3	75
Mathematics	4	3	75
Planning	12	8	67
Computer Science	9	4	44
Engineering	16	7	44
Chemistry	6	1	17
Meteorology	0	NA	NA
Total PS	**65**	**38**	**58**
Publishing	4	4	100
Law	85	58	68
Business	49	33	67
Politics	97	36	37
Anthropology	27	8	30
Economics	55	16	29
Hospitality, Leisure and Tourism Management	29	8	28
Sociology	91	18	20
Total SS	**437**	**181**	**41**
TOTAL	1183	363	31

Extracts 13 and 14: Comparative American Studies and History essays (0003k, 0012d)

13 Comparative American Studies	**14 History**
Introduction	i- Environmental Differences
Argentina: 'a land of exiles'	ii- Technology

Brazil: 'All hail! This samba's going to end in jail'	iii- Attitudes
Mexico: 'A dark Indian, grateful to the party'	Conclusion
Conclusion	Bibliography
Bibliography	

Extracts 15 and 16: Philosophy and Cybernetics essays (0057b, 6101c)

15 Philosophy	**16 Cybernetics**
Introduction	Open Source Software
Quine's case against a complete 'theory of knowledge'	Open Source Hardware
Naturalized epistemology	Patenting
Is naturalized epistemology a subject matter of philosophy?	Feedback
Bibliography	

While most essays have ideational headings at level 1, there are a few in Mathematics, Business and Applied Linguistics which do not:

Extracts 17 and 18: Mathematics and Business Essays (0049a, 0072a)

17 Maths	**18 Business**
Section 1	I. Introduction
Section 2	II. Main part
Section 3	A. The Efficient Market Hypothesis (EMH)
Section 4	B. Questioning the EMH – the concept of "Noise Trader Risk"
Section 5	C. Exploiting market inefficiency – setting up a portfolio strategy
Section 6	III. Conclusion
Section 7	Directory List of Sources
Section 8	1. Books
Section 9	2. Journals and newspapers
	3. Internet sources

Sections 1–9 of Extract 17 have level two headings (not shown here) of *Proof* and *Hypothesis* which foreground textual information but also have little ideational content. This is unusual. More often when the first level headings are not ideational, there are headings which foreground ideational meaning at the second or third level, as in Extract 18.

If we examine References and Appendices, we see that most (over two-thirds) essays have a Bibliography or Reference Section, while very few have Appendices.

The term *Bibliography* (737, 36%) is used across all disciplines; *References* (241, 12%) is also used across disciplines, with the exception of English and History which use *Primary* and *Secondary Sources/Texts*, or *Works Cited*. Other general terms used include *Books*, *Articles*, *Essays*, *Journals*, *Websites*, *Internet Sources*, *Electronic Resources*, *E-books*, *E-journals*, *Lecture Notes*, *Newspapers*, *Magazines*, and the more specialized *Cases*, *Treaties*, *Legislation* and *Filmography*.

This examination of section heading and macrostructure in essays in the BAWE corpus provides evidence about the prevalence of section headings in university student writing, and differences across disciplines. Although this does not allow us to distinguish essays from critiques or explanations, or to identify types of essay, it does enable us to group potential essays. We could then analyse their introductions and conclusions further, which should provide strong evidence for genre identification. Following such analysis, we could return to the macrostructure and look for evidence of whether there is a link between specific essay genres (e.g. discussion) and the use of section headings.

11   Discussion and Conclusions

In our development of the notion of assignment macrostructure, and our unique corpus techniques for extracting these from our corpus, we have been able to display, sort and compare the macrostructure of the large number of texts in our corpus in a meaningful way. Our classification of macrostructures (Table 13.1) developed earlier (Gardner & Holmes, forthcoming) has been enhanced in its ability to account for all the texts in the BAWE corpus. It differs importantly from those of Paltridge (2002) and others who studied small numbers of graduate dissertations and theses manually, not only in its additional categories, but also in its claim to be descriptively adequate.

With the notable exception of the small scale studies on thesis macrostructure, section headings have been largely ignored in research, whether using manual or automated techniques. Yet section headings function as interactive resources (G. Thompson, 2001) in very similar ways to frame markers (Hyland, 2005), with their explicit partitioning of text; their marking of sequences, of levels of embedding of ideas, and groupings within text; and their functions of predicting the content of sections, making links with earlier sections, and providing the macrostructure of the assignment as a whole.

In Hyland's study, frame markers in dissertations showed little variation across six disciplines (2.3 per 1,000 words in Biology to 3.5 in Computer Science) (2004: 146), especially when compared with other metalinguistic categories; in contrast our comparison of section headings in assignments by discipline and year (Figure 13.1) suggests greater differentiation in undergraduate student writing. A similar upwards trend to that found here for year of assignment is observed in Hyland's data for frame markers (from 2.1 per

1,000 words in Masters to 3.0 in Doctoral dissertations written by Hong Kong EFL students) (Hyland, 2004: 140, 2005: 55). Future research could usefully examine how the presence or absence of section headings interacts with the prevalence of other frame markers across disciplines and years of study.

We suggest that the reason earlier corpus-based studies have not included section headings resides in their diversity of linguistic realization coupled with the difficulties of searching for headings in text that has not been marked up with this aim in mind. We addressed these difficulties in two ways, by examining the wording in section headings, and by examining the headings themselves by year, discipline and genre.

To examine the words and phrases in headings, we extracted them from the corpus, and then, using frequency counts, were able to identify key indicators of prevalent genres (Tables 13.2 and 13.3) and typical syntactic structures in headings. We also learned from this that *Introduction* is more frequent than *Conclusion*, that *Bibliography* is more frequently used than *References*, and how both vary with discipline and genre.

To explore the macrostructure–genre interface we examined the prevalence and functions of section headings across disciplines, years of study and genre families and have seen how section headings can point to specific genres (Table 13.6). Headings which foreground textual meanings can point to genres such as **lab report**. Those which foreground ideational meaning are found throughout the corpus, but when they are central, they point to genres such as **explanation**, **critique** and **essay**; further automated analysis, perhaps of key words in the assignments themselves rather than just the headings, might enable us to identify the genre more precisely. Headings which foreground logical meanings, as well as those which foreground interpersonal meanings, are indicative of **exercises**. We also found specific section headings, such as *Executive Summary*, which are good indicators of their genre (**case study**). We have thus understood better the potential and limitations of using macrostructure to expedite genre analysis.

Our third aim was pedagogical. We exemplified macrostructures across all genre families and disciplinary groups, noting clear disciplinary norms within genres (Table 13.5). We provided information on prevalence of headings across genre families (Tables 13.4 and 13.6) and found that assignments typically have one or two levels of headings, with headings of four and five levels being sufficiently rare to suggest they are not advisable in most genres and disciplines. To answer the question about essays in our title, we have to know which discipline, genre and year of study it refers to. Essays in first and second year English and History seldom have headings; unlike other years (Figure 13.1) and disciplines (Table 13.8). In general, the smaller the proportion of essays written in the discipline, and the longer the essay, the more likely they are to have headings. When essays do have headings, they seldom have *Abstract*, often have *Introduction* and *Conclusion*, usually have one or two levels of

ideational headings, and often have *Bibliography* or *References.* Similar accounts of other genres could usefully be developed in further studies.

Acknowledgements

References

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

Dudley-Evans, T. (1999). The dissertation: A case of neglect? In P. Thompson (Ed.), *Issues in EAP writing research and instruction* (pp. 28–36). Reading: CALS.

Gardner, S., & Holmes, J. (2006, July). Section headings and assignment types. Paper presented at *EISFLCW*, Gorizia, Italy.

—(Forthcoming). From section headings to assignment macrostructure in under-graduate 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).

Gardner, S., & Nesi, H. (2008, June). A new classification of genres of university student writing. Paper presented at *Language Issues in English-medium Universities Conference*, Hong Kong.

Gardner, S., & Powell, L. (2006, June). An investigation of genres of assessed writing in British Higher Education: A Warwick-Reading-Oxford Brookes project. Paper presented at annual *Research, Scholarship and Practice in the Area of Academic Literacies Seminar*, University of Westminster, UK.

Halliday, M.A.K., & Matthiessen, C.M.I.M. (2004). *An introduction to functional grammar.* (3rd edn). London: Hodder & Stoughton Educational.

Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing, 13*, 133–151.

—(2005). *Metadiscourse: Exploring interaction in writing.* London: Continuum.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*, 156–177.

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

Miller, D. (2006). From concordance to text: Appraising 'giving' in Alma Mater donation requests. In G. Thompson & S. Hunston (Eds), *System and corpus: Exploring connections* (pp. 248–268). London: Equinox.

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

Paltridge, B. (2002). Thesis and dissertation writing: An examination of published advice and actual practice. *English for Specific Purposes, 21*, 125–143.

Ridley, D. (2000). The different guises of a PhD thesis. In P. Thompson (Ed.), *Patterns and perspectives: Insights into EAP writing practice* (pp. 61–76). Reading: CALS.

Scott, M. (2004). *WordSmith Tools 4.0.* [Computer software]. Oxford: Oxford University Press.

Starfield, S., & Ravelli, L.J. (2006). 'The writing of this thesis was a process that I could not explore with the positivistic detachment of the classical sociologist': Self and structure in *New Humanities* research theses. *Journal of English for Academic Purposes, 5*, 222–243.

Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics, 22*, 58–78.

Thompson, G., & Hunston, S. (2006). System and corpus: Two traditions with a common ground. In G. Thompson & S. Hunston (Eds), *System and corpus: Exploring connections* (pp. 1–14). London: Equinox.

Thompson, P. (1999). Exploring the contexts of writing: Interviews with PhD supervisors. In P. Thompson (Ed.), *Issues in EAP writing research and instruction* (pp. 37–54). Reading: CALS.

Chapter 14

Using the Revision Process to Help International Students Understand the Linguistic Construction of the Academic Identity

Suganthi John

1   Introduction

Writing for an academic discourse community is a crucial activity for student writers as it is a means to enter the community and construct for oneself a visible identity as a competent member. As Pecorari (2002: 63) points out, writing is 'one of the main means by which individuals achieve and maintain a position within the community'. Achieving and maintaining this position is a 'socially constructed process' (Hindman, 1999: 24) involving a series of complex negotiations between students, teachers and the academic discourse community. Primarily, the writing classroom and/or supervision meeting is the place where these 'complex negotiations' occur. It is the place where writers begin their journeys from being *novices* to becoming *initiates* (Thompson, 2001). The journey itself can be characterized by the multiple drafts that writers write as they create and develop their identities as 'fully-fledged member(s) of the discourse community' (Woodward-Kron, 2004: 141).

This chapter investigates the benefits of using the writing process or revision as a means to help postgraduate writers construct for themselves identities as full members of the discourse community. It is a pedagogical study addressing two issues: (a) what linguistic and textual features in a text construe the academic identity; and (b) how these linguistic and textual features may be used to make the academic identity of the writer more or less visible in the text. The results of the study lead into a discussion of the ways in which the revision process can be used by supervisors to heighten writers' awareness of the impact these linguistic and textual features have of the construction and development of their academic identities.

The chapter begins with a definition and general interpretation of the academic identity in a postgraduate dissertation. It then investigates how identity

may be construed in one of the main sections of a traditional dissertation, the methodology. Finally, the paper suggests how supervisors can use the revision process to help postgraduate writers develop academic identities within their texts, thus illuminating Bartholomae's viewpoint that 'the moment of possession (is) not the opening moment but a later one, *where if the writer is present that presence can be seen in the work of revision*' [italics added] (Bartholomae, 1995: 86).

There are two corpora for this study. The first consists of 17 final MA dissertations in applied linguistics and is used to arrive at the linguistic and textual features which construe identity. The second set consists of eight draft and final MA dissertations in the same field and is used for the revision study and for the discussion on pedagogical implications. The contributors are all ESL writers from the Far East. While the corpora are not large, the approach is influenced by the quantitative focus of corpus studies and qualitative focus of discourse studies. The quantitative analysis is useful in providing concrete evidence of revisions having taken place although there is no attempt to equate the frequency of occurrence of realizations of writer identity to the quality of the writing. The qualitative analysis allows a closer discussion of the influence of revision on the identity of the writer and the quality of the writing. Revisions are identified as changes to the linguistic and/or textual features that have occurred between first draft and final version.

2 The Postgraduate Writer and the Academic Discourse Community: A Question of Belonging and a Question of Visibility

The challenge facing writers entering an academic discourse community is recognizing the modes of expression used in these communities and acknowledging that the kind of writing an individual does in the academy is distinct from all other kinds of writing he/she may be familiar with. This can be a complex issue for the writer and involves distinguishing between academic discourse and the language appropriate to other less formal contexts (Hewings, 2001). Moreover, as Bizzel (1992) argues, when basic writers come to college they are faced with a clash of discourse types. Forms of writing which they were used to whilst still at school are no longer 'the ways of winning arguments in academe' (Bizzel, 1992: 165). This lack of familiarity with academic discourse conventions may lead to a loss of self-confidence, with the result that student writers may 'defer to the voice of the academy' (Sommers, cited in Zamel & Spack, 1998: 188). Indeed Bartholomae (1985: 4–5) goes to the extent of saying that they have to 'invent' the university, '[t]hey must speak our language'.

Learning to write for the academy, then, involves taking on a new identity and, as Hyland (2002) points out, this is particularly problematic for second-language students. One reason for this is that these learners already see themselves as competent writers in their L1 and thus have to undergo an identity transformation when writing in the L2 (Hirvela & Belcher, 2001). Many researchers who investigate the complexities L2 writers face in an L1 writing situation highlight the problem of creating an academic identity in terms of having 'an authoritative voice' or 'being visible' in the text. Thus Hyland (2002) notes that novice writers may not establish their identity effectively if they fail to express their 'voice' through the use of first person pronouns. More generally, Ramanathan and Atkinson (1999) argue that second language learners struggle with notions of voice, individualism and textual ownership when they enter L1 academic environments. This accords with the views of Shen (1989) and Li (1996) (reviewed in Stapleton, 2002: 3), who both comment that L2 writers find the notions of 'creating a new identity' or having 'an individual identity' alien. The issue that repeatedly emerges in research is one which questions whether writers can be taught to express their identities in a text – in other words, can they be taught to be visible as 'fully-fledged' academics in their texts? Even if this can be taught, is it always desirable or appropriate to be visible? This is where attention to language becomes crucial as text and textual choices may be viewed as 'identity potentials' (Abasi et al., 2006: 105). As Costley (2008: 81) points out, 'a central part of academic voice is the control that one has over the text'.[1]

For postgraduate writers, much of the writing is read and re-read by the supervisor during supervision meetings. In this chapter I suggest that this opportunity can be harnessed to help the writer develop control over the text, which would help shape the writer's academic identity by allowing the writer to alter their levels of visibility in the text.

The relationship between supervisor and supervisee has been studied by Dysthe (2002), who suggests that the type of relationship affects the way in which drafts are read by supervisors. In particular, Dysthe draws attention to the fact that feedback on drafts involves listening to and 'revoicing' the voice of the student. However, since her study is not a linguistic one, there are no suggestions for how this might be done through attention to writing. Thompson's (2001) work, on the other hand, is text-based and offers more illuminating linguistic advice on how the writing of an ESL writer can be revised and how this might help student writers enter the academic discourse community. His approach involves one-to-one meetings with students, where drafts are discussed and opportunities presented for students to clarify their views and intentions. This then leads to a joint construction of a revised text. While the proposed method may be difficult to apply at all supervision meetings and for all pieces of writing, purely from the perspective of the time and commitment it requires, it can be used on smaller paragraphs of student writing for

illustrative purposes during supervision meetings. In the final section of this paper, using revision data from the dissertations, I will illustrate how this can be done to help writers develop their academic identities.

3   The Academic Identity: Personality and Visibility

The interpretation of identity I propose is one in which identity is defined by the actions of the writer in the text i.e. who the writer *is* in the text is defined by what the writer *does* in the text (John, 2005, 2007). Defining identity using the actions of the writer in the text enables us to connect this abstract concept with more concrete linguistic and textual features. This then allows for a more pedagogically oriented approach, which can help postgraduate writers gain control over the texts they are creating.

Taking into account the notion that the writer in the text is typically multi-dimensional (Cherry, 1998; Ivanič, 1998), I suggest the following model of writer identity in a postgraduate dissertation (John, 2005). The two main identities are that of Person (referring to autobiographical references of the writer to him/herself) and Academic (referring to all scholarly actions relating to the actions of research and research writing for the academic community). This Academic identity is in itself multi-dimensional, as can be seen from the multiple practices that the Academic carries out in the text. These practices can be grouped into two main types: scholarly and discourse-organizing. The Academic in the text may therefore be viewed as using the roles of 'Scholar' and 'Organiser'. This study reports only on the identity of the writer as Scholar, as this is the dominant identity in methodology sections of dissertations.

The linguistic features used to express identity are termed *visibility* choices. The term *visibility,* based on work by Davies (1988), Gosden (1993), Charles (1999) and Martínez (2001), expresses the notion that different language features offer writers a range of visibility options to express their identities in a text. The most visible of these is the first person pronoun. Identity in academic writing can therefore be linguistically understood along two dimensions: *personality* and *visibility,* where *personality* is realized by instances of the use of the first person pronoun construing the identities of Person and Academic (Scholar and/or Organiser) and *visibility* is realized by other linguistic and textual features in the text. This is illustrated in Figure 14.1.

In the next part of this chapter, typical linguistic or textual features associated with the methodology section of the dissertation are investigated and reported in order to enable a pedagogically useful description of the visibility cline of identity in the text. I focus here on how identity is construed through linguistic and textual features other than the first person pronoun, for which there is already ample research (Harwood, 2005; Hyland, 2001, 2002;

Linguistic Resources
which realize a more or less
visible writer in the text

Less visible   **More visible**

<--->
←←←←Visibility choices→→→→

A range of linguistic features   **First person pronoun**
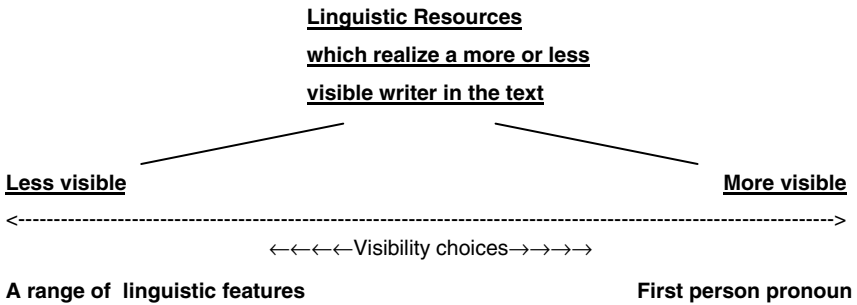
Figure 14.1   Visibility choices of *writer identity*

Ivanič, 1998; Ivanič & Simpson, 1992; Kuo, 1999; Ramanathan & Atkinson, 1999; Tang & John, 1999). Harwood (2005) has an extensive description of first person pronoun use in methodology sections – referred to as the 'methodological *I*'.

4   How the Academic Identity is Construed in Methodology Sections

Two main issues are dealt with in this section. First, what are some of the common textual and linguistic choices available to writers at discourse and clause level when they compose methods sections; and secondly, what do these choices mean for writer visibility?

In methods sections, the writer's main action is to report and justify the chosen methodology and to report and explain the data collection and analysis procedures. These actions construe for the writer the identity of a Scholar. Generally, there are several scholarly activities that take place including:

● descriptions of the data
● explanations of how the data is selected and/or collected
● justifications of why the data is selected and/or collected
● definitions of the methodology/methodologies adopted
● explanations of the methodology/methodologies adopted
● justifications of why the methodology/methodologies are adopted.

4.1   Writer visibility at discourse level

The activities stated above normally occur in a variety of combinations at discourse level. Consider the following example that involves describing the data

(in bold) and justifying the reason for choosing the data (underlined).

Example 1
Subjects for this study were 135 experienced Taiwanese junior high school English teachers who had taught in junior high school for at least five years. These teachers work at different schools in 10 counties. Some of them were friends of mine, and most were colleagues of these friends. <u>A wide spectrum of teachers was covered, in both urban and suburban areas, in order to gain a picture of teacher attitudes and beliefs right across Taiwan, and not just in the main urban centers</u>. (CHL:F:27)

The next example is a combination of describing the chosen methodology (in bold), in this case, questionnaires, and justifying the aim of the questions in the first part of the questionnaire (underlined).

Example 2
The questionnaire was divided into three parts. The first part (Part A) showed the personal data of the respondent. It included sex, major course of study, school level, previous study background and grade point average in English. <u>I need some of these data to examine whether such person categories were of any significance for the student's understanding. For example, I would like to know whether their educational background affects their writing or not. This is because some of them were from vocational schools while some of them were from a non-formal education school which differs much in English teaching methods. Not only educational background, but also sex, major course of study, and grade point average in English may affect their results</u>. (UY:F:15)

These examples suggest that methods sections of these dissertations may be more than instances of 'unevaluated reporting' (Hopkins & Dudley-Evans, 1988: 119) or 'explicit standard academic description' (Salager-Meyer, 1994: 161) as previously described in some research. The justification in the above examples serves to evaluate the data or methodology being used. In Example 1, the writer evaluates her data as having been selected from a **wide spectrum** of teachers. Further, by using the marker **in order to** she justifies the need for this **wide spectrum** by saying that it will give her views from teachers across Taiwan. In Example 2, the writer evaluates her methodology of using particular questions in her questionnaires by giving reasons why she needs personal information from her subjects, using the markers **to examine** and **This is because**. The use of justification is indicative of the writer's sensitivity to the needs of the reader as requiring the reason for the writer's selection of data or proposed methodology – an important practice in any kind of research.

At discourse level, I suggest that there exists a relationship between justification moves and writer visibility since justification moves involve evaluation. According to Thompson and Hunston (2000), one of the main functions of evaluation is the expression of opinion. It can be argued that any instance of the writer's opinion in the text signals their visible presence as the person who holds that opinion. Given that the justification of data or methodology involves the expression of opinion, the use of this move can be held to lead to the writer's visibility in the text.

In the parts of the text where the writer does not evaluate the data and/or methodology, a range of visibility options is still open to the writer. An examination of the data shows that a study of the types of processes used (based on the process types identified in Systemic Functional Grammar, Halliday, 1985), the voice (active or passive) and the agency of the process types (whether the action is carried out by the writer) at clause level reveals insights into the levels of visibility a writer projects. It is beyond the scope of this chapter to attempt a full transitivity analysis of methodology sections but certain aspects of transitivity are selected and adapted to illuminate the visibility of the writer in the text.

4.2 Writer visibility at clause level

Methods sections of dissertations are characterized by the actions of the writer in the scholarly act of reporting about the data and chosen methodology. As stated earlier, these activities typically involve actions of describing, selecting and/or justifying data or methodology. As in this chapter I am interested primarily in the writer's actions and what those actions mean for the visibility of the writer in the text, the focus of the discussion in the next part will be primarily on processes where agency can be attributed to the writer. In functional grammar terms this is where the writer is *Actor* in a material process, *Senser* in a mental process, *Sayer* in a verbal process and *Behaver* in a behavioural process. This also extends to implied agency in passive clauses where the writer is implied *Actor, Senser, Sayer* and *Behaver*. The other two process types, relational and existential, do not lend themselves to the notion of agency. However, later in this section I suggest that selecting these process types over the others also has implications for the visibility of the writer.

Tables 14.1 and 14.2 show the frequency and distribution of the process types in methodology sections from the seventeen final dissertations. The clauses are divided into those that involve (i) data and (ii) methodology. Table 14.1 shows the distribution of process types in clauses that involve data. The passive categories of verbal and behavioural process types are omitted since there are no instances of these in the data. Table 14.2 shows the distribution of process types in clauses that involve methodology.

Table 14.1 Analysis of process types in clauses that involve data

	Mt (A)	Mt (P)	Mn (A)	Mn (P)	V (A)	B (A)	R (Atb)	R (Id)	E
Frequency	157	41	13	1	11	0	22	106	14
+ writer as Actor	56	36	9	1	1	0	–	–	–
% writer as Actor	35.6%	87.8%	69.2%	100%	9%	0%	–	–	–

Key:
Mt (A): Material process active voice
Mt (P): Material process passive voice
Mn (A): Mental process active voice
Mn (P): Mental process passive voice
V (A): Verbal process active voice
B (A): Behavioural process active voice
R (Atb): Relational process (attributive)
R (Id): Relational process (identifying)
E: Existential process

Table 14.2 Analysis of process types in clauses that involve methodology

	Mt (A)	Mt (P)	Mn (A)	Mn (P)	V (A)	B (A)	R (Atb)	R (Id)	E
Frequency	338	195	51	3	21	1	15	148	18
+ writer as Actor	142	174	25	3	7	0	–	–	–
% writer as Actor	42	89.2	49	100	33.3	0	–	–	–

Note: See Table 14.1 for key.

Since the main processes used in methodology sections of dissertations are material, mental and relational, these are the ones that I will investigate in detail in terms of their relationship with writer visibility. Verbal, behavioural and existential processes are either few in number or have no bearing on writer visibility. They are therefore not dealt with here.

Some interesting observations can be made from these tables about the ways in which writers report research activities concerning their data and methodology. Material (active) processes are the highest frequency process type in the data. The reason for this is fairly obvious, since usually these sections report things that have been done (selection, and/or collection of data and selection and use of methodology). From the figures we can see that the percentage of material processes with the writer as Actor are more frequent in sections reporting methodology than data. This is because most methodology sections have longer stretches of text describing the methodology employed than the selection and/or collection of data. Material processes in passive voice where the writer is the implied Actor are the next most frequent process types. However, if we compare the percentages of the writer as Actor in material active processes and the writer as implied Actor in material passive processes, a

striking difference emerges. Writers use more passive material processes than active material processes to signal their involvement in the actions in the text. Since the choice of the passive suggests that the writer does not wish to be seen visibly as the Actor in the clause, this has implications for the writer's visibility in the text. I will return to this point later in the chapter. Mental (active) and mental (passive) processes are generally not very frequent. Again this finding has implications for writer visibility as I will show later.

The next set of processes to be dealt with are relational (attributive) and relational (identifying) processes. These have no agency but it is interesting to note that identifying relational processes are more common than attributive relational processes. This is important since identifying relational processes can tell us about 'the broader concerns and the values of the writer' (Thompson, 2004: 98). So using these processes may show an evaluation on the part of the writer. The fact that relational processes are common and involve no agency is relevant for the following discussion on writer visibility.

In the next part of this section, I suggest that the selection of process type (material, mental or relational), voice (active or passive in material and mental clauses) and agency (marked or implied) affect the visibility of the writer in the text. Consider the following examples:

Example 3
The subjects in this research were students in the second year of the Bachelor's degree course in the English department of the Rajabhat Institute in Bangkok, Thailand. (UY:F:19)

Example 4
The main method of data collection was carried out through audio recorded participant observation. (SY:F:2)

Example 5
I will analyse some elements in the narrative structure of Sherlock Holmes stories taking the chosen story as a case study. (JW:F:1)

These three examples display a progressive increase in the visibility of the writer's academic identity. The relational process used in Example 3 offers the writer the least visibility. In Example 4, the material process in the passive, 'was carried out', offers more visibility because the writer is construed as the implied agent performing the action of 'carrying out'. In Example 5, the active voice option, 'I will analyse', offers the most visibility since the writer can be identified in her role as Scholar explicitly through the use of the first person pronoun (Harwood's (2005) 'methodological *I*').

A further point on visibility can also be made about material and mental clauses, voice and agency. Compare the following examples of material and mental clauses in the active form where the *Senser* or *Actor* of the clause is

marked with the first person pronoun 'I'.

> **Example 6** (mental process: *think*, Writer: *Senser*)
> I do think that in Conan Doyle's detective fictions, the author also pays attention to the description of individual characters. (JW:D:30)

> **Example 7** (material process: *prepared*, Writer: *Actor*)
> I prepared tape-recorders, negotiations simulation cards, etc, carefully before I met them. (ZJH:F:1)

Where subjective interpretation is involved, as in many mental processes (**think** in Example 6), there is greater visibility for the writer than in material processes which involve no interpretation (**prepared** in Example 7). This applies to the passive forms of these processes too, as shown below.

> **Example 8**
> Three factors have been considered in order to justify the reliability of the research: corpus size, topics covered and work and time limit. (MD:F:18)

> **Example 9**
> A questionnaire (see Appendix One) was designed to gather some general information about the subjects. (ZJH:F:3)

Figure 14.2 summarizes our discussion thus far on the visibility of the writer's academic identity at clause level. The lowest visibility is indicated by Relational processes as they have no agency; slightly more visibility is achieved by using passive processes in which agency is implied, with mental processes higher than material processes; active processes give the highest visibility, again with mental processes at the top of the cline.

In summary, the linguistic and textual features which affect the visibility of the writer at discourse and clause level in the text are:

- the justification move
- the use of first person pronouns

Less visible   More visible
←--→
Re <–> Mat (passive)   <–> Ment (passive) <–> Mat (active) <–> Ment (active)

No agency   Agency: Writer   Agency: Writer

Key:

Rel: Relational processes   Mat: Material processes   Ment: Mental processes
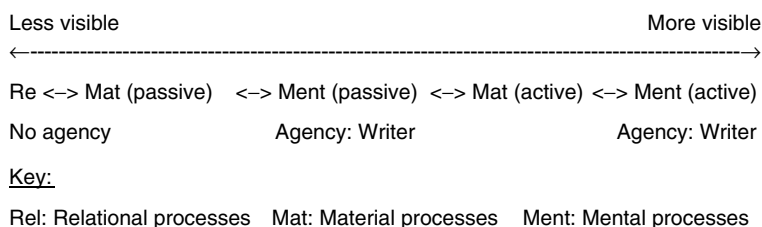
FIGURE 14.2   Linguistic choices and the visibility of the academic identity in methodology sections

- the types of processes
- the indication of agency

Supervisors can help writers understand how to construct a more or less visible academic identity by suggesting revisions that involve these linguistic and textual features. For instance, asking a postgraduate writer to include a 'justification move' when discussing methodology would lead to more visibility for the writer. Likewise, at certain points in a methodology section, it may be more useful for the writer to focus on the data or the choice of methodology and not their visibility as the person who has chosen/collected the methodology/data and in these cases the supervisors may suggest options for the writer to reduce their visibility. It is not always obvious to student writers when it is desirable and appropriate in academic discourse to increase or reduce visibility. Revision presents opportunities to use language in different ways to see what would be the best and most appropriate option for the text and the discourse community. To summarize, postgraduate writers need to be equipped with the linguistic and textual features that will enable them to create and manipulate their academic identities in a text. Some examples are provided in the next section to illustrate this point.

5   The Revision Process as a Pedagogical Tool

I suggest that the revision process can be used by supervisors to scaffold the development of postgraduate writers' academic identities by helping them make informed choices about their visibility in the text. In other words, the issue is not only about how to create a visible identity in the text but also about how to question the visibility in the text (e.g. Would it be an advantage/disadvantage to increase/decrease your visibility here?).

By considering four examples from draft and final versions in detail, I illustrate how revision can be used by the supervisor to shape the academic identity and visibility of the writer through careful attention to language. I provide possible rewrites to show alternative final versions as illustrations of the possibilities that exist if revision is maximally utilized as an opportunity to help students better understand the notion of an academic identity. The rewritten versions contain information that is present in both the draft and the final versions.

5.1   Case 1

The first example illustrates a common challenge faced by ESL writers – how to express disagreement. In this case, the writer has deleted the disagreement altogether in the final version, although arguably it is important to her

methodology and to her visibility as a Scholar.

Example 10

JW (Draft)	JW (Final)
I will point out the different character styles in Sherlock Holmes detective stories and the specific roles in the specific story *The Dancing Man*. Furthermore, I **do not agree** that specific characters in a Sherlock Holmes story is only the 'agent' of a certain character type, carrying a certain functions. **I do think** that in Conan Doyle's detective fictions, the author also pays attention to the description of individual characters. (JW:D:28)	I will point out the different character styles (Actant) in Sherlock Holmes detective stories and the specific roles in the specific story (Acteur) *The Dancing Men*. My analysis is based on Greimas's theory and Propp's 'seven-character style' theory. (JW:F:28)

In the draft, the writer's academic identity is very visible due to the use of mental process verbs in combination with the first person pronoun. However it is not simply visibility in the text that is important for the creation of an academic identity, but also the nature of the academic audience addressed. Explicit or overt disagreement of this nature is not commonly seen in writing for an academic audience (Hunston, 1993).

In this case, the supervisor would need to work with the student to produce a text which would still contain the element of disagreement but expressed in a way that is more appropriate for the discourse community. Attention to language will help the student achieve this, as illustrated in the rewritten examples and discussion which follows. Underlining is used to show the most important parts of the rewrite and bold for individual features that are commented on.

Example 11 (rewrite of Example 10)
I will point out the different character styles (Actant) in Sherlock Holmes detective stories and the specific roles in the specific story (Acteur) *The Dancing Men*. Previously (references), specific characters in Sherlock Holmes stories have been analysed as the 'agent' of a certain character type, carrying a certain function. **I will present an alternative view. I will illustrate** how the author in Conan Doyle's detective fictions pays attention to the development of individual characters. My analysis is based on Greimas's theory and Propp's 'seven-character style' theory.

This alternative version still emphasizes, and indeed heightens, the visibility of the writer's identity as a Scholar through the use of the first person

pronouns, while retaining the element of disagreement with previous research (**I will present an alternative view**). We have, however, seen that the writer has access to a range of linguistic and textual features other than the first person pronoun that affect his or her visibility in the text. Consider the following revision which would also be an option open to the writer.

> **Example 12** (modified version of Example 11)
> I will point out the different character styles (Actant) in Sherlock Holmes detective stories and the specific roles in the specific story (Acteur) *The Dancing Men*. My analysis is based on Greimas's theory and Propp's 'seven-character style' theory. <u>Previously (references), specific characters in Sherlock Holmes stories have been analysed as the 'agent' of a certain character type, carrying a certain function.</u> **This dissertation considers** <u>an alternative view. It is argued that the author in Conan Doyle's detective fictions pays attention to the development of individual characters</u>. My analysis is based on Greimas's theory and Propp's 'seven-character style' theory.

By rewriting the sentences, the proposed revisions show that the writer can retain her visibility of the Scholar. In Example 12, the two sentences that have been revised carry writer agency although in a less visible manner. The third sentence has the noun group **this dissertation** which is associated with the writer and **considers** as a mental process type attributable to the writer. The fourth sentence has the anticipatory *it*-pattern with the mental process **argued** which is once again attributable to the writer. This revision therefore retains the high visibility mental processes but adjusts the levels of visibility of the writer by exercising other linguistic options open to her. In this way, the writer retains the scholarly argument she was attempting to make in the draft.

These examples show how supervisors can scaffold the development of the academic identity of the writer by being aware of the linguistic and textual features that construe identity and suggesting possible rewrites that help the writer retain this academic identity in the text. The next example is a further illustration of this point.

5.2   Case 2

Example 13

JW (Draft)	JW (Final)
The reason I choose Sherlock Holmes stories as my analysis corpus is that first . . . Second, some scholars on	The reason I have chosen a Sherlock Holmes story as my analysis is that first . . .

narrative structure have regarded Sherlock Holmes stories as the analysis corpus, such as Shkovsky and Todorov. It **will be a convenience for me to develop my** research on the basis of the theories of these scholars. (JW:D:1)

Secondly, some scholars of narrative structure, such as Shkovsky and Todorov, have analysed some Sherlock Holmes stories using narrative structure. (JW:F:2)

In this case, the deleted processes are relational (**be**) and material (**to develop**). Agency in this sentence can be attributed to the writer based on the two markers: **me** and **my**. While to some extent the deletion of this sentence and the processes in them have resulted in lower visibility for the writer, the revision was quite possibly due to (a) the grammatically problematic nature of this sentence and (b) the cited reason of '**convenience**' which may not be seen as a sound scholarly reason for employing a particular methodology. However, this final version does not reflect the meaning conveyed in the draft, where the writer states that there are two reasons for her choice of a Sherlock Holmes story. Here we are concerned only with the second reason, which is that, because it has been studied before, she can build on the research of other scholars. In the final version, this second reason appears incomplete, because the writer states that other scholars have used narrative analysis, but does not comment further. This raises the question – why is that important for the current study? An alternative revision (only of the second reason) along the lines of Example 14 could be suggested.

Example 14 (modified version of part of Example 13)
Secondly, some scholars of narrative structure, such as Shkovsky and Todorov, have analysed some Sherlock Holmes stories using narrative structure. This is useful for the methodology **adopted** in **this research** since it can be **developed** along similar lines.

This revision shows how the scholarly explanation the writer was attempting in her draft can be incorporated in her text. It also shows how the writer's visibility can be raised through the noun phrase **this research** which is associated with the writer, and the processes **adopted** and **can be developed** which can be attributed to the writer. Further, this sentence could also be interpreted as a justification of her methodology which, as previously shown, raises the writer's visibility as Scholar. For many of the writers in the data, language proficiency is a problem and quite often during revision problematic sentences are deleted. This, unfortunately, can be at the expense of the scholarly argument being developed and the writer's visibility in the text.

5.3   Case 3

Example 15

ZJH (Draft)	ZJH (Final)
They all have at least bachelor degrees or equivalent (four years full time university education), although majors might not be in business administration. It is really difficulty **to find** so many MBA students who would like to be the subjects. The reasons for this **are obvious**. Firstly, they are busy with their studies and may not be interested in my investigation. Secondly, **I do not know** most of them at all. Thirdly, face-saving is very important to Chinese, they do not want anybody unconcerned to criticise or even know their skills or language. In order to collect my data . . . (ZJH:D:1)	They all have at least bachelor degrees or equivalent (four years full time university education), although their majors might not be in Business Administration. In order to collect my data . . . (ZJH:F:23)

The revision in this case has a considerable number of processes that have been deleted in the final version. I will restrict my discussion to those where agency can be attributed to the writer (in bold): in the second sentence, the process **to find** and in the fifth sentence the mental process **do not know,** which has the writer as *Senser* marked by the first person pronoun. There are also other signals of the writer's presence in this sentence. The use of **my investigation** expressing writer possession and the evaluative adjective **obvious** used after the relational (attributive) process **are**. These have been deleted in the final version reducing the visibility of the writer in the text. Admittedly, some of what is said in this section could be expressed more accurately; however, I see the deleted material as valuable to the dissertation since it concerns problems with collecting data. The rewritten version that follows is a possibility that the supervisor could explore with the writer.

Example 16 (modified version of Example 15)
While there **is** sufficient data for analysis, the process of data collection **was** difficult for three reasons. Firstly, the participants were not always available or interested in the investigation. Secondly, since the participants are volunteers and did not know me personally, they had no obligation to be part of this study. Thirdly, the cultural factor of face-saving **is** important to the Chinese and some of the participants were reluctant to make known their

language weaknesses publicly. Nevertheless, in spite of these difficulties, some steps **were taken** in order to collect sufficient data for the study.

The suggested revision above would still retain the important element of any scholarly discussion of data collection – the challenges involved in getting it. The revision also results in the introduction of relational process types and a passive material process which are attributable to the writer and would allow the writer to retain her academic identity in the text, although not as visibly as in the draft.

5.4　Case 4

The next example illustrates how a successful revision can be used to raise students' awareness of the identities that they project in a text. Draft and final versions can be presented to groups of students, for example, in an academic writing class for international students.

Example 17

SY Draft	**SY Final**
However, where Eggins and Slade (1997) stress on giving generic labels for the stages, this study will label its stages using textual labels. **As I am a novice in the field of genre analysis**, coining my own generic label is felt to be too ambitious for me to undertake. In addition, using textual labels, it would be easier not only for myself but for readers to understand what the stage's communicative purpose is. (SY:D:5)	However, whereas Eggins and Slade (1997) emphasise the importance of giving generic labels to the stages, this study will label its stages using textual labels. **By using** textual labels, it would be easier not only for myself but for readers to understand what the stage's communicative purpose is. **By defining** the communicative purposes of the stages and sub-stages, it is hoped that what may seem like textual is a generic label instead. (SY:F:16)

Most students are able to point out that it is not usual in academic writing to refer to oneself as an inexperienced student researcher, as this would affect their credibility as competent researchers. However, it is true that many of them do feel inexperienced. By showing students how the writer has used justification moves (By using . . . By defining . . .), we can illustrate that the writer can remove the focus on him/herself and place it on the chosen methodology. In the last two sentences of the final version, this writer still shows that her methodology will benefit both herself and her readers.

6 Conclusion

This chapter has shown how writer identity can be made concrete through identifiable linguistic and textual features in methodology sections of dissertations. It has also attempted to raise awareness of the writing process as a means through which writers can alter the visibility levels of their identities by manipulating the language resources available.

Several issues still remain problematic: for example, how do we decide whether having an increased or reduced visibility is appropriate or desirable in a text? A still more challenging question is whether issues concerning identity and its appropriacy and desirability can be taught to students. This chapter has suggested that one way forward is through linking identity to linguistic and textual features. Currently, academic writing courses tend to address issues of identity mostly in passing. Usually, identity is related to issues of academic style and almost always to the use of first person pronouns. However, when students ask, 'Can I use the *I* in writing?', this simple question opens up a highly complex area, involving not only the use of the first person pronoun, but also other linguistic and textual features, such as those discussed here. These aspects of identity are seldom covered in academic writing courses.

This chapter has also presented revision as an opportunity to explore language. The use of first and final drafts and joint revision of text can be effective in helping postgraduate writers to develop an understanding of the concept of academic identity and the corresponding issues of visibility. Being aware of both how language construes identity and how revision can affect the visibility of that identity have important implications for writers as well as supervisors in an institution of higher learning.

The issue of identity in academic writing continues to receive attention in research (see current exchanges between Matsuda and Tardy, 2007, 2008 and Stapleton and Helms-Park 2008). This shows that knowledge about identity and writing remains important for teachers of writing and supervisors as both are concerned with encouraging and scaffolding an identity change in writers from *novice* to *initiate*.

Notes

[1] The term *voice* is frequently used interchangeably with the term *identity*. *Voice* usually refers to the writer's ownership of ideas or the writer's construction of authority in a text.

References

Abasi, A., Akbari, N., & Graves, B. (2006). Discourse appropriation, construction of identities and the complex issue of plagiarism: ESL students writing in graduate school. *Journal of Second Language Writing, 15*, 102–117.

Bartholomae, D. (1985). Inventing the university. In M. Rose (Ed.), *When a writer can't write: Studies in writer's block and other composing-process problems* (pp. 134–165). New York: Guilford Press.

Bartholomae, D. (1995). Responses to Bartholomae and Elbow. *College Composition and Communication, 46*, 84–87.

Bizzel, P. (1992). *Academic discourse and critical consciousness.* Pittsburgh: University of Pittsburgh Press.

Charles, M. (1999). Aspects of the author's voice in three academic research articles. In P. Thompson (Ed.), *Issues in EAP writing research and instruction* (pp. 71–83). Reading: CALS.

Cherry, R.D. (1998). Ethos versus persona: Self-representation in written discourse. *Written Communication, 5*, 251–276.

Costley, T. (2008). 'You are beginning to sound like an academic.' In C.P. Casanave & X. Li (Eds), *Learning the literacy practices of graduate school: Insiders' reflections on academic enculturation* (pp. 74–89). Ann Arbor: University of Michigan Press.

Davies, F. (1988). Reading between the lines: Thematic choice as a device for presenting writer viewpoint in academic discourse. *ESPecialist, 9*(1/2), 173–200.

Dysthe, O. (2002). Professors as mediators of academic text cultures: An interview study with advisors and master's degree students in three disciplines in a Norwegian university. *Written Communication, 19*, 493–544.

Gosden, H. (1993). Discourse functions of marked theme in scientific research articles. *English for Specific Purposes, 11*, 207–224.

Halliday, M.A.K. (1985). *An introduction to functional grammar.* Baltimore: Edward Arnold.

Harwood, N. (2005). 'I hoped to counteract the memory problem, but I made no impact whatsoever': Discussing methods in computer science using *I*. *English for Specific Purposes, 24*, 243–267.

Hewings, M. (2001). Introduction. In M. Hewings (Ed.), *Academic writing in context: Implications and applications. Papers in honour of Tony Dudley-Evans* (pp. 9–16). Birmingham: University of Birmingham Press.

Hindman, J.E. (1999). Inventing academic discourse: Teaching (and learning) marginal poise and fugitive truth. *Journal of Basic Writing, 18*(2), 23–46.

Hirvela, A., & Belcher, D. (2001). Coming back to voice: The multiple voices and identities of mature multilingual writers. *Journal of Second Language Writing, 10*, 83–106.

Hopkins, A., & Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes, 7*, 113–121.

Hunston, S. (1993). Professional conflict: Disagreement in academic discourse. In M. Baker, G. Francis & E. Tognini-Bognelli (Eds), *Text and technology: In honour of John Sinclair* (pp. 115–134). Amsterdam: Benjamins.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes, 20*, 207–226.

—(2002). Options of identity in academic writing. *ELT Journal, 56*, 351–358.

Ivanič, R. (1998). *Writing and identity: The discoursal construction of identity in academic writing.* Amsterdam: Benjamins.

Ivanič, R., & Simpson, J. (1992). Who's who in academic writing? In N. Fairclough (Ed.), *Critical language awareness* (pp. 141–173). London: Longman.

John, S.P. (2005). The writing process and writer identity. Investigating the influence of revision on linguistic and textual features of writer identity in dissertations. Unpublished Ph.D. thesis, University of Birmingham, UK.

—(2007). Meeting the challenge of developing an academic identity: A textual approach. In P. Teo & C. Ho (Eds), *Discourse in the modern world: Perspectives and challenges* (pp. 28–52). Singapore: McGraw-Hill.

Kuo, C.H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes, 18*, 121–138.

Martínez, I.A. (2001). Impersonality in the research article as revealed by analysis of the transitivity system. *English for Specific Purposes, 20*, 227–247.

Matsuda, P.K., & Tardy, C. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes, 26*, 235–249.

—(2008). Continuing the conversation on voice in academic writing. *English for Specific Purposes, 27*, 100–105.

Pecorari, D.E. (2002). Original reproductions: An investigation of the source use of postgraduate second language writers. Unpublished PhD thesis. University of Birmingham, UK.

Ramanathan, V., & Atkinson, D. (1999). Individualism, academic writing, and ESL writers. *Journal of Second Language Writing, 8*, 45–75.

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes, 13*, 149–170.

Stapleton, P. (2002). Critiquing voice as a viable pedagogical tool in L2 writing: Returning the spotlight to ideas. *Journal of Second Language Writing, 11*, 177–190.

Stapleton, P., & Helms-Park, R. (2008). A response to Matsuda and Tardy's 'Voice in academic writing: The rhetorical construction of author identity in blind manuscript review.' *English for Specific Purposes, 27*, 94–99.

Tang, R., & John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes, 18*, S23–S39.

Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics, 22*, 58–78.

—(2004). *Introducing functional grammar* (2nd edn). London: Arnold.

Thompson, G., & Hunston, S. (2000). Evaluation: An introduction. In S. Hunston & G. Thompson (Eds), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 1–27). Oxford: Oxford University Press.

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*, 139–161.

Zamel, V., & Spack, R. (1998). Negotiating academic literacies: Teaching and learning across languages and cultures. Mahwah, NJ: Erlbaum.

Afterword

John M. Swales

There are a number of genres associated with an edited scholarly collection. One is the occluded genre of the reviewer's report of the manuscript mandated by the publisher; this tends to contain, amongst other things, a blow-by-blow account of the strengths and/or weaknesses of each of the individual chapters. In other words, a text-based assessment. A second genre is the Foreword or Preface when written by somebody other than the editor or editors of the volume. In my understanding, authors of Forewords have responsibilities to introduce readers to the subsequent volume in ways that will underscore the volume's value and relevance. Such authors may indeed write in general terms, but they cannot stray far beyond the content of the upcoming material. The intertextual links remain strong, but this time evaluation will be strongly positive. The third genre is the Afterword, which, again in my understanding of these things, allows their authors some greater selective latitude to pick and choose the themes they might want to single out. This is particularly so when the editors responsible for pulling a scholarly collection together offer – as in this case – an insightful introduction explicating both the rationale and the coherence of the volume. In consequence, I have here chosen to focus my comments on just three such themes: The volume's topic of academic written English; ongoing uncertainties in the relationships between discoursal and corporist analytic approaches; and what might be summed up as issues of quantification and accountability.

It is a fact now universally acknowledged that English has emerged in recent decades as the premier international vehicle for the communication of scholarship and research and for advanced post-graduate education. While both the causes and consequences of this pre-eminence are controversial, the present stranglehold of English is in little doubt, as there is little doubt that the amount of scholarly communication in English has been rapidly increasing. (Although, once again, the causes and consequences of this increase are controversial.) Such a phenomenon puts native or near-native speakers of English in a position of undeserved privilege, as well as in a position of having a responsibility to respond proactively to that privilege. And, indeed, this volume can be seen as part of that response as its contributors search for ways to render the rhetorical structures and textual patterns of English academic prose more transparent to themselves and more visible to others. Even more

laudably, many do so with the interests of junior researchers (both speakers of English as an L1 and those with English as an additional language) very much in mind. For three small-scale but telling examples, who could not be struck by Bondi's discussion of the role of chrononyms (a new concept for me) such as *on the eve of, in the wake of and in the aftermath of* in the periodization of history; or by Charles' demonstration of the rhetorical importance of ostensibly straight-forward adverbs such as *simply* and *merely* in both politics and materials science; or of Flowerdew's disentanglement of the functions of the three noun-forms *research, studies* and *study* in applied linguistic literature reviews?

The three foregoing examples by themselves indicate something of this volume's rich coverage of disciplinary discourses, further indications of which come from Pecorari's choice of biology, Shaw's of literature, and Holmes and Nesi's of physics. There is also breadth in writer status, some authors focusing on undergraduate texts, others on post-graduate ones, and still others on those appearing within the covers of scholarly journals. While most contributors focus on the common and established genres of these three levels (undergraduate essays and reports, post-graduate theses and published research articles), Koutsantoni investigates key aspects of research grant proposals, and Hewings et al. make a foray into emerging genres such as e-conferencing. The discoursal approaches adopted are somewhat eclectic, with several following Hallidayan categorizations, others the established EAP move analysis line, but beyond that we can find clause-relations, ethnography and text-based interviews; in some contrast, the corporist techniques used prioritize analyses at the lexical and phraseological levels.

This variation exhibited in this volume can be compared to another major attempt to explore synergies between discoursal and corpus linguistic analyses – that expounded in the 2007 *Discourse on the move* volume from Biber, Connor and Upton. There, the emphasis is narrowed to seeing what benefits might result from an exhaustive amalgam of Biber's multi-dimensional methodology and move analysis (hence the volume's punning title). As far as I can see, the marriage works pretty well, particularly where the genres have become highly conventionalized, but has yet to succeed with academic spoken genres. It might be thought – at least at first sight – that such singularity in a complexly-authored volume would turn out to be a consequence of a North American penchant for rigorous hypothesis testing, given the fact that all the authors are either based in the United States or had done their doctoral work there. However, I do not think any such conclusion is warranted, because *Discourse on the move* is an exceptional volume in both its intention and its execution. This book apart, it does not strike me that there are substantial differences in EAP corpus work on either side of the Atlantic. After all, of the five authors of the highly successful and corpus-based *Longman grammar*, three are American and two are European. However, if the North American volume permits a relatively straightforward assessment, *Academic writing: At the interface of corpus and*

discourse does not, because much will depend on particular readers' particular disciplinary and generic interests, on their methodological preferences, and on their degree of involvement in academic writing instruction. Different readers then will have their different favourites.

Since the marriage metaphor has been invoked, we might now consider the opening sentence to Hyland's chapter: 'Corpora and discourse approaches are perfect bedfellows'. Easily said perhaps, but not always so easy to put into practice, as I know to my personal cost. Corpus linguists these days can – and do – construct written corpora with considerable facility and then apply to them software tools such as keywords and n-grams to reveal underlying propensities and regularities. However, a list of lexical items that occur comparatively frequently or infrequently does not tell us much, without attention to both their placements and functions, and a list of raw lexical bundles, with no attention to whether they have compositional or semantic coherence, may also be unenlightening. Of course, quantification can be very valuable, but it can also be a cover for lack of intuition and conceptualization. In this volume, however, a model essay showing precisely how to move from quantification, to explication, and then on to revelation is the one by John Flowerdew. Another chapter from the opening section of great interest in this regard is Diane Pecorari's 'Formulaic language in biology', partly because of her selection of a topic-specific corpus – indeed one devoted to the single yeast *Candida albicans*, the genome sequence of which is now available. We learn from this chapter that a narrow, highly technical corpus may produce longer phraseological units (indeed up to 12 words) than we have otherwise seen. This finding, as Pecorari notes, may have implications for corpus design and for providing EAP materials. And, I might add, it may also have implications for the way we view plagiarism/patch-writing.

The other aspect of quantification that I would like to raise at the interface of corpus and discourse can be summed up in one of Einstein's aphorisms 'Sometimes that which counts cannot be counted and that which can be counted does not count'. We are faced with some dilemmas here, especially when it comes to comparing the prose of aspirant academic writers and those adjudged to be competent. At present, certain things seem to be beyond our grasp, or at least beyond our capacity to express them in commensurable terms – such things as the nature and character of a '*mature* academic style', the ingredients (if any), across the disciplinary board, of 'successful' writing, or indeed, again across the board, the extent to which *style* matters. Faute de mieux, we explore, as a number of papers in this collection do, contrasts between learner and reference corpora of items and elements of over- and under-use (cf. Gardezi & Nesi; Shaw; Granger & Paquot). The first two of these chapters focus on what are typically called sentence-connectors (prototypically *however*). As it happens, adversativeness can be expressed in a multitude of ways, as pointed out more than 30 years ago by Eugene Winter in 1977 with his

'Vocabulary 3', that is the use of nouns such as 'reason' 'contrast' and 'conse-quence' to express intersentential relations. Advanced writers may have more resources to express those relations, but that does not imply, in my view, that more limited writers should be dissuaded from employing (wherever appropri-ate) the resources they have available. As an instance of the complexity here, consider, for a moment, my very recent use of the possibly-unusual adversative, *As it happens*. At least for the American doctoral students that I have recently been working with, this is for them a connective 'with an attitude'; one student glossed it as 'however I really do know better'. And, to boot, any connotation of superiority, or indeed of snobbishness, is one that – as a mere English person – I had never realized!

This volume indeed represents a series of state-of-the-art explorations of the interface of corpus and discourse. Although, in the last two paragraphs, I have been suggesting that the complementarity of the two approaches is more likely to be a hard-won achievement rather than an easily-assumed precondition, the marital difficulties are reduced – as the editors point out in their intro-duction – whenever we conceive of discourse analysis and corpus linguistics as forming a continuum rather than standing in firm opposition to each other. That said, as we go forward in our attempts to further characterize the intrica-cies of academic prose, we will often see that it is typically somewhat easier for discourse analysts to incorporate corpus linguistics than for corpus linguists to expand their textual horizons to encompass the discoursal plane.

References

Biber, D., Connor, U., & Upton, T.A. (2007). *Discourse on the move: Using corpus ana-lysis to describe discourse structure*. Amsterdam: Benjamins.

Winter, E.O. (1977). A clause-relational approach to English texts. *Instructional Science, 6*, 1–92.

Author Index

This page intentionally left blank

Subject Index