Writing Difficult Texts

A thesis submitted for the degree of Doctor of Philosophy
Lancaster University
September 1999

Candidate:	Christopher Tribble	BA Hons. (American Studies & History) MA (American Literature) MA (Applied Linguistics)
Supervisor:	Dr Greg Myers	
Department:	Linguistics And Modern English Language	

ABSTRACT

This thesis uses concepts and techniques associated with genre analysis, corpus linguistics and discourse analysis to offer some solutions to problems in writing instruction – in particular the problem of learning to write into a new or unfamiliar genre. Two major corpus linguistic analytic frameworks are used in elaborating these solutions. The first is the multivariate / multifunctional approach proposed in Biber D, 1988; the second draws on the notion of keywords developed by Scott M, 1996. These frameworks are used in a detailed analysis of a research corpus of an example genre – fourteen Project Proposals (112,000 words) submitted in bids to win contracts issued by European aid and development agencies,.

The thesis has four major sections. In the first, an interpretation of "writing", "difficult" and "texts" is provided as a way of framing the later discussion. This is followed by a survey of current issues in teaching writing, and an introduction to the Project Proposal corpus (including comments on technical problems of corpus construction). In the second section, *An account of the texts*, a detailed analysis of the texts in the Proposals Corpus is presented in three chapters: *Grammar and Style*, *Lexical Dimensions*, and *Organisation*. This analysis is preceded by a summary of some of the problems which face researchers attempting to replicate Biber 1988. In a small innovation for this kind of corpus study, the single chapter in the third section: *Talking with writers* goes beyond a strictly corpus approach, and reports case-studies and interviews with writers from the organisations which originally provided the research data. This has proved to be a particularly valuable initiative as it demonstrates the importance of not only depending on corpus evidence in developing an understanding of a genre.

The final section of the thesis: *Implications for pedagogy* reviews a set of issues and questions which were raised in Chapter 2. During this chapter, these questions are used as a way of offering a set of practical proposals for integrating genre analysis and corpus linguistic techniques into writing pedagogy.

ACKNOWLEDGEMENTS

First I would like to thank my supervisor, Dr Greg Myers, for his consistent help and support throughout the writing of this thesis. He has offered advice when I have been unable to find a voice for what I have wanted to say, and has ensured that I have not made a fool of myself when I have wanted to speak before thinking! I could not have asked for better help. Amongst the many who have given their time to listen to me talking about this project or who have provided practical assistance, I would also like to thank:

- Professor Charles Alderson for suggesting that I undertook this research at Lancaster – I could not have had better advice
- Professor Doug Biber for reading early drafts of Chapters 3 and 4 and offering clarification when I was confused
- Nick Smith for advice on working with corpora in early stages of the project
- Martin Wynne and his colleagues in UCREL for their invaluable assistance in POS tagging my corpus
- Rachel Fligelstone in the Computer Support Service at Lancaster for advice on statistical matters
- Professors Henry Widdowson and Michael Stubbs for supervising the earlier incarnation of this project at the Institute of Education in London University
- Professor Michael Hoey and Dr Tony McEnery for their helpful and perceptive comments during the viva for this PhD

- and the British Association for Applied Linguistics for a research grant which I received in 1990 when I was involved in this earlier incarnation. I hope that the committee feels that this final outcome justifies their investment…

Finally, I have to thank Susan Maingay for her willingness to be my first reader throughout this project, this in spite of the ever growing pressure of her own work, and my friend and colleague Dianne Wall in the Institute for English Language Education at Lancaster for sharing the process of doing the Ph.D. – knowing that she was out there going through similar pain and pleasure has been a huge support.

I hope that others who set out to complete a similar task have such good teachers, friends and colleagues to help them on their way.

Christopher Tribble: London, Lancaster, Warsaw, Colombo 1988 – 1998

This thesis is dedicated to the memory of my parents Phyllis and Albert Tribble

TABLE OF CONTENTS

ABSTRACT ..... **I**

ACKNOWLEDGEMENTS ..... **III**

SECTION ONE: SETTING THINGS UP **1**

CHAPTER 1: WRITING DIFFICULT TEXTS 1
1. What I am attempting to do in this thesis *1*
2. Writing *4*
3. Difficult *7*
4. Texts *14*
5. Signpost *18*
CHAPTER 2: TEACHING WRITING 20
6. Introduction *20*
7. Pedagogic models of writing *21*
8. Process *23*
9. Genre *28*
10. Approaching writing *33*
11. English for business / professional communication *36*
12. English for academic or study purposes *45*
13. English for international examinations *58*
14. Conclusion *65*
CHAPTER 3: APPROACHING THE DATA: DEALING WITH GENRE 69
15. Introduction *69*
16. Are project proposals a genre? *70*
17. Using corpora in genre analysis *78*
18. The PP Corpus *80*
19. A methodology for using corpora in genre analysis *82*
20. Genre, corpora, writing *85*

SECTION TWO – AN ACCOUNT OF THE TEXTS **88**

CHAPTER 4: WHAT ARE WE LOOKING FOR? 88
21. Introduction *88*
22. The Biber multi-function / multi-dimensional model *89*
23. Working with the Biber framework *93*
24. Applying the Biber 1988 analytic framework *96*
25. Replicating the tools used in the Biber 1988 study *100*
26. Improving the research design *107*
27. Revised results *111*
28. Conclusion *116*
CHAPTER 5: GRAMMAR AND STYLE 120
29. Introduction *120*
30. Style and stylistics *121*
31. Provisional identification of prominent linguistic features of Project Proposals *123*
32. Analysis of prominent features *126*
33. Attributive adjectives *126*
34. Nominalisations *136*
35. Phrasal coordination *139*
36. Predictive modals *145*
37. Low frequency items *148*
38. Adverbs *149*
39. Third person pronouns *151*
40. Private verbs *152*
41. Conclusion *155*
CHAPTER 6: LEXICAL DIMENSIONS 158
42. Introduction *158*
43. Is frequency enough? *160*

44. Keywords *162*
45. Finding keywords *166*
46. Does the reference corpus matter? *169*
47. Keywords in stylistic profiling *173*
48. Interpreting the PP keyword lists *179*
49. Positive keywords *181*
50. Negative keywords *184*
51. Looking at but *186*
52. What lexical patterns is the word part of? *195*
53. Lexical patterns: collocates *196*
54. Lexical patterns: clusters *201*
55. Lexical patterns: patterns *203*
56. Lexical patterns: conclusion *204*
57. Does the word regularly associate with particular other meanings? *205*
58. What structure(s) does it appear in? *211*
59. Is there any correlation between the word's uses / meanings and the structures in which it participates? *215*
60. Is the word associated with (any positions in any) textual organisation? *217*
61. Conclusion *220*
CHAPTER 7: ORGANISATION 222
62. Introduction *222*
63. Macro structures *226*
64. Moves *234*
65. Minimum Discourses *237*
66. Conclusion *245*

SECTION 3 – TALKING TO WRITERS 247

CHAPTER 8: WRITING PROJECT PROPOSALS 247
67. Introduction *247*
68. The organisations *249*
69. The interviewees *250*
70. Interview structure *252*
71. Interview results: Question Set 1 – Starting *254*
72. Interview results: Question Set 2 - Writing the technical proposal *259*
73. Text Analysis Task results *263*
74. Conclusion *268*

SECTION FOUR – IMPLICATIONS FOR PEDAGOGY 269

CHAPTER 9: HELPING LEARNERS WRITE DIFFICULT TEXTS 269
75. Introduction *269*
76. Frameworks for teaching writing *271*
77. Question 1 ..... *272*
78. Question 2 ..... *276*
79. Working with corpus data *279*
80. Questions 3 and 4 ..... *285*
81. Question 5 ..... *289*
82. Some conclusions *292*

BIBLIOGRAPHY, TABLES & FIGURES 295

BIBLIOGRAPHY 295
LIST OF TABLES 312
LIST OF FIGURES 314

LIST OF APPENDICES 315

SECTION ONE: SETTING THINGS UP

Chapter 1: Writing Difficult Texts

Renée Magritte
© C. Herscovici, Brussels – Belgium

1. What I am attempting to do in this thesis

1.1 I chose the title for this thesis at the beginning of the research project, thinking that it would be provocative and interesting – and it has been useful in helping me focus on why I am doing what I am doing, and in providing an explanation of the project. The title has also presented me with a problem, as it could commit me to dealing with several very large research areas and their associated literatures – a task which is beyond the scope of what I want to do in this particular piece of writing. What I propose to do in this opening section is, therefore, to state as clearly as I can the limits of what I am doing, and to situate what I *am* trying to do in an appropriate research context. If I can get this clear from the outset, there is a much better chance that what follows later will be seen as appropriate – and will, I hope, make sense.

1.2 This thesis has a practical aim – to help myself and (I hope) others to work more effectively as teachers of writing. I started the research project with a

general question which had taxed me as a writer and as a teacher for several years: "How is it that some pieces of apparently well-formed writing work better than others?" Coulthard refers to this problem when he talks about the way in which any given text can be seen "*as just one of an indefinite number of possible texts, or rather* ***possible textualizations****, of the writer's message* …." (Coulthard M 1994:1, emphasis in the original). And because I have taught English to speakers of other languages throughout my career, I have also had in mind a more specific question: "How can I help learners write better?" – especially when matching writing style to writing purpose (Tribble 1985).

1.3 The main way in which I have approached this problem has been to draw on insights developed over the last decade in empirical linguistics and corpus research (Biber 1988, Biber D, S Conrad & R Reppen 1994, Botley S, Glass G, McEnery T & A Wilson 1996, Granger S & C Tribble 1998, Hoey M 1997a, McEnery T & A Wilson 1996, Oostdijk N 1988, Stubbs M 1995, Stubbs M 1996, Tribble C 1997, Wilson A & T McEnery 1994) in order to see how useful they might be when it comes to dealing with the problems which arise when someone, especially a foreign language student, is writing a text that he or she finds difficult (what I mean by *difficult* being something that I will discuss below). It is only recently that we have reached a point where it has been possible to apply the results of corpus research to language education (Aston G 1996, Flowerdew J 1996, Johns T 1988, Johns T and P King, eds. 1991, Murison-Bowie S 1996, Stevens V 1995, Tribble C 1991, Tribble C & G Jones 1997), and it is in this area that I would like this thesis to make its contribution.

1.4 The corpus linguistic component of the thesis centres on the analysis of 14 Project Proposals written during the period 1994-1996. Twelve of these were submitted to the European Union's PHARE[1] programme, and two to the UK government's Overseas Development Administration (as it was then called). These texts have been chosen because writing such documents is considered to be a demanding task by professional writers with experience in the field, and a daunting undertaking for the apprentice. This is not to say that the end product of this thesis will be a "How to" handbook for writing project proposals. The texts are simply useful examples of the kinds of factual writing problem that many professionals have to be prepared to address – and which, as a teacher, I want to be able to help my students learn to write. As such, they offer a basis for assessing the usefulness of corpus linguistic approaches in the development of writing instruction programmes for learners wishing to write into other complex genres. A more detailed account of PPs and the texts in the PP Corpus will be given in Chapter 3: *Approaching the data: dealing with genre*.

1.5 While I feel that corpus linguistics offers teachers and students a major – and very exciting – set of insights and tools, I also recognise that it is unlikely that a corpus approach will offer *all* the pedagogic answers. For this reason, the thesis will also contain a number of small case studies (see Section Three: Talking to writers) which reconnect the texts in the research corpus with the organisations and writers who composed the proposals. By considering the texts both as linguistic products and as the result of socially situated practices, I hope to provide a more adequate response to the main motivation for

[1] The European Union's main development programme for countries in Central and Eastern Europe

undertaking this research – the desire to help apprentice writers to write more effectively.

1.6 As with any other project, having decided my aim, I need to be clear about the outputs which will be necessary to its achievement. These outputs can be summarised as follows:

Output 1	a statement of the grounds on which subsequent arguments will be developed. This statement will be mainly concerned with what I understand by *writing*, by *difficult* and by *texts* and will take up the rest of this chapter.
Output 2	a summary of what I consider to be key issues (and restrictions) in current approaches to the teaching of writing. This will form the second chapter in the thesis and will make it possible for later discussions of pedagogy to be situated in an understanding of current writing instruction.
Output 3	an account of possible applications of empirical linguistics in the development of writing instruction materials. This account will form the major part of the thesis and will focus on the needs of writers of one particular kind of text. It will contain several chapters in which the methodology I have adopted will be explained, and the findings I have obtained will be discussed.
Output 4	an account of the ways in which expert writers view a) the texts that have been the subject of the corpus analysis, and b) the writing processes required for the production of these texts.
Output 5	a report of findings and recommendations for possible pedagogic applications.

The next sections in this chapter will, therefore, focus on *writing, difficulty* and *texts*.

2. Writing

2.1 Grabe and Kaplan's distinction between "writing with composing" and "writing without composing" (Grabe W & R Kaplan 1996:4) neatly summarises the basic view of writing that is taken in this thesis:

(now including countries in the former Soviet Union)

> "… one may distinguish writing which involves composing from writing which does not; this distinction is useful because most of what is referred to academically as writing assumes composing. Composing involves the combining of structural sentence units into a more-or-less unique, cohesive and coherent larger structure (as opposed to lists, forms, etc.). A piece of writing which implicates composing contains surface features which connect the discourse and an underlying logic of organisation which is more than simply the sum of the meaning of the individual sentences." (Grabe W & R Kaplan 1996:4)

Such an understanding of writing has the merit of simplicity and relevance to what proposal writers do; it would only be in a moment of extreme cynicism that a proposal writer would say that his or her task was one which involved writing *without* composing! It also has the merit of being a reasonable focus for EFL pedagogy. There are limits to the kinds of project that writing teachers can take on, and the kinds of responsibility they can assume (especially in the foreign language). To help learners write in the sense outlined above – especially if you only have contact once a week for one or two semesters – is realistic. To claim to do much more is probably not.

2.2 Clearly Grabe and Kaplan's account of writing has its limitations. Witte has made a more comprehensive statement of what writing is, and also indicated the problems that teachers and researchers face in dealing with this fuller understanding of the object of their professional activity:

> " …. any conceptualization of writing must be able to accommodate not only the production and use of extensive alphabetic texts but also the production and use of minor (e.g., lists, labels, notes) forms of "writing" and texts such as engineering proposals, guidebooks to indigenous plants, and scholarly articles, all of which typically employ more than one symbol system. Second, […] any conceptualization must be able to account for both the meaning constructive and social-constructive dimensions of writing regardless of whether writing be viewed as a process or a product and regardless of whether the writer traffics in linguistic or nonlinguistic symbols. Third, […] any conceptualization must be able to

> account for both the protracted and the collaborative nature of composing regardless of the symbol system the "writer" might employ at a given time. In my view, none of the perspectives by which meaning-making has been treated in writing research and none of the perspectives into which writing research is currently collapsed meets these three criteria, in large part because those perspectives all presuppose verbal language as the only sign system relevant to the study of writing." (Witte S 1992:249)

2.3 In this understanding of writing, when writing teachers set out to help learners to write, they should not only aim to help their students to use the "minor texts" and more extended forms familiar to EFL teachers (see Chapter 2: *Teaching Writing* below), they should also help learners engage with the "meaning constructive and social-constructive" dimensions of writing, and with the complex interactions that arise around writing in the workplace and the academy. This is a much broader task than helping learners to develop skills in "writing as composing", but can still usefully be borne in mind by teachers and materials writers who are working with apprentice writers.

2.4 We will, in fact, see in Chapter 2 that in areas such as business communication and English for academic purposes EFL materials writers do take on some of these broader cultural and interactional issues suggested by Witte. We will also see that this can be fraught with difficulty, raising as it does questions about which social context text is constructing. Is it an anglophone and geographically discrete institution such as a British university, or a multinational corporation (such as Volvo) whose corporate language may be English but whose headquarters are in Sweden and which employs staff or has affiliates in most countries on the planet? Or is it an amorphous polyglot bureaucracy such as the PHARE office in Brussels? When an apprentice writer is asked to build Babel, what assistance can the teacher offer?

2.5 Witte himself indicates that the questions he raises have not been satisfactorily researched, let alone answered. In this present writing I shall restrict myself to the more rough-and-ready understanding of "writing" outlined by Grabe and Kaplan. Thus, when I say "writing", I am referring to writing as composing, and when I refer to helping students to write better, I am concerned primarily with ensuring that learners of English as a foreign language are given more effective support as they develop the language system knowledge and the cognitive skills they need in order to compose texts that are appropriate to the contexts in which they will find themselves writing.

What then of *difficult*?

3. Difficult

> "… thirty years ago, my older brother, who was ten years old at the time, was trying to get a report written that he'd had three months to write, which was due the next day. We were out at our family cabin in Bolinas, and he was at the kitchen table close to tears, surrounded by binder paper and pencils and unopened books on birds, immobilized by the hugeness of the task ahead. Then my father sat down beside him, put his arm around my brother's shoulder, and said, 'Bird by bird, buddy. Just take it bird by bird.'" (Lamott A 1994:19)

3.1 Part of me (my remembered self at such an age, my own self at this moment of writing) can feel some envy for the boy in the passage quoted above. I can regret that I have no wise mentor beside me now, ready to place a paternal arm around my shoulder and provide me with such advice. 'Of course! Bird by bird! Why didn't I think of it before…?'. Part of me, however, is concerned at the offer of this particular counsel to an apprentice writer, because, all too often, it just isn't like that. I would not deny that part of the difficulty which faces us writers is the problem of working our way through the job in hand,

bird by bird, and that we do discover what we want to write by writing. However, it is not only extent which can daunt. Yes, there is the difficulty of getting through the task, but there can also be problems which arise from a lack of understanding of the problem being addressed, or a lack of familiarity with texts like the one you are writing which creates the problems, and so forth.

3.2 Martin touches on these other kinds of difficulty in a discussion of (some pleasingly apposite) example of texts written by young children in the Australian state education system:

> **Text A** 'Birds live up in a tree. If they don't eat they die. Redbirds blackbirds any colored birds. Dark birds light birds. Some are small and others are big.'
> **Text B** 'My bird lived up in a tree. It ate so it wouldn't die. It was a black bird and it was small' (Martin J 1989:7)

3.3 Martin reports that he found young children were often asked to write in the style of Text B (which he classes as DESCRIPTION), but less frequently in that of Text A (REPORT), and that they often find this latter presents them with problems because they have not been taught how to go about the task. He makes the distinction between these, and two other 'proto-genres' on the basis of differences of focus and generality:

	particular	**general**
event focus	RECOUNT	PROCEDURE
thing focus	DESCRIPTION	REPORT

(Martin J 1989:8)

In such a scheme, Text A is public, open to challenge and refutation, Text B is personal – a challenge to the text also implying a challenge to the writer's own experience. (Tribble C 1997:49). Martin's thesis in this instance was that if we want children to have the capacity to write effectively across a range of

genres – and to have the necessary command of appropriate lexico-grammatical resources to complete such tasks, we need to give them appropriate instruction. Martin concludes by commenting:

> "… by focusing children's attention on the fundamental differences between Descriptions and Reports, teachers could help children to write more consistently in each genre." (Martin J 1989:7)

3.4 Underlying Martin's argument is a view that some approaches to teaching writing – particularly those arising from progressivist educational theory (Bowers C A 1987, Ellsworth E 1989, Giroux H 1988), and which depend heavily on what has been called a "process" approach to writing[2] (Murray D M 1982, White R & V Arndt 1991, Zamel V 1983) – may inhibit learners from taking on a range of powerful social roles which depend on explicit knowledge of language and a range of literacy practices which are not necessarily acquired easily. In so far as it has been critical of process writing and the cultural relativism associated with this moment, part of the genre literacy agenda has, therefore, been to give students access to a range of factual genres (Cope B & M Kalantzis 1993, Martin J 1989) which take them beyond personal writing and into a world where written language is used to make things happen. It would seem that difficulties in writing might not necessarily be solved by just taking things one step at a time, and working through the task from beginning to end – be it black birds or red birds.

3.5 I have proposed elsewhere that writers require four sets of knowledge if they are to have a reasonable chance of succeeding in achieving particular writing objectives. These are:

[2] see Chapter 2

content knowledge	knowledge of the concepts involved in the subject area
writing process knowledge	knowledge of the most appropriate way of preparing for a specific writing task
context knowledge	knowledge of the social context in which the text will be read, and co-texts related to the writing task in hand
language knowledge	knowledge of those aspects of the language system necessary for the completion of the task
	(Tribble C 1997: 43)

Table 1 - What writers need to know

3.6 Viewed from this perspective, the issue of difficulty becomes more approachable, as it is evident that writers may face problems along any of these dimensions. Halliday and Hasan's discussion of the relationship between text and context underscores this:

> "To think of text structure not in terms of the structure of each individual text as a separate entity, but as a general statement about a genre as a whole, is to imply that there exists a close relation between text and context [….]. The value of this approach lies ultimately in the recognition of the functional nature of language. [….] there cannot be just one right way of either speaking or writing. What is appropriate in one environment may not be quite so appropriate in another.
>
> Further, there is the implication that an ability to write an excellent essay on the causes of the Second World War does not establish that one can produce a passable report on a case in a court of law. This is not because one piece of writing is inherently more difficult or demanding than the other, but because one may have more experience of that particular genre." (Halliday MAK & R Hasan 1985:6)

3.7 When one faces difficulty in writing in a particular genre, then, this difficulty can arise from a lack of many different kinds of knowledge, including:

- a lack of knowledge of compositional processes appropriate to writing in this genre
- a lack of knowledge of the social relations in play in a genre

- a lack of knowledge of aspects of the language system required in order to develop an appropriate wording of the required text by the genre

- a lack of knowledge of the text structures normally required in order to instantiate meaning in this particular communicative context

3.8 Myers draws attention to these kinds of problem when he discusses the contrasting narratives which are required by texts developed for an audience of peers as opposed to a general, popular, readership:

> "Textual differences in narrative structure, in syntax, and in vocabulary help define two contrasting views of science. The professional articles create what I call a *narrative of science*; they follow the argument of the scientist, arrange time into a parallel series of simultaneous events all supporting their claim, and emphasize in their syntax and vocabulary the conceptual structure of the discipline. The popularizing articles, on the other hand, present a sequential *narrative of nature* in which the plant or animal, not the scientific activity, is the subject, the narrative is chronological, and the syntax and vocabulary emphasize the externality of nature to scientific practices. (Myers G 1990:142)

If a writer is asked to take a text originally developed for a scholarly journal and re-word it so that it will be appropriate for a news article, he or she should be aware of the contrasting demands of the different narratives required by these two kinds of media. Similarly, writers of project proposals have to be aware of the constraints of the genre, the kinds of narrative that are most likely to be appropriate for their readership, and the aspects of the language system that they will need to draw on in order to achieve the textual effects they require.

3.9 In order to achieve such textual effects authors have to have a broad repertoire of ways of writing – or they may find themselves restricted to the roles in

which they are competent. Myers demonstrates this in his example of the shift in the titling of an article as it was transferred from a specialised science journal to a popularising one:

> "For instance, the title of an article by Geoffrey Parker in Evolution, a specialized scientific journal, was '*The reproductive behaviour and the nature of sexual selection in Scatophaga stercoraria* L. (Diptera. Scatophagidae), IX. Spatial distribution of fertilization rates and evolution of male search strategy within the reproductive area.' The editor gave Parker's New Scientist article the title 'Sex around the cow-pats'. (Myers G 1994:180)

3.10 The writer of the popularising article has had to make a range of content and lexical selections in order to arrive at this contextually felicitous re-wording (i.e. "The reproductive behaviour and the nature of sexual selection " ➔ "sex"; " within the reproductive area" ➔ "around the cow-pats"). Similar issues of knowledge and competence arise when we consider the way in which a piece of writing is developed and the difficulties which can face a writer as she or he prepares, composes and revises. Process writing research (see Chapter 2) has demonstrated that writing is not just a step by step activity and a specific experience exemplifies this. At the time of writing these words in this paragraph (07/03/98 13:07) Microsoft Word told me I had been editing the chapter for 2,366 minutes – nearly 40 hours. Although the figure of 40 hours is misleading as it refers only to the amount of time during which the document has been 'open' on my computer – I might have been spending most of the time having cups of tea in the kitchen – it does start to give an indication of how long this particular piece of text has taken to develop. And the fact of the matter is that although I began to compose this first chapter in February 1998 – that is when I first opened and named this file – and I planned to finish it before the end of the month, the whole the process started during a

conversation with Henry Widdowson and Mike Stubbs over ten years earlier in the autumn of 1987! So this "first" chapter has taken over ten years to write and it has a relationship with the tens of thousands of words I've written during that time which relate to these topics; other chapters which follow in this thesis, but which were written before this chapter; and the two books and several articles and papers which developed themes I am re-visiting in this dissertation. Apart from the difficulty of working out what a PhD thesis might be *about*, other difficulties I have faced here have been to do with time, with technology[3], with my own understanding of the literature and my capacity to organise and marshal ideas and arguments.

3.11 Some of my ways out of difficulty have been found in what Britton has called "shaping at the point of utterance" (Britton J 1983); other escape routes have been down to my experience of planning and developing other long texts. However, given that *writing* can be *difficult* in so many ways, I again want to set a limit to what I have to say in this thesis. The kind of difficulty which I shall be considering will relate primarily to an aspect of knowledge of the texts. I propose to focus on a narrow set of problems which can face writers who are working in what is for them a new genre - namely those that arise from a lack of knowledge of texts like the one in hand, and a lack of knowledge of what readers expect from such texts. I shall, therefore, be focusing on the co-textual aspect of context knowledge and will attempt to find ways of helping the writer who does not know what the genre requires of them in terms of text organisation and wording.

[3] Shifting from Windows 3.1 to Windows 95, the delights of Word master documents, coordinating files between lap-tops and desktops, fitting graphics into the text ….

3.12 Such an approach does not, however, imply that difficulty is uniquely determined by the demands of the reader. There are many reasons why a particular text might be difficult to write.

- Does your job depend on your text closing the deal? – e.g. when an advertising company bids for a new account

- Do you have enough time to complete the task? – e.g. writing under examination conditions

- Is the task of a particularly demanding level of complexity (both logistically and intellectually)? – e.g. writing a PhD thesis

- Are you having to make the shift from primarily oral communication to communication that is primarily written? – e.g. the dilemma which might face a skilled interviewer having to devise a written questionnaire for the first time

I have chosen to narrow my definition of difficulty to the "co-textual aspect of context knowledge" because this is a kind of difficulty that a teacher can help with. As a writing teacher, I cannot take responsibility for the experience and the content knowledge that the learner needs in order to fulfil real-world writing tasks. What I can do, however, is to offer pedagogic pathways into unfamiliar texts so that that which was difficult because it was unfamiliar becomes less strange, and thereby, easier.

4. Texts

> "To study writing is, over and above all else, to study acts for making meaning that are mediated through 'texts'". (Witte S 1992:237)

4.1 So far we have concluded that the main task of this thesis will be to consider the needs of learners involved in the activity of combining "structural sentence units into a more-or-less unique, cohesive and coherent larger structure" (*writing*), in settings which involve what is for them a new genre (*difficult*). The last element of the title of this thesis ought to be unproblematic (we all know what a text is, don't we?), but – as with Magritte's apple in the epigraph to this chapter – it is probably better not to jump to conclusions on the basis of appearances. A representation of an apple is not an apple. What is a text today might not have existed the day before. For Graddol and Goodman:

> " new forms of text [….] have appeared in English as a consequence of technological innovation and social change [….] new kinds of text reflect and help construct changing identities and social relations." Graddol D & S Goodman 1996:1)

4.2 "Texts" can include films, cartoons, wine-labels and advertisements, and if we consider our own experience of the world we can add the cereal packets, junk-mail, jotted notes and shopping lists which clutter the kitchen tables of most households, along with the e-mail printouts and doctoral dissertations which clutter those of the less fortunate amongst us. Texts (like apples) can be many things, some of them more slippery than others.

4.3 Such a view of text accepts that texts do not have to be written. Given below is a series of examples of different definitions of text. Although a rather long list, they are offered here as they all share a feature which will be important for the rest of my argument:

- Bakhtin – a text was "any coherent complex of signs." (Bakhtin MM 1986:103)

- Halliday – the essential feature of text:

> "is that it is interaction. The exchange of meanings is an interactive process, and text is the means of exchange: in order for the meanings which constitute the social system to be exchanged between members they must first be represented in some exchangeable symbolic form, and the most accessible of the available forms is language. So the meanings are encoded in (and through) the semantic system, and given the form of text." (Halliday MAK 1978:139-140)

- De Beaugrande and Dressler – a text is a:

> "COMMUNICATIVE OCCURRENCE which meets seven standards of TEXTUALITY. If any of these standards is not considered to have been satisfied, the text will not be communicative." (de Beaugrande R & W Dressler 1981:3)

De Beaugrande and Dressler's seven standards (1981:3-11) were:

Cohesion	Intentionality	Informativity
Coherence	Acceptability	Situationality
		Intertexuality

- Brandt:

> "… the text appears as an instrument of social interaction, conveying multiple messages about the social world in which it has been developed" (Brandt D 1986:93)

- Stubbs refers to texts as:

> "an instance of language in use, either spoken or written …" (Stubbs M 1996:4).

- Clark and Ivanić follow Fairclough in taking text as either spoken or written, but insist on seeing it as arising from specific social conditions. They comment on Figure 1 below as follows:

> "One advantage of this diagram [after Fairclough 1989] is that it illustrates graphically how the words themselves – 'text' – are embedded in the social forces that produce them. Fairclough shows how a text (written or spoken) is inextricable from the process of production and interpretation that create it, and that these processes are in turn inextricable from the local, institutional and socio-historical conditions within which the participants are situated." (Clark and Ivanić 1997:11)

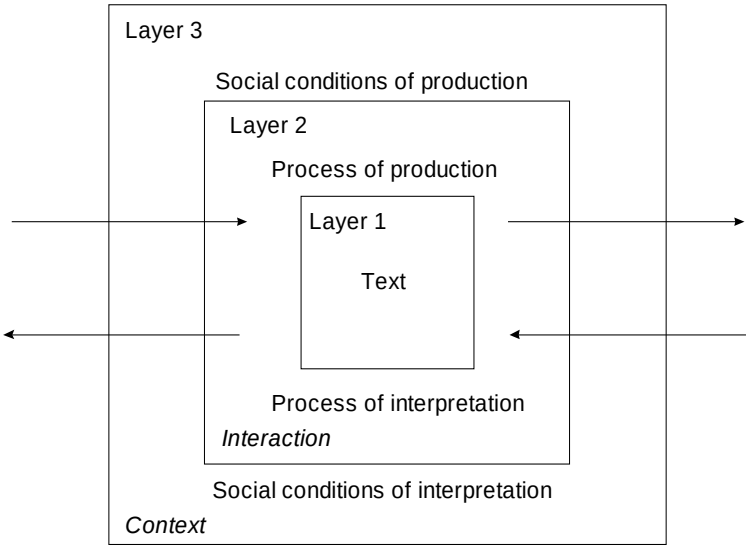


Figure 1 Clark and Ivanić 1997:11

- Grabe and Kaplan provide a useful summary of many of these positions when they offer the following definition of a text as:

> "… a structural equivalent of language in real use which conveys meaning in all four senses of Hymes's (1972) communicative competence (whether a text is possible, feasible, appropriate, and performed), and which suggests a topic of discourse (however minimal) [….] A text occurs when the discourse segment is identified as possible, feasible, appropriate, and performed, and has a topic" (Grabe W & R Kaplan 1996:40)

4.4 These various interpretations of text all emphasise its interactive and communicative dimensions – with Grabe and Kaplan usefully stressing that "a text occurs when the discourse segment is identified as possible, feasible, appropriate, and performed, and has a topic". It is this on this basis that I shall select and discuss the *texts* which will be considered in this thesis. My difficult texts will be examples of written language in use. They will be

identifiable discourse segments, and, they will be drawn from real-life instances which are feasible, appropriate, performed and have a topic.

5. Signpost

5.1 Everything that I have said so far in this chapter about writing difficult texts has been relatively abstract. In paragraph 1.6, I presented 5 major outputs which will be required in order to achieve the overall aim of this thesis – to help myself and others work more effectively as teachers of writing. Having, I hope, fulfilled my obligation to account for the *writing*, *difficult*, and *texts* of my title, I now feel that I should provide the reader with a more explicit map of what follows.

5.2 In Chapter 2, I shall provide an overview of writing theory and review some of the practical problems which arise in the teaching of writing. This review will provide me with a basis for the recommendations for writing instruction that I shall make in the closing chapter of this thesis.

5.3 In Chapter 3, I shall give a more detailed account of the data I shall be using in the corpus study section of the thesis, and also comment on why I feel it constitutes a discrete text genre. For the moment, I hope that the reader will accept that I have chosen these texts because they are interesting examples of a relatively unusual genre, and present significant difficulties for those unfamiliar with the demands such a genre makes of the writer.

5.4 In Chapters 4, 5, 6 and 7, I shall present the findings from my analysis of the PP corpus. This will first require (Chapter 4) a detailed account of the difficulties that are faced when applying one of the major corpus based models of written and spoken language variation to a different data set from that used

in the original study (Biber D 1988). The following chapters will address the different aspects of the PP corpus that I consider significant for this study – that is, Grammar and Style, Lexis, and Organisation.

5.5 Following this linguistics section, I shall report on the findings of a survey in which the results obtained from the empirical study of the research corpus are matched against the understanding of what Project Proposals are that are held by writers in proposal writing organisations. This survey has been undertaken in order to gain insights into how writers went about composing the proposals, and why they chose to write in the particular ways exemplified in the corpus.

5.6 The final chapters in the thesis will offer some recommendations for the possible pedagogic application of the approach adopted in this thesis, and an indication of future research directions. The pedagogy chapter has been included as I feel it important to flag some of the advantages and limitations of what I have done in this study, and to offer an agenda for teachers of writing who are interested in drawing on the sorts of tools and techniques which empirical linguistics offers to the practical language teacher.

SECTION ONE – SETTING THINGS UP

Chapter 2: Teaching Writing

"… text is choice." (Halliday MAK 1975:123)

6. Introduction

6.1 As the stated purpose of this thesis is to make a contribution to the teaching of writing in English as a Foreign Language pedagogy, I want to situate the research I shall be reporting and any comments I have to make in the context of what is currently happening in the mainstream of EFL writing pedagogy. As I believe that a review of practice given in isolation from the theory which informs it will not provide a useful or persuasive account, I shall also provide a summary of some of the major themes which have occupied theory formers with an interest in EFL writing pedagogy. In doing this, I shall not be attempting to replicate the work done by Grabe W & R Kaplan (1996), as they have provided a more exhaustive account of theory and practice in writing than I could hope to undertake here. What I present in the first sections of this chapter should, nevertheless, provide a useful contextualisation for the review of practical teaching which will follow.

6.2 In this review I shall focus on specific instances of contemporary teaching materials that have been published for the UK and European ELT markets (or those markets where UK based publishers have a leading market share – e.g. large parts of East and South East Asia, and Latin America). This will give me an opportunity to review the ways in which the teaching of writing is currently approached in many foreign language teaching classrooms, and to indicate any limitations which I feel such materials have. I have already undertaken a similar survey of writing instruction materials (Tribble C 1997a)

and have found it helpful in identifying those areas where a corpus linguistic / genre orientation may make a contribution to the teaching of writing. This chapter will draw on and extend this earlier work.

7. Pedagogic models of writing

7.1 In discussions of the teaching of writing, it has been common to state that, unlike speaking, writing is not an innate skill or capacity (Halliday MAK, 1989 and Grabe W & R Kaplan 1996). It is a technology that has to be learned. Although there are those like Lytton Strachey who, according to his biographer, "… write very slowly, and in faultless sentences …" (Holroyd M 1971: 868), such individuals are rare – and we would have problems proving that they were *born* that way. For most of us, our induction into writing happens well after we have acquired our mother tongue, and may never happen in the foreign language. One of the problems, therefore, for EFL pedagogy has been in deciding how and when to help learners write effectively and appropriately – and for which communicative purposes.

7.2 The cognitive dimension of writing is of particular importance given its social functions which contrast strongly with those of reading. In the mother tongue the development of effective writing skills enables individuals as writers to position themselves in social relations in very different ways from those who only read. As Foggart comments with regard to the attitudes of earlier conservative educationalists:

> "[for them, it has been] desirable that the majority should read – in that way they can be given instructions and can be educated into a particular ideology; writing, on the other hand, assumes the giving of instructions and the *formation* of views about society." Foggart 1993:6)

7.3 Thus the ability to write – and to write effectively (even though it is not at the Lytton Strachey level) – can have a profound influence on the way in which individuals relate to the world around them, and act upon that world. Kress and Stubbs both refer to this change of positioning and potential:

> "Command of writing gives access to certain cognitive, conceptual, social and political arenas. The person who commands both the forms of writing and speech is therefore constructed in an entirely different way from the person who commands the forms of speech alone." (Kress G 1989:46)

> "Written language makes a radical difference to the complexity of organisation that humans can manage, since it changes the relation between memory and classification, and it allows many forms of referencing, cataloguing, indexing, recording and transmitting information […..] The mere fact that something is written conveys its own message, for example of permanence and authority. Certain people write, and certain kinds of things get written." (Stubbs M 1987:20-21)

As the role of English in the world changes a capacity to write in English as well as in the mother tongue is becoming more important, and not only for elites (Graddol D & S Goodman 1996). Although learners may be able to take on a full range of literacy roles in a *local* language, if they wish to enter a global community, they also need a capacity to write in one of the *global* languages.

7.4 Given the social importance of writing, it is surprising that, up until recently, it appears to have received relatively little attention from educational researchers. Kress has commented on the:

> "startlingly massive discrepancy between the amount of work which has been done on reading, compared to the work on writing. The number of books on the learning of reading is vast; by contrast there are few books on the learning of writing." (Kress G 1993:3).

He ascribes part of the reason for this discrepancy to the dominance through the 1960s and 1970s of theories of language which:

> "focus on structures below the sentence, on decontextualised sentences, on meaning as inherent in the individual linguistic item, on reading as a decoding skill ….". (Kress G 1993:3).

7.5 In such an environment, writing was viewed as a secondary problem, easy to crack once the problem of reading had been solved. In addition, Kress considers that the decoding / encoding metaphor itself has had a negative influence on the development of research in writing, predisposing workers in the field to consider meaning as existing independently of language and, more problematically, creating a false view of the 'code' itself as something empty and neutral (Kress G 1993:4-5).

7.6 In writing pedagogy two major strands of research have, however, been elaborated, one often referred to under the broad heading of *process approach*, the other as the *genre* approach.

8. Process

8.1 The first of these strands (and the one which has had the greatest impact to date on EFL instruction) has its origins in composition studies in the US (Grabe W & R Kaplan 1996:18-23), and has focused on writing processes. The need for this focus arose from a reaction (largely motivated by a change in the kinds of student entering higher education in the US after 1960) against a traditional approach to writing instruction which depended on:

- the three or five-paragraph model
- simplistic assumptions about the organisation and ordering of information
- the typical one-draft writing assignment
- the assumption that each student should be working alone, or only with the instructor on summative feedback

- reliance on grammar/usage handbooks and lectures
- the linear composing model based on outlining, writing and editing
  (Grabe W & R Kaplan 1996:86)

Although it is possible to argue that models developed in response to specific conditions in the US have little relevance in other settings, "process writing" has had such a strong influence on EFL writing teaching (White R & V Arndt 1990) that I feel it is appropriate to present the following summary of its origins and impact.

8.2 Using methods established in cognitive linguistics (Emig J 1983, Graves D 1984), an influential body of research in composition processes was published in the early 1980s, largely based on protocol studies of expert and apprentice writers (Hayes J & L Flower 1983). This led to the elaboration of a model of the composing processes of the writer which has had a major impact on writing instruction in the English speaking world. In the Flowers and Hayes model (Figure 2), composing is seen as having three major components, *the composing processor*, *the writer's long term memory*, and *the task environment*. The composing processor is described as having three operational processes – *planning*, *translating* and *reviewing* which are controlled by a *monitor*. The task environment and the writer's long term memory provide resources and stimulus which are drawn on by the composing processor. Here, ideas in the planner are turned into language on the page by the interpreter and then reviewed.

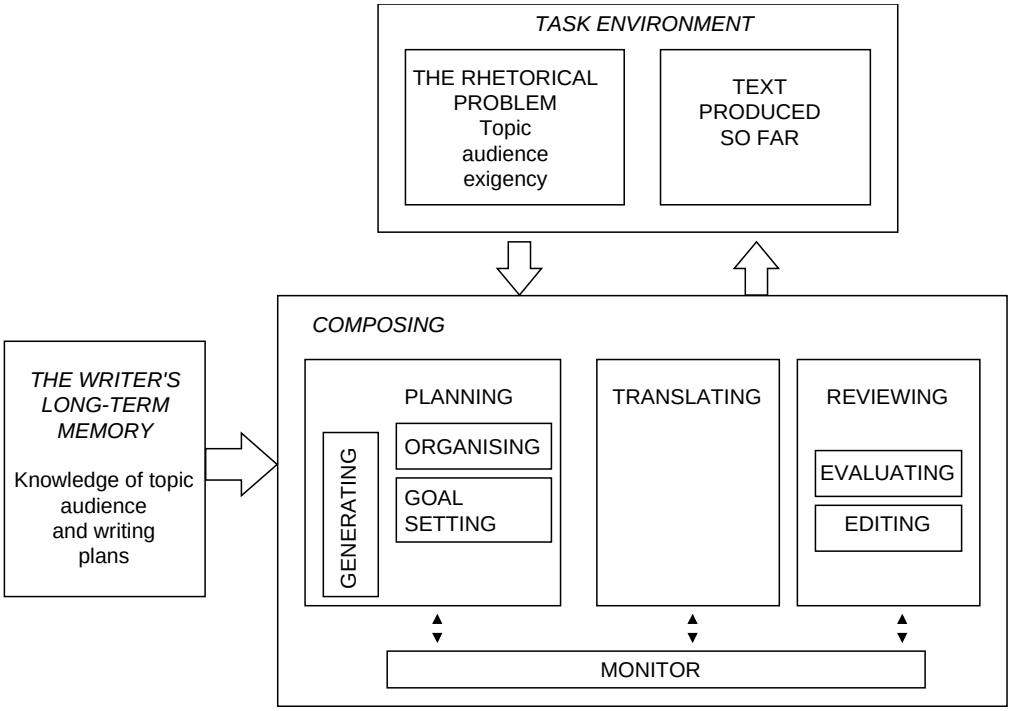


Figure 2 - Flowers & Hayes writing process model (Grabe W & R Kaplan 1996:92)

8.3 Work in ESL and EFL writing instruction (Hedge T 1988, Raimes 1993, White R & V Arndt 1991, Zamel V 1983) has drawn on this model and elaborated a persuasive view of composition as a recursive process which has superseded earlier assumptions about the linear nature of composition. During the 1980s and early 1990s this view came to dominate instructional materials in second and foreign language teaching (Raimes A 1993, Tribble C 1997a) and, when allied with a desire to respect the authenticity of the individual writer's voice (Faigley L 1986, Mayer J 1990, Spack R & C Sadow 1983), provided a powerful agenda for those with an interest in challenging received views of the nature of (in particular) academic writing.

8.4 Later developments in process research have led to refinements of this model, notably in the work of Bereiter and Scardemalia (1987, 1993) and the distinction they make between writing as *knowledge-telling* and writing as *knowledge-transforming*. This development of theory has made it easier to elaborate "explicit hypotheses relating audience and genre differences to

writing task difficulties." (Grabe W & R Kaplan 1996:125). In like manner, drawing on work in critical language studies (Fairclough 1989, 1992, Cooper M 1989) Clark and Ivanić (1997) have developed their own version of the Flower and Hayes model - something which they call an "alternative representation of the process of writing as a social practice" (Figure 3).

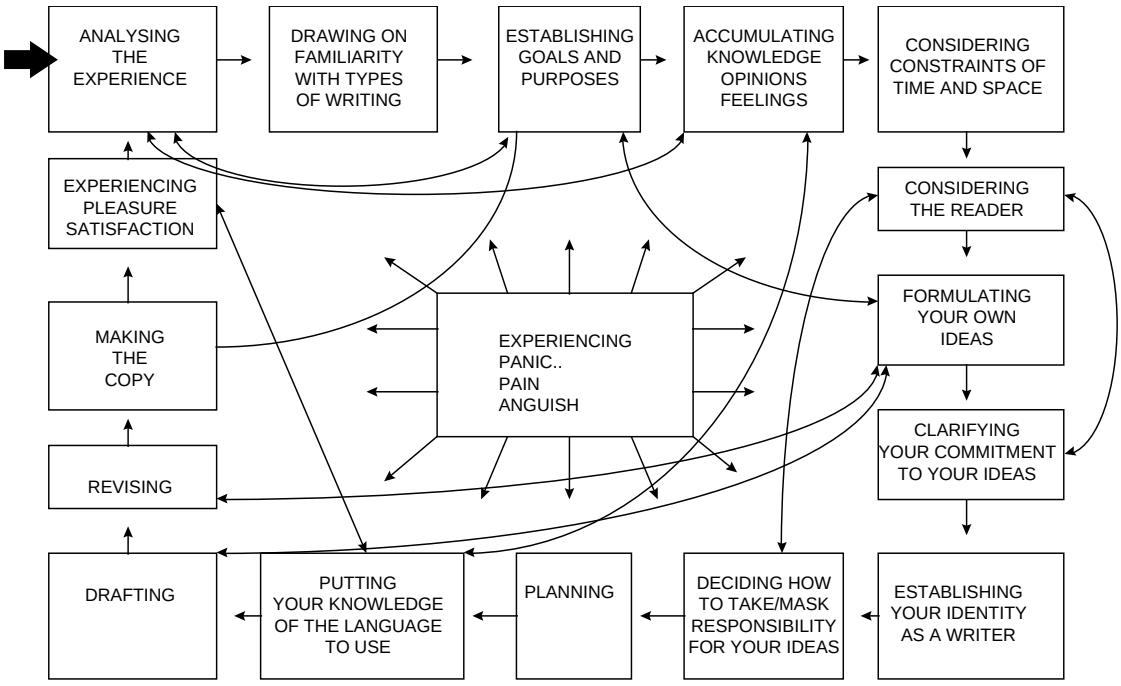


Figure 3 - An alternative representation of the process of writing as a social practice (Clark & Ivanić 1997:98)

8.5 Clark and Ivanić are not attempting to develop a *model* – something which purports to have generative potential (such as that of Flowers and Hayes):

> "We deliberately avoid using the term 'model' for this view of what is involved in writing, because it suggests a fixed, predetermined and hence prescriptive route through the process and does not allow for differences in practices." (Clark R and R Ivanić 1997:94)

The "representation" does, however, constitute both an interesting pedagogic tool and a view on writing processes and concerns in higher educational settings. By presenting the components of composition in a largely non-sequential way (though there is a degree of idealisation in the sequence following 'planning') and showing the recursive inter-relations between these

boxes, Clark and Ivanić catch many of the things that writers need to do, and, importantly in their perspective, also situate these activities in a real world:

> "… we think it is crucial to include and emphasise the socio-political dimensions to the writing process … as expressed in particular in the components 'clarifying your commitment to your ideas', 'establishing your socio-political identity' and 'deciding how to take responsibility'."

8.6 Grabe and Kaplan attempt to capture many of the complex interrelations implied by these various interpretations of writing processes in their model of writing as communicative language use (Figure 4). They argue that:

> "… such a model, or some similar model, provides a way to integrate the three major concerns for a theory of writing: a writer's cognitive processes, the linguistic and textual resources which instantiate the writing task, and the contextual factors which strongly shape the nature of writing." (Grabe W & R Kaplan 1996:229)

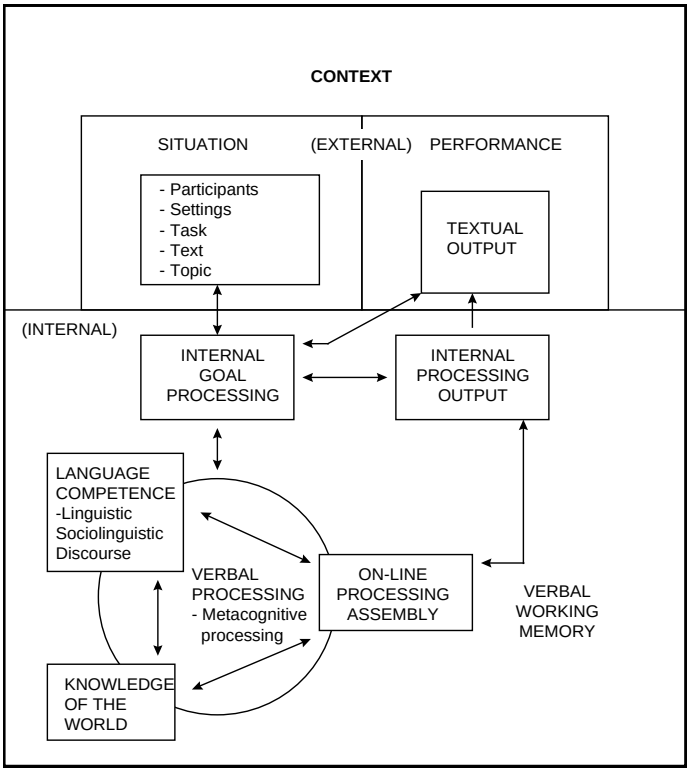


Figure 4 - Writing as communicative language use (Grabe W & R Kaplan 1996:226)

8.7 Such views of the writing process, explicitly connecting as they do the social dimensions of text with the processes of text production, imply a shift from the

strong writer-centredness of some of the more extreme exponents of "process" approaches. Writing begins to be seen as a socially situated activity in which the writer as composer must contend with other texts and other epistemologies as they establish their own voice and identity in their writing

9. Genre

9.1 Grabe and Kaplan's account of writing as communicative language use provides a handy bridge to the second area where research into writing pedagogy has taken place – work undertaken by those who are sometimes described as 'social-constructivist' (Coe RM 1985, Johns AM 1990). Key to this approach to writing are the notions of *discourse community* and *genre*.

9.2 The idea of a discourse community developed from work in social anthropology and social constructivism (Kuhn TS 1970, Fish S 1980, Foucault M 1972, Geertz C 1983) and has been taken up by researchers with an interest in written communication. Swales considers a discourse community to be an institution which depends on texts for its existence, and asserts that it has six key characteristics:

> *A discourse community has a broadly agreed set of common public goals …*
>
> *A discourse community has mechanisms for intercommunication among its members …*
>
> *A discourse community uses its participatory mechanisms primarily to provide information and feedback …*
>
> *A discourse community utilises and hence possesses one or more genres in the communicative furtherance of its aims …*
>
> *In addition to owning genres, a discourse community has acquired some specific texts …*
>
> *A discourse community has a threshold level of members with a suitable degree of relevant content and discoursal expertise …* (Swales 1990:24-27)

9.3 This view of the nature and purpose of discourse communities (and in particular the implications of *Characteristic 6* above) is felt to be contentious by teachers of writing who question the right of institutions to make linguistic demands of aspiring members (Bizzell P 1987, Clark R & R Ivanić 1997). It represents, however, a view which has been taken on and elaborated by researchers with an interest in academic and professional communication (Bazerman C 1994, Swales J 1981, Smart G 1993, Spilka R 1993, Swales J 1990, Swales JM & P Rogers 1995) and accepted by many authors of ESL / EFL instructional materials for English for Academic Purposes and professional communication (Doherty M, L Knapp & S Swift 1987, Hamp-Lyons L & B Heasley 1987, Tribble C 1997, Turk C & J Kirkman 1989, White R & D McGovern 1994). It also fits the kind of organisations that are engaged in the design, tendering, bidding and management of large social and technical development projects. These have "*a broadly agreed set of common public goals*" – often declared in organisational mission statements; they have "*mechanisms for intercommunication*" and "*provide information and feedback*" through intranets, websites and gazettes[4]; and they possess "*one or more genres in the communicative furtherance of its aims*" in this instance, for example, "Invitations to Bid" and "Project Proposals". Not only do they have these common goals and genres, at an organisational level, they also have "a threshold level of members with a suitable degree of relevant content and discoursal expertise" which is established through the management procedures

[4] A visit to homepages such as http://www.undp.org or http://www.dfid.gov.uk makes clear the considerable overlap between these large development agencies in their communication strategies and the tools that they are developing to ensure information flow between the agency, stakeholders and potential suppliers

that control the proposal writing process. (see Chapter 8: *Writing project proposals*)

9.4 If *discourse community* has proved to be a contentious term, *genre* has also given rise to debate and confusion. I propose to use an account elaborated by Hasan (Halliday MAK & R Hasan 1985) as my main working definition:

> "1. A genre is known by the meanings associated with it. In fact the term 'genre' is a short form for the more elaborate phrase 'genre-specific semantic potential'.
> 2. Genre bears a logical relation to CC [Communicative Context] , being its verbal expression. If CC is a class of situation type, then genre is language doing the job appropriate to that class of social happenings.
> 3. Genres can vary in delicacy in the same way as contexts can. But for some given texts to belong to one specific genre, their structure should be some possible realisation of a given GSP [Generic Structure Potential].
> 4. It follows that texts belonging to the same genre can vary in their structure; the one respect in which they cannot vary without consequence to their genre-allocation is the obligatory elements and dispositions of the GSP." (Halliday MAK & R Hasan 1985:108)

9.5 I find this approach to genre particularly useful as it allows for higher or lower levels of genre specificity as a function of the relative specificity of the Communicative Context (CC). Through the breaking down of the CC into *situation types* and the Genre Specific Potential (GSP), Hasan offers an explanation of the way in which some textual realisations of a genre are considered more appropriate than others – i.e. are allowable contributions – and also of the way in which some genres are more narrowly specifiable than others. It also dissolves the potential contradiction between Swales' apparently narrow account of genre[5], and accounts of "protogenres" offered by theorists such as Martin (Martin JR 1989:7). Thus one CC may be highly

[5] "*The rationale behind a genre establishes constraints on allowable contributions in terms of their content, positioning and form* Established members of discourse communities employ genres to realize communicatively the goals of their communities. The shared purposes of a genre are thus recognized – at some level of consciousness – by the established members of the parent discourse community; they may be only partially recognized by apprentice members; and they may be either recognized or not recognized by non-members." (Swales J 1990:53)

specific and have a very narrowly restricted GSP – e.g. Swales' example of the Journal/Newsletter of the Hong Kong Study Circle of stamp collecting enthusiasts (Swales 1990:27); or it might be of a more general nature – e.g. "good-news" and "bad-news" letters in administrative correspondence (Swales 1990:53) with a correspondingly broader, though still identifiable GSP; or it might be so general that it bears no relation to a particular discourse community – e.g. Martin's categories of RECOUNT vs. PROCEDURE / DESCRIPTION vs. REPORT (Martin JR 1989:3-8), although experienced readers will still be able to identify allowable contributions. Thus whether it is specific or general, each CC has associated with it a GSP which can be realised through a (specifiable) range of allowable texts.

9.6 This view provides the basis for future discussion of genre in this thesis. Such a dynamic interaction between contexts and texts is partially summarised in Swales' model of the schematic structure of genres (Figure 5):

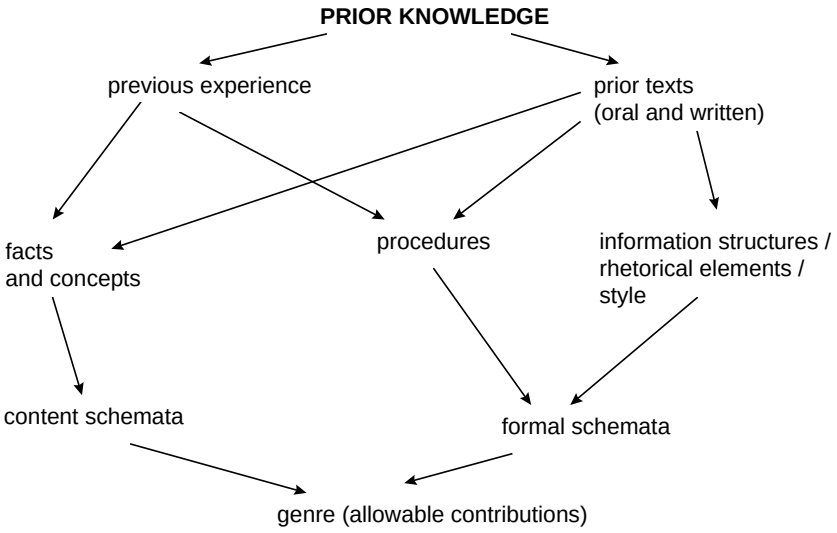


Figure 5 - Swales 1990:84

9.7 Here, if one works back from an individual instance of an allowable contribution within a genre – "a possible realisation of a given GSP" (Halliday

MAK & R Hasan 1985:108) – one can see its formation in terms of the results of a series of decisions that are informed by:

- prior knowledge of the world which informs content schemata
- knowledge of prior texts
- knowledge of procedures (how one goes about the writing in a particular genre)

The one limitation of this particular model is that it fails to take into account the knowledge of *social* relations which is also required if one is to ensure the acceptability of a contribution, or take the risk of knowingly flouting genre expectations in order to achieve another rhetorical purpose. Nevertheless, it is a useful visual representation of the kind of knowledge base which writers draw on during the development of a text.

9.8 Whatever the problems associated with the notion of genre, it has been widely adopted (Benson JD & WS Greaves 1980, Bhatia VJ 1993, Halliday MAK 1978, Martin JR 1985, Swales J 1990, Tribble C 1997a) and has provided the basis for an alternative, socially situated view of writing. This has become increasingly influential in recent years, and is now referred to as a *genre approach* to writing (Bamforth R 1993, Cope B & M Kalantzis 1993, Gee S 1997). Curricula and instructional materials originating from a genre approach typically consider text as simultaneously constructing and resulting from social processes. Such a view is exemplified by Smart's account of genre development in the Canadian National Bank, where he presents an account of situated writing in which:

> "… four major contextual influences – the executives' issue-resolving mandate, several mental constructs that inform their thinking, a tacit understanding between the executives and the research staff about the delegation of problem-solving responsibility, and the intertextual resonance of prior written discourse – shape the executives' reading practices and consequent expectations. These expectations, conveyed and given authority through the institutional hierarchy, in turn exert a compelling force on the composing processes of the staff as they collaborate in preparing documents for the executives' use, thereby giving rise to a body of typified discourse." (Smart G 1993:125)

9.9 In the banking context Smart was studying the texts which were used in policy formation within an organisation which depended on documents for this process. In such a setting the interaction is not only between a group of powerful readers and the bank's technical and research staff. It is also an interaction between current texts and prior texts, and between the problem solving strategies of some individuals and the solution needs of others. For the purposes of this discussion, the example is useful because it underscores the socially motivated and purposeful nature of writing in organisations. So, while the writing that I will consider in this thesis will remain an activity that involves the "combining of structural sentence units into a more-or-less unique, cohesive and coherent larger structure." (Grabe W & R Kaplan 1996:4), I will also consider *writing* as a socially motivated activity – an activity in which someone is trying to do something which requires them to write (rather than to speak or to gesture).

10. Approaching writing

10.1 The literature on writing might give the impression that there has been a heated debate going on in the teaching profession between two warring camps – one waving the banner of process and the other that of genre. However, a

consideration of mainstream ELT writing materials paints a less divided picture, indeed a picture in which theoretical debate rarely seems to raise its head. Apart from a few recent specialised textbooks such as *Academic writing for graduate students* (Swales J & CB Feak 1994) we will see that the materials used by many teachers of English as a foreign language have been developed much more pragmatically than the research literature might lead one to believe. Rather than depending on a single theoretical position, authors and publishers tend to draw on aspects of the theory that look as if they will be useful. As for teachers, they are less interested in theory than in having practical materials which will help them teach groups that more often than not have a low level of motivation to develop writing skills.

10.2 The relatively low level of heat in the professional debate amongst teachers of writing in EFL settings arises for several reasons: Firstly, much of the original research and the professional motivation for the elaboration of the "process writing" approach has arisen from the needs of teachers working in further and higher education in the USA. As "freshman comp" and the tradition of explicit instruction in rhetoric are not common outside the US, the more partisan aspects of the debate have had little significance for EFL teachers. Thus most EFL teachers of English with an interest in writing have accepted as useful a book like *Process Writing* (White R & V Arndt 1991) without then feeling a need to take sides in a debate. Secondly, English language teachers working in universities and those working in EFL language schools have strongly contrasting concerns. The former have a professional interest in the needs of advanced writers and a concern to publish theoretical articles; in general, the professional concerns of the latter are focused on the needs of

general students and practical matters of classroom teaching. An additional reason why the teaching of writing sometimes receives less attention in EFL classrooms may be that many English language teachers (mother-tongue or speakers of other languages) are often intimidated by what they perceive as one of the more demanding aspects of foreign language teaching – most of us are not confident writers. It is, therefore, often a low priority for learners, and something that receives little attention in general teaching and little interest from publishers who do not see large sales for writing skills books.

10.3 In this review of practical classroom materials, I shall follow an approach developed in Tribble 1997 where I considered teaching materials that had been developed for three classes of learner:

- students learning English for business or professional purposes
- those with academic or study needs
- those preparing for international EFL examinations.

A review of the publishers' catalogues confirms that this still covers the majority of publications related to writing instruction. The only other area where writing becomes a focus is in general English language teaching, but in this setting it is more often the case that students are "writing to learn" rather than "learning to write" (Tribble C 1997:72) – i.e. writing without composing, where the pen is put to paper as a handmaid to learning, not as an end in itself.

10.4 In the rest of this chapter I shall, therefore, discuss examples of currently available published teaching materials that have been prepared in order to help these three different groups of learners. During this review I shall draw

attention to a series of issues in writing instruction which I feel an adequate pedagogy of writing should address. These will be flagged in the text in separate **ISSUE** boxes, and will provide the basis for recommendations for the future development writing of pedagogy that I shall make in Chapter 9: *Helping learners write difficult texts.*

11. English for business / professional communication

11.1 Two main groups of student are catered for in materials marketed as "English for Business" or "English for professional communication": qualified business practitioners, and students preparing to become business professionals. The consensus amongst materials writers appears to be that the two groups have very similar needs, that is:

- a need for information on business practice in the international English speaking business community
- a need for opportunities to develop a command of the target language in a motivating context.

11.2 Working on this premise, authors use the behaviour, concepts, activities and concerns of the international business world as a context for the development of language skills - including writing skills. They also give prominence to the importance of cultural factors in professional communication (whether this is spoken or written) and to the learner's need to become aware of the ways in which speech and writing styles can have a major impact on business success.

11.3 Such an emphasis produces course materials with contents pages such as those reproduced in Table 2.

BBC Business English	International Business English
The factory and the firm 1 A tour of the factory 2 The firm **People and Jobs** 3 Appointments and applications 4 Selecting staff 5 Going on a training course **The daily routine** 6 Office management 7 Paperwork 8 Data processing etc..... (Owen R 1992 :3)	1 Face to face 2 Letters, telexes and memos 3 On the phone 4 The place of work 5 Import and export 6 Money matters etc ... (Jones L & Alexander R 1989:iii)

Table 2 - Business English materials (Contents)

11.4 Typically, teaching units in such materials are contextualised either by the creation of an imaginary company and its business activities (as in *BBC Business English*) or through a set of business themes similar to those found on technical and vocational business training courses (*International Business English*). These are then used:

- to create a communicative context which requires some form of response from the learner and then

- to present language skills and knowledge that are relevant to that context and which, it is hoped, will enable the learner to perform effectively in the present (simulated) setting and in future (analogous) ones.

11.5 When this approach is realised in teaching materials two major problems become apparent. The first is that the materials frequently fail to offer students a satisfactory analytic framework on which to base judgements they are asked to make about example texts. The second is that they do not offer learners sufficient data on which to make reliable judgements. The result of these limitations is that all too often students are invited to imitate a model without having a clear understanding of why the model is (or is not)

appropriate. Such problems are exemplified in these two extracts from currently available teaching materials.

11.6 In *BBC Business English* (Owen R 1992) contextualisation for a teaching unit is provided through small scenarios such as the one below:

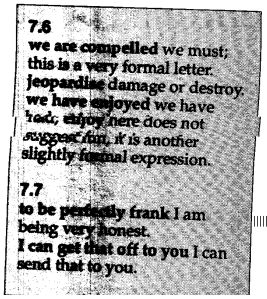
> We are still at the offices of Bookmart Publishing Services. In this unit we see how to deal with problems of credit control; how to conduct an employee assessment interview; and how to investigate and report an accident.
> **CHARACTERS**
> George Harvey
> Terry Cabe
> Frank Penny
> Mr Harris owns a bookshop in York
> Mr Martinu is General Manager at Martinu Books
> Andy Brumshaw is the warehouse foreman at Bookmart
> Alice Perkins is an office cleaner at Bookmart
> *All these characters are British*
> (Owen R 1992:3)

Once the scene has been set, the authors select communicative activities that would most typically be associated with that context. Thus the "employee assessment interview" is used as a starting point for oral skills development, while accident reporting and credit control are used as contexts for writing skills development.

11.7 The Credit Control subsection began with the activity reproduced in Table 3, and was completed with a task in which students were asked to write an equivalent letter on the basis of a set of parallel information.

Document Study: A formal letter

George Harvey had to write a letter to another bookshop, Martinu Books. They have not paid anything since the end of March. Read through the letter. Notice that it is very formal.



7.6
we are compelled *we must;* **this is a very** formal letter.
jeopardise damage or destroy.
we have enjoyed we have had, enjoy here does not suggest fun, it is another slightly formal expression.

7.7
to be perfectly frank I am being very honest.
I can get that off to you I can send that to you.

BOOKMART
PUBLISHING
SERVICES LTD

Mr G F Martinu
General Manager
Martinu Books
Eastgate
York YO1 1DX

13 June 1991

Dear Mr Martinu

Our statements dated 31 March, 30 April and 31 May 1991

As we have received no reply to our letters of 4 May and 30 May, *we are compelled* to draw your attention once more to your non-payment of the sums owed to us. The total amount outstanding is now £15,872.87.

We regret that we cannot continue to allow credit terms as long as this debt remains uncleared. Until further notice, therefore, goods will be supplied to you only on receipt of cash against a pro forma invoice.

I shall be obliged if you will contact me personally within the next seven days to discuss means by which the amount outstanding can be cleared. We are anxious to avoid any action which might *jeopardise* the good business relationship *we have enjoyed* in the past.

Yours sincerely

G A Harvey
Sales Manager

Table 3 - Business English Materials (Owen R, 1992:59)

11.8 The problem with this activity (and similar examples in other publications) is that it offers too much and too little. The example letter is a potentially rich source of language information, but no guidance is offered (here or elsewhere in the book) on *how* to analyse such a text in order to learn from it. The injunction to "Notice that it is very formal", in combination with the "Language Study" box, is supposed to guide learners towards an understanding of the linguistic features which signal differing degrees of polite distance. It is unlikely, however, that such an injunction would enable a learner to obtain sufficient information to go on to elaborate an equivalent text independently.

Issue 1 The materials reviewed here do not offer an account of language in use that will help students to carry out a useful analysis of data samples. As a result, text data may have limited value, and confuse rather than assist learners.

11.9 When compared with the first example, Doherty M, L Knapp & S Swift (1987) provide an analysis which might allow learners to move towards independent composition in a new communicative context. They do this by:

a) giving an explicit discussion of the power relations that exist between readers and writers in business settings and the way in which language use changes as a result of different reader-writer relationships – e.g.:

UNIT 2

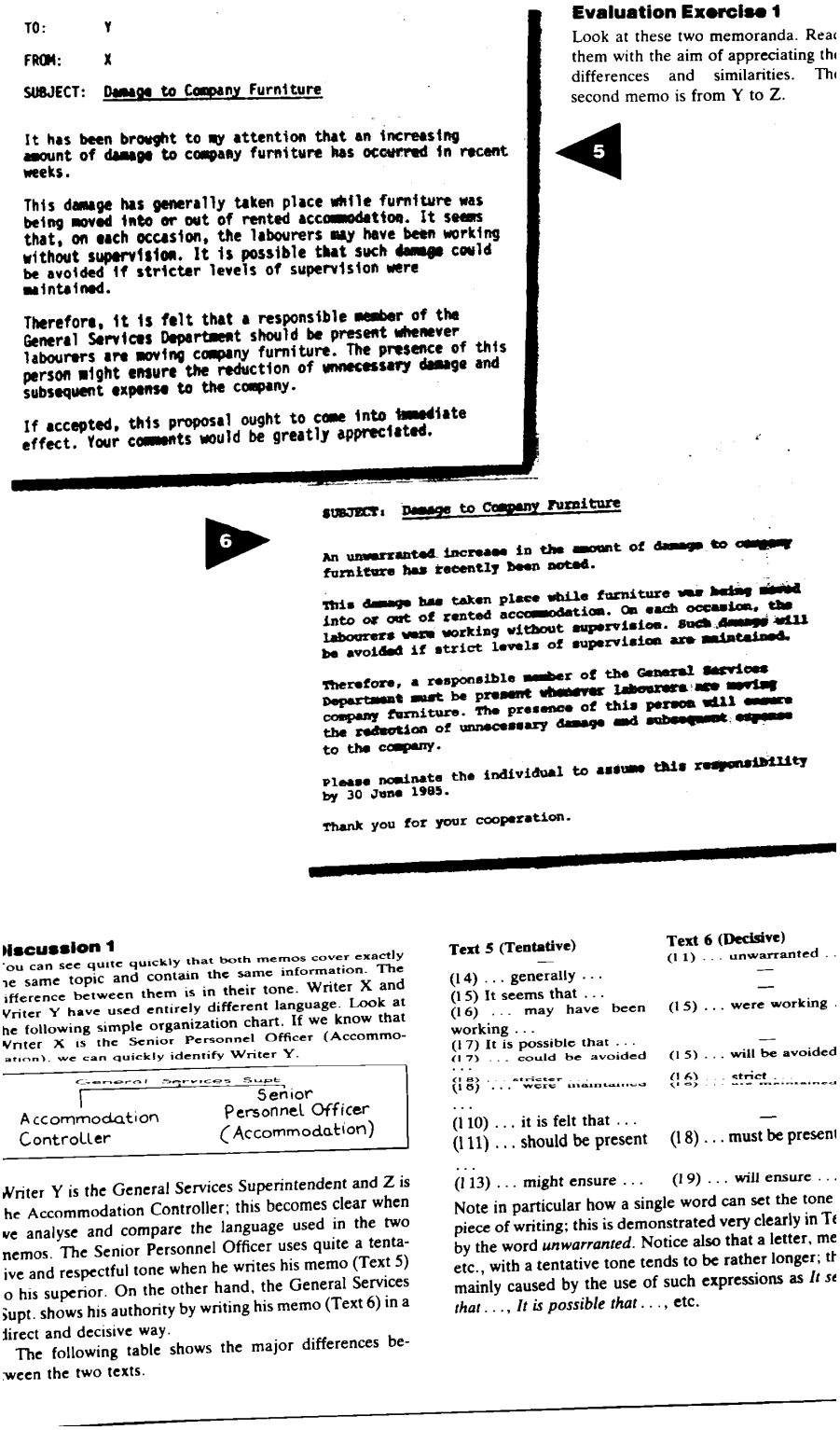
Achieving a Tone to Suit the Business Reader

Good writers in the business world know how to choose their words. The communication must achieve its goal or time and energy are wasted. If you wish to be a good writer, good grammar and spelling are necessary but not sufficient. Your words must be chosen with care and tested for their suitability for each particular writing task. Good writers can write **upwards** and **downwards** without offending their readers. They use the most suitable expressions for their position and particular business.

The notion of tone is very important to the business writer. Since readers can be insulted not by the message but by the manner of expression, an otherwise perfect memo or letter could quite possibly cause a breakdown in communication. It is therefore important for you to be able to analyse your relationship with your reader(s) and then to choose the appropriate language and tone for your message.

b) then offering an opportunity to compare expert performances from differently positioned members of the same organisation, along with

c) a clear analysis of some of the key linguistic features that the writers have selected in order to achieve their contrasting purposes (see example material in Figure 6.)



Evaluation Exercise 1

Look at these two memoranda. Rea
them with the aim of appreciating th
differences and similarities. Th
second memo is from Y to Z.

5

TO: Y

FROM: X

SUBJECT: Damage to Company Furniture

It has been brought to my attention that an increasing amount of damage to company furniture has occurred in recent weeks.

This damage has generally taken place while furniture was being moved into or out of rented accommodation. It seems that, on each occasion, the labourers may have been working without supervision. It is possible that such damage could be avoided if stricter levels of supervision were maintained.

Therefore, it is felt that a responsible member of the General Services Department should be present whenever labourers are moving company furniture. The presence of this person might ensure the reduction of unnecessary damage and subsequent expense to the company.

If accepted, this proposal ought to come into immediate effect. Your comments would be greatly appreciated.

6

SUBJECT: Damage to Company Furniture

An unwarranted increase in the amount of damage to company furniture has recently been noted.

This damage has taken place while furniture was being moved into or out of rented accommodation. On each occasion, the labourers were working without supervision. Such damage will be avoided if strict levels of supervision are maintained.

Therefore, a responsible member of the General Services Department must be present whenever labourers are moving company furniture. The presence of this person will ensure the reduction of unnecessary damage and subsequent expense to the company.

Please nominate the individual to assume this responsibility by 30 June 1985.

Thank you for your cooperation.

Discussion 1

You can see quite quickly that both memos cover exactly the same topic and contain the same information. The difference between them is in their tone. Writer X and Writer Y have used entirely different language. Look at the following simple organization chart. If we know that Writer X is the Senior Personnel Officer (Accommodation), we can quickly identify Writer Y.

Writer Y is the General Services Superintendent and Z is the Accommodation Controller; this becomes clear when we analyse and compare the language used in the two memos. The Senior Personnel Officer uses quite a tentative and respectful tone when he writes his memo (Text 5) to his superior. On the other hand, the General Services Supt. shows his authority by writing his memo (Text 6) in a direct and decisive way.

The following table shows the major differences between the two texts.

Text 5 (Tentative)	Text 6 (Decisive)
(l 1) . . . unwarranted . .	
(l 4) . . . generally . . .	—
(l 5) It seems that . . .	—
(l 6) . . . may have been working . . .	(l 5) . . . were working .
(l 7) It is possible that . . .	—
(l 7) . . . could be avoided . . .	(l 5) . . . will be avoided
(l 8) . . . stricter . . . (l 8) . . . were maintained . . .	(l 6) . . . strict . . .
(l 10) . . . it is felt that . . .	—
(l 11) . . . should be present . . .	(l 8) . . . must be present
(l 13) . . . might ensure . . .	(l 9) . . . will ensure . . .

Note in particular how a single word can set the tone piece of writing; this is demonstrated very clearly in Te by the word *unwarranted*. Notice also that a letter, me etc., with a tentative tone tends to be rather longer; th mainly caused by the use of such expressions as *It se that . . .*, *It is possible that . . .*, etc.

Figure 6 - Doherty Knapp & Swift 1987:13-15

11.10 Although the account of the topic is not exhaustive, it is *pedagogically* useful, and sufficient for learners to draw on in future writing tasks. In this sense the materials are more likely to achieve their purpose – to enable apprentice

writers to compose texts that are appropriate to their professional needs (a purpose that will be shared by the authors of the first example). Having worked through these materials learners should have a better understanding of how choices from the grammatical and lexical systems of the language have an impact on the way in which writers and readers interact, and to make use of this understanding in their own writing..

11.11 The problem with the materials (and here they suffer from the same limitations as all currently available published materials) is that they simply do not offer students enough data. A single example – genuine or fabricated – is not a sufficient basis for making decisions about the way in which a text genre is constructed. We will discuss possible solutions to this problem in more detail in the closing chapters, but an example will be useful at this point. Remaining in the area of business communication, consider a letter such as the one reproduced below (Table 4 – xxx indicates information removed for reasons of confidentiality) .

xxx xxx
xxx School of Art and Design
xxx, Falmouth, Cornwall

Dear xxx,
Thank you for your letter enquiring about the possibility of xxx xxx funding for a xxx citizen, xxx xxx, to enrol in one of your BA courses. I am afraid that our Scholarships programme for the xxx xxx is currently limited to support for post graduate studies, so we cannot offer any assistance in this matter. Unfortunately I am also unable to suggest to you any alternative sources of funding. The Open Society Foundation in xxx sometimes offers scholarships for studies in the arts, though I know that they have many calls on their resources and I would not therefore like to hold out too much hope.

I am sorry not to be able to be more helpful.

With best wishes,

Table 4 - Letter of rejection (Corpus of administrative correspondence – personally held)

11.12 If learners were presented with such a letter, a certain number of learning points might be made – e.g. :

- **Genre** – the letter is one of rejection / refusal (produced by a grant awarding organisation)

- **Discourse** – in this particular (British) culture letters of rejection:
  - open by acknowledging the recipient (Thank you for your letter …)
  - offer face-preserving reasons for the rejection (our Scholarships programme is currently limited to support …)
  - offer alternatives (The Open Society Foundation sometimes offers…)

- **Language –** a number of key words and phrases are typically used in such letters:
  I am afraid that … / Unfortunately I am unable … / I am sorry not to be able etc.

11.13 Unfortunately, some of these conclusions could be specifically misleading – or at least miss a main point. If a large collection of letters from the same organisation is considered (this time contrasting letters of OFFER / ACCEPTANCE and letters of REJECTION), systematic patterns of use can be identified which call into question the conclusions based on the single instance. The first thing that you notice is that there is no simple opposition between *pleased* and *afraid* – the one identifying OFFER and the other REJECTION. In fact, *afraid* is a poor identifier of REJECTIONS and, in the case of the data under review, was much more common in letters which were

apologising for a failure. While *pleased* was a strong indicator of OFFER / ACCEPTANCE, the best identifier of REJECTIONS was *unfortunately*.

11.14 Of more importance than this, the data for these two words (see Table 5) gives a strong indication of a major contrast between OFFER / ACCEPTANCE and REJECTION which would not have been noticed in a study of the first example. The contrast lies in the fact that in OFFER / ACCEPTANCE the actor is usually the first person singular (13/19) and in REJECTION it is the first person plural (19/21). On the basis of observation of a large selection of texts from a single organisation, it becomes clear that a) when offers are made, the writer wishes to associate him/herself with the act, and that b) when a rejection is required, there is a tendency in this writing culture to distance oneself from the act, implying that responsibility rests with an institution rather than the individual.

```
idered your application and I am pleased to inform you that the Xxxxxx Xxxxx
sting and valuable one, and I am pleased to learn of the contribution from t
t the Edinburgh Festival.   I am pleased to be able to offer you a contribut
Dear Professor Xxxxxxxxx,   I am pleased to be able to assist with the fundi
Festival on behalf of xxxX. I am pleased to tell you that we will be able to
my of Art later this month. I am pleased to tell you that we will be able to
 their performance in Oslo. I am pleased to tell you that the Xxxxxx Xxxxxxx
 our scholarship programme. I am pleased to be able to tell you that you hav
 our scholarship programme. I am pleased to be able to tell you that you hav
 Sakala Street, Tallinn.    I am pleased to know that you are already in tou
 of your time in the UK.    I am pleased to be able to tell you that the Dir
to the interview last week. I am pleased to tell you that we have sent your
is award earlier this year. I am pleased to be able to tell you that you hav
ch a successful partnership, and pleased to be able to tell you that the Xxx
ing her visit.  We would also be pleased to host a meeting for her to talk a
 Professor Xxxxxx,   You will be pleased to know that we have been able to p
uage Unit - Latvia   You will be pleased to hear that we have now prepared a
member of your Office I would be pleased to attend such a meeting along with
l it would be useful, I would be pleased to talk to you in more detail on my
 are in Tallinn we would be very pleased if you could run a couple of sessio
 can hope to satisfy, and we are unfortunately unable to offer you any finan
fore writing to say that we are, unfortunately, unable to offer you any assi
fore writing to say that we are, unfortunately, unable to offer you any assi
rtment in London about this but, unfortunately, they are unable to help this
  Xxxxxx, the Xxxxxx Xxxxxxxx is unfortunately unable to help you with fundi
nd offer of support. We are not, unfortunately, in a position to establish a
n interesting one, but we do not unfortunately have sufficient budget availa
5, and I am writing to say that, unfortunately, we will not be able to offer
d I am writing to tell you that, unfortunately, we have not been able to of
lf of Xxxxxx Veterinary Academy. Unfortunately, we are unable to help you wi
ve you an application form.   c. Unfortunately, we are unable to help you wi
id, to follow up our phone call. Unfortunately the Xxxxxx Xxxxxxxx will not
g Sustainable Rural Development. Unfortunately, the timing of this course ma
isit the University of Xxxxxxxx. Unfortunately we have far more demands on o
onomies of Transition in Xxxxxx. Unfortunately, as you will realise, we have
```

hraseology to be held in Leeds.	Unfortunately we are not able to contribute
r any assistance in this matter.	Unfortunately I am also unable to suggest t
n number of Baltic participants.	Unfortunately, we are unable to enter into
Xxxxxxxx Fellowship programme.	Unfortunately we have already completed rec
nvolved in the Health project.	Unfortunately our resources are limited and
tive of the aims of the project.	Unfortunately, however, I do not currently
sible to waive this requirement.	Unfortunately, however, we never received t
ndustry in the Baltic States.	Unfortunately we are unable to offer you an
he EDEN conference in Tallinn.	Unfortunately we have far more demands on o
urse to be delivered in Vilnius.	Unfortunately we are unable to provide any

Table 5 - pleased / unfortunately

11.15 Thus, while it might be possible to identify some of the important features of the texts that are required by a communicative context by analysing a single instance, such an analysis can be significantly misleading. A study of the example letter given above would not have revealed the important information about the different roles speakers take on in contrasting genres that can be gained from a study of REJECTION and OFFER / ACCEPTANCE in a corpus of administrative correspondence[6].

Issue 2 Single instances of language in use do not provide a sufficient basis for conclusions about the linguistic specification of a genre. In order to make such generalisations, multiple examples are needed – ideally from several comparable sources

12. English for academic or study purposes

12.1 Business English courses and English for academic purposes (EAP) courses differ to a surprising extent and present contrasting pedagogic problems. The former are designed for practitioners or those preparing to enter an identifiable professional field, and who are frequently unfamiliar with the dominant market oriented culture of the global business community. Writers of "business English" courses, therefore, appear to feel justified in offering a degree of content and context knowledge as a way of framing the language

system and process knowledge that they provide. Most courses in EAP, by contrast, have to cater for the needs of a group whose needs are a great deal more varied..

12.2 First it is impossible to assume that learners will share a common content knowledge (they may be taking courses ranging from applied linguistics to zoology), still less that English teachers will know more about their students' subjects than the students do. Context knowledge can be equally problematic. Course participants might be on vocational, undergraduate or post-graduate courses; they might be foreign students already in English speaking countries, students preparing to enter such institutions, students who will take some courses through the medium of English while living in their own countries – and, given the current rapid rate of change in academic institutions and texts, their teachers' knowledge of these contexts may be extremely out-of-date. This presents materials writers and publishers with major problems when it comes to designing and marketing EAP courses.

12.3 Broadly speaking, authors have found two solutions to this dilemma. One, common to all published EAP materials, has been to avoid content knowledge and to give a strong emphasis to language system and writing process knowledge, and, to a lesser extent, context knowledge. The other – and it is here that EAP books differ most strongly one from the other – has been to opt for one of two contrasting views of what makes up the context of academic

[6] This is a collection of administrative correspondence collected from colleagues when I was working in such an organisation. For reasons of confidentiality, they are not publicly available. I would stress here that if learners were working with this collection they would be advised that any conclusions they reached on the basis of this data should be viewed as interim and would require testing against equivalent data sets from other sources.

writing and how this context and the language use it generates can best be described.

12.4 The first view of the context of academic writing is based on a long-standing tradition in the English speaking world, and in many European countries. This view assumes a common intellectual framework for *all* academic discourse - a common academic context. In this tradition the modes of classical rhetoric are taken as the starting point for instruction in academic writing skills, and it is assumed that students have a primary need to gain a mastery of these rhetorical modes if they are to become competent writers in their chosen disciplines. The list below provides an example of the content that is commonly taught in writing courses of this kind.

- **Exposition**
  - Examples
  - Process
  - Cause and effect
  - Comparison and contrast
  - Definition
  - Division and classification
- **Description**
- **Narration**
- **Argumentation and persuasion**

(Langan J, 1993 p 113)

Table 6 - EAP Programme

12.5 In teaching materials that are built around these categories, it is common for these rhetorical modes to be chapter headings (as in Table 6). Examples of language use and writing exercises are then provided to help students to express these rhetorical modes in appropriate language. Once learners are considered able to use the exponents at sentence level, they move on to paragraph and text length exercises. I shall call this approach to teaching academic writing "intellectual / rhetorical".

12.6 The second view takes as its starting point the notion of the DISCOURSE COMMUNITY (Berkenkotter C & TN Huckin 1993, Berkenkotter C & TN

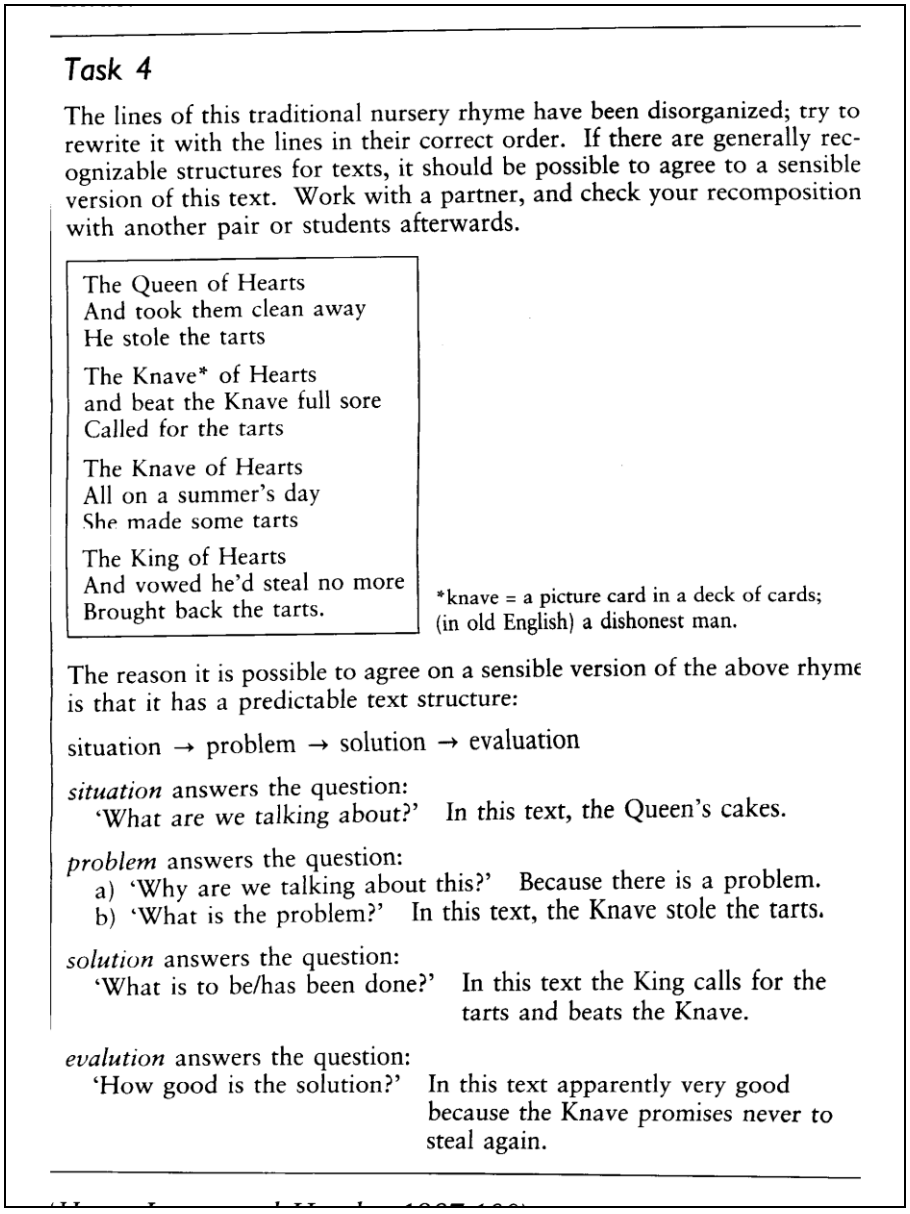
Huckin 1995, Clark R & R Ivanić 1997, Flower L 1994, Freedman A & Medway P 1994, Grabe W & R Kaplan 1996, Miller C 1984, Reither J 1985, Swales J 1990) and identifies not just one but a range of possible academic contexts. Swales describes a discourse community as being made up of writers and readers with "common goals, participatory mechanisms, information exchange, community specific genres, a highly specialised terminology and a high general level of expertise" (Swales J 1990:29). In this account, "common goals" refers to the objectives of the scholars in any community - these can include accounts of replicable experimental procedures, the reporting of new knowledge, the examination of students, and so forth. In the case of writing, the "participatory mechanisms" are the texts that are associated with the genres of a particular discipline. These may be the academic journals produced by specific disciplines, the text books that academics prepare as a means of teaching their subject to their students, or the examinations that students have to write in order to proceed through their courses.

12.7 Teachers who work with this *social* view of the context of academic writing typically ask their students to discover how their own specific discourse communities function and how this affects the way in which members of that community write. In this process of discovery learners will look closely at the relationships that exist between different readers and writers and in doing so their primary data will be the texts themselves. These can be analysed, imitated and challenged, and – as the learner becomes an expert practitioner – may well be transformed. We should note that such an investigative approach will often include a consideration of modes of rhetoric, though not as models

for imitation, but rather as examples of ways of writing that students should be aware of. I call this approach to writing "social / genre".

12.8 In the following paragraphs I will consider examples of published EAP materials from these two perspectives – "intellectual / rhetorical" and "social/genre". I have discussed elsewhere the ways in which several components of an academic writing syllabus – text structure and organisation, argumentation, style etc. – can be delivered (Tribble C, 1997). For the purposes of this chapter I shall restrict myself to a discussion of one of the features most commonly taught in English for Academic Purposes materials – text organisation and argument development – as this usefully exemplifies some of the major contrasts between intellectual / rhetorical and social / genre approaches.

12.9 The two following examples (Figure 7 and Figure 8) are both concerned with problems of text organisation, but one is a fairly extreme example of the rhetorical / intellectual approach to EAP and the other is one of the few published examples of an EAP text book which uses a well elaborated social/genre approach.



Task 4

The lines of this traditional nursery rhyme have been disorganized; try to rewrite it with the lines in their correct order. If there are generally recognizable structures for texts, it should be possible to agree to a sensible version of this text. Work with a partner, and check your recomposition with another pair or students afterwards.

The Queen of Hearts
And took them clean away
He stole the tarts

The Knave* of Hearts
and beat the Knave full sore
Called for the tarts

The Knave of Hearts
All on a summer's day
She made some tarts

The King of Hearts
And vowed he'd steal no more
Brought back the tarts.

*knave = a picture card in a deck of cards; (in old English) a dishonest man.

The reason it is possible to agree on a sensible version of the above rhyme is that it has a predictable text structure:

situation → problem → solution → evaluation

situation answers the question:
'What are we talking about?' In this text, the Queen's cakes.

problem answers the question:
a) 'Why are we talking about this?' Because there is a problem.
b) 'What is the problem?' In this text, the Knave stole the tarts.

solution answers the question:
'What is to be/has been done?' In this text the King calls for the tarts and beats the Knave.

evalution answers the question:
'How good is the solution?' In this text apparently very good because the Knave promises never to steal again.

Figure 7 - Hamp-Lyons L & B Heasley 1987:100

1 Model Report

Before we look at different ways of writing up a report, we will look at a model report and the language used in it.

The report is about an experiment to find the resistivity of iron. The resistivity is a constant which is a measure of the electric resisting power of a substance. It is defined as the resistance offered by a one metre cube of the material. The resistivity is represented by the symbol p and is given by the formula

$p = \frac{RA}{l}$ where R is the resistance of a uniform conductor of length l and cross-sectional area A.

Read through the report and look for answers to the questions which follow it.

[Full text of laboratory report follows.]

Questions

1. What are the headings in the report?
2. What are the aims of the different sections of the report? Choose the best alternative in the following:
   (a) The introduction is concerned with:
      (i) the theory of the experiment
      (ii) the method for carrying out the experiment; or
      (iii) the theory of the experiment and the method for carrying it out.

Figure 8 - Dudley-Evans T 1985:5-8

12.10 The Hamp-Lyons & Heasley extract appears to assume that the best way to help students learn about text organisation is to offer them highly generalisable accounts of major patterns of discourse – it then becomes the learner's responsibility to implement this understanding in his or her own subject area. Such an approach, essentially an intellectual/rhetorical one, has many advantages. One is that it makes it possible for the course book to be used with many different kinds of student; another is that the learner should be

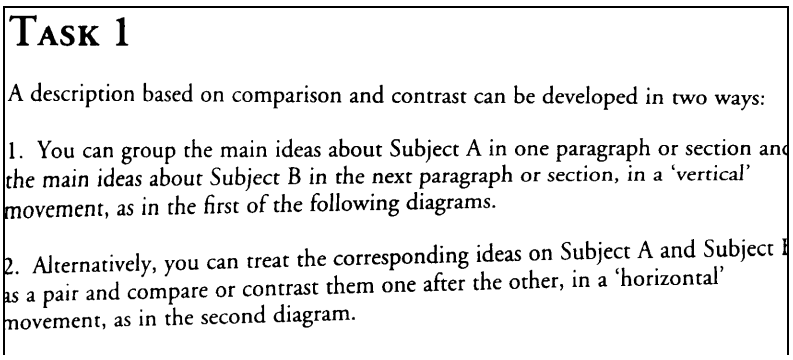
enabled to engage autonomously with new textual problems as and when they arise. Thus Hamp-Lyons and Heasley do not attempt to give specific information about how, say, historians, language students or earth scientists organise their texts. Rather, they draw the learner's attention to common rhetorical patterns that can be found in many texts - General to Particular, Problem / Solution and so forth. While this ought to have the benefit that learners will develop an understanding of how texts work that will stand them in good stead for the rest of their professional careers, there can also be problems. Most importantly, students can feel that the examples of texts presented in the materials, along with the activities that they are asked to carry out, are so distant from their immediate needs or interests they are unwilling to suspend disbelief for the purpose of the task in hand.

12.11 The second extract demonstrates a social / genre approach. It provides learners with highly specific information about the forms of text required by the genre of reporting laboratory experiments. This way of working also has advantages. First, learner motivation can be high as students with the right interests can see an immediate return on their investment. Secondly, the function of grammatical and lexical features in the development of a text can be shown more clearly, thereby making explicit the relationship between meaning, communicative purpose and language form. This link can be further clarified (as, in fact, is done in the present case) by organising the instructional materials around chapter headings such as: "Title and Aim, Procedure, Results and Discussion of Results, Comparison of Results .... ". The disadvantages of such an approach is that a book like *"Writing Laboratory Reports"* will *only*

appeal to a narrow range of learners – although it makes sense pedagogically, it is never going to be a big earner[7]

Issue 3 Authors of published EAP materials (along with course developers and teachers of writing) face a dilemma when choosing the context from which they will draw pedagogic examples. If the context is too narrow, it will only appeal to a very small group of students. If the context is too general there are (at least) two possible problems. The first is that learners will not be willing to relate to the texts they are being asked to exploit. The second, and more dangerous, problem is that the examples chosen will mislead learners into making generalisations that will not hold when they start to write within their own disciplines.

12.12 The next two examples offer contrasting approaches to teaching learners to develop an argument in EAP. The first is taken from White R and D McGovern 1994:



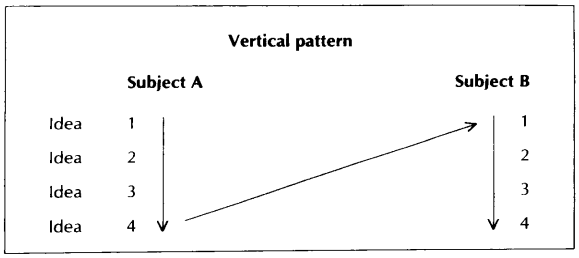
TASK 1

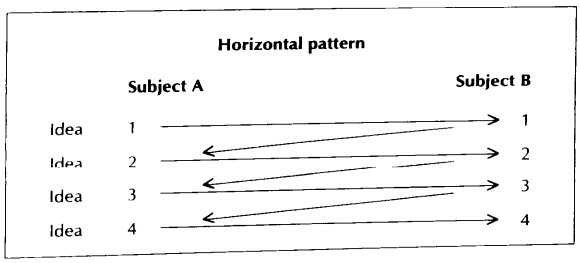
A description based on comparison and contrast can be developed in two ways:

1. You can group the main ideas about Subject A in one paragraph or section and the main ideas about Subject B in the next paragraph or section, in a 'vertical' movement, as in the first of the following diagrams.

2. Alternatively, you can treat the corresponding ideas on Subject A and Subject B as a pair and compare or contrast them one after the other, in a 'horizontal' movement, as in the second diagram.

[7] As was the case with *Writing laboratory reports*! (private communication from Tony Dudley-Evans 1994)





Whether you choose the 'vertical' or the 'horizontal' pattern depends on the kind of text you are writing, its purpose and your own preference. Some writers and readers find the 'horizontal' pattern clearer because it repeatedly reminds them of the comparison or contrast relationship. Others prefer the 'vertical' pattern because of its relative simplicity. The 'horizontal' pattern is often more suitable for a longer piece of writing. Both patterns are commonly used in descriptions involving comparison and contrast.

Step 1

1.1 Read Text 3.1.

1.2 Consider the following questions:

(a) Is the writer's description based on comparison or contrast – or both?
(b) Has the writer used a 'vertical' or 'horizontal' pattern of organisation?
(c) Which words are used to indicate a relationship of comparison or contrast? Can you think of similar words or phrases? (See Appendix 1.)

Text 3.1

(1) Concepts for the disposal of highly radioactive wastes have been around for more than 40 years. (2) The most authoritative early work was a 1957 report by the US National Academy of Sciences. (3) This recommended that rock-salt would be the most suitable medium in which to emplace wastes because it represented a dry geological environment. (4) In countries with large salt deposits, such as West Germany and the US, the concept of a 'dry' repository took hold. (5) Furthermore, this was the basis of research into disposal until the late 1970s.

(6) In other European countries, however, large salt domes do not exist. (7) From the mid-seventies, having ignored the disposal problem for about two decades, these countries chose instead to look at crystalline rock-types. (8) Invariably these were water-bearing.

Adapted from 'The folly of perfection' by Dr Frans Berkhout, *The Independent*, 1 May 1989

Figure 9 - White R & D McGovern 1994:22-4

12.13 The White and McGovern materials work in a top-down way, giving learners an overview of typical organisational patterns for this type of description and then giving them an opportunity to work as analysts and editors. In the process of moving from text to writing activity students are under a continuous obligation to discuss and share experiences – reference materials which provide instances of language use associated with comparison and contrast

being supplied in an appendix. An advantage of this activity is that, at this point in the unit of work, the learners are not under any pressure to be inventive. Instead, they are able to gain confidence to operate on the written language that is appropriate to a particular kind of text. The materials draw on an intellectual / rhetorical tradition of instruction, but face validity tends to be higher for learners than it is with e.g. the Hamp-Lyons / Heasley materials as the texts used are closer to those that students may be asked to write. Once the analysis has been undertaken learners are expected to apply ideas learned in a teaching unit to writing projects which they should be maintaining in parallel to the published course materials.

12.14 The material from Jordan 1992 in the next example approaches the problem from a different starting point and uses a more bottom-up approach (Figure 10). Here learners are given data to work on and are then asked to complete a series of gapped sentences. This done, they are asked to complete further exercises with a progressively more and more open structure – gapped text to notes to free writing - again of sentence length. The materials, with their movement from writing activity to text, are perhaps more typical of a rhetorical / intellectual approach to writing instruction, using a 'presentation, practice, production' cycle in order to help learners gain control of the skills they need in the target language.

Stage 1 Comparison

1 Look at Tables 1 and 2.

Table 1: The Longest Rivers in the World

1 The Nile (Africa)—4,160 miles (6,695 kilometres)
2 The Amazon (South America)—4,080 miles (6,570 kilometres)
3 The Mississippi-Missouri (North America)—3,740 miles (6,020 kilometres)
4 The Yangtse (Asia: China)—3,430 miles (5,520 kilometres)

Table 2: Temperatures and Rainfall in Beijing, China

Month	J	F	M	A	M	J	J	A	S	O	N	D
Temperature (°C)	−4.7	−1.5	5.0	13.7	19.9	24.5	26.0	24.7	19.8	12.5	3.6	−2.6
Rainfall (cm)	0.2	0.5	0.5	1.5	3.6	7.6	23.9	16.0	6.6	1.5	0.8	0.2

Note:
a the first month is January (J), the last is December (D).
b the temperature is measured in Centigrade, and is an average.
c the rainfall is measured in centimetres, and is also an average.

top the Nile (by satellite)
bottom the Amazon (aerial view)

Now complete the following sentences. If necessary, look at *Appendix 7: Comparisons*. Put one or more words in each space.

Table 1:

1 The Nile is _____ the Mississippi–Missouri.
2 The Amazon is _____ long _____ the Nile.
3 The Nile is _____ river in the world.
4 The Mississippi–Missouri is _____ the Amazon.
5 The Yangtse is _____ river in China.

Table 2:

6 In Beijing, January is a _____ month _____ _____ December.
7 July is a _____ month _____ June.
8 There is _____ rain in May _____ in March.
9 July has the _____ rain; in other words, July is _____ month.
10 August is _____ warm _____ July.
11 December and January are _____ months.
12 The rainfall in February is _____ in March.
13 April is _____ wet _____ October.
14 The rainfall in November is _____ in May.
15 July is the _____ month, and also _____ month.

2 Look at Table 3.

Table 3: The Highest Mountains in the World

1 Everest (Nepal/Tibet)—29,028 feet (8,848 metres)
2 K2 (Kashmir/Sinkiang)—28,250 feet (8,611 metres)
3 Kangchenjunga (Nepal/Sikkim)—28,168 feet (8,586 metres)
4 Makalu (Nepal/Tibet)—27,805 feet (8,475 metres)
5 Dhaulagiri (Nepal)—26,810 feet (8,172 metres)

Figure 10 - Jordan RR 1992:49-50

12.15 It is possible that with learners who have limited control of sentence level grammatical relationships there are advantages in using the bottom-up approach favoured by Jordan. Students (and teachers) are given a secure and systematic means of controlling a set of rhetorical modes, and comprehensive keying of the exercises makes it possible for learners to work with the materials independently. This way of working also makes it possible to plan classes in which a discrete set of information is dealt with in a controlled way.

A disadvantage of the approach may be that insufficient emphasis is paid to complete texts so that learners do not have the opportunity to develop their ability to edit and revise texts with a view to communicative quality. In other words, work at this level may not ultimately transfer to the higher level of organisation of texts discussed previously.

12.16 Such a transfer, on the other hand, may be more readily achieved in the White and McGovern materials since they focus on discourse organisation and help learners to match language use to communicative purpose at such an early point in the book. This does, of course, presuppose a fairly high degree of sophistication on the learner's part - but this is not an unreasonable presupposition in higher education settings. The remaining tasks in this teaching unit are consistent with this philosophy, requiring discussion, investigation and the discovery of rules or patterns. Although such a way of working may be demanding for teachers and learners alike, it does require learners to do the things that they will eventually have to do as effective writers, and also provides them with strategies for co-operative editing and peer reviewing that will be useful to them in their later professional careers. In short, such an approach allows for an integration between this level of argumentation and the overall level of organisation of texts in specific genres.

12.17 The approaches to text organisation and argument development that are exemplified above are representative of two major tendencies in writing instruction for foreign language students. One combines process and genre approaches, encouraging learners to discover patterns and rules of use as they become more familiar with the genres in which they will need to write. The other tendency ("intellectual-rhetorical") is more bottom-up and has the

strength of a long tradition – and a high degree of acceptability to many teachers. It works on the premise that learners need to develop an awareness of the linguistic options that exist at sentence level before they embark on the composition or transformation of complete texts.

Issue 4 Most teachers will accept that there is no single "correct" methodology for teaching writing. It will always be important to find an appropriate a balance between approaches which encourage learners to develop a mastery of language system knowledge appropriate to the kinds of writing they want to do, and approaches which focus on knowledge of co-texts and context. Teachers will have to make choices in relation to the level of their students' knowledge of the language system, the writing demands that they will face and the resources that are available – and also in relation to preferred learning styles.

13. English for international examinations

13.1 You might reasonably assume that there would be a marked difference between the situation of teachers preparing students for international EFL examinations such as the Cambridge Certificate in Advanced English (CAE) and teachers of students on Business English or EAP programmes. Teachers of examination classes should know the demands their students have to face: the examination syllabus is explicit – as are the marking criteria. Students preparing for an examination such as the globally dominant University of Cambridge Local Examinations Syndicate First Certificate in English (FCE) or Certificate in Advanced English (CAE) should know why they are taking a writing course: they want to pass the exam, and writing is a compulsory paper.

13.2 In such circumstances, materials writers and teachers might appear to have a straightforward task – and one which allows relatively little room for variation. At first sight this analysis is supported by the high degree of similarity between the different books that are published with titles along the

lines of *Writing skills for the xxx examination*. In order to see if this apparent simplicity is borne out in teaching materials, we will consider an example of a recent examination preparation book from a leading UK publisher, Cambridge University Press, *CAE Writing Skills* (O'Dell F 1996).

13.3 The official specification of tasks in the CAE written component lists the following texts that candidates may be asked to write:

announcements	articles (newspaper / magazine)
directions	formal and informal letters
instructions	leaflets
notices	personal notes and messages
reports	reviews

and, as is the case in the other exam preparation books on the market, this specification provides the basis for the contents of *CAE Writing Skills*:

Map of the book *To the student* *To the teacher* Foundation unit Writing based on a reading task Letters	Articles Narratives Reports Writing about work Notes, notices and announcements	Instructions and directions Reviews Brochures Competition entries *Answer key* (O'Dell F 1996:iii)

13.4 After the general front matter, two opening units are devoted to a review of the examination syllabus and marking scheme, and to introducing learners to issues in question interpretation and examination technique. The rest of the book is designed to help students to prepare for the different kinds of text which they may be asked to write in the CAE exam. This all looks unproblematic. We all know what letters, articles and brochures are, don't we? The job of the author and teachers who use the book will be to demonstrate to learners what "letters" and "brochures" are, and then find ways of assisting these learners to write this kind of text under examination conditions.

13.5 When individual teaching units are considered, however, the problems faced by authors of exam preparation materials become apparent – along with the still tenuous relationship between the theory of teaching writing and pedagogic realities. In, for instance, the chapter on "Brochures" the aim of the unit is stated thus:

> "The aim of this unit is to answer the following questions:
>
> - What kind of brochure might CAE ask you to write?
> - What are the characteristics of brochures and similar literature?
> - What style is appropriate when writing a brochure?
> - How can you make sure you correct your work efficiently?
> - How can you effectively write promotional or publicity literature?
> - How can you use idioms effectively?" (O'Dell F 1996:85)

13.6 Here, the questions about "the characteristics of brochures and similar literature" and the "style appropriate when writing a brochure" indicate that a "social / genre" approach is being applied to the teaching task, and that "brochure" is a reasonably discrete genre category. The concern about writing processes and language system knowledge indicates that a balance is being struck between the needs of the learners *qua* language learners and their needs as examination candidates. However, several difficulties arise when the tasks in specific teaching units are considered in detail. We will take the unit on *Brochures* (*Leaflets* in the CAE specification) as an example of the kinds of problem that authors face when writing teaching materials to help students trying to prepare for CAE[8].

13.7 The first problem is that the text label that O'Dell has chosen to use, *brochure,* seems to be inappropriate in relation to the texts to which it is applied. The

[8] Out of fairness to O'Dell, I should stress that the example given here was selected at random from the small number of recently published books designed to help learners preparing for the CAE examination. The weaknesses I indicate in the materials are weaknesses shared by many books on writing (including my own – see Tribble C 1989, Hopkins A and C Tribble 1989!)

Longman Dictionary of Contemporary English (2nd Edition) defines *brochure* as:

> a small thin book with a paper cover, esp. one giving instructions or details of a service: *a holiday brochure; an advertising brochure*" (Summers D & M Rundell 1987:122)

… and the idea that brochures are mainly booklets designed to sell or promote services is further supported by a random sample of 50 instances in the BNC (see Appendix 2: Brochure (BNC search results)). Yet in the teaching unit, students are asked to accept the examples given below as also fitting into the category *brochure*:

> *(1) Your local area is keen to encourage foreign tourists. You have been asked to write a brochure to send to other countries in order to promote the area. You have been requested to pay particular attention to at least three of the following – landscape; local food and drink; leisure facilities; places of historic interest; transport. Write the brochure.*
>
> *(II) You feel particularly strongly about an issue that is causing considerable discussion in your area. You decide to publicise your views by producing a pamphlet in why you lay out clearly what the issue is, why you feel as you do and why those with opposing views are, in your opinion, mistaken. Write the pamphlet.*
> (O'Dell F, 1996:87)

13.8 The problem is, in the light of the earlier discussion in this chapter (Paragraph 9 ff.), we would not consider the two writing contexts / texts presented above to be members of the same genre – they share neither communicative purpose, audience, authorship, nor, in fact, name. The only thing that they seem to have in common is that the texts which result from such communicative contexts are sometimes printed on pieces of folded paper, and they sometimes include other visual or typographically distinct information. Physically the textual realisations of these contexts would not be understood to be brochures;

generically they are non-equivalent. However, in spite of this, the examination puts these writing contexts into the same genre category.

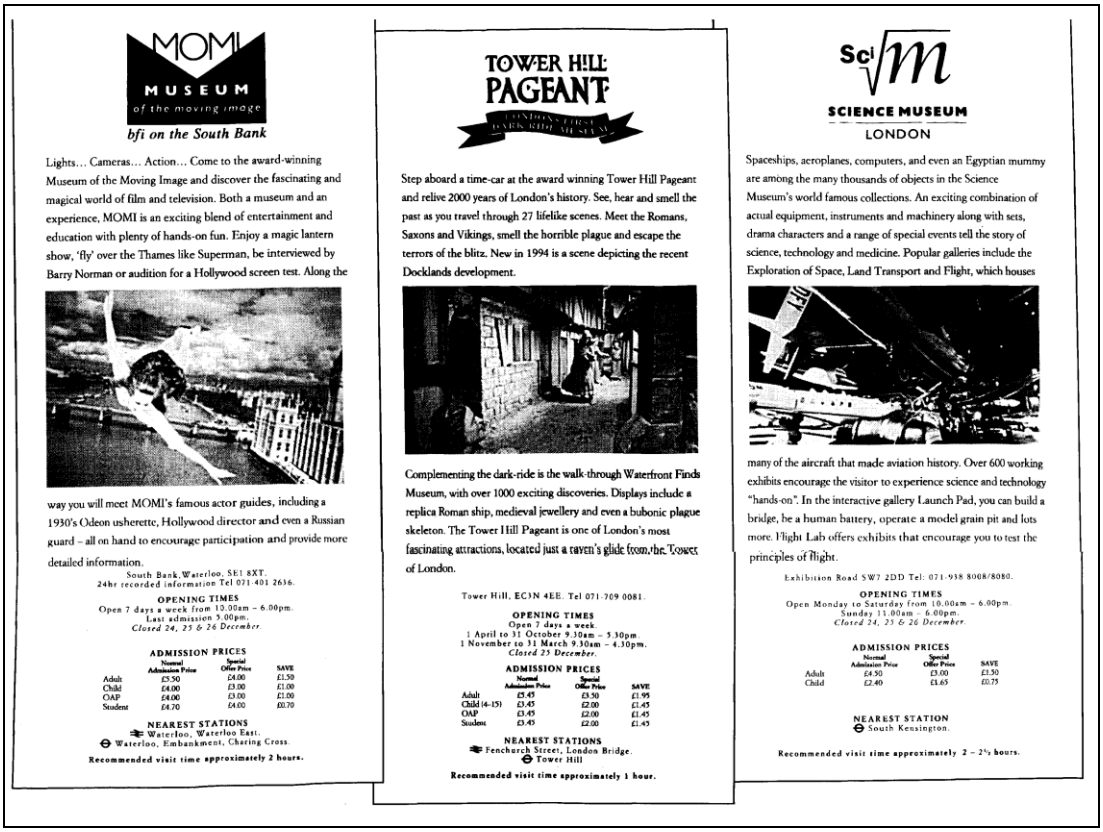
13.9 O'Dell tries to overcome this difficulty by introducing sub-categories (leaflet, information sheet, pamphlet) and then identifying the features they have in common:

> "As the exam tasks on the previous page demonstrate, there are different kinds of brochure and *they can even be called by different names* [my emphasis]. What they have in common is that they:
>
> - are short
> - aim partly to inform people
> - aim also to attract customers or supporters" (O'Dell F 1996:86)

Unfortunately a large number of other kinds of text – e.g. advertisements and CVs – can also be short and aim to inform and attract. *Brochure* remains elusive on the basis of this specification.

13.10 The problems continue when O'Dell goes on to invite comment on a London Transport brochure (Figure 11 below). She asks the following questions:

"- What information is provided in the brochure?
- What is it trying to attract people to do?
- Which parts of the text are simply giving information?
- How did the writer of this brochure try to interest readers? Think about general presentation and sentence structure in particular.
- Underline any words or expressions that are used with the aim more of attracting customers than simply informing." (O'Dell F 1996:86)



MOMI
MUSEUM
of the moving image
bfi on the South Bank

Lights… Cameras… Action… Come to the award-winning Museum of the Moving Image and discover the fascinating and magical world of film and television. Both a museum and an experience, MOMI is an exciting blend of entertainment and education with plenty of hands-on fun. Enjoy a magic lantern show, 'fly' over the Thames like Superman, be interviewed by Barry Norman or audition for a Hollywood screen test. Along the way you will meet MOMI's famous actor guides, including a 1930's Odeon usherette, Hollywood director and even a Russian guard – all on hand to encourage participation and provide more detailed information.

South Bank, Waterloo, SE1 8XT.
24hr recorded information Tel 021-401 2636.

OPENING TIMES
Open 7 days a week from 10.00am – 6.00pm.
Last admission 5.00pm.
Closed 24, 25 & 26 December.

ADMISSION PRICES

	Normal Admission Price	Special Offer Price	SAVE
Adult	£5.50	£4.00	£1.50
Child	£4.00	£3.00	£1.00
OAP	£4.00	£3.00	£1.00
Student	£4.70	£4.00	£0.70

NEAREST STATIONS
Waterloo, Waterloo East.
Waterloo, Embankment, Charing Cross.

Recommended visit time approximately 2 hours.

TOWER HILL
PAGEANT
LONDON'S FIRST LIVING HISTORY MUSEUM

Step aboard a time-car at the award winning Tower Hill Pageant and relive 2000 years of London's history. See, hear and smell the past as you travel through 27 lifelike scenes. Meet the Romans, Saxons and Vikings, smell the horrible plague and escape the terrors of the blitz. New in 1994 is a scene depicting the recent Docklands development.

Complementing the dark-ride is the walk-through Waterfront Finds Museum, with over 1000 exciting discoveries. Displays include a replica Roman ship, medieval jewellery and even a bubonic plague skeleton. The Tower Hill Pageant is one of London's most fascinating attractions, located just a raven's glide from the Tower of London.

Tower Hill, EC3N 4EE. Tel 071-709 0081.

OPENING TIMES
Open 7 days a week.
1 April to 31 October 9.30am – 5.30pm.
1 November to 31 March 9.30am – 4.30pm.
Closed 25 December.

ADMISSION PRICES

	Normal Admission Price	Special Offer Price	SAVE
Adult	£5.45	£3.50	£1.95
Child (4-15)	£3.45	£2.00	£1.45
OAP	£3.45	£2.00	£1.45
Student	£3.45	£2.00	£1.45

NEAREST STATIONS
Fenchurch Street, London Bridge.
Tower Hill

Recommended visit time approximately 1 hour.

Sc/m
SCIENCE MUSEUM
LONDON

Spaceships, aeroplanes, computers, and even an Egyptian mummy are among the many thousands of objects in the Science Museum's world famous collections. An exciting combination of actual equipment, instruments and machinery along with sets, drama characters and a range of special events tell the story of science, technology and medicine. Popular galleries include the Exploration of Space, Land Transport and Flight, which houses many of the aircraft that made aviation history. Over 600 working exhibits encourage the visitor to experience science and technology "hands-on". In the interactive gallery Launch Pad, you can build a bridge, be a human battery, operate a model grain pit and lots more. Flight Lab offers exhibits that encourage you to test the principles of flight.

Exhibition Road SW7 2DD Tel: 071-938 8008/8080.

OPENING TIMES
Open Monday to Saturday from 10.00am – 6.00pm.
Sunday 11.00am – 6.00pm.
Closed 24, 25 & 26 December.

ADMISSION PRICES

	Normal Admission Price	Special Offer Price	SAVE
Adult	£4.50	£3.00	£1.50
Child	£2.40	£1.65	£0.75

NEAREST STATION
South Kensington.

Recommended visit time approximately 2 – 2½ hours.

Figure 11 - O'Dell F 1996:87

13.11 O'Dell's answers to these questions are given in the key to this chapter, but unfortunately the information that is offered is, again, more likely to be misleading rather than useful to the learner. For example, on the matter of style, O'Dell comments:

> "4. The writer uses a number of techniques: pictures; addressing the reader directly by using *you* and imperatives; vivid vocabulary to make the places sound interesting; lists e.g. *Lights …Camera …Action* and the sentence beginning *Enjoy a magic lantern show* … The sentences tend to contain a lot of adjectives and try to pack in as much information as possible without become too long and complex" (O'Dell F 1996:119)

13.12 Such comments are not internally consistent – and could be equally true of other kinds of text. Counts for some of the features O'Dell mentions (*you, imperatives, vivid vocabulary, lists, high adjective count, sentences "not too long and complex", information level high*) are given below, alongside those

for a non-comparable text: a biographic essay on Mahatma Gandhi taken from a CD-ROM encyclopaedia[9]:

MOMI	TOWER HILL PAGEANT	SCIENCE MUSEUM	GANDHI TEXT
words = 101 sentences = 4 Average sentence length = 25.25 clauses = 12 Average clauses per sentence = 3 adjectives = 8 imperatives = 4 you = 1	words = 105 sentences = 7 Average sentence length = 15 clauses = 11 Average clauses per sentence = 1.5 adjectives = 5 imperatives = 6 you = 1	120 words 6 sentences Average sentence length = 20 words clauses = 9 Average clauses per sentence = 1.5 adjectives = 7 imperatives = 0 you = 2	words = 109 sentences = 4 Average sentence length = 27.5 clauses = 10 Average clauses per sentence = 2.5 adjectives = 13 imperatives = 0 you = 0

Table 7 - Brochure counts

13.13 Although the main text for each item is of comparable length, there is a wide range of variation from one to the other across most of the features O'Dell mentions. Most significantly, there are no instances of imperatives in SCIENCE MUSEUM, and a strikingly high imperative count (6) in TOWER HILL PAGEANT. Similarly there is a major difference between MOMI and the other texts in terms of the number of clauses per typographic sentence[10]. The problem with the characterisation O'Dell offers is further underlined when the counts for the Encarta extract are reviewed. GANDHI has sentence length and clause numbers similar to the MOMI article – and it has a higher adjective count than any of the "brochure" articles. Yet GANDHI is not a brochure – or a leaflet, pamphlet or information sheet. It would seem to be the case that while O'Dell offers a useful insight in the way in which second person address may characterise certain kinds of short persuasive texts, this point is lost amongst other, less significant, features and would not be consistently relevant to all the text categories that are brought together under the heading brochure / leaflet.

[9] see Appendix: Brochure Texts for full details

Issue 5 A social/genre approach to teaching writing becomes highly problematic if you cannot specify the genre you are working with. The problem with the CAE category "leaflet" (and also "letter" etc.) is that it is *not* a genre. As Gee says:

"Genres are firstly social processes and, as such, entail participation, that is, the social interaction of their participants. Further, each genre is a purposeful process and therefore the participation is meaningful. Genres are staged which means that they consist of different stages or steps which lead to the achievement of a goal." (Gee S 1997:27)

Thus the Technical Component of a Project proposal is amenable to a social / genre pedagogic approach because it *is* a clearly identifiable genre[11]. Similarly, the discussion section in one MSc dissertation can be seen as standing in clear analogous relation with other discussion sections in other MSc dissertations[12]

A letter, however, is not a genre; neither are brochures, leaflets, and pamphlets. LETTER, LEAFLET etc. may be "text types" (Stubbs M 1996:10-12) and as such describable in terms of certain common organisational features (e.g. "headline / body" in newspaper articles / "block-letter" format in commercial correspondence), but that is probably the limit to the pedagogically useful account that you can make of them. The moment you start to make strong generalisations about other instances of the same text type – especially when you do this on the basis of a single instance – the more likely you are to face the kinds of problem we have outlined in the last section.

14. Conclusion

14.1 Five issues in writing instruction have been identified during the course of this review of teaching materials, each of which raises a question for teachers of writing. The issues and associated questions can be summarised as follows:

Issue 1 Text data may confuse rather than assist if learners do not have an adequate theoretical framework to shape their analysis of examples.

Question: What constitutes a theoretically adequate pedagogic framework for

[10] a more useful measure of a text's "writtenness" see Halliday MAK 1989
[11] see Chapter 3:Approaching the data: dealing with genre
[12] see Dudley-Evans T 1994 for a fuller discussion of Discussion Sections

teachers who want to help learners to write difficult texts (as we have defined them here)?

Issue 2 Single instances of language in use do not provide a sufficient basis for conclusions about the linguistic specification of a genre. In order to make such generalisations, multiple examples are needed – ideally from several comparable sources.

Question: What kinds of examples do teachers and learners need when approaching the problem of writing into a new genre?

Issue 3 Authors, course developers and teachers of writing face a dilemma when choosing the context from which they will draw pedagogic examples. If the context is too narrow, it will only appeal to a very small group of students. If the context is too general there are (at least) two possible problems. The first is that learners will not be willing to relate to the texts they are being asked to exploit. The second, and more dangerous, problem is that the examples chosen will mislead learners into making generalisations that will not hold when they start to write within the genres that matter to them in their own disciplines.

Question: Is there any help for teachers who cannot get hold of examples of the kinds of texts their students need to write?

Issue 4 There is no single "correct" methodology for teaching writing. It will always be important to find an appropriate a balance between approaches which encourage learners to develop a mastery of language system knowledge appropriate to the kinds of writing they want to do, and approaches which focus on knowledge of co-texts and context. Teachers will have to

make choices in relation to the level of their students' knowledge of the language system, the writing demands that they will face and the resources that are available – and also in relation to preferred learning styles.

Question: What practical guidance can be given to help teachers develop appropriate methodologies for the writing courses they offer their students?

Issue 5 When using a social/genre approach to teaching writing it is particularly important not to confuse *text types* (which can be grouped together on the basis of common organisational features) with written *genres* (i.e. texts which are components of the GSP of a clearly specified communicative context). Though useful at certain stages in a writing syllabus, the kinds of generalisations that you can make about e.g. letters *in general* (format, address conventions, opening salutation, closing formulae) are different from the kinds of generalisations that you can make about e.g. letters of appointment or letters of rejection.

Question: What practical means are there to help teachers decide whether or not the texts they are dealing with are exemplars of an identifiable genre?

14.2 These issues and associated questions will be re-visited in Chapter 9: *Helping learners write difficult texts*. It will be possible by this point to identify the extent to which the theoretical and practical frameworks for the analysis of difficult texts developed thus far provide a basis for practical classroom instruction. What I shall be looking for is a framework for teaching writing which:

- offers criteria for the selection of appropriate pedagogic examples
- offers a practical basis for analysing these examples

- enables students to develop useful hypothesis about language use in genres that are important to them, so that they will be better able to write texts which will achieve their various purposes.

14.3 However, I am in advance of my argument. Before conclusions can be reached, I will first have to give an account of the data I shall be working with (Chapter 3: *Approaching the data: dealing with genre*), and then offer a reasoned basis for my analysis. I will then have to make sure that I have the right tools for the job and analyse my corpus. These task will be the focus of the four chapters in Section Two: An account of the texts.

SECTION ONE – SETTING THINGS UP

Chapter 3: Approaching the Data: dealing with genre

> "Both Firth and Malinowski believed that meaning in language arises primarily out of speakers' and listeners' recognition of conventional social situations which are associated with linguistic choice" (Couture B, 1986:1)

15. Introduction

15.1 In Chapter 1: *Writing Difficult Texts*, one of my main tasks was to outline my understanding of "writing" as a socially situated activity. In addressing this topic I began a discussion of genre, and this was further developed in Chapter 2: *Teaching Writing* where genre approaches to writing instruction were contrasted with the current dominant model of the process approach.

15.2 In these opening chapters we have seen that there is a growing literature which either uses the idea of genre to connect language in use to its social context (Miller C 1984, and Bakhtin MM 1986) or focuses on the interrelation between written language, social context, and the teaching of writing (Bamforth R 1993, Cope B and Kalantzis M 1993, Freedman A Adam C & G Smart 1994, Halliday MAK & R Hasan 1985, Halliday MAK 1989, Jacobs SE 1989, Samraj BR 1989, Smart G 1993, van Leeuven T 1993). In addition, and building on earlier work (Benson JD & WS Greaves 1980, Martin JR 1989, Swales J 1981), two extensive and influential studies have developed arguments for using genre as a framework for the teaching of writing in foreign language education (Swales J 1990 and Bhatia VK 1993). Swales' work particularly has begun to have a significant impact on the teaching of writing in EFL settings, an influence that can be traced in recent work done by

work by Bazerman C 1994, Smart G 1993, Spilka R 1993, Swales J & P Rogers 1995, Tribble C 1997.

15.3 While nothing in this body of literature constitutes a fully satisfactory *theory* of genre – i.e. it is not yet possible to identify a set of rules and procedures for the study of genre, and most work in the field remains more concerned with practice rather than principles or methods – there is a sufficient amount of work clustering around the notion itself for it to be useful for my present purposes. Although fuzzy, the notion of *genre* remains helpful as it makes possible the uncovering of connections between texts, communicative purposes, and the lexico-grammatical resources that writers draw on when making texts in contexts.

15.4 In this chapter, I shall consider the extent to which the texts I have selected as an example of "difficult texts" can be seen as a genre. I shall then discuss the possibilities of using corpus linguistic techniques in the analysis of genres, and the implications such analyses might have in helping foreign language learners and apprentice writers come to grips with the difficult texts they need to write.

16. Are project proposals a genre?

16.1 We have seen in Chapter 2 how Hasan's notion of "genre-specific semantic potential" and Generic Structure Potential (GSP) can be used as a starting point for explaining and working with genre. For Hasan, a genre bears a direct relation to a communicative context and can be seen as a verbal expression of a class of situation type – "genre is language doing the job appropriate to that class of social happenings" (Halliday MAK & R Hasan 1985:108). We also noted within her definition: "that texts belonging to the same genre can vary in

their structure; the one respect in which they cannot vary without consequence to their genre-allocation is the obligatory elements and dispositions of the GSP" (Halliday MAK & R Hasan 1985:108). Such a view of genre harmonises with Swales' 1990 extended definition in which he emphasises the primacy of communicative purpose:

> "Communicative purpose is both a privileged criterion and one which operates to keep the scope of a genre as here conceived narrowly focused on comparable rhetorical action ..... In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience". (Swales 1990:58)

On these criteria, are Project Proposals a genre?

16.2 First, Project Proposals (PP) are associated with a number of analogous, communicative contexts involving contractors who require a professional service, and suppliers who are competing with one another to provide this. The PPs that I have included in the research corpus have been prepared by professional consultancy groups or larger institutions, all of which have expertise in the areas that are required for the implementation of a proposed reform process. The PPs are, thereby, associated with the cycle of activities and texts that have been required in the tendering processes for European Union and other development agency projects in Eastern and Central Europe[13]. This cycle has included:

- notification of forthcoming projects (through the official gazettes and – more commonly now – via the Internet)

[13] See Chapter 8 *Writing Project Proposals* for comments on the cycle of work within consultancy organisations bidding for EU projects

- the submission of an expression of interest by organisations which consider themselves competent to undertake the management of the project

- a shortlisting process (in the case of the EU this is usually managed co-operatively by the Project Management Unit within a Ministry and the relevant office in Brussels)

- the issuing of an Invitation to Bid (ITB), Terms of Reference and Guidance Notes for potential contractors. These are usually issued at the same time as the ITB and may be subsumed under that text

- the submission of Project Proposals by agencies bidding for the management of the project in question

- the selection procedure and, if all goes to plan, the final issuing of contracts.

16.3 Not only are the PPs in the research corpus associated with analogous communicative contexts and communicative purposes, they also have similar structures. Typically, a Project Proposal submitted to the EU (and other development agencies such as ODA/DFID) has the following major components:

PROJECT PROPOSAL

TECHNICAL PROPOSAL

Part 1: Administrative

- Statements and Terms of Reference
  - A statement of intention to provide the services in conformity with this tender dossier by the Tenderer or his duly authorised agent.
  - The Terms of Reference for the services as they appear in the tender dossier, initialled on each page by the Tenderer.
  - A copy of the first page of the General Conditions for Service Contracts financed from PHARE Funds, signed by the Tenderer.
  - A statement concerning the sub-contracting envisaged for parts of the services, if sub-contracting is envisaged, signed by the Tenderer.
  - A statement concerning the bank account to which payments may be made.
  - A signature by the Tenderer or his duly authorised agent

Part 2: Technical Component (Organisation and Methods)

- A precise indication concerning the total amount of *man-days / man-months / man-years* proposed for each expert
- Logistics Plan:
- Risk Analysis:
- Time Schedule:
- Remarks, comments and suggestions which the Tenderer may consider it advisable to raise.

Part 3: List of Staff, including *Curriculum Vitae*

FINANCIAL PROPOSAL

- Fees
- Allowances
- Direct Costs
- Reimbursables

16.4 In addition, the PPs share a range of other features: *theme*, c*ommon issuing agency*, *geographical location*, *revenue and professional implications*:

16.4.1 THEME – texts in the PP Corpus were written as bids to manage extensive projects concerned with the restructuring of social or financial institutions (See Appendix 36: PP Themes for a detailed summary of the theme of each PP)

16.4.2 COMMON ISSUING AGENCY – four donor agencies with a commitment to reform processes in the former Communist countries of Eastern and Central Europe were involved in issuing the Invitations to Bid for these projects:

EU PHARE (EU Assistance Programme for Central Europe)	11 proposals
EU TACIS (EU Assistance Programme for Former USSR)	1 proposal
UK Overseas Development Administration (ODA) – now known as the Department for International Development (DFID)	1 proposal
The Know-How Fund (a development programme for former Warsaw Pact countries provided by the UK ODA/DFID)	1 proposal

Table 8 - Proposals: break-down by agency

At first sight, it might seem that the proposals submitted to the UK agency should not have been included in the corpus as they would skew the sample. However, following discussions with professional writers in the organisations which provided the texts I decided to retain them for the following reasons:

- ODA / DFID ITBs and TORs for projects in ECE have been designed to ensure maximum transparency, and are very similar to those issued by the EU PHARE office

- PPs submitted to ODA / DFID had closely analogous structures to those presented to the EU organisations

Although in one instance a writer commented that he differentiated between writing for DFID and writing for the EU (taking care not to write in an overcomplicated way for an audience which might contain no British readers[14]), other writers commented that they felt there were only superficial differences between proposals to the two organisations. Given the problem of obtaining proposal data, it was felt that these comments mitigated the risk of distorting the sample sufficiently to warrant their inclusion.

16.4.3 GEOGRAPHICAL LOCATION – the projects in question were located in 6 different countries (although the Czech and Slovak Republics were treated as a

[14] See Chapter 8: Writing Project Proposals

unit in one of the ITBs as it was issued prior to the break-up of Czechoslovakia).

Poland	7
Bulgaria	2
Hungary	2
Russia	2
Czech and Slovak Republics	1

Table 9 - Proposals: location

Although the geographical location of a proposed project will not make a significant difference to the wording of the project proposal, the fact that all 14 proposals focus on activities in adjacent countries in the former communist block is likely to have some influence on the themes and issues that the proposals face (e.g. privatisation, collapse of state structures) and helps maintain the comparability of the texts in question.

16.4.4 REVENUE and PROFESSIONAL IMPLICATIONS – each of the agencies involved in these bids has a commercial interest in winning the project. They will gain revenue by managing the project and without this revenue they cannot operate as businesses. Winning or losing the project will have an impact beyond the immediate satisfaction of winning an argument. Additionally, each agency also stands to benefit in terms of future earning capacity if it wins the management of a project. The stronger your track record in a field, the greater your chance of winning further contracts. Apart from this financial incentive, bids are also made within a professional culture which requires openness and transparency in its dealings. Part of this requirement is met by the process of anonymous sealed bids commonly used by aid and development organisations. It is also responded to by the discourse practices of those involved in bidding for contracts. Most proposal writers

come from an academic background and appear to derive their literacy practices from those of the academic discourse community. Thus, although PP writers want to make their proposals as persuasive as possible, they also feel a need a) to warrant the claims they make (although as we shall see in the results of the analysis of the language of PPs this warranting is often in relation to the prior experience of the bidding organisation rather than against a published literature) and b) to demonstrate their professional integrity through the way in which they present their case (as we shall see in the results of the interviews reported in Chapter 8). It would seem that writers have plenty of reasons for producing good quality texts when they are preparing PPs.

16.5 From the above, we can conclude that the PPs in the corpus share sufficiently similar communicative contexts, communicative purposes and other similarities "in terms of structure, style, content and intended audience" for us to see them as a distinct genre. Following Hasan (Halliday MAK & R Hasan 1985) I, therefore, consider the Project Proposals under consideration to be an exemplar of a genre within the Generic Structure Potential (GSP) of a communicative context we can call *Identifying Project Implementation Suppliers*. Such a typification situates the PP alongside other genres associated with the process of identifying project implementation suppliers, including: Invitations to Bid, Terms of Reference, Expressions of Interest and Contracts for suppliers, and, as we have seen, these genres are in turn closely connected to analogous genres required by other contractor/contractee relationships.

16.6 The main features which characterise such communicative contexts are:

- the organisation that invites the proposals has a specific social transformation goal in view whose successful implementation depends on the provision of appropriate specialist expertise and management skills

- the organisation preparing the proposal has to demonstrate that it has such appropriate expertise and skills and that it is sensitive to the needs and concerns of the donor agency and the recipient government

- there is an institutional requirement to ensure transparency in the contractor / contractee relationship

- there is an obligation to achieve "best value-for-money" through the use of market mechanisms.

16.7 It is worth noting that in the context of Eastern Europe in the period after 1991, the introduction of such tender processes had a profound influence on the ways in which individuals and government departments in recipient countries worked. By becoming involved in partnerships with organisations such as the EU, United Nations Development Programme or the UK Government's Know-How Fund, government officials in the newly independent states of the former Soviet Union or former Warsaw Pact countries had to meet head-on a completely different approach to planning, change management and resource procurement. I am not aware of research into the way in which a sudden obligation to develop a capacity to write into new genres can trigger changes in social relations – but it would appear to be a potentially important area to investigate.

17. Using corpora in genre analysis

17.1 In Chapter 1, I referred to four sets of knowledge that a writer requires in order to have a reasonable chance of satisfactorily completing a writing task– *content knowledge*; *context knowledge*, *language-system knowledge* and *writing process knowledge*. One way of describing the professional problem faced by teachers of writing is, then, that they first have to identify those areas of the knowledge base where they can most usefully assist their students, and then find pedagogic means for helping learners to extend that knowledge base. As we saw in Chapter 2, it is rarely appropriate for the *language* teacher to focus heavily on the content knowledge requirements of language learners on writing courses – they bring this with them, and in plentiful supply. Where writing teachers can make a contribution is in relation to the other three sets of knowledge – but, in order to make this contribution, teachers themselves require resources.

17.2 One of the main resources that teachers of writing need is implied in Swales' comment on the characteristics of a genre (quoted above – para 16.1). "In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience". We have already seen (Chapter 2) the problems that students and teachers can face if they only work with single instances when they are approaching a new genre. In order to identify the "patterns of similarity" Swales talks about, teachers and learners need access to *many* relevant examples of the genre they are attempting to write into, rather than one.

17.3 If teachers and students need multiple examples, we will also see in this chapter that they also need appropriate computational resources to reveal this patterning. Without such computer tools, it is extremely difficult (if not impossible) to identify regularities across a set of texts, or *dissimilarities* in relation to other text sets. McEnery and Wilson comment on the way in which frequency creates one such kind of patterning that can be revealed by corpus analysis:

> "There are certain kinds of language data that can only be gathered accurately from a corpus. Human beings have only the vaguest notion of the frequency of a construct or a word. Natural observation [using a computer] of data seems the only reliable source of evidence for such features …" (McEnery T & A Wilson 1996:12)

17.4 There is nothing new about using a corpus of relevant exemplars as a starting point for genre analysis. Swales' pioneering essay *Aspects of Articles Introductions* (1981) was based on manual counts of a small collection of instances of a genre and the principles underlying this approach have been followed through to recent publications such as Stubbs' *Text and Corpus Analysis* (1996), where, in spite of the huge computer resources now available, the starting point for the analysis is still counting patterns in texts.

17.5 A reasonably extensive corpus of appropriate examples of the genre in question has, therefore, a unique potential to provide the writing teacher (and writing learner) with the kind of resource they require. With such a corpus, teachers and students can develop the language system knowledge and (co-textual) context knowledge that they need in order to approach a difficult text. While the corpus will not provide all of the learner's needs, it can provide the best starting point for the other questions that they will need to ask in order to

gain a fuller appreciation of the context and the writing skills needed to fulfil a demanding task.

18. The PP Corpus

18.1 Even though we may have established that PPs are a genre and that a corpus of PPs will be a useful resource for writing teachers and students, we are faced with a problem regarding those elements of PPs which should be included in the research corpus. The account of a typical PP structure given in paragraph 16.3 showed that PPs typically have two main sections, the first of which is frequently divided into three sub-sections.

TECHNICAL PROPOSAL
Part 1: Administrative
Part 2: Technical Component (Organisation and Methods)
Part 3: List of Staff, including *Curriculum Vitae*
FINANCIAL PROPOSAL

18.2 It also showed that each of these sections and sub-sections has specific characteristics – some being more useful in a study designed to help learners develop their competence as writers than others. The first sub-section of the Technical Proposal, for example, is mainly a replication of information provided in the ITB and Terms of Reference. As such, it is not the product of the proposal writing team – often being photocopied and directly bound into the PP. It has not, therefore been included in the research corpus.

18.3 For different reasons, the second main section (Financial Proposal) has also not been included. Although it can be critical to the success of the bid, the financial proposal was felt by all the donating consultancy groups to be too commercially sensitive to release for this study. Fortunately, from a language teaching point of view, this lack is not critical. From my own professional

experience, I know that although financial proposals are subject to very specific generic constraints (they are tabulated figures in the main), they present in the main a *technical* problem in which important strategic and professional experience is called on, rather than a *language* problem. Lists of Staff and CVs are also extremely important in terms of the overall value of the proposal. However, the need to ensure confidentiality renders them more or less meaningless when they are anonymised (see Appendix 38: Preparing the PP Corpus), and again, they do not present a major linguistic challenge once a few complete instances have been studied. They too will be excluded from the corpus based study.

18.4 Given these constraints, I have decided not to include the whole of each PP in the research corpus. In the light of the discussion above, it should be now be clear that the PP subsection which will be the main focus of this study will be the "Technical Component" (sometimes known as the Organisation and Method section). This is usually the most extensive section in the Project Proposal, and, as it contains the response to the ITB is therefore, the most professionally (and linguistically) challenging section to write (despite the importance of the numbers in the Financial Proposal). So long as they have been able to get the numbers right, it is in the Technical Component that writers have to elaborate their response to the ITB and TOR, and it is here that they establish the case for winning the bid.

18.5 While it is appropriate that the Technical Component of the PP should be the focus of the corpus study, this is not to say that the sections that have been excluded are unimportant. They are not. As we shall see in Chapter 7: *Organisation*, any apprentice proposal writer will be well advised to study the

way in which the first section of the PP is organised, and also to study examples of Financial Proposals, Lists of staff and CVs. Rather, it is that I have included in the corpus study those sections of the PPs which would be most helpful to those who are unfamiliar with this genre and who have an interest in learning how to write into it.

19. A methodology for using corpora in genre analysis

19.1 In Halliday's seminal essay on style in William Golding's novel, *The Inheritors*, he comments on the use of quantitative methods in the analysis of style, arguing that "a rough indication of frequencies is often just what is needed" (Halliday MAK 1973:117). This justification for drawing on empirical techniques in literary analysis now has much greater force given the recent massive increase in the sophistication and availability of computerised tools for the analysis of texts. Although the figures of themselves still do not "constitute an analysis, interpretation or explanation of the style" (Halliday MAK 1973:117) this increase in sophistication does mean that the figures can form a much greater part of an analysis than was the case in 1973. It is, in fact, now possible to argue that the figures can provide the starting point for the analysis. The computer can reveal "patterns that would not otherwise be visible." (Tribble C & G Jones, 1990:9)

19.2 If we decide to give the figures a more prominent part in the analysis of texts, then there is a need for a methodology which will allow a reasoned use of the numbers, and which makes it possible to compare the data obtained in one study with the results of work done by others in the field. Such an approach

should constitute what Stubbs (1994) calls a corpus linguistic "meta-method": a way of working with texts which is:

> "based on explicit comparative descriptions of substantial corpora of data, in which clearer relations are drawn between lexicogrammar and text, and which incorporates a theory of probabilistic grammar." (Stubbs 1994:205)

19.3 The meta-method I plan to use in this study will be most useful for teachers and learners if it meets Stubbs' criterion of drawing on "comparative descriptions of substantial corpora of data". Fortunately I will be able to benefit from work that has already been carried out in this field and can make use of existing methodologies and published results in developing my own analysis.

19.4 In an earlier study (Tribble C, 1985), I used a literary stylistics framework derived from Leech G & M Short (1981) to study differences in wording between referential texts which shared a common content focus. Based on manual counts of short extracts with the common theme of "photography" it was possible to demonstrate the ways in which different communicative purposes and reader / writer relations required contrasting grammatical realisations. For the present study I intend to continue in this tradition, using work reported in Biber D 1988 and which connects with the corpus referenced tradition pioneered by Leech and Short. I shall draw largely on Biber D (1988) for the main framework for analysis, and also refer to Biber D (1989a), Biber D (1989b), and Biber D (1990) and particularly Biber D & E Finegan (1989) where a reduced set of variables is introduced. Biber has many advantages for this study. Stubbs draws attention to one of the most important of these when he says:

> "Many linguistic features are of potential stylistic interest, and several overlapping lists are proposed, e.g. by Fairclough (1989: 110-11), Fowler (1991b:68ff.), and Myers (1992); and Biber (1988) provides a list of lists. However, these lists summarise experience and intuition, and as Fowler (1991b:90) admits: 'apart from Biber's they are unformalized and only partly useable in computer assisted studies'."

The multi-variate analysis in Biber 1988 provides an accessible and established framework in which to exploit the tools that experience in corpus linguistics has made available to the text analyst.

19.5 A further reason for drawing on Biber 1988 is the significant progress that has been made since his research in automatic text parsing at the University of Lancaster (Garside R, Leech G & G Sampson, 1987; Black, Carling & G Leech, 1993) with the development of the CLAWS 7 part-of-speech (POS) tagging program, and other corpus processing tools associated with the British National Corpus project (Burnard L, 1995). This has greatly simplified the job of corpus preparation[15] (see Appendix 38: Preparing the PP Corpus) and has meant that I have been able to concentrate on the problem of applying the approach developed in the Biber study rather than having to develop a POS tagging system – a job which would have been well beyond my computing skills.

19.6 Carrying out this kind of study in the late 1990s rather than the 1980s has also meant that it has been possible to extend the Biber framework by taking advantage of recent developments in tools for corpus analysis. The groundbreaking work done by Scott (Scott M 1996, 1997b, 1997c, 1998) has made it possible to undertake a sophisticated analysis of an unannotated corpus and produce results which are of great potential benefit to the learner. I

shall include the findings of this extension of Biber in Section Two of this thesis: *An account of the texts*.

20. Genre, corpora, writing

20.1 The utility of including a genre approach in programmes of writing instruction has begun to gain wider acceptance in a range of educational settings (see e.g. Flowerdew J, 1993 for an interesting example from teaching professional writing skills, and essays in Cope B and Kalantzis M, 1993, for accounts of the experience in the context of the Australian secondary school system). Projects to teach writing from the point of view of genre – and in that process focusing on the linguistic features of the texts which are members of the genre in question – also harmonise well with Bazerman's argument in favour of a theory of written language which "posits a framework for constructing an understanding of prior events and creating new ones." (Bazerman C, 1994:165). Like Bazerman, I am arguing for an approach to the teaching of writing which gives importance to a knowledge of other "relevant written texts (intertextuality) along with the better established notions of audience, self, reality and text" (ibid.: 163).

20.2 In order to achieve this purpose I shall, however, be obliged to address a wider set of themes than the corpus linguistic one alone. We have seen in Chapter 2 that teachers of writing need to take into account more than one kind of knowledge when addressing the needs of writing students. I, too, shall have to ensure that any proposals I make here for teaching written communication have more than a single focus. The corpus is not going to give all the answers.

---

[15] Many thanks to Martin Wynne and Nick Smith for their help in undertaking this task

Swales and Rogers' comment that an analysis of linguistic features alone "would in the end do little more than underscore the limits and limitations of a purely textual approach to genre analysis", (Swales J & P Rogers, 1995:233) is particularly apposite. I have already indicated (Chapter 1) that I will, therefore, also be taking a "contextual turn" (Swales J & P Rogers, 1995:233) once the corpus linguistic section of the thesis is complete. Using structured interview procedures (Fielding N 1993) I plan to achieve the following main objectives:

- to test conclusions reached in the corpus study
- to identify with authors how they and their organisations have responded to the requirements of the ITB and organised the process or writing PPs
- to identify with writers the reasons they have for preferring particular salient wordings in the texts they have written, and in analogues to those texts

20.3 The chapters which follow this opening section: *Setting things up*, will have, therefore, three different kinds of focus. They will be:

- CORPUS FOCUS (Chapters 4,5,6,7): to develop techniques of corpus description which will be of relevance to teachers of writing. These techniques will be applied to the research corpus in order to identify the language-system knowledge which is required by learners wishing to develop a capacity to write PPs (or any other genre which is analysed using the same tools).

- CASE STUDY FOCUS (Chapter 8): to gain insights into the ways in which contextual knowledge (knowledge of audience, of other texts) causes writers of PPs to make particular selections from the meaning potential of English.

- PEDAGOGIC FOCUS (Chapter 9): to propose learning / teaching strategies which will help foreign language learners (or apprentice writers in general) to write difficult texts.

SECTION TWO – AN ACCOUNT OF THE TEXTS

Chapter 4: What are we looking for?

Polonius: ..... What do you read, my lord?
Hamlet: Words, words, words
(Hamlet: Act 2, scene 2)

21. Introduction

21.1 The first Section of this thesis was devoted to "setting things up". In the three chapters in Section One I presented the major themes which I consider relevant to my overall purpose – to help learners write difficult texts. Having set out my understanding of *writing*, *difficult* and *texts* in Chapter 1, in Chapter 2, I raised five issues associated with the teaching of writing. I shall re-state two of these issues here, as they are of particular relevance to this first chapter in Section Two.

> **ISSUE 1:** Text data may confuse rather than assist if learners do not have an adequate theoretical framework to shape their analysis of examples.
>
> **ISSUE 2:** Single instances of language in use do not provide a sufficient basis for conclusions about the linguistic specification of a genre. In order to make such generalisations, multiple examples are needed – ideally from several comparable sources

21.2 These two issues are central to this second Section of the thesis, as it is here that I wish to establish a pedagogically useful means of analysing examples of a genre. If I am correct that learners benefit from having access to more than one instance of an exemplar text when learning how to write into a new genre, and that they need a rationale on which to base their analysis and develop generalisations which can guide their future writing, then it is in this Section that I have to justify my assertion. In this first chapter I shall, therefore outline the main theoretical framework I propose to draw on as the starting point for

learner analysis of examples of difficult texts. In doing this, I shall also have to give an account of some of the practical problems that are faced by anyone wishing to use this framework – and some of the solutions that I have found to these problems. This will require a fairly extensive account of some of the pitfalls that can lie in wait for anyone doing an empirical study of language in use. This account may go beyond the interest of the reader who has a primary interest in the results of this study rather than the means whereby those results were obtained. I make no apology, however, for I feel it important to include this study at this point in the text as it allows me:

- to indicate the problems that may be faced by other researchers who wish to use the Biber 1988 framework

- to share my solutions to these problems so that they can be tested and further refined by others with similar interests to my own.

21.3 The account that I present in the first chapter will, therefore, focus on some of the practical problems of corpus study. Once I have clarified my own resolution of these problems I shall report the results that I have obtained from the computational analysis of a genre in the next three chapters of Section Two.

22. The Biber multi-function / multi-dimensional model

22.1 With the original intention of establishing a linguistic basis for differentiating between spoken and written language production, Biber used a corpus based, multivariate statistical procedure (Biber D 1988:61-97) to identify six

"dimensions" which could be used as a means for specifying the relationships between texts. These dimensions were:

Factor 1: Involved versus Informational Production – "a dimension marking high informational density and exact informational content versus affective, interactional and generalised content. Two separate communicative parameters seem to be involved here: (1) the primary purpose of the writer / speaker: informational versus interactive, affective and involved; and (2) the production circumstances: those circumstances characterised by careful editing possibilities, enabling precision in lexical choice and in integrated textual structure, versus circumstances dictated by real-time constraints, resulting in generalised lexical choice and a generally fragmented presentation of information" (Biber D 1988:108).

Factor 2: Narrative versus Non-narrative Concerns – distinguishes "narrative from other types of discourse. It might also be considered as distinguishing between active, event-oriented discourse and more static, descriptive or expository types of discourse." (Biber D 1988:109).

Factor 3: Explicit versus Situation Dependent Reference – "corresponds most closely to the distinction between endophoric and exophoric reference" (Biber D 1988:110). Typically, texts with high exophoric reference are associated with spoken, unplanned, or markedly informal discourse.

Factor 4: Overt expression of persuasion – ".... this dimension marks the degree to which persuasion is marked overtly, whether overt marking of the speaker's own point of view, or an assessment of the advisability or

likelihood of an event presented to persuade the addressee." (Biber D 1988:111)

Factor 5: Abstract versus Non-abstract Information – ".... marks informational discourse that is abstract, technical and formal, versus other types of discourse...." (Biber D 1988:112-113)

Factor 6: On-line Informational Elaboration – "seems to distinguish discourse that is informational but produced under real-time conditions from other types of discourse" (Biber D 1988:114).

22.2 The importance of Biber's 1988 study for language teaching is that he demonstrated that it was not only possible to distinguish between spoken and written production by identifying the way in which examples of different genres were distributed along these textual dimensions – it was also possible to provide a coherent account of linguistic variation across the different text genres exemplified in his corpus. Having isolated the linguistic features which were closely associated with particular modes of production, he showed it was also possible to see how these same factors were associated with production in contrasting genre settings.

22.3 A later study (Biber and Finegan, 1989) looked at the way a smaller set of the 1988 factors (1/3/5) can be particularly helpful in differentiating between modes of written production. Specifically:

- Factor 1 ("A" in Biber & Finegan 1989) distinguishes between texts which have a high information loading as opposed to texts (such as unscripted conversation as a polar extreme) which have the primary purpose of

relationship building. (see also Halliday MAK 1989 for background to this argument).

- Factor 3 ("B") Biber and Finegan comment that this factor "seems to distinguish between highly explicit, 'context-independent' reference, on the one hand, and non-specific, 'situation-dependent' reference on the other" (Biber & Finegan: 1989:492).

- Factor 5 ("C") has the capacity to distinguish between Abstract and Non-abstract styles, and is, again, an important means for distinguishing between genres of written communication.

22.4 Biber and Finegan comment that with reference to the broad range of genres in the LOB corpus: "It is possible to characterize particular genres as relatively LITERATE or ORAL, where 'literate' refers to language produced in situations that are typical for writing, and 'oral' refers to language produced in situations typical for speaking. Conversation is a stereotypically oral genre, and academic prose a stereotypically literate genre." (ibid:493). It is this aspect of Biber's approach which makes it so potentially important for teachers of writing. He offers a way of seeing how certain well formed texts may be perceived as more acceptable than others – the problem stated at the beginning of this thesis.

22.5 I shall, therefore, focus particular attention on the results gained for these three factors when I come to analyse texts in the PP Corpus. Although I have collected data for all six of the Factors used in Biber 1988 in order to provide a comprehensive review of the generic relations the PP texts enter into, the factors of greatest interest to teachers of writing will be at the centre of the

study. In their 1989 study Biber and Finegan used dimensions 1, 3, and 5 in an analysis of diachronic change in written style – their premise being that since the 18th Century there has been a shift across a range of written genres from a more formal and archetypically "written" style to a style that is more informal and "orally" oriented. While I am not attempting to trace diachronic shift, these three factors – and the notions of "literate" and "oral" texts that are elaborated in Biber & Finegan 1989a – provide an effective means of identifying the extents to which the texts in my corpus are similar or dissimilar along specific dimensions, and the extent to which they can be considered to be linguistically differentiated from other written genres. The same categorisation was used in Tribble (1985:56) in a study of texts with related topics but different readerships (i.e. expert → expert vs. expert → non-expert).

23. Working with the Biber framework

23.1 One of the charms of Biber 1988 was that his starting point is very simple. In line with Hallidayan notions of the situationally determined nature of language use, Biber predicts that texts produced under different conditions and for different purposes will have contrasting grammatical realisations. In order to test this hypothesis he then identifies a list of sixty-seven grammatical features and find ways of counting their distribution across a large number of different kinds of text. Having obtained his raw data (no mean feat, especially given the development of computing at the time of the study), the much more sophisticated aspect of Biber's study lies in the statistical procedures which he used in order to make reasonable generalisations about different modes of production on the basis of his counts.

23.2 The grammatical / lexical categories which are used in Biber 1988 are summarised in the table below. They were grouped by means of a weighting system allocated to each linguistic feature, the weightings being established by calculating the co-occurrence of each of 67 text features using factor analysis procedures. This analysis provided the means whereby the features were allocated to a particular factor. The process of allocating features to factors involved:

- identifying those features with the largest weighting scores (either positive or negative), and
- allocating them to a particular factor through a process of elimination in which a linguistic feature was allocated to *one* factor only.

This procedure allowed the identification of the six factors that will be used in this study.

23.3 The linguistic features associated with the factors which will be used in this study, are given in the table below:

dimension +1	present-tense verbs first-person pronouns second-person pronouns pronoun IT demonstrative pronouns indefinite pronouns DO as pro-verb WH questions BE as main verb WH clauses sentence relatives adv. sub. – cause hedges amplifiers emphatics discourse particles possibility modals private verbs contractions THAT deletion stranded prepositions non-phrasal coordination analytic negation	**dimension +3**	nominalizations WH-relatives: subj. position WH-relatives: obj. position WH-relatives: pied pipes phrasal coordination
dimension -1	nouns prepositions attributive adjectives type/token ratio word length	**dimension -3**	place adverbials time adverbials adverbs total
dimension +2	past tense verbs perfect aspect verbs third person pronouns present participle clauses public verbs synthetic negation	**dimension +4**	infinitives adv. sub. – condition necessity modals predictive modals suasive verbs split auxiliaries
		dimension +5	agentless passives BY-passives past participle adverbial clauses past prt WHIZ deletions adv. sub. – other conjuncts
		dimension +6	THAT verb complements THAT adj. complements THAT relatives: obj position demonstratives

Table 10 - Biber 1988: Linguistic Features

I have in turn attempted to use these linguistic features in the analysis of the Project Proposal corpus and they have provided the means for establishing a characterisation of each text in terms of its position relative to the other genres in LOB+.

23.4 Biber describes three key stages in developing a multidimensional analysis of texts: "Preliminary analyses; Step 1: Factor Analysis; and Step 2: Factor scores as operational representatives of the textual dimensions" (Biber D. 1988:64). As I have already mentioned, Biber's 1988 study was initially designed to see if it was possible to make a linguistic differentiation between

spoken and written language. During the development of his analytic model he also found that it was possible to provide a linguistic characterisations of all of the genres used in his study through this account of their textual dimensions.

23.5 As my study has not had the same starting point as Biber's – the desire to identify the differentiation between spoken and written production – it has not been necessary to pass through the first two stages of this sequence. Instead, I have been able to draw on the factor analysis already undertaken, and to use Biber's results as the starting point for establishing factor scores for individual exemplars in my own research collection and for the Project Proposal corpus as a whole. However, parallel to Biber's first two stages I have had to identify and count the linguistic features that are needed to establish Factor scores for individual texts or for the whole PP Corpus. It has been during this preliminary analysis that I have had to come to terms with some of the problems that were mentioned in my opening remarks. While "a rose is a rose is a rose", a word might not always be a word when seen through the tunnel vision of a computer.

24. Applying the Biber 1988 analytic framework

24.1 Geoffrey Leech's apparently straight-forward comment that "a corpus can be stored in a 'raw' orthographic form, that is with the texts in the same form as they would have on the printed page, with words represented by strings of characters separated by spaces" (Garside R et al 1987:8) in fact says a great deal about the problem that faces anyone who tries to use a computer to analyse texts. When a confused or contrary Hamlet responds to a question

about what he is reading, he is at least able to give the plausible, though uncooperative, response: "words". If a computer is asked the same question it may only reply, "strings, strings, strings". These strings have no meaning for the computer, they are merely – "characters separated by spaces" – until a human intervention informs the machine that *this* string means *this* word, and *this* string means another.

24.2 This point has been brought home to me while attempting to use the analytic framework outlined in Biber 1988. The basic problem is that your way of counting Class A words may not be the same as mine. We then find that although we both appear to be counting the same Class A words in the same text, our results do not tally. What happened during the initial analysis of the PP corpus, happened because of the differences between what Biber was doing in the late 80s and what I did in the late 1990s. The basic difference between our two studies can be summarised under the headings *hardware*, *software*, and *corpus mark-up*. These differences have had a significant impact on the results I have obtained – and I have had to undertake major revisions of the search algorithms that I have used in order to ensure as much as possible that the "words" and word classes that I have been counting in the late 1990s are as close as possible to those that Biber was counting ten years earlier.

- *Hardware* – during the decade between Biber's original study and my current work there have been huge technological changes – which mean, for example, that the analysis that I have done has been carried out away from the university on relatively simple computing equipment (by today's standards). While this development is one which I welcome, it means that I am not able to replicate exactly the manner in which Biber carried out his preliminary

analysis. In fact, as things stand today, neither he nor I are able to reproduce his results as (if still usable) the IBM tapes which recorded the original results can only be used on a now obsolete generation of IBM computers, and no paper printout of this original data has been retained (personal communication from Doug Biber – 22 July 1997[16])

- *Software* – in parallel with these changes in the technology, there has been an increase in the sophistication of the software available for researchers. For the present study I have used four programs for text searching and analysis:

- WordSmith Tools, a suite of text analysis programs developed by Mike Scott of Liverpool University and published by OUP (Scott M, 1997)

- Windows Grep (Millington H 1995, 1997), a commercial program for the Windows PC environment

- Microsoft Word 97

- Microsoft Excel 97.

The easy availability of such sophisticated software tools has been of enormous significance to the recent development of corpus studies, as extensive corpus research can now be undertaken by individuals who do not have access to major institutional resources – and who do not have the programming knowledge needed by earlier researchers (Tribble C 1997a). It also means, however, that in spite of Biber's providing a detailed listing of

[16] "I did not include modals in the verb counts as far as I can remember, but the bottom line is that I really have nothing left from that analysis to check concretely. This is really unfortunate! I guess the one thing that I do have is some old IBM MVS tapes, that have everything on them (I think??) -- if you're able to process those (and if they are still readable??), I'd be willing to send them off to you -- but I'm not optimistic that that would really work out." Personal communication from Doug Biber

algorithms he developed during the course of his study (Biber D 1988: 211-245), extensive "reverse engineering" has been required to approximate the specific search procedures used in Biber 1988.

- *Corpus mark-up* – the third, and most significant, difference between my study and Biber 1988 is the contrast between the part of speech (POS) tagging system which he uses and that which has been available to me in the development of my own analysis. During the period 1983-1986 Biber built a set of dictionaries and computer programs which made it possible to allocate POS tags to words in his research corpus with around 90 percent reliability (Biber D 1988:217). While Biber was able to produce results which were internally consistent and sufficient to his purpose using these tools, a significant drawback to Biber's study is that it seems to be impossible to replicate it. Although Biber 1988 worked within the same probabilistic paradigm as, for example, researchers at Lancaster and Leeds Universities (Garside et al, 1987) his tag set remains unique. The tag set that I am working with has the advantage of being more accessible than Biber's and of being used on major projects such as the British National Corpus (Burnard L 1995). CLAWS7, the latest version, (Wynne M 1996) is a development from earlier work at Lancaster (Garside R et al, 1987). It provides 148 tags (see CLAWS7 Tagset list in Appendix 6). Its disadvantage is that it is *more* comprehensive than Biber's and also *more* accurately applied through the CLAWS tagging program used at UCREL at Lancaster University. In other words, it gives different results from Biber's, even when the same "words" are being counted.

24.3 The procedure I followed in putting together the PP corpus is described in "Supplementary issues: Corpus Preparation". This provided me with a data set

capable of being analysed in a way which paralleled (as much as possible) Biber's 1988 study. The next task was to prepare and test a set of algorithms which would mimic Biber's published set. This proved to be more difficult than I had at first thought would be the case.

25. Replicating the tools used in the Biber 1988 study

25.1 We have already seen that 67 linguistic features were searched for in Biber 1988. Biber found that not all of these were required in order to establish the six textual dimensions mentioned above, and the final set that is required for practical analysis contains the 58 essential features listed above (para 23.3), plus 3 other features (predicative adjectives, existential THERE, downtoners). These latter are only included in the counts as a way of excluding them from the final analysis.

25.2 In order to identify these linguistic features it is necessary to have a set of search algorithms which specify the strings you wish to find in a way which makes sense in the context of the corpus mark up being used and the software available for searching and counting. A full list of the first set of algorithms I prepared is provided in Appendix 40.

25.3 Some of these algorithms look unambiguous and unproblematic; others look more complex. The instructions that are issued to the search software will depend on the computer resources that are available. Take for example Algorithm 1 "Private verbs" – a class derived from Quirk *et al* (1985:1180-1) which expresses "various intellectual states (e.g. *believe*), or non-observable intellectual acts (e.g. *discover*)." (Biber D 1988:242).

```
1   56 private verbs          C:\WSMITH\SEARCH\V-PRIVAT.TXT Context  WS
                              Word 1 right [V*
```

The algorithm can be interpreted as follows:

> *Using WordSmith Tools, search for all instances of the strings in the lemmatized list of private verbs stored in the file V-PRIVAT.TXT where they are immediately followed by a verb POS tag*

25.4 The first version of the algorithm only required the creation of a list of the full lemmas of all of the private verbs listed in Biber (1988:242), saving this as a text file and then using this as a search reference in WordSmith. Such an algorithm should produce completely dependable results with any set of plain text files in so far as it will identify and count the strings that are specified. The weakness of the algorithm was that it could not distinguish between verbal uses of the strings and other uses which may be, for example, nominal (e.g. *Our only enemy is fear itself* …). I, therefore, improved the algorithm by requiring a tag for any verb form ([V*) in the immediate right context. In this way only those strings in the corpus that have been identified as verbs will be selected. The impact of this refinement is considerable. The total count for Private Verbs for the PP Corpus using the wordlist alone is 533. Using the refined algorithm the count is 382 – a 28% difference. Output is presented in the following form (counts and file references are automatically provided by the programme):

```
1   VVI] successfully [RR] . [.] We [PPIS2] believe [VV0] that [RG] mu sq-vlad.gre
2    [NN] that [CST] has [VHZ] already [RR] demonstrated [VVN] serious sq-vlad.gre
3   [NN2] arise [VV0] from [II] : [:] – [-] demonstrating [VVG] new [J sq-vlad.gre
4   AT] factors [NN2] which [DDQ] will [VM] determine [VVI] its [APPGE sq-vlad.gre
5   ] will [VM] need [VVI] to [TO] be [VBI] found [VVN] to [TO] decent sq-vlad.gre
6   IO] these [DD2] initiatives [NN2] – [-] indicates [VVZ] that [CST] sq-vlad.gre
7   [NN1] , [,] but [CCB] do [VD0] not [XX] know [VVI] it [PPH1] well  sq-vlad.gre
8   CST] businesses [NN2] do [VD0] not [XX] know [VVI] what [DDQ] dema sq-vlad.gre
9   ved [VVN] are [VBR] likely [JJ] to [TO] recognise [VVI] that [CST] sq-vlad.gre
10  that [CST] it [PPH1] is [VBZ] also [RR] recognised [VVN] that [CST sq-vlad.gre
```

25.5 The problem here is that it is not clear whether or not a similar refinement has been implemented in Biber 1988 (see Biber D 1988:242) as no mention is

made of the need to qualify the search in such a way. When Biber is counting Private Verbs is he counting *his* understanding of Private Verbs or *mine*? More of this anon.

25.6 If Private Verbs were an apparently simple case that became complicated, another class in the sample above started complicated and stayed that way. "THAT deletion" presents us with the major problem of counting something that isn't there. In an earlier study (Tribble 1985) I identified post-modifying participle clauses as a feature which appeared to distinguish expert ➔ expert discourses from expert → non-expert written exchanges. In that very restricted study all counts were manual, so, although it was tedious, it was possible to identify unambiguously any instances of present or past participiles immediately following a subject.

25.7 Biber 1988 addresses the problem of counting *that* deletions by using 3 separate algorithms:

- PUB/PRV/SUA + (T#) + demonstrative pro/SUBJPRO – i.e. count all verbs in the public/private/suasive lists + a tone marker + either demonstrative pronoun or subject pronoun (used in searching transcriptions of spoken production)

- PUB/PRV/SUA + PRO/N + AUX/V i.e. count all verbs in the public/private/suasive lists + all pronouns or nouns + all auxiliary verbs or verbs in the dictionary

- PUB/PRV/SUA + ADJ/ADV/DET/POSSPRO + (ADJ) + N + AUX/V i.e. count all verbs in the public/private/suasive lists + adjective or adverb or determiner or possessive pronoun + (optional adjective) + noun + auxiliary or verb.

Within the constraints of his own system of corpus mark-up and search software, this produced an acceptably accurate count of the feature in question.

25.8 Working with CLAWS marked up text and my own toolkit I had to develop two new algorithms in order to replicate Biber's search patterns as closely as possible. These were:

- [P* * [V*/N* * [V*

  Context word [V*/~[CST*/~[?]/~[VD*/~[VM*/~[VH*/~[VVN*/~[VVG* 2

  LEFT Manually edit to identify private/public/suasive words

  *– i.e. count all instances of either a PRONOUN + VERB, or NOUN + VERB – NOT immediately preceded by VERB or THAT (as conjunction), or PUNCTUATION, or DO, or MODAL, or HAVE, or PAST PARTICIPLE, or *ING verb and then sort to identify PUBLIC/PRIVATE/SUASIVE VERBS*

- C:\WSMITH\SEARCH\V-PPS.TXT Context Word [J*/[R*/[AT*

  – i.e. count all instances of any PUBLIC/PRIVATE/SUASIVE VERBS immediately followed by ADJECTIVE + VERB / ADVERB + VERB, or ARTICLE + NOUN + VERB

25.9 The algorithms I had developed seemed, therefore, to be catching the strings I was looking for, but with the way I had designed my research I could not be sure. I was, perhaps, following the method used in Biber 1988, or perhaps I was not. My algorithms might appear to be internally consistent, but I could not guarantee they would produce reliable results for anybody else. Although the numbers that were produced looked reasonable – and certainly positioned the texts in the PP Corpus in relation to other texts in LOB+ in ways which

were predictable – I was not sure that they were really counting the same things. The charts below summarise results for this first attempt[17].

Factor 1

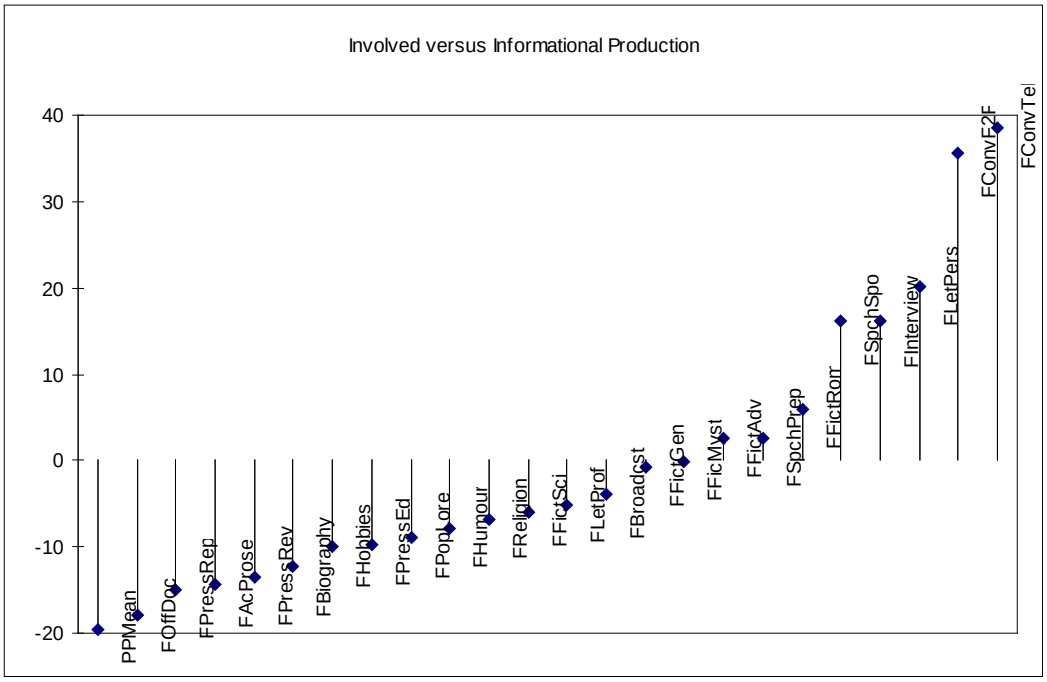


Figure 12 - Involved vs. Informed: first attempt

Factor 2

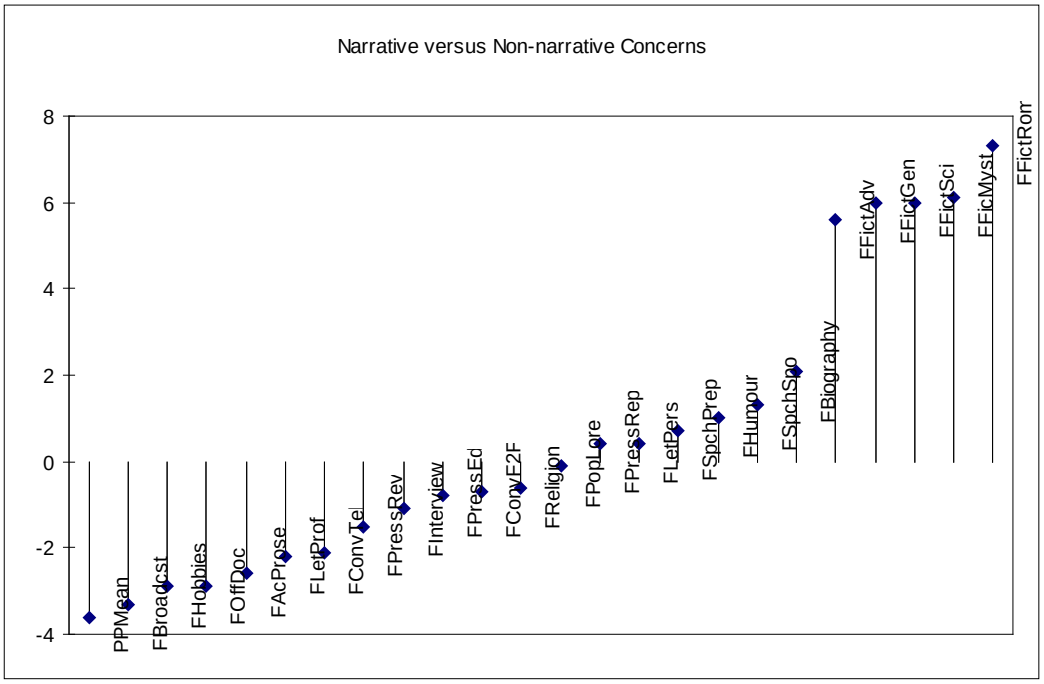


Figure 13 - Narrative vs. Non-narrative: first attempt

[17] In the charts PPMean refers to results from the PP corpus when plotted against the 23 genre categories Biber 1988 includes in LOB+

Factor 3

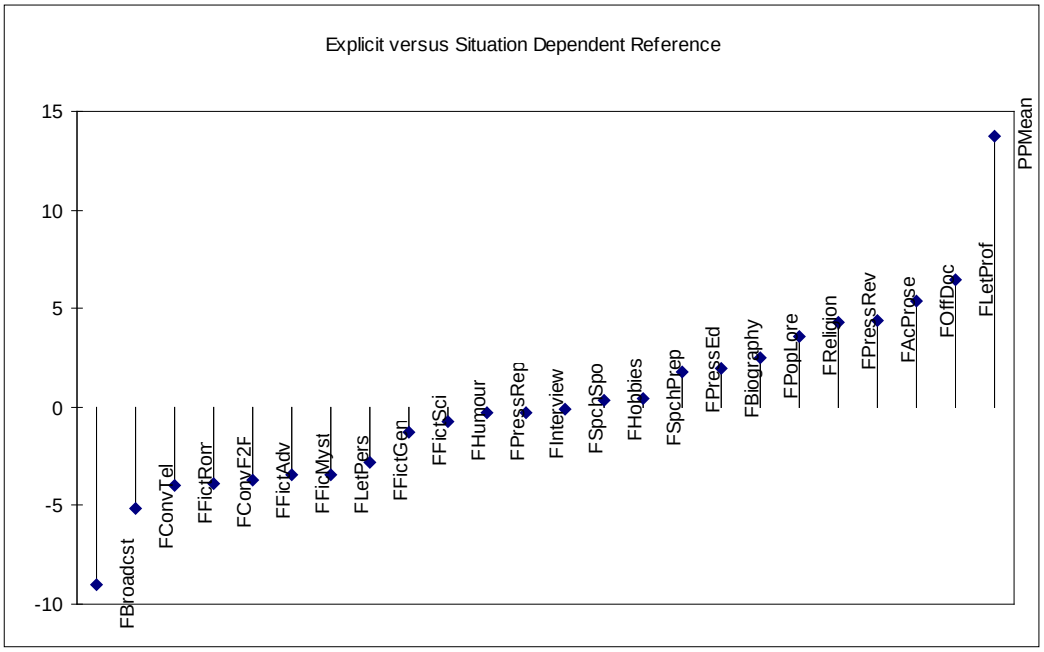


Figure 14 - Explicit vs. Situation Dependent: first attempt

Factor 4

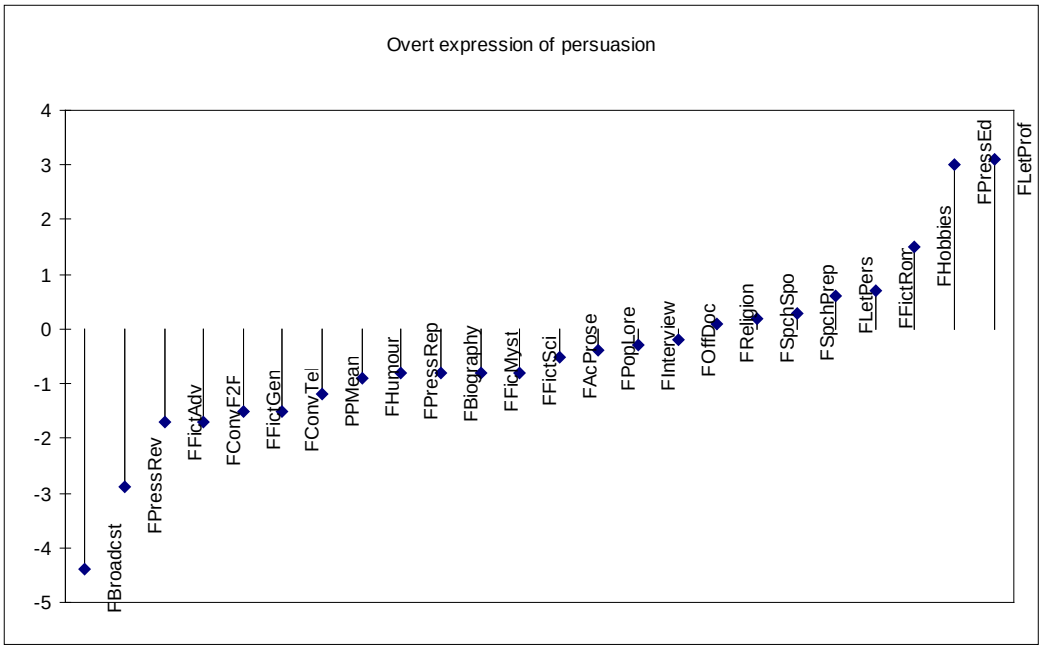


Figure 15 - Overt expression of persuasion: first attempt

Factor 5

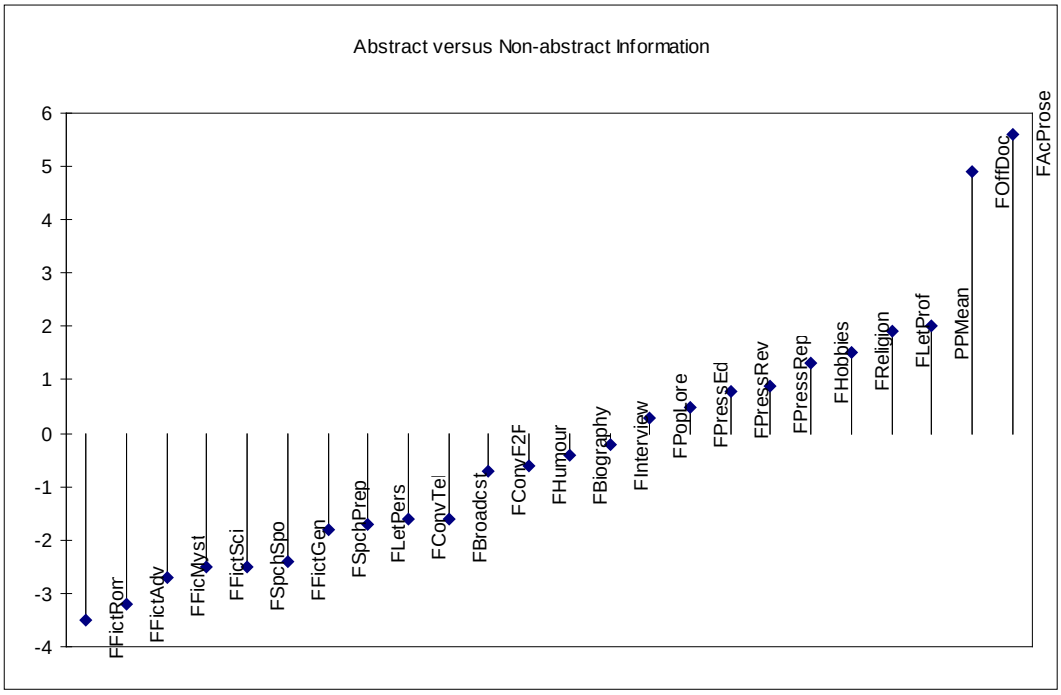


Figure 16 - Abstract vs. non-abstract: first attempt

Factor 6

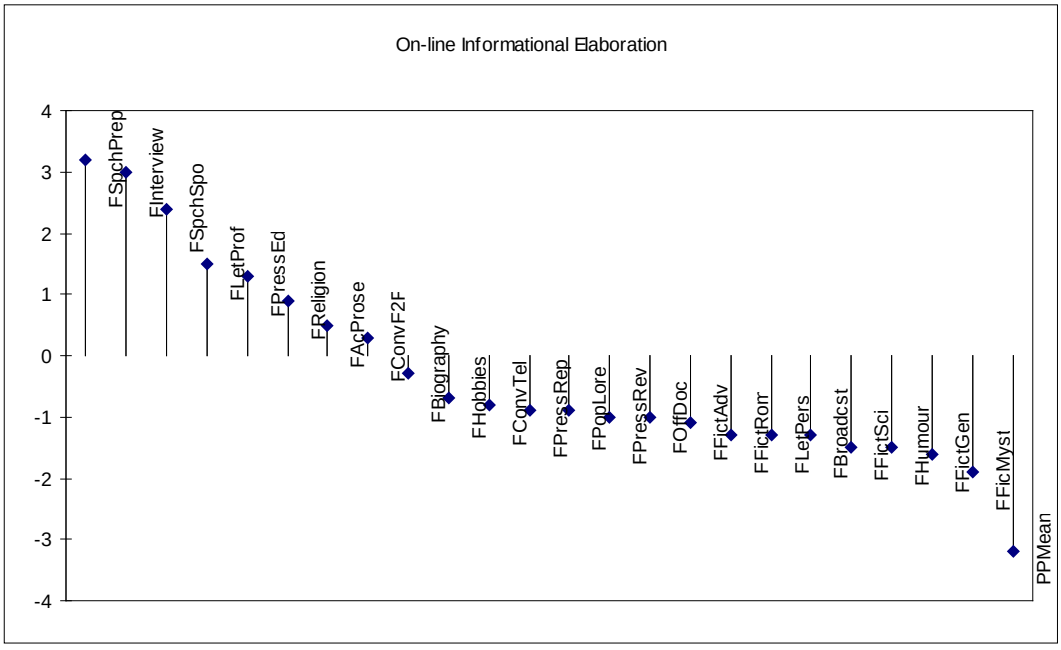


Figure 17 - On-line informational elaboration: first attempt

25.10 The important fact about each of these charts is that they were completely unsurprising. In each area which Biber and Finegan (1989) consider critical to the differentiation of written texts (Factors 1,3,5) the Project Proposals are

situated at the "written" extreme of a cline between "spoken" and "written" styles. The five closest texts[18] in each of these factors are:

Factor 1		**Factor 3**		**Factor 5**	
Pressrev	13.5	Pressrev	4.3	Religion	1.5
Acprose	14. 4	Acprose	4.4	Letprof	1.9
Pressrep	14.9	Offdoc	5.4	PPmean	2
OffDoc	17.9	Letprof	6.5	Offdoc	4.9
PPmean	24.5	Ppmean	13.9	AcProse	5.6

Table 11 - Neighbouring Genres

25.11 In this first analysis then, Project Proposals emerge as giving a strong emphasis on information in their orientation towards the reader, extremely explicit and placing little dependence on the extra-textual context for their interpretation, and addressing predominantly abstract information. Predictably, they share strongly similar characteristics with birds of similar feathers. (Printouts of the datasheets underlying the charts above are provided in the Appendices as Excel Spreadsheet 2 - Corpus Results Original Counts (XL Corpus Results Original Counts.xls)

26. Improving the research design

26.1 The key resources which I had assembled at the beginning of this research exercise were:

- the results reported in Biber 1988 and Biber and Finegan 1989
- two electronic corpora of 14 Project Proposals – one with POS mark up, the other as plain text.
- a set of search algorithms for use with the two main software tools I had available for analysis.

[18] Pressrev = Press Reviews / Acprose = Academic Prose / Pressrep = Press Reports / Offdoc =

26.2 These resources had initially appeared to be adequate, but once I had done my first batch of counting and summarised the results I faced two questions. One was "So what?". I had managed to produce results which confirmed what was pretty predictable – Project Proposals are formal, written texts and it is not surprising that they share many of the characteristics of other formal written texts. To what extent would such a (laboriously achieved) analysis help resolve any of the five issues I have identified Chapter 2? The other question was the one that I discussed in the first part of this chapter. Was I counting the same things as Biber?

26.3 In order to answer the first question, I needed to go to a level lower than the summary of the data and to look at individual counts of specific strings. It was here that it will be possible to find pedagogically useful insights into the way in which the PP corpus differs from its apparent analogues in LOB+. However, in order to feel safe making generalisations based on these numbers, I also needed to answer the second question and feel sure that my counts matched those of Biber's. Without this certainty my comparisons would remain dangerously speculative. We seem to have been here before …

26.4 The way out of the problem was to take a set of data from the original Biber study – i.e. a genre from the LOB+ corpus – and to repeat my analysis on this data. If the results I obtained through this process replicated those of Biber I would feel safer about my own algorithms. If they did not, then I could re-design and improve my algorithms in order to get better results, or, in the event that there was a fundamental problem with Biber's numbers (e.g.

resulting from the lower accuracy of his tagging), to make a statistical adjustment to my data to make it more reliably comparable with his.

26.5 In order to implement this addition to my study I selected one of the genres used in LOB+ – Romantic Fiction LOB Category P. Part of my reason for doing this was that it would provide me with the means of making more interesting statements about the style of the PP corpus in later discussions if I was working with a fairly extreme comparator. I arranged with Martin Wynne at UCREL to have the whole genre POS tagged with the CLAWS7 tag set and then took the first 30,000 words – the amount of text included in LOB+ (Biber D 1988:67 & Biber D 1988:209) [19]. This has proved to be a slight problem as Biber's original results for the corpus were based on means of totals for each component text rather than for totals of the whole text. Given the laborious process involved in building the counts I have decided for the moment to stick with the results arrived at from a total rather than for means. Even with this potential source of inconsistency, the results were illuminating.

26.6 The set of counts obtained with the first algorithms are given in Table 12. The areas where there seemed to be a major discrepancy between my own counts for Romantic Fiction and those obtained by Biber are summarised in Table 13. I have included counts for text types closely associated with PPs in order to provide a basis for comparison. The column labels stand for:

Corp Mean	**PP Mean**	**Press Rev**	**Off Doc**	**Ac Prose**	**Fict Rom**	**ROM FIC**	**% diff.**
Mean of all LOB + scores	*Mean PP Corpus scores*	*LOB Press Reviews*	*LOB Official Docments*	*LOB Academic Prose*	*LOB Romantic Fiction*	*My counts for LOB Romantic Fiction*	*Difference between Fict ROM and ROMFICT*

[19] There is a slightly confusing inconsistency in Biber 1988 in that at page 67 it is stated that there are 13 texts from this category in the LOB+ corpus and at page 209 Biber reports that there are 14 (the first 30,000 words).

Linguistic feature	CorpMean	PPMean	PressRev	OffDoc	AcProse	FictRom	ROMFIC
present-tense verb	***77.7***	***37.7***	***70.90***	***59.10***	***63.70***	***65.80***	***35.01***
first-person pronoun	27.2	5.5	7.50	10.00	5.70	32.40	28.84
second-person pron	9.9	0.2	1.20	1.40	0.20	18.60	16.40
pronoun IT	10.3	3.0	7.90	3.20	5.90	9.80	12.54
demonstrative pro	***4.6***	***15.3***	***1.90***	***1.10***	***2.50***	***2.60***	***13.70***
indefinite pronouns	1.4	0.0	1.00	0.20	0.20	2.30	3.95
DO as pro-verb	3.0	0.5	1.10	0.60	0.70	3.70	2.70
WH questions	0.2	0.2	0.00	0.00	0.00	0.20	0.39
BE as main verb	***28.3***	***7.3***	***25.50***	***16.50***	***23.80***	***28.10***	***16.11***
WH clauses	0.6	3.6	0.40	0.20	0.30	0.50	4.69
sentence relatives	0.1	0.1	0.00	0.10	0.00	0.00	0.00
adv. sub. - cause	1.1	0.1	0.20	0.10	0.30	3.20	0.64
hedges	0.6	0.0	0.40	0.00	0.20	0.10	0.51
amplifiers	2.7	0.9	2.00	0.90	1.40	2.20	2.86
emphatics	6.3	1.4	6.50	4.00	3.60	6.80	6.75
discourse particles	1.2	0.0	0.20	0.00	0.00	0.00	0.96
possibility modals	5.8	2.2	3.50	5.00	5.60	6.50	7.59
private verbs	18.0	3.5	11.30	7.80	12.50	24.20	22.09
contractions	13.5	0.0	1.90	0.00	0.10	19.00	19.68
THAT deletion	3.1	0.1	0.50	0.80	0.40	5.20	4.15
stranded preposition	2.0	0.0	0.50	0.30	1.10	1.50	0.26
non-phrasal coordin	4.5	0.6	1.80	1.20	1.90	2.80	1.86
analytic negation	***8.5***	***0.7***	***6.00***	***3.40***	***4.30***	***12.70***	***5.14***
nouns	180.5	314.8	208.30	206.50	188.10	146.80	140.08
prepositions	110.5	141.1	119.30	150.90	139.50	82.00	86.68
attributive adjectiv	***60.7***	***77.5***	***82.30***	***77.90***	***76.90***	***41.90***	***27.07***
type/token ratio	51.1	47.2	56.50	47.80	50.60	52.90	41.66
word length	4.5	5.4	4.70	4.90	4.80	4.10	4.09
Involved versus Informational Production							
past tense verbs	40.1	6.1	18.20	16.20	21.90	83.70	82.95
perfect aspect verbs	8.6	5.6	6.80	7.90	4.90	13.60	15.63
third person pronour	29.9	5.6	33.60	10.10	11.50	78.50	79.48
present participial cl	1.0	3.3	0.50	0.30	1.30	4.50	4.63
public verbs	7.7	1.5	4.90	4.90	5.70	8.60	7.33
synthetic negation	1.7	0.1	2.00	1.50	1.30	2.50	1.93
Narrative versus Non-narrative Concerns							
nominalizations	19.9	71.4	21.60	39.80	35.80	8.50	9.03
WH-relatives: subj.	2.1	1.5	3.50	2.70	2.60	0.80	0.58
WH-relatives: obj. p	1.4	0.4	2.60	2.00	1.30	0.40	0.23
WH-relatives: pied p	0.7	1.0	1.50	3.00	2.00	0.10	0.55
phrasal coordination	3.4	21.3	6.50	1.30	4.20	3.20	5.27
place adverbials	3.1	2.1	1.90	2.10	2.40	3.60	4.50
time adverbials	5.2	1.2	4.30	3.40	2.80	6.80	6.59
adverbs	65.6	18.0	60.80	43.70	51.80	78.40	64.91
Explicit versus Situation Dependent Reference							
infinitives	14.9	15.7	11.60	13.40	12.80	19.00	19.58
adv. sub. - conditic	***2.5***	***0.5***	***1.10***	***1.00***	***2.10***	***0.80***	***3.83***
necessity modals	2.1	1.0	1.10	2.20	2.20	1.90	2.38
predictive modals	5.6	13.5	3.00	4.90	3.70	8.50	8.68
suasive verbs	2.9	1.9	1.90	5.20	4.00	2.60	2.93
split auxiliaries	***5.5***	***1.7***	***4.80***	***5.70***	***5.80***	***6.00***	***3.34***
Overt expression of persuasion							
agentless passives	9.6	9.6	8.60	18.60	17.00	5.00	5.11
BY-passives	0.8	2.0	1.40	2.10	2.00	0.00	0.13
past participial adve	0.1	0.6	0.20	0.50	0.40	0.10	0.26
past prt WHIZ deleti	2.5	2.1	3.80	7.50	5.60	0.10	0.26
adv. sub. - other	***1.0***	***0.7***	***0.80***	***0.90***	***1.80***	***0.20***	***1.06***
conjuncts	1.2	1.4	1.20	1.20	3.00	0.10	0.48
Abstract versus Non-abstract Information							
THAT verb complen	3.3	1.0	1.80	1.40	3.20	2.50	3.02
THAT adj. complem	0.3	0.3	0.10	0.20	0.40	0.40	0.93
THAT relatives: obj	0.8	0.1	0.90	0.70	0.80	0.20	1.00
demonstratives	***9.9***	***2.5***	***8.70***	***9.60***	***11.40***	***7.30***	***3.38***
On-line Informational Elaboration							
SEEM/APPEAR	0.8		1.40	0.40	1.00	0.80	

Table 12 - Counts (First Algorithms)

Linguistic feature	*Corp Mean*	***PP Mean***	*Press Rev*	*Off Doc*	*Ac Prose*	***Fict Rom***	***ROM FIC***	*% difference*
present-tense verbs	77.7	**37.7**	70.90	59.10	63.70	**65.80**	**35.01**	***46.79%***
demonstrative pronouns	4.6	**15.3**	1.90	1.10	2.50	**2.60**	**13.70**	***-426.92%***
BE as main verb	28.3	**7.3**	25.50	16.50	23.80	**28.10**	**16.11**	***42.67%***
analytic negation	8.5	**0.7**	6.00	3.40	4.30	**12.70**	**5.14**	***59.53%***
attributive adjectives	60.7	**77.5**	82.30	77.90	76.90	**41.90**	**27.07**	***35.39%***
adv. sub. – condition	2.5	**0.5**	1.10	1.00	2.10	**0.80**	**3.83**	***-378.75%***
split auxiliaries	5.5	**1.7**	4.80	5.70	5.80	**6.00**	**3.34**	***44.33%***
adv. sub. – other	1.0	**0.7**	0.80	0.90	1.80	**0.20**	**1.35**	***575.33%***
demonstratives	9.9	**2.5**	8.70	9.60	11.40	**7.30**	**3.38**	***53.70%***

Table 13 - Major discrepancies

26.7 The nine features focused on here are those where there is a major difference between my counts and Biber's counts for the same data set (expressed as a percentage in the final column). Having established that there was indeed a major difference between my results and Biber's 1988 results I had a clear motivation for reviewing all of the algorithms I had been using – focusing in the first instance on those where there seemed to be gross discrepancies.

26.8 One of the lessons learned from this process has been how easy it is to produce internally consistent and plausible results from this sort of counting exercise. Moreover, as we shall see in the following discussion, there is still some question over what is being counted in the published data. The most depressing thing for me has been that in many instances discrepancies have not arisen from my failure to design an algorithm correctly (although this has also been the case) but for simple mechanical reasons – usually typing the right data into the wrong cell on the spreadsheet! A lesson in humility if nothing else …

27. Revised results

27.1 As mentioned above, in order to check if the variance between my counts for RomFict and those in Biber 1988 were the result of inaccuracies in my algorithms (most likely case) or in his data (least likely case), I double-

checked and, where necessary, re-worked each algorithm in my set (even in those cases where there was a very close correlation between Biber's results and mine – e.g. Agentless passives: Biber – 5.00 per 1000 / Tribble – 5.11 per thousand). This produced a significant improvement in the counts for the "problem" set and also improved the overall match of my numbers with those in Biber 1988.

27.2 In the light of the earlier discussion (para 25.10) it was interesting to note the relatively small overall impact of these corrections to the data on the positioning of the PP corpus in relation to LOB+. Once again I give the environments in which PPs are placed for the three key Factors (Table 14) and the charts which plot the relationship between PP and other genres in the LOB+ Corpus (Factors 1 to 6, Second Version – see below). It is immediately obvious that although there has been a major improvement in the quality of my algorithms and a corresponding reduction in the mismatch between my counts and Biber's (see Table 15 below), this improvement has not had a significant impact on the overall profile of the PP Corpus. The flawed counts had been "good enough" to permit the categorisation of the PP Corpus and to demonstrate the (expected) linguistic contrast between PP and RomFict.

Factor 1		**Factor 3**		**Factor 5**	
no change from Version one results (other than a reduction in the distance between PP and Official Document)		*no change + still a considerable distance between PP and Professional letter*		*no change + still a considerable distance between PP and Professional letter*	
PressRev	-13.5	PressRev	4.3	Religion	1.5
AcProse	-14.4	AcProse	4.4	LetProf	1.9
PressRep	-14.9	OffDoc	5.4	PPMean	2
OffDoc	-17.9	LetProf	6.5	OffDoc	4.9
PPMean	-19.6	PPMean	13.7	AcProse	5.6

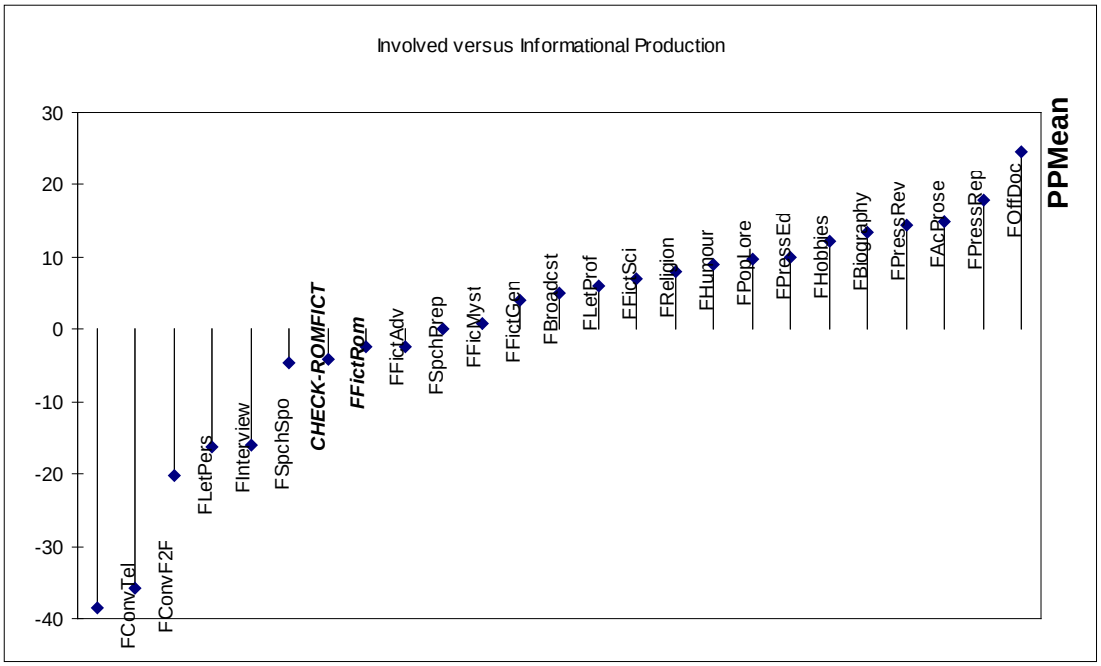
Table 14 - Neighbouring texts

Linguistic feature	*Corp Mean*	***PP Mean***	***Fict Rom***	***ROM FIC***	*current % difference*	*previous % difference*
present-tense verbs	77.7	**57.10**	**65.80**	**55.51**	***15.64%***	*46.79%*
demonstrative pronouns	4.6	**2.50**	**2.60**	**3.44**	***-32.31%***	*-426.92%*
BE as main verb	28.3	**16.40**	**28.10**	**27.05**	***3.74%***	*42.67%*
analytic negation	8.5	**0.73**	**12.70**	**13.70**	***-7.87%***	*59.53%*
attributive adjectives	60.7	**77.5**	**41.90**	**27.07**	***35.39%***	*35.39%*
adv. sub. – condition	2.5	**0.5**	**0.80**	**3.83**	***-378.75%***	*-378.75%*
split auxiliaries	5.5	**1.59**	**6.00**	**3.54**	***41.00%***	*44.33%*
adv. sub. – other	1.0	**0.76**	**0.20**	**1.35**	***575.33%***	*575.33%*
demonstratives	9.9	**8.61**	**7.30**	**8.17**	***-11.92%***	*53.70%*

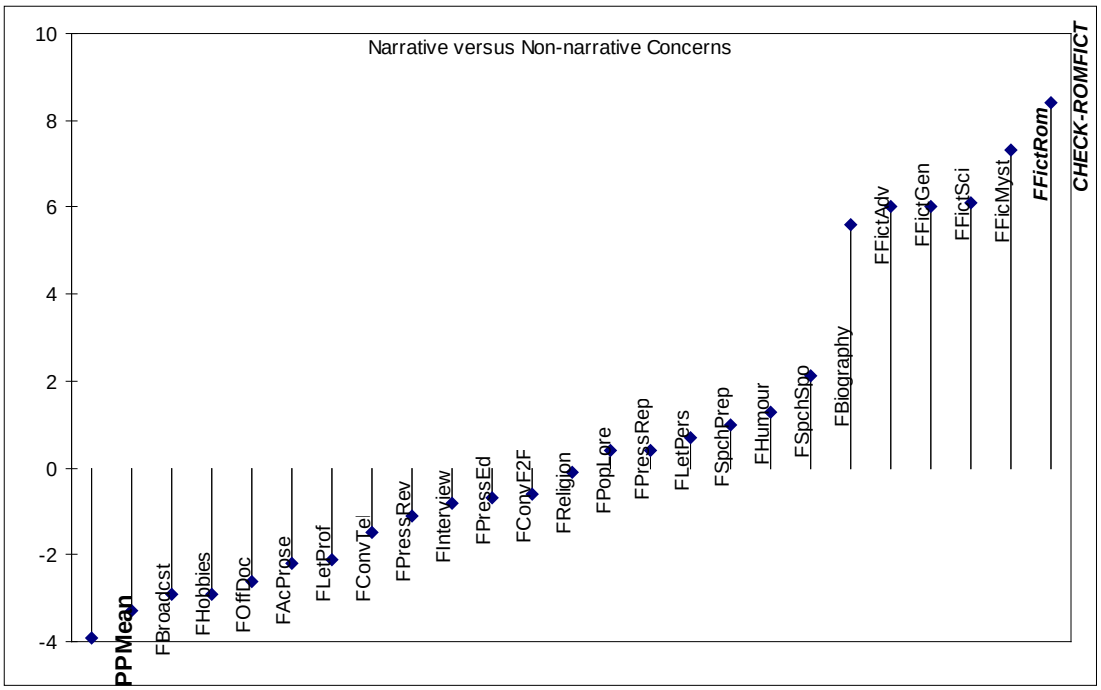
Table 15 – Discrepancies

27.3 Charts summarising the revised Factor scores are given below:

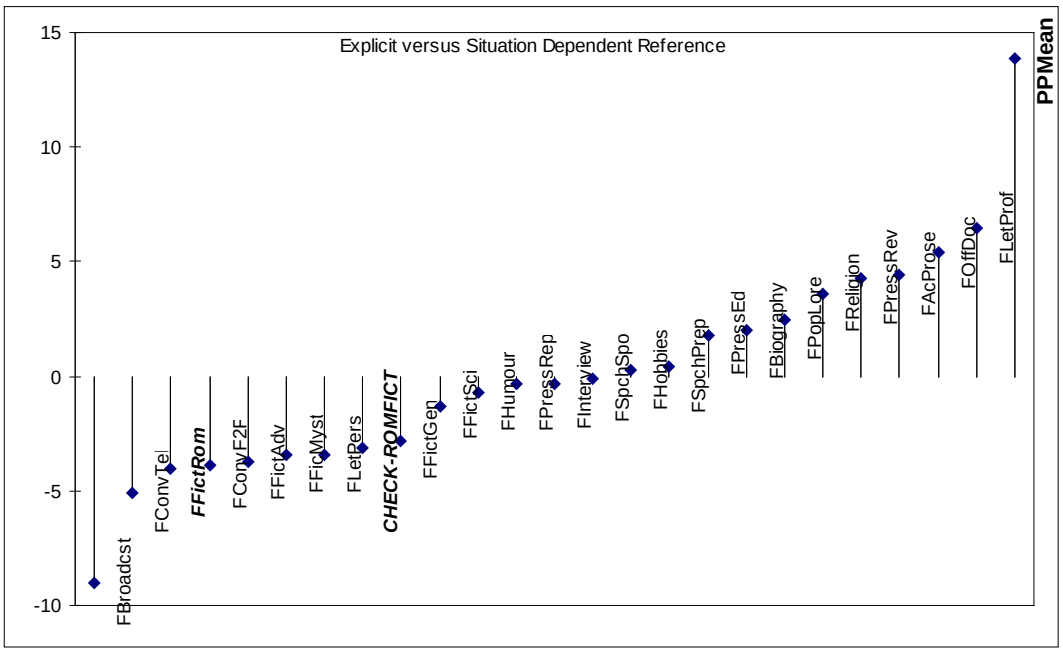
Factor 1 – Version 2



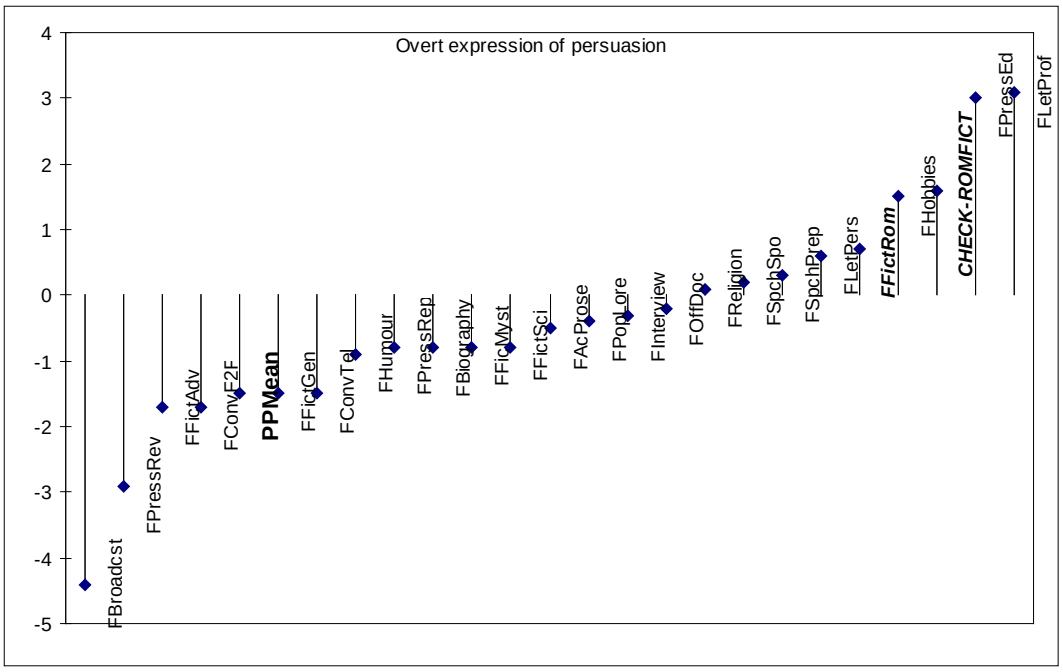
Factor 2 – Version 2



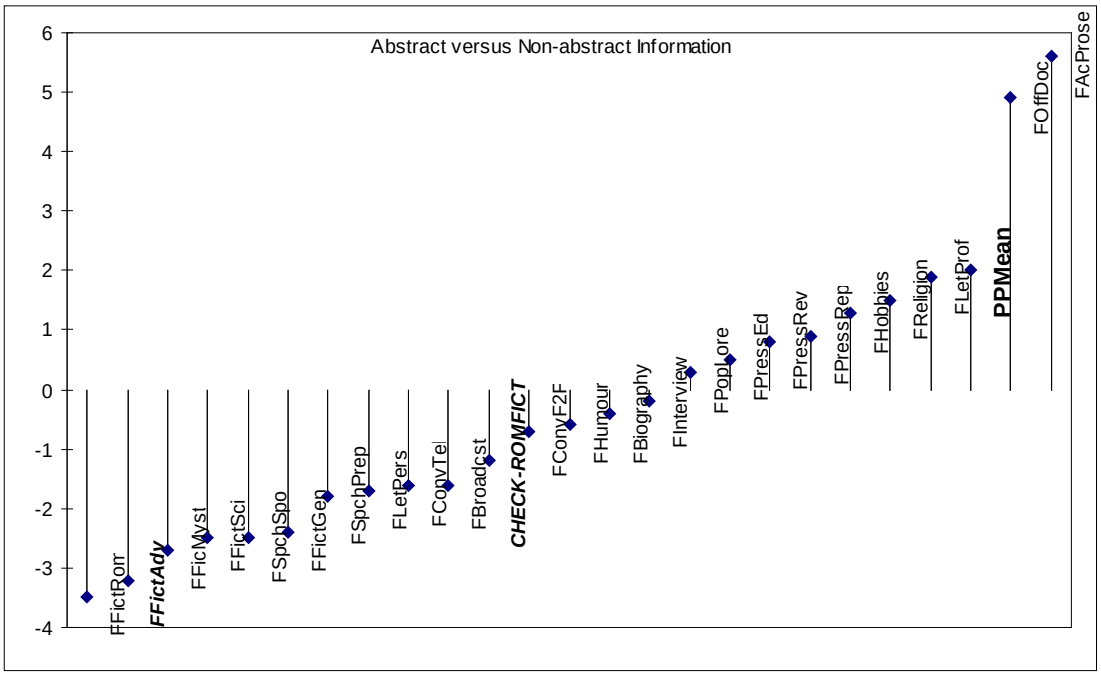
Factor 3 – Version 2



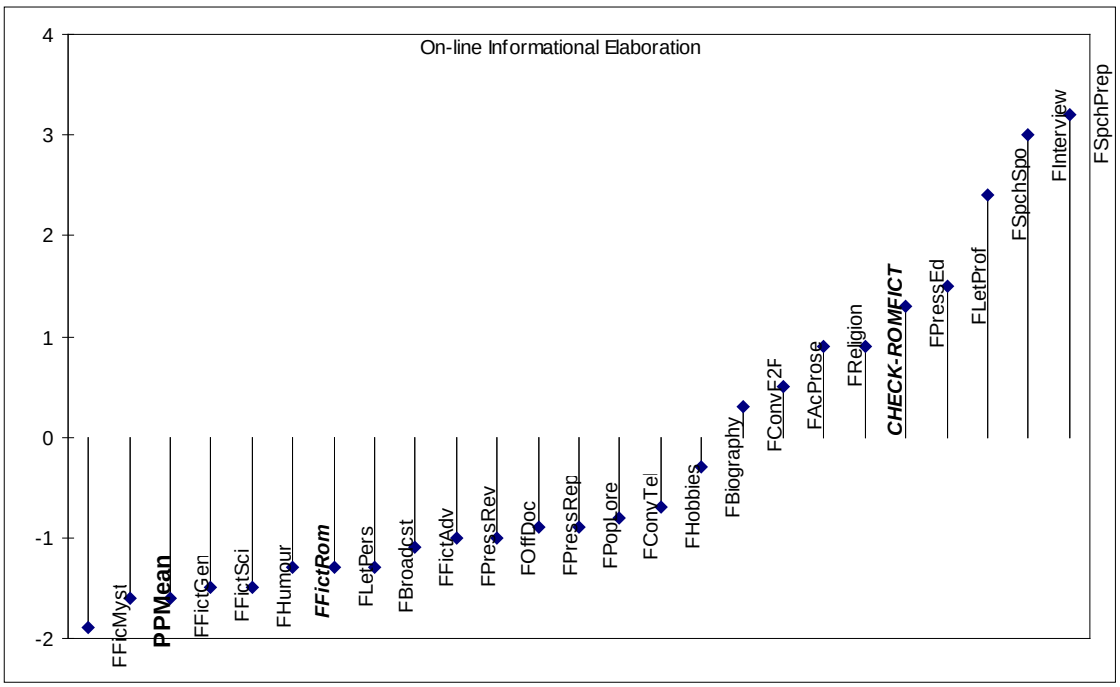
Factor 4 – Version 2


Overt expression of persuasion
4
3
2
1
0
-1
-2
-3
-4
-5
FBroadcst
FPressRev
FFictAdv
FConvF2F
PPMean
FFictGen
FConvTel
FHumour
FPressRep
FBiography
FFicMyst
FFictSci
FAcProse
FPopLore
FInterview
FOffDoc
FReligion
FSpchSpo
FSpchPrep
FLetPers
FFictRom
FHobbies
CHECK-ROMFICT
FPressEd
FLetProf

Factor 5 – Version 2



Factor 6



28. Conclusion

28.1 If the eventual result of this revision of the search algorithms has been so slight, has it worth the effort? In brief – "yes". I say this for two reasons. The first is that it has given me an insight into the problem of handling such large

amounts of data status of the published data with which I am working. In all but three instances – Conditional adverb subordinator (adv. sub. - condition), Other Adverbial Subordinators (adv. sub. - other), and Attributive Adjectives – it was possible to effect a significant improvement in the quality of the search algorithms and a concomitant improvement in the match of the Biber results and my own. In the case of these two linguistic features, the search algorithms are simple, unambiguous search strings:

• adv. sub. - cause	*count all instances of:* because
• adv. sub. - other	*count all instances of:* since / while / whilst / whereupon / whereas / whereby / such that / so that / inasmuch as / forasmuch as / insofar as / insomuch as / as long as / as soon as

28.2 This being the case, either my Romantic Fiction texts are different from those in Biber 1988 – unlikely, as they are the same as the set Biber specifies (Biber D 1988:209), the first 30,000 words of Category P in LOB, and also there is such a close match between the rest of my counts – or there are errors in the published data in Biber 1988 in the table at page 260. If there are errors in the RomFict results, are there more as yet unaccounted for, and are the results of Biber 1988 open to serious challenge? I think not. For a start, many of the minor differences between Biber's results and mine will arise from the unavoidable differences between our separate studies:

- the specific differences of *hardware, software* and *text mark-up* mentioned above

- Biber's use of means as opposed to my use of totals in presenting the results for RomFict

Where the differences between our results appear to be errors in Biber's reported data, it is unlikely that there are many such mistakes – and part of the great strength of the design of Biber 1988 is that he deals with such a large number of separate factors, and bases his conclusions on the results of statistical procedures which have been employed in order to reduce the risk of distortion and bias in the results which might arise from inaccurate sampling. A conclusion from this argument is that – given the impossibility of exactly replicating Biber's 1988 study for reasons already discussed – those of us who follow his work do not need to be unduly concerned about minor differences between his counts and our counts, so long as we have made an effort to be as scrupulously accurate as possible and have presented our results in such a way that, this time, they can be reproduced and tested by others. The fundamental principal holds – a multifunction, multivariate approach to corpus study has much to recommend it and is robust enough to cope with minor inaccuracies in the counts..

28.3 The second reason for feeling that the revision of the algorithms has been effort well spent is that I feel more confident in following up insights that emerge from an analysis of the counts of the individual features which contribute to the factor scores which have positioned PP at the extreme end of the written / spoken cline. It had never been my purpose to replicate the results from Biber 1988. It has been interesting to discover that it is so difficult to do this, but not important from my point of view. The factor scores for PP were predictable and there was no need to use the Biber tool to conclude that PPs are formal written texts. What is interesting is that the counts for many of the individual items which go to make up these scores

were surprising and not expected. The predictable positioning of PPs in relation to other texts in LOB+ is the outcome of specific choices that writers have made about the words they want to use. These choices are the things that interest me, and it is the surprising counts of the strings which mark the trace of these choices that will form the basis for the next chapter.

SECTION TWO: AN ACCOUNT OF THE TEXTS

Chapter 5: Grammar and Style

> "A rough indication of frequencies is often just what is needed: enough to suggest why we should accept the writer's assertion that that some feature is prominent in the text, and to allow us to check his statements. The figures, obviously, in no way constitute an analysis, interpretation or explanation of the style." (Halliday 1973:117)

29. Introduction

29.1 In the previous chapter we discussed issues that arose from differences between the results of the study carried out in Biber 1988 and this present study. We established that these differences were the result of contrasts between the text mark-up, software and hardware used in the two studies – and also from possible inaccuracies in the published research data. In addition we established that although there were some major differences between the present results and the results in Biber 1988, the data obtained from the analysis of the PP Corpus provide a sufficient basis for an identification of contrasts between the PP Corpus which are potentially salient in an analysis of PPs as a genre.

29.2 In this chapter we will consider the extent to which the analysis of the PP corpus provides information which might have pedagogic value. In the first instance, we will do this by comparing the linguistic profiles of the PP Corpus in relation to the profiles of the six text genres which are located most closely to the PP corpus in each key text dimension in LOB+. These text genres are listed below in Table 16.

Dimension 1	Dimension 3	Dimension 5
Press Review	Press Review	Religion
Academic Prose	Academic Prose	Letters – Professional
Press Reports	Official Documents	Official Documents
Official Documents	Letters – Professional	Academic Prose

Table 16 - Neighboring texts (by Dimension)

29.3 We shall make this comparison in order:

- to identify those linguistic features in the PP Corpus which are statistically prominent, and

- to establish which of these features are salient to a stylistic typification of PPs.

In order to help teachers or other researchers to make use of this approach when studying genres which are of importance to them, I shall also give relatively detailed comments on some of the practical techniques I have used to develop these analyses.

30. Style and stylistics

30.1 The identification of *prominence* and a subsequent exegesis of *salience* is an approach that has been used successfully in literary stylistics (Halliday MAK 1973, Leech G & M Short 1981). It provides a means of identifying those linguistic features which might be considered as having *literary significance* (Leech G & M Short 1981: 48-51) or, in Halliday's terms, which are *foregrounded* in relation to a reader's probabilistically established expectations of textual patterning:

> "…. we are concerned not only with deviations, ungrammatical forms, but also with what we may call 'deflections', departures from some expected pattern of frequency." (Halliday 1973:113).

Those of us with an interest in teaching students to deal with a range of referential texts are also interested in such departures from "some expected pattern of frequency" as it is precisely this sort of understanding that apprentice writers need as they develop a performance repertoire in a particular genre.

30.2 It is also of relevance to this study that these earlier analyses do not attempt to reference their statistical data against a corpus which purports to be representative of the "language as a whole" – indeed Leech & Short argue forcefully against the possibility of such a procedure:

> "Without some clear cut notion, for statistical purposes, of what is meant by "the language as a whole" any sampling procedure is bound to involve subjective decisions. The norm of 'the language as a whole' is not the reality that it seems to be …." (Leech G & M Short 1981:43)

Instead, Leech & Short consider a range of possible norms – the most important of which for our purposes are *relative norms*:

> "Where an absolute norm cannot be relied on, the next best thing is to compare the corpus whose style is under scrutiny with one or more comparable corpuses, thus establishing a relative norm." (ibid:51)

30.3 Using a methodology similar to that elaborated in Biber 1988 (though it lacked the multivariate / multidimensional delicacy of this later study) they demonstrated that it was both possible and interesting to literary scholars to identify contrasts between the linguistic features of three small corpora of work by three authors – Conrad, Lawrence and James (ibid:74-118) – and to use this as the basis for a discussion of stylistic difference between these authors. The procedure involved developing a list of sixty nine textual features and identifying saliently prominent differences between the counts of

these features in the three text collections prior to a discussion of the literary style of the three authors. Such a comparative approach accords strongly with the positions taken in e.g. Fox G 1993 and Stubbs M 1996:127, and constitutes a useful starting point for the following discussion of the differences between PPs and the text genres with which they stand in close relation.

31. Provisional identification of prominent linguistic features of Project Proposals

31.1 The core data with which we shall be working are the means of counts (normalised to counts per 1000) for the 58 linguistic features in Biber 1988 across all the genres under discussion (i.e. PP + its near neighbours in LOB+). These data are given in ranked order in Table 17 below. In order to achieve this ranking the procedure outlined below was followed[20]:

- the minimum, maximum and mean scores for the six genres in the reference set (PressRep / PressRev / Religion / OffDoc / AcProse / LetProf) were calculated (*Excel* MIN, MAX, AVERAGE functions)

- those counts in PP which are greater than the MAX or less than the MIN scores for the reference set were identified (*Excel* OR function)

- percentage values for the difference between the PP mean and the reference set MIN or MAX scores were calculated

- the table was sorted in descending order by the ">MAX" and "<MIN" columns to produce the results in Table 17.

[20] I have given the Microsoft Excel functions here for the benefit of others who may have an interest in replicating parts of this study

31.2 Given below are the ranked scores (PP vs. PressRep / PressRev / Religion / OffDoc / AcProse / LetProf) plus means for all language features implicated in all 6 Factors:

	Linguistic feature	PP mean	REF Max	REF Min	> Max	% diff	< Min	% diff	REF Mean	REF Range
+3	phrasal coordination	21.88	6.5	1.3	1	236.62	0		4.48	5.2
+2	present participle clauses	2.68	1.3	0.3	1	106.15	0		0.57	1
+3	nominalisations	69.4	44.2	19.2	1	57.01	0		31.23	25
+5	past participle adverbial clauses	0.66	0.5	0.1	1	32.00	0		0.27	0.4
-1	nouns	282.93	220.5	172.6	1	28.31	0		197.28	47.9
-1	attributive adjectives	93.24	82.3	59.5	1	13.29	0		72.93	22.8
+4	predictive modals	13.47	11.9	3	1	13.19	0		5.93	8.9
-1	word length	5.33	4.9	4.5	1	8.78	0		4.73	0.4
+1	stranded prepositions	0	1.1	0.1	0		1	100.00	0.50	1
+2	synthetic negation	0.06	2.8	1	0		1	94.00	1.68	1.8
+1	non-phrasal coordination	0.18	2.9	1.2	0		1	85.00	1.93	1.7
+1	THAT deletion	0.08	2	0.4	0		1	80.00	1.08	1.6
+3	WH-relatives: obj. position	0.24	2.9	1	0		1	76.00	2.02	1.9
+1	indefinite pronouns	0.05	1.1	0.2	0		1	75.00	0.67	0.9
+2	public verbs	1.52	12	4.9	0		1	68.98	7.68	7.1
+4	split auxiliaries	1.59	6	4.8	0		1	66.88	5.60	1.2
+6	THAT relatives: obj position	0.21	1.1	0.6	0		1	65.00	0.82	0.5
+1	WH clauses	0.07	1	0.2	0		1	65.00	0.52	0.8
+1	amplifiers	0.33	2	0.9	0		1	63.33	1.43	1.1
+1	emphatics	1.38	7.8	3.6	0		1	61.67	5.02	4.2
+1	DO as pro-verb	0.23	2.6	0.6	0		1	61.67	1.42	2
-3	adverbs total	17.49	60.8	43.7	0		1	59.98	52.17	17.1
+1	private verbs	3.46	17.1	7.8	0		1	55.64	12.27	9.3
+4	adv. sub. – condition	0.55	2.1	1	0		1	45.00	1.52	1.1
+2	past tense verbs	6.12	45.1	10.1	0		1	39.41	23.00	35
+1	analytic negation	0.73	7	1.2	0		1	39.17	4.43	5.8
+2	third person pronouns	5.62	33.6	8.7	0		1	35.40	19.18	24.9
-3	time adverbials	1.32	6.5	2	0		1	34.00	3.68	4.5
+1	second-person pronouns	0.14	15.2	0.2	0		1	30.00	3.67	15
+6	THAT verb complements	1.1	4.3	1.4	0		1	21.43	3.05	2.9
+3	WH-relatives: subj. position	1.6	3.5	2	0		1	20.00	2.83	1.5
+4	necessity modals	1.01	2.2	1.1	0		1	8.18	1.93	1.1
+1	pronoun IT	3.02	9.6	3.2	0		1	5.63	6.58	6.4
+5	adv. sub. –	0.76	2.9	0.8	0		1	5.00	1.37	2.1

	Linguistic feature	PP mean	REF Max	REF Min	> Max	% diff	< Min	% diff	REF Mean	REF Range
+1	other possibility modals	3.11	7.7	3.2	0		1	2.81	5.07	4.5
+1	first-person pronouns	5.56	40.9	5.7	0		1	2.46	15.03	35.2
-1	type/token ratio	46.9	56.5	47.8	0		1	1.88	52.22	8.7
+1	BE as main verb	16.4	30.4	16.5	0		1	0.61	23.98	13.9
-1	prepositions	141.69	150.9	116.6	0		0		127.15	34.3
+1	present-tense verbs	57.1	94.7	55.8	0		0		70.60	38.9
+4	infinitives	14.31	24.1	11.6	0		0		15.10	12.5
+5	agentless passives	10.16	18.6	7.3	0		0		12.85	11.3
+6	demonstratives	8.61	13.5	7.4	0		0		10.43	6.1
+2	perfect aspect verbs	5.66	10.3	4.9	0		0		7.33	5.4
+1	demonstrative pronouns	2.5	3.6	1.1	0		0		2.18	2.5
+5	past prt WHIZ deletions	2.05	7.5	1.3	0		0		4.12	6.2
-3	place adverbials	2	4.7	1.6	0		0		2.50	3.1
+4	suasive verbs	1.94	5.2	1.9	0		0		3.45	3.3
+5	BY-passives	1.84	2.1	0.6	0		0		1.47	1.5
+5	conjuncts	1.32	3	0.6	0		0		1.73	2.4
+3	WH-relatives: pied pipes	0.97	3	0.6	0		0		1.48	2.4
+6	THAT adj. complements	0.29	0.5	0.1	0		0		0.25	0.4
+1	adv. sub. – cause	0.12	2	0.1	0		0		0.62	1.9
+1	sentence relatives	0.06	0.1	0	0		0		0.02	0.1
+1	hedges	0.02	0.4	0	0		0		0.18	0.4
+1	WH questions	0.01	0.1	0	0		0		0.02	0.1
+1	discourse particles	0	0.2	0	0		0		0.08	0.2
+1	contractions	0	4.7	0	0		0		1.72	4.7

Table 17 - Ranked scores

31.3 The rank orderings and counts given here are purely descriptive statistics and will have to be tested later from an inferential point of view. However, as a starting point they help identify those areas in which there may be interesting divergencies from relative norms – particularly if we remember Halliday's comment (given in the epigraph to this chapter) on the value of counting in stylistics.

31.4 Given this caveat, I propose then to examine the stylistic significance of those features which have a relatively high percentage difference from the maximum

score in the neighbouring texts *and* where there is a reasonably large number of counts in the overall corpus. Thus I shall be considering e.g. Phrasal Coordination (236.62% / 21.88 per 1000 mean) and Nominalisations (57.01% / 71.3 per 1000 mean), but not Present Participle clauses where although there is a high percentage difference (106.15%) the count is so relatively low at 2.68 per 1000, that it is improbable that any major stylistic impact could be identified for this feature.

32. Analysis of prominent features

32.1 Using the criteria outlined above, seven prominent features of PPs have been identified. They are listed in Table 18 and Table 19 along with a reference to the Factor with which they are associated (Column 1). The rest of this chapter will be devoted to a review of their potential salience to a typification of PPs.

Factor	Linguistic feature	PP mean	REF Max	REF Min	% diff
1	attributive adjectives	93.24	82.3	59.5	13.29
3	nominalisations	69.4	44.2	19.2	57.01
3	phrasal coordination	21.88	6.5	1.3	236.62
4	predictive modals	13.47	11.9	3	13.19

Table 18 - high frequency items

Factor	Linguistic feature	PP mean	REF Max	REF Min	% diff
3	adverbs total	17.49	60.8	43.7	59.98
2	third person pronouns	5.62	33.6	8.7	35.40
1	private verbs	3.46	17.1	7.8	55.64

Table 19 - low frequency items

33. Attributive adjectives

33.1 In order to get counts of individual attributive adjectives it was necessary to run the WordSmith Tools Wordlist program with Clusters set at 4 to create a 4 "word" wordlist for the PP Corpus. This list contained combinations of both "Word + POS-tag + Word + POS-tag" and "POS-tag + Word + POS-tag +

Word". This latter set was then be eliminated by means of the following procedure:

- Convert the WordSmith Wordlist to a plain .txt file
- Use WordSmith Text Converter to
- replace all tab codes with single spaces
- replace all single spaces with tab codes
- Open this tab-delimited file in Excel and format it so with the following headings and columns (the # column used the Excel LEFT function as a way of identifying the base form of tag in the adjacent column):

Attrib test	*No.*	*Word 1*	*#*	*TAG 1*	*Word 2*	*#*	*TAG 2*	*COUNT*	*char*	*2*	*3*	*4*	*5*
	1	LONG	J	JJ	TIME	N	NNT#	5	2	2	0	0	0

- Use Excel text and logical formulae to identify all instances of 2, 3, 4, and 5 character strings in the TAG 1 and TAG 2 columns (See Appendix 1 for a full listing of the formulae used and the results of this procedure) and test for those instances which contain attributive adjectives.
- This "word + tag + word + tag" list was then available for use in the discussion of attributive adjectives and other features in the PP. The procedure outlined above was also applied to the RomFict data set for purposes of comparison. It appears to be a robust and easily accessible means for identifying and quantifying strings in POS tagged texts.

33.2 Biber describes attributive adjectives as one of the devices "used for ideas integration" (Biber D 1988:237) and categorises attributive adjectives

themselves as being "highly integrative in their function" (ibid:237). It seems, however, that this view of attributive adjectives only accounts for one aspect of the function of these terms in the development of discourse as there appears to be a significant contrast between the roles of the two main classes of attributive adjectives – *qualitative* and *classifying*.

33.3 In the discussion in Sinclair et al 1990:64-72 attributive adjectives are divided into two main classes – *qualitative* adjectives (which identify a quality that something has) and *classifying* adjectives (which identify the class something belongs to). Taking those attributive adjectives which occur in the PP Corpus with frequencies of 20 or more across all 14 texts (see Table 20), the majority (33:4) are classifying adjectives. What is more, within the very small count for qualifying adjectives, "short" is part of the pair "short-term" (having been separated from its partner in CLAWS processing), and would in other circumstances count as a classifying adjective. An informal survey of this trend confirms that it is sustained throughout the set (see Table 21 below for a list of the last ten items in the 852 types accounted for in the attributive adjective set in the PP Corpus).

Word 1	TAG 1	Class	Word 2	TAG 2	COUNT
WIDE	JJ	***Q***	RANGE	NN#	25
SHORT	JJ	***Q***	TERM	NN#	24
CONSIDERABLE	JJ	***Q***	EXPERIENCE	NN#	23
KEY	JJ	***Q***	ISSUES	NN#	20
BRITISH	JJ	***C***	*{institution name}*	NNJ	134
TECHNICAL	JJ	***C***	ASSISTANCE	NN#	131
ENVIRONMENTAL	JJ	***C***	EDUCATION	NN#	131
SLOVAK	JJ	***C***	REPUBLICS	NN#	55
INSTITUTIONAL	JJ	***C***	DEVELOPMENT	NN#	48
PUBLIC	JJ	***C***	ADMINISTRATION	NNJ	39
FOREIGN	JJ	***C***	INVESTMENT	NN#	39
LOCAL	JJ	***C***	SPECIALISTS	NN#	37
ECONOMIC	JJ	***C***	DEVELOPMENT	NN#	36
HUMAN	JJ	***C***	RESOURCES	NN#	34
SOCIAL	JJ	***C***	WELFARE	NN#	34
REGIONAL	JJ	***C***	DEVELOPMENT	NN#	33

NATIONAL	JJ	*C*	EDUCATION	NN#	33
EASTERN	JJ	*C*	EUROPE	NP#	32
TECHNICAL	JJ	*C*	PROPOSAL	NN#	32
INWARD	JJ	*C*	INVESTMENT	NN#	31
ENGLISH	JJ	*C*	LANGUAGE	NN#	31
SOCIAL	JJ	*C*	POLICY	NN#	27
SENIOR	JJ	*C*	ADVISER	NN#	25
VOCATIONAL	JJ	*C*	EDUCATION	NN#	25
FINANCIAL	JJ	*C*	MANAGEMENT	NN	25
ENVIRONMENTAL	JJ	*C*	PROTECTION	NN#	24
POLISH	JJ	*C*	CONSULTANTS	NN#	23
LOCAL	JJ	*C*	GOVERNMENT	NN	23
HUMAN	JJ	*C*	RESOURCE	NN#	23
CIVIL	JJ	*C*	SERVANTS	NN#	23
INTERNATIONAL	JJ	*C*	SPECIALISTS	NN#	23
CENTRAL	JJ	*C*	EUROPE	NP#	21
EUROPEAN	JJ	*C*	UNION	NNJ	21
PUBLIC	JJ	*C*	ADMINISTRATION	NN	20

Table 20 - top 34 out of 852 attributive adjectives

Word 1	**TAG 1**	**Class**	**Word 2**	**TAG 2**	**COUNT**
TOTAL	JJ	*C*	TRAINING	NN#	2
TRAINED	JJ	*C*	STAFF	NN	2
TRANSNATIONAL	JJ	*C*	NETWORKS	NN#	2
UNEMPLOYED	JJ	*C*	PERSONS	NN#	2
URGENT	JJ	*Q*	NEED	NN#	2
VISUAL	JJ	*C*	AIDS	NN#	2
VOCATIONAL	JJ	*C*	TEAM	NN	2
VOLUNTARY	JJ	*C*	SECTOR	NN#	2
WESTERN	JJ	*C*	INVESTMENT	NN#	2
WORKING	JJ	*C*	DAYS	NNT#	2

Table 21 - bottom 10 out of 852 attributive adjectives

33.4 In order to get an impression of the cumulative stylistic effect of this preponderance of classifying adjectives it is instructive to review the role of attributive adjectives in the corpus set which has been taken as a comparator for this study (RomFict – see the preceding chapter: *What are we looking for?*, for a discussion of the reason for choosing this comparator). Table 22 summarises all counts for attributive adjectives in this data set. In this set 16 out of 29 instances are qualitative adjectives:

Word 1	**TAG 1**	**Class**	**Word 2**	**TAG 2**	**COUNT**
LONG	JJ	*Q*	TIME	NNT#	5
DAMP	JJ	*Q*	PATCH	NN#	3
WEALTHY	JJ	*Q*	WOMAN	NN#	3
YOUNG	JJ	*Q*	MAN	NN#	3
ANGRY	JJ	*Q*	TEARS	NN#	2

COOL	JJ	***Q***	DRINK	NN#	2
DEEP	JJ	***Q***	BREATH	NN#	2
HIGH	JJ	***Q***	SPIRITS	NN#	2
LITTLE	JJ	***Q***	WHILE	NNT#	2
LONG	JJ	***Q***	WAY	NN#	2
NEW	JJ	***Q***	PLAY	NN#	2
NICE	JJ	***Q***	RED	JJ	2
OLD	JJ	***Q***	HENRIETTA	NP#	2
PRETTY	JJ	***Q***	GIRL	NN#	2
SMALL	JJ	***Q***	TABLE	NN#	2
SWEET	JJ	***Q***	THING	NN#	2
CHINESE	JJ	***C***	QUARTER	NN#	3
RIGHT	JJ	***C***	THING	NN#	3
WHITE	JJ	***C***	CAT	NN#	3
BARDIC	JJ	***C***	ROBES	NN#	2
EURASIAN	JJ	***C***	SEWING	NN#	2
FRENCH	JJ	***C***	GIRL	NN#	2
FRONT	JJ	***C***	DOOR	NN#	2
FULL	JJ	***C***	MOON	NN#	2
GREY	JJ	***C***	ROCK	NN#	2
OTHER	JJ	***C***	ROOM	NN#	2
OTHER	JJ	***C***	WAY	NN#	2
PREVIOUS	JJ	***C***	EVENING	NNT#	2
RED	JJ	***C***	DRESS	NN#	2

Table 22 - Attributive Adjectives

33.5 The three qualifying adjectives which occur in the high frequency group of attributive adjectives in the PP Corpus (Table 23, Table 24, Table 25) are used in the main at moments when the authors are establishing either their own authority or the authority of team members. *Key* is interesting as it is the most widely used attributive adjective in the PP Corpus associated with moments when authors are making a commitment to the value of an aspect of their technical response. I shall return to these moments in which authority or commitment are expressed in my specific consideration of the lexical typification of PPs.

```
Office of Xxxxxxxxxxxxxx which has considerable experience in organising an
f client satisfaction. We also have considerable experience in delivering tr
hare. Experience Xxxxxxxxxxxxxx has considerable experience relevant to this
s. – Project Implementation We have considerable experience of project imple
United Kingdom. Xxxxxxxxxxxxxx have considerable experience of supporting te
formation and training. He also has considerable experience in the problems
```

Table 23 - Considerable

```
ent with a contractor to supply a wide range of high-quality expertise in
al Studies. We have carried out a wide range of strategic, policy and anal
```

ntre which will offer access to a wide range of information on occupationa
ining provision in Eastern Poland wide ranging consultancy skills capabili
rtant that the report should have wide support among stakeholders. The dra
luded a survey of 600 firms world-wide to assess their attitudes to invest
nd Slovak Project Unit, she has a wide understanding of the training needs
cluding: inspection of sites in a wide variety of heavy industry; transpor
ia, and as sub-contractors with a wide variety of networked partners mostl

Table 24 - Wide

hould identify target markets and key actions to increase awareness and pr
ntify objectives, target markets, key actions and resource requirements. T
work in Poland and in the EL. The key activities during Phase Two are like
ct identification, involvement of key actors in the region, packaging to a
on which follows, we summarise the key areas of expertise of XXX Limited an
This list encompasses most of the key aspects of economic transformation,
ble at this stage to identify the key beneficiaries of the Programme and t
e business training including for key client groups, the unemployed and gr

Table 25 - Key

33.6 The role and effect of the qualifying adjectives in the PP Corpus stands in predictable contrast with the qualifying adjectives in RomFict. For the authors of these representational texts qualifying adjectives are a basic tool for establishing character and mood. Interestingly, apart from *right* and *front* the classifying adjectives in RomFict are also implicated in the same aspects of text development – character and mood – with the emblematic values of red/white/full/grey with reference to dresses and other apparel being exploited to the full.

Qualifying				
angry	cool	damp	deep	high
little	long	new	nice	old
pretty	small	sweet	wealthy	young
Classifying				
right	white	bardic	Eurasian	French
front	full	grey	red	

d worn at school in Paris, a soft grey wool which blended with the rock ag
trousers, or swathed herself in a red blanket! There was n't a man in the
s richly furnished, with handsome red brocade curtains, and a carved mahog
always a powerful rival. The deep red colour of this dress would surely pr
get the chance of wearing a nice red dress instead of your old blue trous
all mirror in the kitchen. A nice red dress, Tom had said. Probably that w
through, tall and elegant in his white dinner jacket. " Hullo, you 're lo
cked her nylons towards the high, white sandals. They waded into the shall
le strapless dress of a green and white silky cotton. Her shoulders and fa
of the sunshine. She wore a full white skirt of some silky material and a

Table 26 - Colours

33.7 Yet is the rhetoric of project proposals entirely free of affect? Traditionally, adjectives – as one of the constituents of "purple prose" – have been viewed as something to treat with great care in formal modes of discourse:

> "Don't use extravagant adjectives. Don't say *magnificent* when a thing is merely pretty, or *splendid* when *excellent* or some other word will do. Extravagance of this kind is never in good taste." (Censor 1880/1982:67)

If the writer of sober prose should eschew the adjective, why is it that there is such a high preponderance of attributive adjectives in the PP Corpus?

33.8 This apparent flouting of conventions can be partly justified be recognising that the majority of attributive adjectives in the PP Corpus are classifying rather than qualifying – and hence less marked as evaluative. We have already commented on this contrast between classifying and qualifying adjectives and noted that the attributive adjectives in the PP Corpus are mainly in the former category. As we have seen, one of the things that attributive adjectives does allow you to do is to integrate information in the noun phrase and to express ideas economically:

```
s capability undertaking local and international assignments     Curricula
able value in raising national and international awareness about investmen
ojects:     collaboration with the International Baltic University;     TE
15.     The company operates on an international basis from offices in the
onment on a regional, national and international basis through:     89
studies (Task A8) and the proposed international benchmarking exercise (Ta
des consultancy services to assist international business partnership and
ssist individual companies seeking international business projects, to ass
```

Table 27 - International

33.9 However, there is a very large number of these classifying adjectives, and – because PPs are prose constructs like any other – these adjectives participate in the creation of cumulative effects across a span of text: at the end of the day, classifying does not have to be a strictly neutral activity. Taking the most

frequent classifying adjectives + noun pairs *(Table 28 - Top PP classifying adjectives*), some (*British, Slovak*) are clearly neutral, the adjective being the necessary other half of an institutional name, and others are closely linked to the theme of the proposal itself (e.g. environmental education). Many pairs, however, appear to be constitutive elements of a common language for the organisations which are bidding for the management of project funds in post-communist Europe. In this closed universe, assistance is technical, development is economic, institutional or regional (#16 in Table 20), and specialists are either local or international (#29 in Table 20).

technical	environmental	Slovak	institutional
public	foreign	local	economic
human	social	regional	national
eastern	technical	inward	English
social	senior	vocational	financial
environmental	Polish	local	
civil	international		

Table 28 - Top PP classifying adjectives

The adjectives are not flowery and the prose is not purple – but an effect of competent professionalism is produced by their use, with specific adjectives becoming almost talismanic. In this particular nexus, *international*, for example, can come to equal *good* and our category "classifying adjective" has become ambiguous.

We have designed programmes worth over ECU 300 million for **international** *agencies* over the last five years.
We have carried out a wide range of strategic, policy and analytical studies for leading **international** *agencies*.
We are currently managing projects worth ECU 42 million on behalf of **international** agencies.
Objective 2 Set the criteria for both the **international** and local *technical assistance team specialists* and assist in their recruitment and selection.
About half of the activities of XXX International are implemented in consortia with European organisations and with private **international** *consultancy firms*.
Some of the recent and present **international** *activities* are listed below.
Task D6: Promotional Events 3.46 The XXX has recently drawn **international** *attention* to its activities by hosting an OECD Conference on Foreign Investment for Small and Medium Sized Enterprises in XXX.
Promotional activities such as these have considerable value in raising national and **international** *awareness* about investment opportunities in Bulgaria in addition to the immediate benefits of the events themselves.

All texts from PP Corpus (XXX indicates anonymisation)

Table 29 - International

33.10 Viewed purely syntagmatically this quality may not appear obvious. It is when *international* is seen paradigmatically in its relation with other "positive" adjectives that the positive force it takes on in this setting becomes clearer (see Paragraph 53.3 below for a more extended discussion).[21] Such a phenomenon can be seen as a local exploitation of the potential *semantic prosody* of a particular word. The term was proposed by John Sinclair in 1988 (Louw B 1993:158) and refers to a mechanism whereby the meaning of a word is "coloured" by the company that it keeps across large numbers of instances. While such an effect will not be perceivable in a single text, if many instances can be viewed (as in a KWIC concordance of a large corpus) the semantic prosody of an item becomes immediately apparent.

33.11 Accepting the existence of such a phenomenon, I would contend that a significant aspect of genre formation is the way in which semantic prosodies are exploited locally. When *international* is reviewed in a large corpus, although it does not always take on the specifically positive prosody we have noted above, that tendency is apparent. Given below are the results of a search across the BNC taking 50 random samples from 22,215 instances.

```
id-1980s.     HL4 2965 Major international affiliations: ECOWAS; ICO;
ages between vast numbers of international and transnational actors.
tional (0624 821212), Albany International Assurance (0624 823262) an
las Information System Ltd's International Banking system: the PRO-IV
also provide vital access to international capital and credit markets
ed in the 1990 Wine Magazine International Challenge.     CBX 1607 Ll
ent.     AR5 520 To gain the International Champion title, it is nece
rimarily under the impact of international competition on its motor v
ey awaited the outcome of an international conference which met at Ge
, the prospects for peaceful international cooperation will become st
 training began with a major international cosmetics company, and he
perators is co-ordinating an international effort to improve operatin
 800 Unix line.     CR9 2488 International equity offers are booming.
onal feature, is now such an international feature that it has led to
L 2365 &lsqbFor extension of international force presence in Turkey t
ich is illustrated in Leaf v International Galleries   2          Pag
```

[21] I owe this insight into the interpretation of semantic prosody to Professor Michael Hoey. The clarification was offered during the viva of this PhD. thesis.

eep faith with England youth international goalkeeper Ian Walker when
ethical rules into line with international guidance where a former pa
high priority is creating an international hospitality and tourism fi
ment of monopoly capital and international imperialism. CS5 69 En
England had probably lost in international importance during the fift
volunteers have also exerted international influence in recent years.
the substantial barriers to international integration, prices have n
the fruitful development of international law in the control of forc
pt outof a rule of customary international law inter se , they cannot
t odds with the realities of international law which was prepared to
9 As would be expected in an international market, product life does
telock, who is a lecturer in International Marketing at Salford have
allenge. CBX 1607 Lloyds International Money Market Fund(0481 724
he Almeida, while the London International Opera Festival brought Udo
rly January aid agencies and international organizations warned that
rate. G37 413 But I'm an international performer in an internatio
Chairman of Stoddard Sekers International PLC, praised the efforts o
as taken by the regional and international press in the course of the
had been able to circumvent international regulations and build or e
ozloduy and bring them up to international safety standards. HBM
on the ground that they gave international sanction to an undemocrati
general allow that war is an international sphere. CHC 853 The th
nternational performer in an international sport on an international
an international sport on an international stage. GV5 1386 The wo
ision status. ANX 952 An international team of astronomers estima
the country of origin and an international ticket in two of the other
ference of the Convention on International Trade in Endangered Specie
ld have a monopoly of Polish international trade. BPF 1373 Marque
n staff at the new Ramenskoe International Training School in Moscow.
mber States has an effect on international transactions in that withi
s of a Convention regulating international transactions. EEL 818
(Source: BNC)

Table 30 - international (BNC)

33.12 In this sample, the main potential of "international" (although there are exceptions to this tendency) does appear to be *positive* to *neutral* e.g.:

- "the prospects for peaceful <u>international</u> cooperation will become…"
- " …. brings UK ethical rules into line with <u>international</u> guidance…"
- "But I'm an <u>international</u> performer in an <u>international</u> sport on an <u>international</u> stage."

Given the existence of this phenomenon, we can begin to explain the cumulative effect of the large number of attributive adjectives identified in the PP Corpus: in the context of PPs, the potential positive semantic prosody of "international" has been exploited so that the use of an apparently neutral classifying adjective provides the writer with a form of warrant. By stressing the *international* aspect of an organisation's experience and connections, the

proposal writer is able to represent the organisation as "good" without making unallowable claims to that effect.

34. Nominalisations

34.1 Nominalisations are the second most prominent feature distinguishing the PP Corpus from its neighbours – 69.4 counts per thousand and 57.01% more frequent than the highest neighbouring mean (these counts having been obtained with an algorithm able to mimic the results for RomFict in Biber 1988 with good accuracy – 8.50 Biber / 8.91 Tribble). Biber reports that nominalisations "tend to co-occur with passive constructions and prepositions and thus interprets their function as conveying highly abstract (as opposed to situated) information" (Biber 1988:227). This view is supported by Halliday's account of the reasons why nominalisations tend to be so numerous in formal written texts, ascribing their high frequency to the need that writers have for a concise and economical way of getting abstract entities into informationally key regions of texts – theme and grammatical subject position.

> "… In other words, even things that are not expressed as nouns have to **behave like** nouns in order to gain their appropriate status in the thematic and information structure." (Halliday 1989:74)

34.2 As the PP Corpus is not tree-parsed, it is impossible to identify grammatical subjects and themes automatically. It is, however, possible to get an indication of the numbers of nominalisations which occur in situations where they have a high likelihood of being grammatical subjects or themes – i.e. immediately after clause limiting punctuation and/or immediately before a verb. A search of the corpus using these criteria produces the following results:

34.2.1 *immediately preceded by clause limiting punctuation*: 513 instances (6.26% of all nominalisations)

- Analysis: Many of these examples are section titles. This could explain one of the reasons for the higher nominalisation counts in PP when compared with the neighbouring genres. Section titles are an important means for ensuring "reader management", and also serve to make the text as explicitly and transparently organised as possible. Also, they are often an implicit requirement of the Terms of Reference for the proposal as bidding organisations are requested to ensure that specific topics are addressed and that certain headings are used as a rubric.

```
N      : [:] Information [NN1] and [CC] Research [NN1] 37 [MC] . [.] Sele
1      : [:] Implement [NN1] the [AT] Worker [NN1] Protection [NN1] Progr
2      , [,] Employment [NP1] and [CC] Social [JJ] Protection [NN1] . [.]
3      . [.] Institutions [NN2] and [CC] their [APPGE] organisational [JJ
4      , [,] complementarity [NN1] and [CC] the [AT] respective [JJ] role
5      , [,] Community [NNJ] and [CC] environment [NN1] " ["] 65 [MC] , [
6      : [:] Organisation [NNJ] and [CC] method [NN1] ( [(] Annex [NN1] B
7      , [,] equipment [NN1] and [CC] a [AT1] small [JJ] research [NN1] f
8      , [,] coordination [NN1] and [CC] liaison [NN1] between [II] team
9      , [,] equipment [NN1] and [CC] training [NN1] . [.] – [-] Instrume
10     , [,] organisation [NN1] and [CC] experience [NN1] of [IO] these [
11     , [,] cooperation [NN1] and [CC] full [JJ] partnership [NN1] of [I
12     : [:] Organisation [NNJ] and [CC] method [NN1] I [MC1] I [PPIS1] S
13     , [,] management [NN] and [CC] control [NN1] and [CC] strategic [J
14     ) [)] ORGANISATION [NNJ] AND [CC] METHOD [NN1] 1.2.1 [MC] BACKGROU
15     , [,] commitment [NN1] and [CC] continuity [NN1] offered [VVN] by
```

Table 31 - Nominalisations: punctuation

34.2.2 *Followed by a finite verb within two words*: 818 instances (9.98% of all nominalisations)

- Analysis: This set is more unambiguously implicated in grammatical subject / theme roles. The sample of output from a concordance of the PP Corpus also shows that nominalisations are not predominantly involved in passive structures. There is a maximum of 263 instances (32% of nominalisations + verb) in the corpus of nominalisations as subjects in passive clauses (i.e.

Nominalisation + Verb BE within an horizon of 5 strings to the right) – though not all of these are passives as the search algorithm used to identify these strings cannot automatically disambiguate passives from copula uses of BE. In most instances the nominalisation is the subject of an active verb: e.g. "*This combination will bring … These organisations emerged …* " (the lexical verbs which associate with these nominalisations are an interesting set in themselves and will be studied in more detail in subsequent chapters).

hich [DDQ] a [AT1] regional [JJ] business [NN1] centre [NN1] might
[NN1] 9401 [MC] . [.] This [DD1] combination [NN1] will [VM] bring
J] interest [NN1] – [-] an [AT1] administration [NN] that [CST] has
eds [NN2] . [.] In [II] the [AT] section [NN1] which [DDQ] follows
N1] – [-] the [AT] economic [JJ] situation [NN1] is [VBZ] difficult
from [II] the [AT] regional [JJ] administration [NN] and [CC] made
[AT1] important [JJ] early [JJ] question [NN1] will [VM] be [VBI]
Z] , [,] the [AT] regional [JJ] government [NN] works [VVZ] to [TO
rivate [JJ] sector [NN1] of [IO] business [NN1] services [NN2] is [
system [NN1] face [NN1] by [II] business [NN1] is [VBZ] complete [
gives [VVZ] us [PPIO2] the [AT] opportunity [NN1] to [TO] take [VV
w [RGQ] favourable [JJ] the [AT] conditions [NN2] are [VBR] in [II]
Europe [NP1] , [,] similar [JJ] conditions [NN2] have [VH0] been [
] trusts [NN2] . [.] These [DD2] organisations [NN2] emerged [VVD]
A [AT1] separation [NN1] of [IO] functions [NN2] could [VM] be [VBI

Table 32 - Nominalisations: verbs following

34.3 Although the informational role of the nominalisations in the PP Corpus cannot always be identified, the semantic sets they encompass give a strong indication of the overall tenor and emphasis of the PP Corpus. The full set of nominalisation types and their counts are given in Appendix 25 – the top 20 are given in the table below:

business	414	administration	131	studies	69
education	389	university	118	preparation	66
implementation	225	quality	106	addition	64
information	187	countries	102	section	63
activities	172	agencies	90	operation	62
organisation	165	inception	85	opportunities	62
evaluation	132	companies	71		

Table 33 -Top 20 nominalisations

34.4 It is possible to begin to group these nominalisations into distinct sets: e.g.

- directly implicated in the themes of the proposals – (*business*, *education*)

- language that is used in association with the elaboration and management of projects per se (*preparation, inception, implementation, activities, evaluation*)

- language associated with the internal organisation of the proposal (*section*)

… and in so doing, to begin to identify key lexis for different aspects of PPs, not only in the context that is specific to the proposals in the corpus, but also in proposals that have been written for different kinds of project and for different agencies. By identifying prominent aspects of the lexicogrammar, it is possible to identify salient aspects of the lexis of the texts in question. This has implications for pedagogy which will be discussed in the final chapters of this thesis.

35. Phrasal coordination

35.1 Phrasal coordination involves the symmetrical pairing of grammatical equivalents e.g. adjective & adjective / verb & verb / adverb & adverb. It is positioned in Factor 3 *Explicit versus Situation Dependent Reference* and is accounted for in Biber 1988 as having an integrative function and being used for ideas expansion (Biber D 1988:245). In the LOB+ corpus there is a relatively wide range of variation across the means of the genres, with some genres having zero instances. In the PP Corpus, although the range and standard deviation give higher values, no text has a lower count that 12.9 per 1000 – above the maximum for LOB+. The algorithm + corpus markup I have used on the test genre RomFict do give a higher count for phrasal coordination than the results in Biber 1988 (Biber 3.20 per 1000 / Tribble 5.27 per 1000) so there may be a need to adjust my results, but if this were done it is improbable that there would be a major reduction in the overall count for the PP Corpus

Linguistic feature	LOB Mean	Corp Min	Corp Max	Range	Corp S/D	PP Mean	PP Min	PP Max	Range	PP S/D
phrasal coordination	3.4	0.0	12.0	12.0	2.7	21.9	12.9	28.7	15.8	4.5

35.2 In the PP Corpus the patterns for this structure are – in order of frequency:

NN* AND NN*	1923
JJ* AND JJ*	452
V* AND V*	175
R* AND R*	11

35.3 Although the least frequent, the adverbs have a certain appeal – demonstrating as they do the writers' attempts to persuade the reader of the quality and extent of their technical approach. This phenomenon is not unique to a single particularly keen-to-persuade author, the examples being spread across seven different proposals from the three different organisations who provided texts for the corpus.

```
n [MC] years [NNT2] ago [RA] and [CC] now [RT] employing [VVG] 50 [MC
lp [VVI] build [NN1] on [RP] and [CC] further [RRR] develop [VV0] exp
ssessed [VVN] regularly [RR] and [CC] continuously [RR] improved [VVN
 [TO] act [VVI] quickly [RR] and [CC] decisively [RR] on [II] potenti
ation [NN1] efficiently [RR] and [CC] effectively [RR] . [.] Our [APP
VBI] met [VVN] promptly [RR] and [CC] efficiently [RR] . [.] However
vices [NN2] effectively [RR] and [CC] efficiently [RR] , [,] the [AT]
 used [VVN] effectively [RR] and [CC] efficiently [RR] through [II] t
[RR] , [,] economically [RR] and [CC] financially [RR] sound [VV0] .
y [RR] , [,] nationally [RR] and [CC] horizontally [RR] across [II] t
being [VBG] financially [RR] and [CC] technically [RR] self-sustainin
```

Table 34 - coordination: adverb

35.4 Verb coordination gives a different insight into the development of the rhetoric of the PPs. Most instances only occur once, but in many cases they are part of a pattern which emphasises either the management processes which will be required by the project under discussion or the management competencies of the bidding agency. A full concordance for this set is provided in the appendices (Appendix 35), but the sample below illustrates the tendency:

```
N2] have [VH0] been [VBN] adapted [VVN] and [CC] published [VVN] by
 be [VBI] speedily [RR] addressed [VVN] and [CC] resolved [VVN] . [
volved [JJ] in [II] administering [VVG] and [CC] monitoring [VVG] s
 [CC] Austria [NP1] – [-] advised [VVD] and [CC] signed [VVD] a [AT
esponsible [JJ] for [IF] agreeing [VVG] and [CC] reviewing [VVG] th
nsultants [NN2] and [CC] analysed [VVD] and [CC] worked [VVD] on [R
N2] required [VVD] to [TO] assess [VVI] and [CC] improve [VVI] perf
oordinator [NN1] will [VM] assist [VVI] and [CC] strengthen [VVI] c
[)] that [CST] both [DB2] attract [VV0] and [CC] deter [VV0] foreig
] resources [NN2] to [TO] attract [VVI] and [CC] retain [VVI] high
] delivery [NN1] of [IO] briefing [VVG] and [CC] training [VVG] to
] This [DD1] experience [NN1] can [VM] and [CC] will [VM] be [VBI]
 [RR] longer [RRR] to [TO] change [VVI] and [CC] are [VBR] dependen
orientation [NN1] may [VM] change [VVI] and [CC] develop [VVI] quit
articularly [RR] in [II] coaching [VVG] and [CC] mentoring [VVG] sk
PHS2] will [VM] be [VBI] collated [VVN] and [CC] produced [VVN] by
[IO] data [NN] is [VBZ] collected [VVN] and [CC] analysed [VVN] : [
es [NN2] , [,] to [TO] commission [VVI] and [CC] manage [VVI] a [AT
se [NN1] I [ZZ1] to [TO] consider [VVI] and [CC] agree [VVI] a [AT1
eted [VVN] to [II] : [:] consider [VV0] and [CC] evaluate [VV0] the
s [NN2] : [:] to [TO] consolidate [VVI] and [CC] develop [VVI] exis
2] will [VM] be [VBI] constructed [VVN] and [CC] incorporated [VVN]
RR] be [VBI] self [NN1] contained [VVN] and [CC] written [VVN] in [
mination [NN1] will [VM] continue [VVI] and [CC] intensify [VVI] be
ed [JJ] bodies [NN2] coordinating [VVG] and [CC] implementing [VVG]
N] involved [JJ] in [II] defining [VVG] and [CC] implementing [VVG]
, [,] and [CC] is [VBZ] delivered [VVN] and [CC] facilitated [VVN]
] have [VH0] been [VBN] described [VVN] and [CC] explained [VVN] in
 [MC] years [NNT2] to [TO] design [VVI] and [CC] establish [VVI] na
 [MC] years [NNT2] to [TO] design [VVI] and [CC] establish [VVI] na
] to [TO] FNP [NP] to [TO] design [VVI] and [CC] implement [VVI] an
opean [JJ] countries [NN2] design [VV0] and [CC] write [VV0] up [RP
[NN1] will [VM] be [VBI] designed [VVN] and [CC] delivered [VVN] by
2] We [PPIS2] have [VH0] designed [VVN] and [CC] delivered [VVN] tr
S2] have [VH0] also [RR] designed [VVN] and [CC] delivered [VVN] a
ertaking [NN1] has [VHZ] designed [VVN] and [CC] delivered [VVN] sp
[NN1] will [VM] be [VBI] designed [VVN] and [CC] delivered [VVN] du
.] We [PPIS2] have [VH0] designed [VVN] and [CC] delivered [VVN] a
S2] have [VH0] also [RR] designed [VVN] and [CC] delivered [VVN] a
1] We [PPIS2] have [VH0] designed [VVN] and [CC] delivered [VVN] tr
tion [NN1] systems [NN2] designed [VVN] and [CC] established [VVN]
```

Table 35 - coordination: verb

35.4.1 The verb coordinations with more than one count also focus on this managerial function; they are:

design	and	establish
designing	and	delivering
develop	and	design
develop	and	pilot
develop	and	implement
giving	and	managing
identify	and	exploit
meet	and	be
reporting	and	monitoring

35.5 Unlike Verb & Verb combinations, Adjective & Adjective coordination exhibits a striking level of repetition of particular strings, Of a total of 465 counts, 149 are related to geographical locations or institutional titles e.g.

```
 [NN2] . [.] Work [NN1] in [II] central [JJ] and [CC] eastern [JJ] Eur
] Hungary [NP1] and [CC] the [AT] Czech [JJ] and [CC] Slovak [JJ] Rep2
tish [JJ] Council [NNJ] in [II] eastern [JJ] and [CC] central [JJ] Eur
 [IO] assisting [VVG] both [RR] Italian [JJ] and [CC] foreign [JJ] comp
 the [AT] Ministry [NNJ] of [IO] Labour [JJ] and [CC] Social [JJ] Welfa
aining [NN1] needs [NN2] of [IO] Polish [JJ] and [CC] Hungarian [JJ] m
```

with the remaining pairs appearing to fulfil a range of purposes, although the largest groups are either implicated in condensed accounts of previous professional experience or forward plans, or in reinforcing the impression of the "quality" of the proposal.

```
  N1] in [II] addressing [VVG] economic [JJ] and [CC] physical [JJ]
 [NN2] including [II] the [AT] economic [JJ] and [CC] physical [JJ]
he [AT] projects [NN2] in [II] economic [JJ] and [CC] physical [JJ]
 [JJ] of [IO] supporting [VVG] economic [JJ] and [CC] physical [JJ]
ical [JJ] skills [NN2] in [II] economic [JJ] and [CC] physical [JJ]
] CCC [MC] – [-] regional [JJ] economic [JJ] and [CC] physical [JJ]
ng [VVG] ever [RR] more [RGR] effective [JJ] and [CC] efficient [JJ]
to [TO] ensure [VVI] an [AT1] effective [JJ] and [CC] efficient [JJ]
 are [VBR] not [XX] only [RR] effective [JJ] but [CCB] efficient [JJ]
31] relation [II32] to [II33] effective [JJ] and [CC] efficient [JJ]
MU [NN1] to [TO] ensure [VVI] effective [JJ] and [CC] efficient [JJ]
```

35.5.1 As we have already noted, adjective phrase coordinations are a much larger set than Adverb or Verb coordination with 452 individual instances. One of the implications of this relatively high frequency is that it permits a doubling effect of many of the instances of classifying predicative adjectives mentioned above (paragraph 33.3 ff) as most of the coordinated adjectives are classificatory. In addition, if one examines the context of combinations which are *not* collected by the algorithms used in Biber 1988 (Adjective + and + Adjective) a much large set of doubled classifying adjectives can be found, including Noun + and + Adjective (sometimes the result of failures in CLAWS tagging) and other asymmetric but nevertheless clause-coordinating

combinations such as "in [II] Enterprise [NN1] and [CC] Economic [JJ] Development [NN1]" – e.g.:

r [IF] primary [JJ] education [NN1] and [CC] vocational [JJ] education [NN1]
Environmental [JJ] Education [NN1] and [CC] public [JJ] awareness [NN1] rai
Environmental [JJ] Education [NN1] and [CC] Public [JJ] Awareness [NN1] Cli
] to [II] the [AT] Employment [NN1] and [CC] Social [JJ] Development [NN1] (
O] enhancing [VVG] employment [NN1] and [CC] human [JJ] resources [NN2] deve
] of [IO] the [AT] employment [NN1] and [CC] social [JJ] development [NN1] p
] to [II] the [AT] Employment [NN1] and [CC] Social [JJ] Development [NN1] (
O] enhancing [VVG] employment [NN1] and [CC] human [JJ] resources [NN2] deve
O] enhancing [VVG] employment [NN1] and [CC] human [JJ] resources [NN2] deve
O] enhancing [VVG] employment [NN1] and [CC] human [JJ] resources [NN2] deve
] of [IO] the [AT] employment [NN1] and [CC] social [JJ] development [NN1] p
TECs [NN2] in [II] Enterprise [NN1] and [CC] Economic [JJ] Development [NN1]
on [II] the [AT] environment [NN1] and [CC] public [JJ] relations [NN2] . [
on [II] the [AT] environment [NN1] and [CC] public [JJ] relations [NN2] . [
evels [NN2] of [IO] expertise [NN1] and [CC] international [JJ] experience [
, [,] including [II] Finance [NN1] and [CC] Civil [JJ] Service [NN1] Admini
ithin [II] this [DD1] General [NN1] and [CC] Technical [JJ] appendix [NN1] ,
ithin [II] this [DD1] General [NN1] and [CC] Technical [JJ] appendix [NN1] ,
2] , [,] Economic [JJ] Growth [NN1] and [CC] Public [JJ] Policy [NN1] and [C
2] , [,] Economic [JJ] Growth [NN1] and [CC] Public [JJ] Policy [NN1] and [C
ccountancy [NN1] , [,] health [NN1] and [CC] English [JJ] language [NN1] tra
rengthen [VVI] the [AT] human [NN1] and [CC] technical [JJ] resources [NN2]
rengthen [VVI] the [AT] human [NN1] and [CC] technical [JJ] resources [NN2]
] management [NN] information [NN1] and [CC] financial [JJ] control [NN1] sy
2] at [II] primary [JJ] level [NN1] and [CC] technical [JJ] and [CC] vocatio
2] at [II] primary [JJ] level [NN1] and [CC] technical [JJ] and [CC] vocatio
Centres [NN2] for [IF] Local [NN1] and [CC] Regional [JJ] Authorities [NN2]

Table 36 - Doubling effect

35.5.2 If we accept that one of the rhetorical impacts of the high level of classifying adjectives (paragraphs 33.9 ff) is the establishment and maintenance of a tone of professional authority (**we** are the ones who know how to classify these complex phenomena) this effect appears to be reinforced by examples such as those above. Not only is phrasal coordination exceptionally pronounced in PPs when it is identified by means of the Biber algorithms, it is, in fact, even more pervasive across the corpus than appears to be the case at first sight.

35.6 Noun coordination is the largest component of this set and has a close affinity with Verb & Verb coordination in that many of the instances in the PP Corpus are associated with either the internal organisation of the proposals, or with their specific ideational content. Thus amongst those combinations which occur with high frequency "organisation and method" is a section heading

used across nearly all the proposals, but "business and innovation" only occurs in one text and is specific to that single proposal. A summary of the distribution of the top thirty items is given below – instances are categorised on the basis of whether a string occurs in one text only (1), in 2 only, or in 3 or more cases (>3).

Noun coordination			#	**dist.**
organisation	and	method	28	>3
business	and	innovation	18	1
design	and	implementation	17	>3
techniques	and	skills	16	1
trade	and	industry	16	>3
aims	and	objectives	13	>3
education	and	training	13	>3
safety	and	health	13	>3
science	and	technology	13	>3
assessment	and	analysis	12	>3
strategy	and	business	12	1
development	and	implementation	11	>3
managers	and	staff	11	1
research	and	development	10	>3
design	and	delivery	9	>3
health	and	safety	9	>3
strengths	and	weaknesses	9	2
introduction	and	summary	8	>3
knowledge	and	skills	8	>3
provisions	and	terms	8	>3
reconstruction	and	development	8	>3
training	and	consultancy	8	2
background	and	experience	7	>3
development	and	delivery	7	>3
knowledge	and	experience	7	>3
PCU	and	ITC	7	1
review	and	evaluation	7	1
training	and	development	7	>3
advice	and	support	6	>3
design	and	development	6	>3
		Total	**324**	

35.7 If these combinations are considered in contexts where they are the objects of immediately preceding verbs e.g.:

```
.] – [-] To [TO] achieve [VVI] status [NN1] and [CC] security [NN1] and
they [PPHS2] acquire [VV0] competence [NN1] and [CC] confidence [NN1] wi
evitably [RR] cause [VV0] dislocation [NN1] and [CC] underuse [NN1] of [
[NN2] to [TO] develop [VVI] awareness [NN1] and [CC] appreciation [NN1]
to [TO] contain [VVI] inconsistencies [NN2] and [CC] contradictions [NN2
[VVN] to [TO] combine [VVI] knowledge [NN1] and [CC] experience [NN1] of
VVN] to [TO] develop [VVI] confidence [NN1] and [CC] self-reliance [NN1]
NP] have [VH0] developed [VVN] skills [NN2] and [CC] knowledge [NN1] to
s [NN2] will [VM] dictate [VVI] roles [NN2] and [CC] priorities [NN2] fo
e [VBI] to [TO] enable [VVI] managers [NN2] and [CC] staff [NN] of [IO]
```

it begins to be possible to identify a listing of key verbs associated with moments where stress and emphasis are being given through the use of coordination. Such a listing is given in the table below. This list was arrived at by listing all Noun & Noun combinations with a preceding verb and then reducing lemmas to the stem.

achieve	acquire	cause	change	combine
comprise	contain	continue	cover	develop
dictate	enable	encourage	ensure	establish
exchange	fund	gain	handle	identify
include	increase	indicate	link	maintain
meet	monitor	need	obtain	offer
operate	plan	promote	propose	provide
research	screen	supply	sustain	train
transfer	transmit	trial	use	visit

Table 37 - Verb preceding noun coordination

35.8 Such lists of verbs associated with frequent patterns across a set of exemplars of a genre have a potential pedagogic value and – along with the insights obtained vis-à-vis the noun combinations themselves – provide a means for exploiting the linguistic information accumulated about this genre in teaching settings.

36. Predictive modals

36.1 Predictive modals provide the least interesting set of features in this first group – largely because it was highly predictable that they would be common in texts which are making proposals about future action. There is a normalised count of 13.47 per thousand words for the mean of PPs – as opposed to a

maximum of 11.9 and a minimum of 3 in the 'neighbouring' texts. The point of interest which arises from this does not come from the fact there is a high frequency of predictive modals in the PP corpus, but rather from the words associated with these modals. This is particularly the case with "will", by far and away – and again, predictably – the most frequently used of the two predictive modals in the Biber set (will: 1347 / would: 69)

36.2 Identifying grammatical subjects associated with predictive modals was achieved by the use of the cluster feature in the WordSmith wordlist program to identify all two word clusters in the corpus, and then to reverse sort this list so that one can identify all instances "word + *will*" with a frequency greater than two. Such a list clearly shows the elements in the text which are given theme roles in relation to predictive modal "will". They are:

which will	68	it will	64	this will	59	they will	58
team will	55	we will	55	and will	41	who will	31
leader will	28	programme will	26	project will	26	consultants will	22
participants will	18	consultant will	16	these will	16	adviser will	15
he will	15	manager will	13	{company} will	13	specialists will	13
experts will	12	council will	11	report will	11	reports will	11
{company} will	11	there will	11	training will	11	staff will	9
stage will	9	that will	9	approach will	8	activities will	7
materials will	7	teams will	7	trainers will	7	workshop will	7
plan will	6	PMU will	6	she will	6	strategy will	6
work will	6	assistance will	5	B will	5	group will	5
information will	5	teachers will	5	Warsaw will	5	consortium will	4
course will	4	ii will	4	ministry will	4	modules will	4
objective will	4	office will	4	requirements will	4	seminars will	4
task will	4	tasks will	4	visits will	4	workshops will	4
advisers will	3	advisor will	3	aim will	3	coordinator will	3
education will	3	factors will	3	{company} will	3	{company} will	3
incubators will	3	inputs will	3	manual will	3	{company} will	3
plans will	3	success will	3	support will	3	systems will	3
trialling will	3	two will	3	activity will	2	attention will	2
Bulgaria will	2	{company} will	2	capacity will	2	centres will	2
cluster will	2	courses will	2	criteria will	2	curriculum will	2
details will	2	development will	2	English will	2	exercise will	2
fund will	2	{company} will	2	hand will	2	identification will	2
initiative will	2	input will	2	leaders will	2	management will	2
meeting will	2	negotiations will	2	{company} will	2	output will	2
partners will	2	phase will	2	review will	2	secondments will	2
services will	2	specialist will	2	{company} will	2	supervisor will	2
three will	2	tutor will	2	UK will	2	workbooks will	2
{company} will	2	{company} will	2				

Table 38 - predictive "will"

36.3 One immediate (and, for me, unexpected) result of this search is the identification of deictic and relative pronouns as the most frequent grammatical subject / theme in the context of "will". There are 185 instances of "it" / "this" and "which" + will – with 76 of these being sentence initial (It: 32 / This: 44) and 20 clause initial (it: 6 / this: 2 / which: 12). I had expected that grammatical subjects of "will" clauses would include a large number of abstract or institutional categories, along with the institutional first person plural. I had not expected that relatives and deictics would occupy the top positions. The effect of this predominance can be seen in the following extract where "this + will" serves the function of summarising the previous argument or representing the processes and stages involved in the elaboration of the projects in question as semi-autonomous and necessary. In citation 1, for example, the precise referent of "this" is ambiguous – it is the Business Needs Analysis as a whole *and* the Consortium's 'pinpointing' activities *and* the conclusions which are reached regarding the challenges faced by local business leaders *and* the areas where they are seeking support. (See Appendix 37 for a full listing):

1.	Business Needs Analysis: each Consortium will identify ways of pinpointing the challenges which business leaders feel they are currently facing and in what areas they are potentially seeking external development support. **This** will further clarify the specific training and consultancy needs of local industry.
2.	A further de-briefing meeting will take place on the completion of Phase 4 where each of the plans and frameworks for delivery will be presented. **This** will be the point at which final adjustments can be made to ensure Polish 'ownership' and the full integrity of the whole.
3.	The review team will be invited to comment on the action plan and in particular on the proposed structure and content of the resources both of the teachers' handbook and the students' resources through a consultation workshop. **This** will give feedback to the consultants at an early stage of the project
4.	The idea is to build on existing activities and to make them more effective by working closely with BFIA staff. **This** will involve practical measures to increase awareness and promote Bulgaria 's attractions to foreign investors.

Table 39 - this + predictive "will"

36.4 The other aspect of theme organisation which is brought out by this listing is the very narrow range of lexical items which take thematic status in most

instances in these predictive statements. Lemmatising the list to show those word forms with the highest frequencies gives the results (excluding 'this', 'it' and 'which') shown in Table 40. The proposals are texts in which institutions and roles (*company name, teams, consultants, leaders, participants, advisers, specialists, managers*) and processes (*programme, project, training, activity, work*) play the dominant roles in what *will* happen.

team	62	they	58
{company name}	56	we	55
consultant	38	who	31
leader	30	programme	26
project	26	adviser	21
participant	18	he	15
specialist	15	manager	13
expert	12	training	11
activity	9	work	6

Table 40 - words in Theme

36.5 Considering the predictive modals from the point of view of the verbs associated with them (Table 41 - *will* + verb), WordSmith Wordlist can be used to identify all two word or three word clusters with *will* at the head. This provides further corroboration of the way in which predictive *will* is strongly implicated in predictions about the management of the project process.

will provide	61	will work	36
will have	35	will include	29
will involve	23	will assist	22
will need	22	will take	22
will ensure	20	will deliver	16
will be developed	15	will focus	15
will be undertaken	13	will form	12
will identify	12	will participate	12
will advise	11	will be able	11
will be based	11	will be led	11
will be provided	11		

Table 41 - will + verb

37. Low frequency items

37.1 As we noted above (para. 32.1) the set of linguistic features which have strikingly *low* frequency in PP is:

Linguistic feature	PP mean	REF Max	REF Min	% diff
adverbs total	17.49	60.8	43.7	59.98
third person pronouns	5.62	33.6	8.7	35.40
private verbs	3.46	17.1	7.8	55.64

The algorithms used to identify these features also appear to be safe – scores for RomFict and check-RomFict being:

Linguistic feature	RomFict	check-RomFict
adverbs total	78.40	75.83
private verbs	24.20	22.09
third person pronouns	78.50	79.50

38. Adverbs

38.1 The words studied in this set exclude "all those adverbs counted as instances of hedgers, amplifiers, downtoners, place adverbials and time adverbials". In order to see how the few adverbs that are in PP are used, it has proved useful to compare them with those in RomFict. A set of the most frequent adverbs in both corpora is given below (Table 42).

PP			RomFict	
	particularly	57	really	28
	currently	46	suddenly	14
	closely	29	quickly	13
	mainly	22	obviously	12
	effectively	20	probably	12
	monthly	20	quietly	11
	quarterly	20	certainly	10
	specifically	20	gently	9
	clearly	17	surely	7
	directly	17	deeply	6
	primarily	15	finally	6
	TOTAL	283	TOTAL	128

Table 42 - adverbs in PP and Romfict

38.2 The most striking aspect of the contrast between PP and RomFict is that while there is such a large distance between the normalised counts for adverbs in PP (17.49) and RomFict (75.83), the top 14 PP adverbs counts for 15.61% of the 2030 instances in PP, while in the case of RomFict this set only accounts for

5.98% of the 2358 instances in this text collection. In other words there is a much greater ratio of types to tokens for this particular word class in RomFict than in PP. When it comes to the specific adverbs that are used, there are also major differences between PP and RomFict (Table 43 and Table 44 below). Specifically, if adverbs of time are discounted (their use as discourse organisers is obvious) the adverbs in PP are used as a means of organising the argument; they signal *how* and *why* things will be done. In RomFict, with its necessary focus on human interactions, the majority of the adverbs are about how people seem and how they behave (with the odd infelicitous exception such as the *damask deeply lace-edged cloth*…), along with the (over) frequent use of "really" as an intensifier in dialogue.

arly in its inception phase, must clearly take account of existing developme
e years of operation. The work is closely integrated with a local enterprise
onsultancy activities relate very closely to these areas of specialism. Its
ion institutions. They also work closely with government departments, profe
Development Plans will be written directly based on the outputs from Phase 3
and consultancy markets, who will directly compete for client work. Analysis
familiar with and use these tools effectively prepare tailored in-compan
. Financial resources can be used effectively and efficiently through this h
sector experience at working effectively with institutional partners, b
45. At present REEC 's act mainly as information points and as a focu
each part of the project, made up mainly of practising teachers. The teams w
ucation. The review teams will be mainly ordinary practising teachers. T
been contracted by other clients, mainly overseas governments and developmen
spective on the Polish situation, particularly for the Enterprise Developmen
ence. The Polish consultants, particularly from PBLCA, will brief the EU
est in the field. It is therefore particularly important to develop pilot pr
Trade and Industry. The project. particularly in its inception phase, must
nd develop the project with them, particularly in the inception phase. They
nds critically on innovation. and particularly on effectiveness in exploitin
ncrease in the private sector and particularly the growth of new small priva
ironmental Policy of Poland. More particularly, the programme aims to establ
taff in CSFR municipalities. More particularly, the project developed, imple
manager and business development primarily through a portfolio of open lear
se has been developed since 1991, primarily through KHF funding, and involvi
ther since the foundation of WBS, primarily within the terms and funding of
pproach is to establish an agency specifically to: attract foreign d
roject. These have been developed specifically for the Polish market accordi
nable in-company programmes to be specifically tailored to identified needs

Table 43 - Adverbs in PP

e caused the accident? There was certainly a strange tension in the air. " You
face was as white as the damask deeply lace-edged cloth spread over the refec
etray 1him. He loved Sandra too deeply to ruin her future happiness. Had ever
to marry and obviously loved so deeply. He doubted that Philip would betray h
ssed that I care for you – very deeply. I 've loved you ever since you were a
like Alice in Wonderland, " Doc gently mocked her, " and I 'm glad you 're be
he previous evening. Yet it was probably no more than gossip that Mrs Henders
e, not even to Fergus, Diana ran quietly across the hall to the garden-room an
ou had guessed, Rob? " she asked quietly. He smiled into the darkness. " Yes,

```
you allow me to choose for you a really French meal? " " Of course. " As she g
n to work now, had n't we? It 's really only a few personal letters. " " I 'm
! a new-style Sir Lancelot! – 's really terribly funny when you think about it
ot buttered scones to Gregory. " Really, we do n't see much of her at all thes
```

Table 44 - Adverbs in RomFict

38.3 The differentiation between the two extremes – PP and RomFict – appears to arise from the strongly contrasting agendas of the writers. The writers in PP are aiming for a style which has minimum affect – hence the minimal use of adverbs, their preference for nouns over verbs, and the high frequency of nominalisations in subject position. Where they feel a need to add rhetorical emphasis, this tone is masked by the favouring of adjectival modification and clause coordination rather than adverbial qualification – a choice which is also to some extent imposed as it is a corollary of the high noun to verb ratio across the PP corpus (see Table 45 for the counts of normalised means of verbs and nouns in the two corpora).

Linguistic feature	Pp mean	FictRom
present-tense verbs	57.10	65.80
past tense verbs	6.12	83.70
perfect aspect verbs	5.66	13.60
total verbs	68.88	163.1
nouns	282.93	146.80

Table 45 - verb:noun

When there are so many more nouns than verbs in a text, changes in emphasis and orientation are necessarily going to occur in noun phrases more than in verb phrases. By the same token, the writers of more explicit fictions have recourse to the adverb as their preferred means of meaning modification.

39. Third person pronouns

39.1 Once again, it is not surprising that third person pronouns – PP Mean 5.62 / Ref Max 33.6 / Ref Min 8.7 / per cent difference 35.40% have low frequency in PP. They are considered to be indicators of whether a text has predominantly narrative or non-narrative style (Biber D 1988:225) and PPs

clearly fall into the less narrative oriented group. In the PP Corpus, third person pronouns are most frequently used in moments when past experience is being referred to – this is often in statements of technical competence – or when future responsibilities are being outlined. The majority of third person pronouns are plural in PP (426:210), although many instances of "their" refer anaphorically to single individuals or single institutional referents (see *Table 47 - their*). One aspect of third person pronouns that is of some interest arises from the insight they offer into the gender balance amongst key staff involved in project management and consultancy in this sector (although bearing in mind that CVs have not been included in the PP Corpus). The counts for masculine/feminine pronouns is given in *Table 46 - male/female pronouns*. In this set at least, the singular boys seem to have a 2:1 advantage!

he	118	she	47
him/himself	3	her/herself	11
his	27		
Total	**148**	**Total**	**58**

Table 46 - male/female pronouns

1.	… companies looking for Bulgarian companies that can supply materials and components or undertake sub-contract manufacture on **their** behalf.
2.	… {company name} have been chosen for the recruitment team because of **their** breadth of experience in recruitment both within the ESNRO and in other UK organisations.
3.	… demonstrates full commitment to **their** Business School and is a fluent communicator in English Formation of Faculty Teams
4.	… focused on business performance improvement and thereby of direct practical benefit to the individual and **their** business.
5.	… assisting local business to identify land for the expansion of **their** business.
6.	Each AS and MPS participant will have a personal tutor (who may be **their** class teacher) with whom they will have monthly 30 minute tutorials.

Table 47 - their

40. Private verbs

40.1 Private verbs offer a more interesting set of absences (PP 3.46, REF Max17.1, Ref Min 7.8, percent difference 55.64%). They belong to the set of

FACTUAL verbs designated in Quirk et al 1985:1180 as describing "factual or propositional information" – i.e. making statements – and differ from Public Verbs in that they refer to states "that are not observable: a person may be observed to *assert that God exists*, but not to *believe that God exists*. Belief is in this sense 'private'." (Quirk et al 1985:1181). The function of private verbs as a means of indicating the status of information and the thought processes of individuals is clearly exemplified from the conversational and narrative contexts in Romantic Fiction given below in Table 48.

Gavin on the terrace could hardly believe that she was the same girl, m
French, a very well known model I believe. She only arrived today. That
really. " " Dull? " She could n't believe it. Self-deprecation was not
or a burial? " she asked, hardly believing he could be serious. " Sure
ood irresolute for a moment, half decided to go back and join Lilyan an
ould simply blow a feather, " she decided, rather envying the gentleman
ss-roads was unsettling, she also decided. One felt as though one had n
on the seat by early October, Di decided, as she collected her various
know for certain and told you, I doubt if you would be able to catch u

Table 48 - private verbs – RomFict

40.2 Given the very different social function of PPs, the overall low count for these verbs in this corpus is not surprising. However, the *way* in which the verbs are actually used in this contrasting context is interesting as it gives an important insight into some of the strategies for "reader management" that are available to writers in formal settings. There are 71 private verbs given in Quirk et al 1985:1180 (see Appendix 39 for a full listing), 26 of which occur in one form or another in the PP corpus[22] (*Table 49 - PP private verbs*).

know	61	estimate	8
see	36	hearing	8
show	35	assume	7
anticipate	31	feel	6
indicate	28	realise	6

[22] One comment needs to be made at this point regarding the accuracy of CLAWS tagging for this set of verbs. A manual analysis of the concordances for *know* and *means* has revealed that many instances of these have been counted as verbs when they are in fact being used nominally in the text. This problem is not restricted to these two items as it also 'infects' the results for other items. Thus the count for Private verbs in the PP corpus is even more markedly lower than the count for neighbouring texts. This observed inaccuracy for PP (in spite of the attempt to weed out nominal uses by incorporating [V* POS tags in the search algorithm) also raises questions over the accuracy of the original counts in Biber 1988.

recognise	26	decide	5
believe	25	learn	5
determine	25	hope	4
find	25	assure	3
demonstrate	21	conclude	3
mean	21	prove	3
understand	21	remember	2
think	11	implies	1

Table 49 - PP private verbs

40.3 Private verbs in PP appear to be implicated in three major functions which are not specifically addressed in Biber 1988 or Quirk et al 1985 – the signalling of *vagueness*, (e.g. anticipate / assume / estimate / hope), *professional integrity* (believe / demonstrate / recognise / think / understand) and *reader guidance* (indicate / see / shown). All of these sets are interesting from a pedagogic standpoint as they occur across all 14 texts in the PP corpus. The examples of instances related to *vagueness* given below (*Table 50 - vagueness*) show the prevalent use of *we* and dummy subject *it* in these rhetorical moments. The *commitment* examples (*Table 51 - integrity*) show, amongst other features, how *think* can be used as a way for the writer to nail his or her organisation's professional colours to the mast (the *we* subject is critical in these instances).

```
 stable in both Republics we can anticipate a considerable increase whe
 the middle of Month Two, and we anticipate that FNP will review it ov
erm specialists is available. We anticipate a need for short-term expe
ntation from their home base. We anticipate that the whole inception p
e strength, depth and breadth we anticipate for the Delivery Stage.
gues in other departments. It is anticipated that this will involve cl
ommunicator in English     It is anticipated that each individual will
events.     UK Placements: It is anticipated that senior trainers ' sk
s] available for purchase. It is anticipated that each Consortium will
they will be interwoven; and we assume assistance will be more intensi
     5     Method of work     We assume in this account that the proje
 in Bulgaria in recent years. We assume there is at least some basic d
 plus the inception phase. It is assumed that month 1 of the project w
in detail the production process assumed and the investments plan
planned consultancy days used to estimate the cost of the project. ---
 By the end of the project it is estimated that at least 200 primary s
lation     –     completeness of estimated costs and exactness of thei
cuments;     a bar chart showing estimated duration and timing of assi
iled SWOT analysis (Task A6), we hope to undertake these studies joint
ocal experts in this work and we hope that several members of BFIA 's
```

Table 50 - vagueness

the terms of reference, and	believe that they are sensible. In particular
, and there is no reason to	believe that inspectors in the NLI will embra
situation outlined above we	believe the support should consist of: –
As we set out above, we	believe that refinement of the implementation
e long-term advisor; We	believe that it is essential that the HRDF PM
each will focus on, but we	believe that it is important they work togeth
earning Resource Centre. We	believe it is essential that the local person
mples of project experience	demonstrate not only a strong background in r
s of projects which we feel	demonstrate our combined capability to undert
vided as are further CVs to	demonstrate the strength, depth and breadth w
ources. The intention is to	demonstrate good practice and an approach tha
e likely to be necessary to	demonstrate this to regional organisations an
tion phase. In addition, we	recognise that working closely with counterpa
the local conditions We	recognise the importance of being responsive
re procedures; While we	recognise the need for speedy and effective I
and the media. It has also	recognised that adults need a clear understan
e, however, that it is also	recognised that the scope of the work will ch
possible suspicion, must be	recognised from the outset, and taken into ac
of knowledge. This has been	recognised in the recent policy statement of
orting arrangements are, we	think it equally important that the Implement
es be linked together as we	think exchange of information, periodicals, s
part in this project but we	think it premature to define their role at th
d be provided in-country we	think that the opportunity to participate in
then supported delivery. We	think this approach is necessary for this cou
e staff. It is important to	understand at the outset the expectations tha
ur view, it is important to	understand clearly the particular requirement
ure of the BFIA in order to	understand the constraints within which a bus
l. The second enables us to	understand the way in which businesses and th
e in schools, we 34	understand that teachers are given considerab
s to Manage the HRDF We	understand that the process of recruiting the
tract inward investment. We	understand than some industrial sector work h
: Supporting the PMU We	understand that the PIP PMU seeks support fro

Table 51 - integrity

41. Conclusion

41.1 In our discussion so far we have considered in some detail the stylistic impact of the high and low levels of use of seven linguistic features out of over seventy (*Table 52 - linguistic features*). Six of these features have been strongly implicated in the positioning of the PP Corpus at the extreme of the range of distribution of text types in the three key text dimensions identified in Biber & Finegan 1989 as significant in a classification of written texts along a "spoken" to "written" cline. One feature (predictive modals) was also highly significant in pushing PP texts away from the profile of texts with which it was otherwise a "neighbour" (Dimension 4: Overt Expression of Persuasion). What conclusions can we then come to?

above mean		**below mean**	
Dimen. 1	attributive adjectives	Dimen. 3	adverbs total
Dimen. 3	nominalisations	Dimen. 2	third person pronouns
Dimen. 3	phrasal coordination	Dimen. 1	private verbs
Dimen. 4	predictive modals		

Table 52 - linguistic features

41.2 First, it has to be remembered that we have been dealing here with *prominent* features in the PP corpus. If a teacher wishes to give students insights into the linguistic difference between spoken and written production, all of the features in the three key dimensions (1/3/5) will be worth considering. Once learners have become aware of how texts in general can vary in terms of their relative "writtenness and "oralness", those features which are particularly prominent in a genre can become the focus of pedagogy. One way of using the Biber study is, therefore, immediately available for teachers and students with an interest in the linguistic construction of different kinds of text production – and this without any prior analysis of a corpus of examples.

41.3 Secondly, if one has an interest in gaining a detailed insight into the ways in which a particular genre is realised linguistically, Biber does offer a useful framework for analysis. Although it requires painstaking coding and counting, Biber 1988 can be used to begin to see not only how the structure of a set of exemplars of a written genre differs from that of other comparable written genres. Information gained in such a study can also be used to develop an understanding of a much wider range of phenomena than the grammar of a text. Thus the insights we have gained into the role of coordinating conjunctions and the kinds of local semantic prosodies that exist in PP have only been possible because of the analytic framework that Biber 1988 offers.

41.4 From the point of view of practical pedagogy, however, there is a question as to whether many teachers would ever have the time to do the kind of study

that I have reported in this chapter. My experience tells me that they would not. It remains, therefore, to see if there are any alternatives to Biber; alternatives which may not give us such a detailed picture of the language of our difficult texts, but which will, nevertheless, allow us to make useful pedagogic generalisations.

SECTION TWO: AN ACCOUNT OF THE TEXTS

Chapter 6: Lexical dimensions

Caution is the key word.
(example for "key word" from the British National Corpus)

42. Introduction

42.1 A starting point for the discussion which follows can be found in two recent festschrifts for scholars with contrasting professional interests: J McH Sinclair, long established as an empirical linguist, and HG Widdowson, an applied linguist whose reputation is more closely associated with pedagogy, stylistics and the profession of language education (Baker M, G Francis & E Tognini-Bonelli eds. 1993; Cook G & Seidlhofer G, 1995). Both these collections contain papers which contribute to the argument of an earlier chapter in this present writing. The first of these (Louw B, 1993) introduces the notion of *semantic prosody*. The later paper (Stubbs M, 1995) goes on to develop this idea. That both these collections contain contributions which interconnect so closely gives rise to a moment of reflection on an unexpected convergence of opinions in the study of texts. More significantly, from my point of view, these articles also point to the need for greater precision when accounting for the range of phenomena which are often described under the single term "collocation".

42.2 The thrust of a large part of Louw's and Stubbs' argument is that, although the ideas brought together under *collocation* have long proven valuable in studies of vocabulary, the term is best restricted to an account of the *lexical* co-occurrence of words – what Sinclair refers to as the *idiom principal* (Sinclair J, 1991:110-115). In such a view, collocation should be no more and no less

than an account of "the occurrence of two or more words within a short space of each other in a text" (Sinclair 1991:170). Other categories of co-occurrence require their own frames of analysis.

42.3 Following these two papers (and a later expansion of the idea in Stubbs 1996 in which the notion of semantic prosody is contrasted with that of collocation), Hoey (1997a) has proposed a more comprehensive framework for the analysis of words in texts – and has gone on to elaborate this idea in Hoey M 1997b. Hoey 1997a proposes the following set of questions to be used in the study of words in texts:

> "There are a number of questions that we need to ask of any set of concordance lines. Many of them are questions which we are used to routinely asking, but there is still some value in articulating them:
>
> 1. What lexical patterns is the word part of?
> 2. Does the word regularly associate with particular other meanings?
> 3. What structure(s) does it appear in?
> 4. Is there any correlation between the word's uses / meanings and the structures in which it participates?
> 5. Is the word associated with (any positions in any) textual organisation?" (Hoey M 1997a:1)

42.4 These questions will require further specification before they can be applied to the set of texts which are the focus of this present discussion. They can, however, be taken as a starting point for the elaboration of a framework of analysis to complement Biber 1988. Their main restriction at present is that they only provide a starting point for the study of words in texts in general as they do not provide a rational basis for deciding *which* words you are going to study. What is required as a first step is some means to help students make this decision.

43. Is frequency enough?

43.1 I have outlined elsewhere a simple pedagogic methodology for exploiting electronic texts (Tribble C & G Jones 1997:36), proposing that the most effective starting point in understanding the overall orientation of a text or text collection is a frequency sorted wordlist. Frequency sorted lists of this kind have long been tools for lexicographers and linguists – "Anyone studying a text is likely to need to know how often each different word-form occurs in it." (Sinclair J, 1991:30) – and they can provide insights into where a text is 'coming from'. Table 53 - "PP / Guardian Wordlists" contains such lists. The first shows the 30 most frequent words in the 112,000 word PP Corpus. This list has predictable features, but also holds some surprises. In order to interpret it, it is useful to compare it with a list for a larger corpus – comparison remaining one of the most helpful ways of making sense of this sort of data (Stubbs M 1996). Column two in the same table contains the first 30 words from a much larger (95 million word) Guardian newspaper data set (wordlist provided by Mike Scott – http://www.liv.ac.uk/~ms2928/homepage.html).

Table 53 - "PP / Guardian Wordlists"

* = lexical item

PP Corpus Wordlist (112,000 words)			**Guardian Corpus Wordlist** (95 million words)		
word	**freq.**	**%**	**word**	**freq.**	**%**
the	7,817	6.83	the	6,065,747	6.38
and	5,359	4.68	of	2,718,004	2.86
of	5,249	4.58	to	2,443,312	2.57
in	3,299	2.88	a	2,176,851	2.29
to	2,717	2.37	and	2,115,496	2.23
a	1,767	1.54	in	1,863,642	1.96
for	1,666	1.46	is	980,542	1.03
will	1,347	1.18	for	913,570	0.96
training 1*	1,148	1.00	that	877,191	0.92
be	1,002	0.88	on	704,719	0.74
with	866	0.76	was	701,712	0.74
is	824	0.72	it	701,199	0.74
project 2*	811	0.71	with	613,691	0.65
development 3*	806	0.70	he	574,604	0.60

on	707	0.62	as	569,463	0.60
this	659	0.58	by	546,409	0.57
management 4*	628	0.55	be	534,556	0.56
programme 5*	609	0.53	at	519,039	0.55
as	601	0.52	but	473,281	0.50
by	600	0.52	his	457,268	0.48
we	536	0.47	have	445,852	0.47
has	521	0.46	are	426,059	0.45
that	467	0.41	from	420,488	0.44
are	453	0.40	has	400,174	0.42
team 6*	437	0.38	an	344,534	0.36
which	426	0.37	this	340,598	0.36
staff 7*	425	0.37	they	339,982	0.36
have	416	0.36	I	331,052	0.35
business 8*	414	0.36	which	321,877	0.34
an	412	0.36	had	307,887	0.32

43.2 The most immediately striking difference between the two lists is the occurrence of a relatively large number of lexical items in the PP list (8:50) and the complete absence of lexical items in the top 30 words of the Guardian. While this difference is not surprising to those with some experience of corpus study – word lists for large corpora of mixed texts consistently show a larger number of types of high frequency non-lexical items than specialised corpora – it can be revealing to learners coming to this sort of analysis for the first time. Being able to ask learners to identify the lexical items in wordlists derived from individual texts or small, specialist text collections which have some importance for those learners, already constitutes a leap in the sorts of independent analysis that learners can begin to make. Thus the top 30 (or top 50, top 100 etc.) words in a research corpus can be used to produce wordlists for "most frequent content words" – a starting point for the investigation of lexical patterning.

43.3 However, while easy access to frequency wordlists represents a major step forward in text analysis, it raises almost as many questions as it answers. Although the lexical items in a wordlist appear to provide a starting point for a study of the research corpus, many of the other words in the lists are much more difficult to come to grips with, even though they may have a significant

contribution to make to our understanding of a text. Thus, while some of the non-lexical items – e.g. *and* and *will* (PP: *and* – #2; *will* – #8 / Guardian *and* – #5; *will* – #31) appear to be interesting, especially in the light of the work based on the Biber framework and discussed above, a frequency sorted wordlist does not, of itself, provide a reason for singling out these items for specific study. Neither does the list tell us the whole story about the other less frequent, but possibly important, lexical and non-lexical items in the texts in question. The frequency sorted list cannot provide a way of identifying the *key* words in the texts we are studying. In order to start using Hoey's 'Five Questions' as a framework for further analysis we need a better means for identifying which words in a text "matter".

44. Keywords

44.1 Drawing on the 20 volume Oxford English Dictionary, Williams introduced the notion of *keywords* as a "critical area of the vocabulary" (Williams R 1976:23), taking as his starting point 110 words which he considered had specific cultural or social significance. This study had its limitations. As Stubbs (Stubbs M 1996:166-168) has pointed out, not only was Williams working with the paper (as opposed to the CD-ROM) version of the dictionary, he also had no corpus based frequency data to draw on, and no computerised means of dealing with the multiple ambiguities of many of the words he was analysing (e.g. *class*). William's work was, however, important in indicating the roles that certain words take on within a culture, and has been used in other studies with similar intent – notably the development of the idea of *keywords* in Stubbs M 1996, where he concludes: "The study of recurrent wordings is therefore of central importance in the study of language and

ideology, and can provide empirical evidence of how the culture is expressed in lexical patterns." (Stubbs M, 1996:169). Stubbs goes on to exemplify the potential of this interpretation of keywords in a detailed study of one of the words in William's original list: WORK. His basic methodology involves the analysis of the collocation relations and semantic prosody (following Louw B, 1993) of WORK, using corpus data from the COBUILD data bank at Birmingham University.

44.2 From the perspective of my own needs, although both Williams and Stubbs offer interesting approaches to the study of words in cultural contexts, neither are interested in solving the problem which I have outlined above. They are not asking *which* words should be selected for study in an analysis of the ways in which particular writers word particular texts. When Stubbs goes on to outline a possible *Dictionary of Keywords in British Culture* he does not present statistically warranted or textually driven criteria for those items which should, or should not, be included in the dictionary's headword list. Instead, he proposes starting with the Williams list of 110 keywords, a list based on cultural and historical insights rather than on any quantitative study of *which* words get used in *which* contexts.

44.3 By adopting a radically different approach to deciding which words might be revealing of a text's or text collection's orientation, Scott M R 1997a does provide a means for choosing the words to focus on. The basis of this difference is that, unlike the studies of lexis mentioned above, Scott starts from the position that *texts* are central categories for linguistic study. For Scott (Scott M R, 1997c:234), therefore, key words are neither:

- entities intuited in relation to a specific cultural context;

- nor are they 'key' in the sense that has become common in corpus linguistics, i.e. the KWIC (Key Word In Context) 80 or 110 character printout which is the default display for computer concordancing software

- nor are they 'key' in terms of the two-words-to-the-left / two-words-to-the-right definition of collocation proposed by Sinclair and used in Stubbs M 1996.

44.4 For Scott, keywords are key in relation to a whole *text*, and are identified by making comparisons between one text or collection of texts, and other, larger text collections. Although taking the text as a central category for analysis presents problems if one wishes to use any of the large, publicly available corpora, which for reasons of copyright or principle are usually composed of text fragments (Sinclair 1987; Garside R, Leech G & G Sampson,1987), it has immediate and significant advantages for anyone with an interest in genres and the whole area of language in social context. By developing the Keyword program in WordSmith Tools (Scott M R, 1996), Scott has provided an adequate and robust means for identifying statistically prominent words in a text (or collection of texts). In my opinion, this tool, in combination with Hoey's five questions, will provide a valuable way of coming to an understanding of the lexical dimension of a text or set of texts.

44.5 The way in which the Keyword program operates can be summarised as follows:

- frequency sorted wordlists are generated for a *reference* corpus (a collection that is larger than the individual text or collection of texts which will be studied), and for the *research* text or texts.

- each word in the research text is compared with its equivalent in the reference text and the program makes judgement as to whether or not there is a statistically significant difference between the frequencies of the word in the different corpora. The statistical test evaluates the difference between counts per token and total words in each text and can be based either on a chi-square test for outstandingness or on a log-likelihood procedure (Scott MR 1996, WordSmith Keywords Help File)

- the wordlist for the research corpus is reordered in terms of the *keyness* of each word.

Although criticisms of chi-square statistics have been made by Stubbs M, 1996 and Dunning T, 1993, Scott argues that:

> "Technically there is no reason why one cannot use chi-square to relate the frequency of a word in one text to its frequency in a corpus as long as one is considering the same word in each case. It all depends on what one is trying to *claim* on the basis of the chi-square procedure. The misgivings have to do with the skewed nature of types in a corpus and the very high frequency of singleton items ….. [and because] … very rare items will *not* be spread around all possible corpora very very thinly …. but will crop up occasionally in relation to some sort of topicality or stylistic factor." (Scott M R, 1997a:243)

44.6 As "topicality and stylistic factors" are precisely what we are interested in, the methodology which Scott proposes seems reasonable and appropriate. However, bearing in mind the epigraph for this chapter, we will continue to exercise caution while reviewing the results obtained from the use of this software.

45. Finding keywords

45.1 Unlike a frequency sorted wordlist in which the counts of each word type are absolute for a given text or set of texts, a keyword list is the result of a comparative process, and it is likely that there will be some variation between two lists of keywords generated for a single corpus in comparison with two *different* reference wordlists. Given this likelihood, before commenting in detail on the keywords in the PP corpus I will comment on two keyword lists derived from comparisons with different reference wordlists – the Guardian newspaper list already mentioned above, and a wordlist for the written data set from the "core" BNC collection. My purpose here is twofold. First, I want to see what sort of impact the different reference lists have on the keyword lists generated. Second, I want to test the usefulness of the texts in the "core" BNC which will eventually be distributed on the BNC Sampler CD-ROM. This collection will contain a representative 88 subset of written texts (1 million words) and 300 or so transcripts of spoken texts (1 million words) taken from the British National Corpus. Although it has been due for publication since May 1996, it has been delayed for technical and financial reasons (October, 1997). However, it is likely to be a widely available resource, and could be used in any studies which attempt to replicate or extend this present study. I feel, therefore, that it will be helpful to evaluate how effective a wordlist based on the 1 million word written data set can be in comparison with one based on a much larger but not generally available reference corpus (e.g. Mike Scott's Guardian corpus). In the following section I shall compare the two keyword lists (henceforward PP/BNC and PP/Guardian) and discuss any similarities and differences between them. After this comparison has been made, I shall go

on to comment on how keyword lists can be used in a lexical specification of a particular genre.

45.2 Keyword lists contain two main categories:

- *positive* keywords are those which are unusually frequent in the target corpus in comparison with the reference corpus
- *negative* keywords are unusually infrequent in the target corpus.

45.3 The complete printouts of PP/BNC and PP/Guardian keyword lists are provided in Appendices 23 and 24. PP/BNC contains 401 words (334 positive / 67 negative), and PP/Guardian 563 (484 positive / 79 negative). The lists were generated with the WordSmith Tools Keyword Program using the "Log likelihood" method for keyword identification (Scott M 1997b:69) with the ***p*** value[23] set at 0.00000000000001 in order to reduce the size of the keyword list.

45.4 The extracts below (*Table 54 - PP/BNC Keywords, Table 55 - PP/Guardian Keywords*) show how keyword lists appear after WordSmith Tools has carried out the necessary calculations – each word is accompanied by frequency and percentage information for the target corpus (columns 2 and 3) and the reference corpus (columns 4 and 5). Column 6 gives the score for *keyness* which the program allocates to each wordstring – the score being a function of its frequency in each corpus and the number of words in each corpus.

[23] A gloss of the meaning of the "p" value is that it is:
"… used in standard chi-square and other statistical tests. This value ranges from 0 to 1. A value of .01 suggests a 1% danger of being wrong in claiming a relationship, .05 would give a 5% danger of error. In the social sciences a 5% risk is usually considered acceptable. In the case of key words analyses, where the notion of risk is less important than that of selectivity, you may often wish to set a comparatively low p value threshold such as 0.000001 (one in 1 million) so as to obtain fewer key words. (Scott M 1997b:70)

45.5 I shall focus on the first 30 keywords in the PP/BNC and PP/Guardian keyword lists and in the frequency list for PP. This will enable us to compare the usefulness of the two reference corpora for keyword list generation, and the value of keyword lists in relation to the initial analysis of frequency sorted lists (section 43 above). I shall also consider the 10 most *negative* keywords in each list as, although it is unlikely that they will provide starting points for investigations of secondary patterns (absences do not create strong collocational or colligational traces in texts), they do provide a means for evaluating earlier observations in the Biber framework sections of this dissertation.

PP\BNC

1 WORD	**2** FREQ.	**3** PP %	***4*** *FREQ.*	***5*** *BNC %*	**6** KEYNESS
TRAINING	1,148	1.00	*154*	*0.01*	4478.9
PROJECT	811	0.71	*111*	*0.01*	3153.0
DEVELOPMENT	806	0.70	*296*	*0.03*	2562.2
MANAGEMENT	628	0.55	*140*	*0.01*	2247.2
PROGRAMME	609	0.53	*187*	*0.02*	2028.8
WILL	1,347	1.18	*3,123*	*0.29*	1483.8
TEAM	437	0.38	*172*	*0.02*	1360.6
AND	5,359	4.68	*28,900*	*2.68*	1284.8
EXPERIENCE	389	0.34	*143*	*0.01*	1235.1
EDUCATION	389	0.34	*167*	*0.02*	1179.6
POLAND	300	0.26	*44*		1153.5
TECHNICAL	270	0.24	*45*		1017.6
BUSINESS	414	0.36	*325*	*0.03*	994.6
STAFF	425	0.37	*366*	*0.03*	976.1
IMPLEMENTATION	225	0.20	*12*		963.1
PMU	193	0.17	*0*		905.5
POLISH	228	0.20	*34*		874.4
ENVIRONMENTAL	243	0.21	*54*		869.4
PROJECTS	232	0.20	*46*		848.3
PROGRAMMES	256	0.22	*82*		843.1
SUPPORT	364	0.32	*310*	*0.03*	840.5
PHARE	178	0.16	*0*		835.1
(CONSULTANT NAME)	185	0.16	*5*		822.7
ASSISTANCE	223	0.19	*45*		812.8
PHASE	206	0.18	*39*		759.6
OF	5,249	4.58	*32,656*	*3.02*	734.7
EU	155	0.14	*0*		727.2
(CONSULTANT NAME)	150	0.13	*0*		703.7
CONSULTANTS	161	0.14	*13*		665.5
SKILLS	202	0.18	*76*		636.9

Table 54 - PP/BNC Keywords

PP/Guardian

1	**2**	**3**	***4***	***5***	**6**
WORD	FREQ	PP %	*FREQ.*	*Guardian %*	KEYNESS
TRAINING	1,148	1.00	*15,773*	*0.02*	7091.7
PROJECT	811	0.71	*9,950*	*0.01*	5181.6
DEVELOPMENT	806	0.70	*16,687*	*0.02*	4348.1
MANAGEMENT	628	0.55	*16,497*	*0.02*	3102.5
PROGRAMME	609	0.53	*22,584*	*0.02*	2611.0
PMU	193	0.17	*17*		2477.4
AND	5,359	4.68	*2,115,496*	*2.23*	2415.1
PHARE	178	0.16	*11*		2309.8
POLAND	300	0.26	*3,022*		2027.2
(CONSULTANT NAME)	150	0.13	*38*		1828.0
IMPLEMENTATION	225	0.20	*1,153*		1802.0
EXPERIENCE	389	0.34	*15,811*	*0.02*	1600.0
WILL	1,347	1.18	*297,237*	*0.31*	1597.5
(CONSULTANT NAME)	185	0.16	*743*		1562.6
TECHNICAL	270	0.24	*4,580*		1557.8
STAFF	425	0.37	*23,412*	*0.02*	1507.2
POLISH	228	0.20	*2,788*		1457.0
ASSISTANCE	223	0.19	*2,867*		1403.9
PHASE	206	0.18	*2,344*		1344.5
TEAM	437	0.38	*32,513*	*0.03*	1309.4
PROJECTS	232	0.20	*5,201*		1215.3
EDUCATION	389	0.34	*27,842*	*0.03*	1192.7
PROGRAMMES	256	0.22	*8,301*		1161.7
ENVIRONMENTAL	243	0.21	*6,959*		1159.9
EVALUATION	132	0.12	*482*		1137.1
BFIA	83	0.07	*0*		1116.1
OF	5,249	4.58	*2,718,004*	*2.86*	1040.4
BUSINESS	414	0.36	*40,474*	*0.04*	1038.8
CONSULTANTS	161	0.14	*2,115*		1006.9
SUPPORT	364	0.32	*32,418*	*0.03*	973.1

Table 55 - PP/Guardian Keywords

46. Does the reference corpus matter?

46.1 An edited list of the top 30 positive keywords in PP is given below (Table 56 - BNC / Guardian Keywords – PP Frequency) along with the top 30 words from the original PP frequency list (column 3).

1. PP/BNC keywords		**2. PP/Guardian keywords**		**3. PP frequency**
✠ = non-lexical item		✠ = non-lexical item		* = lexical item
TRAINING	✓	TRAINING	✓	the
PROJECT	✓	PROJECT	✓	and
DEVELOPMENT	✓	DEVELOPMENT	✓	of
MANAGEMENT	✓	MANAGEMENT	✓	in
PROGRAMME	✓	PROGRAMME	✓	to
AND ✠1	✓	AND ✠ 1	✓	a
ASSISTANCE	✓	ASSISTANCE	✓	for
(CONSULTANT NAME)	✓	BFIA	✗	will
BUSINESS	✓	(CONSULTANT NAME)	✓	training 1*
CONSULTANTS	✓	BUSINESS	✓	be
EDUCATION	✓	CONSULTANTS	✓	with

ENVIRONMENTAL	✓	EDUCATION	✓	is
EU	×	ENVIRONMENTAL	✓	project 2*
EXPERIENCE	✓	EVALUATION	×	development 3*
IMPLEMENTATION	✓	EXPERIENCE	✓	on
OF ✠ 2	✓	IMPLEMENTATION	✓	this
PHARE	✓	OF ✠ 2	✓	management 4*
PHASE	✓	PHARE	✓	programme 5*
PMU	✓	PHASE	✓	as
POLAND	✓	PMU	✓	by
POLISH	✓	POLAND	✓	we
PROGRAMMES	✓	POLISH	✓	has
PROJECTS	✓	PROGRAMMES	✓	that
(CONSULTANT NAME)	✓	PROJECTS	✓	are
SKILLS	×	(CONSULTANT NAME)	✓	team 6*
STAFF	✓	STAFF	✓	which
SUPPORT	✓	SUPPORT	✓	staff 7*
TEAM	✓	TEAM	✓	have
TECHNICAL	✓	TECHNICAL	✓	business 8*
WILL ✠ 3	✓	WILL ✠ 3	✓	an

Table 56 - BNC / Guardian Keywords – PP Frequency

46.2 The first feature you notice when comparing keyword lists with a simple frequency list for the same data is that they are usually markedly different from one another. Thus while PP Frequency (Column 3) only contains 8 lexical items, both PP/BNC and PP/Guardian contain 27. Similarly, while the top 5 keywords are all nouns (three of them being nominalisations), the top five in PP Frequency are the definite article, a conjunction and prepositions. However, it is also interesting that all 8 nouns / nominalisations which occur in PP Frequency also appear in a similar sequence in the keyword lists – i.e. the top five words in the keyword lists are identical to the top five lexical items in the frequency list. On the evidence of this count the most frequent lexical items in a text would seem to stand in some relationship with keywords. The problem with a frequency list, however, is that it only reveals a small number of the potential keywords in any particular text or genre, and does not group those that it finds in a "user-friendly" way.

46.3 Focusing specifically on the keyword lists, what is striking here is that the top five words in each are identical, and are in the same frequency order. In order

to see what the similarities between the two lists might be in the second rank of words, the remaining twenty five items in columns one and two in Table 56 - BNC / Guardian Keywords – PP Frequency have been re-sorted into alphabetical order (the original sequence can be found in Table 54 - PP/BNC Keywords and Table 55 - PP/Guardian Keywords). What the re-sorted lists show is that there is very little difference between the two sets of most frequent keywords generated by using core BNC as the reference corpus (40,431 word types / 1,080,072 tokens), as opposed to the much larger Guardian list (206,439 word types / 95,075,856 tokens). Each keyword top 30 list only contains two words that are not included in the other (marked ×):

- BNC based keyword list: *EU* (at position 41 in Guardian keyword list) and *skill* (not included in the Guardian keyword list)

- Guardian based keyword list: *evaluation* (at 48 in BNC keyword list) and *BFIA* (65 in BNC keyword list)

This small degree of difference appears to be maintained through the list – re-sorting demonstrating that many of the items which show up in the BNC based keyword list are also included in the Guardian keyword list. There are simply more items in the Guardian list. This provides a first indication that there is little advantage in using the relatively large Guardian wordlist to identify keywords in a small target corpus (112,000 word for PP). There may, in fact, be advantages in future studies if small specialised corpora such as PP are consistently referenced against a standard corpus such as BNC Core as this will make possible more reliable cross genre comparisons.

46.4 The situation is similar for the most negative keywords – i.e. there is a close match between the results obtained by using the BNC and Guardian collections as reference corpora. (Table 59 - PP/BNC/Guardian Negative Keywords). This adds to the evidence that there is little to choose between the keyword list derived through comparison with a 95 million word corpus and that derived from a 1 million word data set (part of speech information has been edited in by hand and will be used in later analysis).

NEGATIVE Keywords: PP/BNC

1	**2**	**3**	***4***	***5***	**6**
WORD	FREQ	PP %	*FREQ.*	*BNC %*	KEYNESS
HER (pers. pron.)	11		*2,468*	*0.23*	408.3
HE (pers. pron.)	120	0.10	*4,995*	*0.46*	432.9
NOT (negative)	82	0.07	*4,454*	*0.41*	462.6
BUT (conj)	66	0.06	*4,360*	*0.40*	503.2
SAID (verb)	1		*2,615*	*0.24*	514.5
HAD (verb)	14	0.01	*3,663*	*0.34*	620.9
HIS (pers. pron.)	28	0.02	*4,090*	*0.38*	621.5
YOU (pers. pron.)	11		*4,161*	*0.39*	738.9
I (pers. pron.)	91	0.08	*6,315*	*0.58*	747.0
WAS (verb)	113	0.10	*8,362*	*0.77*	1019.5

Table 57 - PP/BNC: negative keywords

NEGATIVE Keywords: PP/Guardian

1	**2**	**3**	***4***	***5***	**6**
WORD	FREQ	PP %	*FREQ.*	*Guardian %*	KEYNESS
I (pers. pron.)	91	0.08	331,052	0.35	347.0
IT (pronoun)	351	0.31	701,199	0.74	372.3
THAT (conj. / det.)	467	0.41	877,191	0.92	418.9
NOT (negative)	82	0.07	391,934	0.41	493.8
HAD (verb)	14	0.01	307,887	0.32	622.5
SAID (verb)	1		263,949	0.28	622.7
BUT (conj)	66	0.06	473,281	0.50	725.0
HE (pers. pron.)	120	0.10	574,604	0.60	725.7
HIS (pers. pron.)	28	0.02	457,268	0.48	880.3
WAS (verb)	113	0.10	701,712	0.74	1013.2

Table 58 - PP/Guardian: negative keywords

* = common to both lists	
PP/BNC	**PP/Guardian**
BUT (conj) *	BUT (conj) *
HAD (verb) *	HAD (verb) *
HE (pers. pron.) *	HE (pers. pron.) *
HER (pers. pron.)	HIS (pers. pron.) *
HIS (pers. pron.) *	I (pers. pron.) *
I (pers. pron.) *	IT (pronoun)
NOT (negative) *	NOT (negative) *
SAID (verb) *	SAID (verb) *
WAS (verb) *	THAT (conj. / det.)
YOU (pers. pron.)	WAS (verb) *

Table 59 - PP/BNC/Guardian Negative Keywords

47. Keywords in stylistic profiling

47.1 Before moving to a more detailed analysis of the keywords in PP, I want to make a brief diversion to discuss an exciting implication of this capacity of WordSmith Tools to identify positive and negative keywords. The procedure derived from Biber 1988 to typify texts is lengthy and complex, and requires carefully marked up corpora. WordSmith Tools, in contrast, is able to give a pedagogically useful profiling of unmarked-up texts on the basis of positive and negative keywords alone. In essence, this means that the program can not only be used to identify the "aboutness" of texts; it also has the potential to provide important stylistic information. While this potential will be immediately recognised by students of literature, it may, perhaps, be less obvious to language teachers and students. It is for this reason that I am including an analysis of the top ten positive keywords and the five negative keywords that WordSmith Keywords identified from the Romantic Fiction (RomFict) set in the LOB corpus (already used for purposes of comparison in earlier chapters) in this chapter. It offers an illustration of how relatively straightforward such an analysis can be, and how much information can be gained about different genres. It also provides a useful introduction to the main analysis which follows.

POSITIVE Keywords: RomFict

WORD	FREQ.	RomFict %	FREQ.	BNC %	KEYNESS
SHE	566	1.83	2,505	0.23	1266.8
HER	559	1.80	2,468	0.23	1253.2
I	656	2.12	6,315	0.58	712.1
HE	575	1.86	4,995	0.46	705.3
YOU	512	1.65	4,161	0.39	675.9
N'T	266	0.86	1,787	0.17	423.9
HAD	373	1.20	3,663	0.34	393.0
HIM	180	0.58	1,199	0.11	288.8
WAS	530	1.71	8,362	0.77	253.2
NIGEL	45	0.15	16		253.0

Table 60 - RomFict: positive keywords

47.2 In contrast with PP, the top keywords in RomFict are personal pronouns, the past forms *had* and *was*, the negative particle *n't* and, last but not least, *Nigel*. First and second person pronouns are associated with factor #1 "Involved versus Informational Production" – a high factor score indicating emphasis on relationship building rather than on factual information – a typical feature of spoken language use. Third person pronouns, synthetic negation and past tense verbs are all associated with the Biber text dimension #2 "Involved versus Informational Production". Although Biber & Finegan 1989a do not make use of this dimension in their classification of texts on a literate ➔ oral cline, it is relevant here given the large amounts of dialogue or reported speech which occur in RomFict. With a sufficiently extensive basis for comparison, it is probable that evidence from the positive keyword counts would indicate that, along two text dimensions, RomFict is significantly unlike many other written texts – sharing more features with spoken communication.

47.3 The findings we can obtain from negative keywords in RomFict are in some senses even more revealing. We will deal with the two biggest surprises – *the* and *of*.

Negative Keywords: RomFict

N	WORD	FREQ.	RomFict %	FREQ.	BNC %	KEYNESS
48.	IN	394	1.27	21,184	1.96	85.2
49.	BY	52	0.17	5,908	0.55	110.2
50.	IS	120	0.39	9,954	0.92	121.1
51.	OF	533	1.72	32,656	3.02	206.6
52.	THE	1,258	4.06	67,075	6.21	271.0

Table 61 - RomFict: negative keywords

47.4 When compared with BNC, the most negative (i.e. the most prominently infrequent) keyword in RomFict is *the*. As the definite article is usually the most frequent word in any general text corpus, it is strongly counter-intuitive to find it occupying this position in the keyword list for a genre. One possible explanation for why this should be is that there is a significantly smaller proportion of nouns in RomFict than in BNC. We have already seen from our discussion of positive keywords that RomFict appears to be more "oral" than "literate" (and this is supported by the position that it occupies in the results of the Biber study where it is consistently positioned alongside spoken or informal epistolary texts). If, therefore, RomFict is more "oral" than many other kinds of writing, we would expect it to have a relatively low proportion of common nouns (Halliday 1989) – and this, allied with large numbers of proper nouns and personal pronouns could well result in a relatively low frequency of definite articles.

47.5 Such a view is supported by making a keyword list for the 1 million word Spoken Component of Core BNC referenced against the 95 million word Guardian corpus. In this list of 1286 keywords, *the* is also the most negative keyword – and, interestingly, *the* does not appear at all in the keyword list resulting from a comparison of the Written Component of Core BNC with the same corpus. The most negative keywords in this instance being *source, Guardian, date, page* and *pounds*.

47.6 The other surprising negative keyword is *of* . We already have an indication that RomFict is different from general text populations from a frequency sorted wordlist derived from the corpus (Table 62 - Romfict: frequency). Rather than occupying the second or third position, as is the case in general collections of written texts, in this instance *of* is #9 and only represents 1.72% of the total words in the corpus.

N	Word	Freq.	%
1	THE	1,258	4.06
2	TO	927	2.99
3	AND	805	2.60
4	I	656	2.12
5	A	633	2.04
6	HE	575	1.86
7	SHE	566	1.83
8	HER	559	1.80
9	OF	533	1.72
10	WAS	530	1.71

Table 62 - Romfict: frequency

47.7 According to Sinclair *of* is typically more than 2% of the words in a corpus "regardless of the kind of text involved" Sinclair J, 1991:84, so there is something odd going on in RomFict (1.72%). Again, the relatively low frequency could be to do with the large amounts of dialogue or indirect speech which occur in romantic fiction. Two random extracts from the original text of RomFict and the PP corpus demonstrate how this contrast might arise. Postmodifying uses are marked thus: **of**; other uses are marked ***of***.

> "There were few passengers on the plane and Gavin was quickly through the customs. 'Gay! Gavin!' The girl and her luggage had disappeared and they were alone together. The porter brought Gavin 's bag out to the taxi. 'Just a moment, darling,' Gavin pressed her hand and smiled. 'I want to check up on the flights back.' Gay went out to the waiting taxi, and then found that in the excitement **of** meeting Gavin she had left her sun-glasses on the veranda. She went quickly back to fetch them. Gavin and the girl who had got off the plane with him were talking. He was writing something in his pocket-book, with a sick feeling **of** despair Gay knew that ***of*** course it was her address. Gavin joined her and at once dispelled her fears. 'That little bit you saw me talking to, her father is a big land agent, she says that he sometimes has farms for lease... you know that's what I want, Gay, a farm and you!'"
> 169 words: LOB Romantic Fiction Collection

> "This first **of** these examples is particularly significant as it is an almost exactly parallel project to the one proposed for the xxx. The Latvian Development Agency was established with assistance from xxx. Over the last two years, we have supported the planning and development **of** its main activities, attracting inward investment and encouraging export development, we have assisted its promotional programme, undertaken a series **of** action-orientated industrial sector studies and provided a programme **of** training activities for key staff. The scale, terms **of** reference and work undertaken by xxx in Latvia are very similar to those proposed for xxx, in a country at a similar stage **of** economic, social and political development and undergoing the same fundamental transition from a centrally planned command economy to an open market economy. Xxx therefore has current and directly comparable experience to bring to this project in Bulgaria."
> 145 words. SQ-BULG.FMT PP Corpus (xxx indicates a proper noun which has been substituted to maintain confidentiality.)

47.8 In the RomFict extract there are only three instances of *of* and of these one (*of course*) is an idiom. All instances occur outside the dialogue and are associated with the elaboration of the description of the emotions of the actors (excitement **of** meeting, sick feeling **of** despair). In the PP there are 6 instances in a slightly shorter passage. All of these are postmodifying qualifications of relatively general superordinate terms – first / development* / series / programme* / terms / stage – two of which (marked *) are

keywords in PP. It is also of interest that *of* is itself a positive keyword in the PP corpus. Sinclair 1991 argues that *of* can be thought of as a partitive or quantifier rather than a preposition (Sinclair J, 1991:87). In this role it is frequently used in N1 + of + N2 patterns, where N2 is best interpreted as the headword of the nominal group as it is the "principal reference point to the physical world" (ibid:87). What we are possibly seeing in the contrasting frequencies of occurrence of *of* in RomFict and PP is a contrast between texts in which there are proportionately fewer nouns – and where there is a low level of need to elaborate their meanings – and informationally dense texts which use such meaning elaboration in the noun phrase in achieving their effect. Again, further confirmation is provided by the fact that *of* is the second most negative keyword in the Spoken BNC list (referenced against Scott's Guardian corpus).

47.9 This brief review of major positive and negative keywords in RomFict gives an indication of the WordSmith program's potential value in genre analysis. While Biber's approach to genre differentiation remains the most comprehensive available (although in need of further refinement), it would be unrealistic to assume that a teacher or student would undertake a similar study in order to come to grips with an unfamiliar genre. WordSmith Tools, by contrast, does appear to offer teachers and students a way into a text or text collection. It also seems to have the potential to help teachers and learners start to come to conclusions about how lexis, grammar and communicative purpose are interacting in specific genres in a way which no other easily available software can do. While these results are not definitive, they do indicate that

there will be a benefit in using the Keyword program in the analysis of spoken / written differences and the stylistic typification of specific genres.

48. Interpreting the PP keyword lists

48.1 In paragraph 42.3, I introduced five questions which Hoey 1997 suggests we need to ask about words in texts if we want to have a reasonable appreciation of how those words interact with their textual environment. I then went on to consider how to establish *which* words we should be studying in a given corpus, as the proposed framework for the analysis of lexis did not provide a solution for this problem. A possible answer to this question has been found in Scott's notion of *keywords*. This provides us with a rational basis for selecting which words to consider when attempting to provide a lexical specification of genre.

48.2 Hoey's five questions are a little cryptic when quoted out of the context of the original article and require some glossing. To this end, they are repeated below, but this time with an explanation for each.

- **What lexical patterns is the word part of?** The lexical patterns referred to here will be kept to the narrow definition of collocation referred to above (paragraph 42.2) – in effect the sort of pattern which can be identified within a span of five words around a "node" word.

- **Does the word regularly associate with particular other meanings?** Here Hoey is referring to the idea of semantic prosody already discussed in the earlier chapter in this dissertation, "Grammar and Style". However, he extends this notion (introduced in Louw 1993) in line with Stubbs 1996 who shows how *cause* and *happen* are associated with negative events, and

Campanelli and Channell's (1994) study of *train as a* which is seen as having semantic prosody with occupation. Hoey proposes, therefore, that semantic prosody should be taken to answer the question: "Does the word regularly associate with other meanings?" (Hoey M 1997a:2) where meaning does not only entail the positive or negative associations identified for a word like *international* in the chapter "Grammar and Style", but can also entail categories like occupation or negative events.

- **What structure(s) does it appear in?** Hoey introduces *colligation* to describe this form of relationship. Drawing on earlier uses of the term in Firth (Firth JR, 1957:13) and Halliday (Halliday MAK 1959:46), Hoey 1997a summarises what is meant here as follows:

> "…. *colligation* can be defined as *the grammatical company a word keeps*. Just as a lexical item may have a strong tendency to co-occur with another lexical item, so also that lexical item may have an equally strong tendency to occur in a particular position or (a separate point) to co-occur with a particular grammatical category of items." (Hoey M, 1997a:4)

Hoey goes on to exemplify this by a consideration of the lexical item *reason* (in the sense of *cause*) showing how (in a large newspaper corpus) it takes on a strong colligational relation with deictics, and in particular that "*whatever* is 5 times more likely than *what*, that *this* only occurs with *reason* if preceded by *for*, and that *that* is thirteen times more common than *this* if reason is accompanied by an adjective." (Hoey M, 1997a:5)

- **Is there any correlation between the word's uses / meanings and the structures in which it participates?** The argument here is that when a word has two or more clearly differentiated meanings (e.g. reason *qua* reasoning faculty, and reason *qua* cause) it is frequently the case that these words will be

found in contrasting linguistic environments. This may well be of more importance for lexicographers than for those with an interest in the elaboration of written texts, but we will retain the question for the moment.

- **Is the word associated with (any positions in any) textual organisation?** This question, on the other hand, has a very strong relevance for investigations of text organisation and patterning. Hoey M (1997b) provides a persuasive example of the ways in which particular categories of lexical item (e.g. names of people, the phrase *as a* when sentence initial) are found in paragraph initial positions, and how this kind of knowledge of the behaviour of words is used by subjects in a text paragraphing experiment. Given that the principal motivation for this present study is to identify possible pathways into difficult texts for debutante writers, the relationship between lexis and text organisation is likely to be a fruitful area for investigation.

48.3 The rest of this chapter will be devoted to an investigation of keywords in the PP Corpus. In order to make this investigation comparable with possible future investigations I will work with the keyword list created with reference to the written component of the BNC Core. As this chapter is concerned both with the elaboration of a lexical specification of the PP genre, and with modelling possible approaches to the analysis of the genre specific corpora, in the first instance, I shall restrict detailed analysis to the thirty most prominent positive and ten most prominent negative keywords for PP.

49. Positive keywords

49.1 Five main word classes can be found in the top 30 words in the PP keyword list. They are:

nouns	common nouns proper nouns nominalisations	n np nm
derived adjectives		adj
conjunction		conj
of		of
predictive modal		pm

Table 63 - Positive keywords (word class)

This breaks down as follows:

AND	conj	1	ASSISTANCE	nm	6
ENVIRONMENTAL	adj	3	DEVELOPMENT	nm	
POLISH	adj		EDUCATION	nm	
TECHNICAL	adj		IMPLEMENTATION	nm	
WILL	m	1	MANAGEMENT	nm	
BUSINESS	n	11	TRAINING	nm	
CONSULTANTS	n		(CONSULTANT NAME)	np	6
EXPERIENCE	n		(CONSULTANT NAME)	np	
PHASE	n		EU	np	
PROGRAMME	n		PHARE	np	
PROGRAMMES	n		PMU	np	
PROJECT	n		POLAND	np	
PROJECTS	n		OF	of	1
SKILLS	n				
STAFF	n				
SUPPORT	n				
TEAM	n				

Table 64 - Positive keywords

49.2 As Scott comments, a useful function of Keywords is that they "usually give a reasonably good clue to what the text is about." (Scott 1997:67). It does this by revealing the lexical items which are outstandingly frequent in the target text. In the case of PP the nouns in the first 30 keywords give an immediate insight into six aspects of the corpus:

- the *institutions* and *organisations* which are important in the PP corpus – (CONSULTANT NAME), EU, PHARE[24], PMU[25]

- the *social areas* in which PPs are involved – BUSINESS, MANAGEMENT, TRAINING, EDUCATION, DEVELOPMENT

[24] PHARE: An acronym (now obscure) used to refer to the European Union programme which administers development aid to the former communist countries of central Europe.
[25] PMU: Project Management Unit - usually in a Ministry.

- the kinds of *approach* and *methodology* being proposed by these organisations – PROGRAMME, PROGRAMMES, PROJECT, PROJECTS, TEAM, IMPLEMENTATION, TRAINING

- the *professional groups* who will be involved in carrying out this work – CONSULTANTS, STAFF, TEAM

- the *geographical* areas in which these activities take place – POLAND

- the *qualities* associated with these activities or those carrying them out – EXPERIENCE.

49.3 The adjectives in the list consolidate this picture (ENVIRONMENTAL, POLISH, TECHNICAL), but the three other words / word classes are more problematic. The future modal *will* has already been discussed in "Grammar and style", where we have seen its role in establishing the future orientation of the proposals in the PP corpus. Its appearance in the top 30 keywords is not, therefore, a great surprise, although it is interesting that its use in PPs is so outstanding and demonstrates the way in which a keyword list can give indications of the stylistic specificity of a genre. *and* would be much more difficult to interpret without the earlier discussion of phrasal co-ordination, but its status as a keyword must be accounted for through its association with this prominent textual feature. *of*, on the other hand, was not identified as a separate feature in the Biber framework, and this may indicate a limitation of the "preposition" category Biber decided to work with. We discussed above (paragraphs 47.6 – 47.8) the reasons for treating *of* as a separate case, and the preliminary findings from the keyword lists for RomFict and BNC Spoken

Core indicate that it may well be an important stylistic discriminator. Further analysis of *of* in the PP corpus, and comparison with the environments in which it is used here with the environments in which it is found in other texts, could be rewarding.

50. Negative keywords

50.1 Although it may be inappropriate to use Hoey's questions as a basis for the analysis of negative keywords (patterns of *absence* are of a different kind from patterns of *presence*?) I will, nevertheless, discuss them briefly here as a balance to what has just been said about positive keywords. I will initially discuss the lists generated with reference to both the BNC corpus and the Guardian (Table 57 - PP/BNC: negative keywords and Table 58 - PP/Guardian: negative keywords) as they provide a way in to an appreciation of the significance of the negative keywords.

50.2 Two features stand out when the keyword lists for PP are reviewed. The first is the designation of some items as negative keywords when they are, in terms of the PP Corpus, relatively high frequency items (though not all in the top thirty in both lists). The top ten negative keyword list referenced against the Guardian contains *it* (ranked 35/5493 by frequency in PP) and *that* (ranked 23/5493 by frequency). These seem surprising given their position on the frequency wordlist for PP, but their status as negative keywords is confirmed in the BNC based list where they appear at position 12 (*it*) and 15 (*that*). Although not so "negative" as *it* and *that*, the relatively high frequency items *he* (ranked 131/5493 by frequency) and *was* (ranked 138/5493 by frequency) also occur in both of the negative keyword lists (BNC and Guardian).

50.3 The second feature of note in the PP negative keyword wordlist is the actual word classes which occur at its extreme end:

BUT	conj
NOT	negative
HER	pers. pron.
HE	pers. pron.
HIS	pers. pron.
YOU	pers. pron.
I	pers. pron.
SAID	verb
HAD	verb
WAS	verb

Table 65 - Negative KW: word class

50.4 The fact that personal pronouns and the verbs *said*, *had* and *was* are negative keywords in the PP corpus confirms the profile for the genre provided by the Biber framework (high distance, dense information, low past orientation, low affect). The status of *it* and *that* as negative keywords in the PP/Guardian list (positions 12 and 15 in PP/BNC) gives further confirmation. Biber reports that a high use of *it* is implicated in texts with high levels of generalisation and inexplicit lexical content, low usage being associated with texts at the opposite end of the spectrum (Biber D, 1988:226). *that* is more difficult to interpret as we are not sure if it is being used as a relative pronoun (associated with informal, orally oriented texts) or as a verb or adjective complement (associated with more literate texts). However, the various forms of *that* are all positive factors in Biber's "Abstract versus Non-abstract", so – put crudely – a comparatively low incidence of *that* in a text will contribute to positioning that text at the non-abstract end of this text dimension. This interpretation is supported in the case of PP, as the evidence from the detailed analysis of text features (see Table 66 - "that" in the PP Corpus) is that when PP is compared with near analogues and against the LOB+ mean, it does present an overall lower incidence of structures containing *that* – both in verb complementation

and relative structures, and is at the mean for THAT adjective complementation.

Abstract versus Non-abstract Information

Linguistic feature	LOB Mean	PP Mean	Press Rep	Press Rev	Relig.	Off Doc	Ac Prose	Let Prof
THAT verb complements	3.3	1.1	3.30	1.80	4.30	1.40	3.20	4.30
THAT adj. complements	0.3	0.3	0.10	0.10	0.20	0.20	0.40	0.50
THAT relatives: obj position	0.8	0.2	0.80	0.90	0.60	0.70	0.80	1.10

Table 66 - "that" in the PP Corpus

51. Looking at *but*

51.1 The fact that *not* is a negative keyword is again confirmation of the Biber profiling. Analytic negation is a component of Dimension 1: Involved versus Informational Production, and a low incidence of *not* is associated with written, information oriented production. *but* presents a problem as it was not included in Biber's original study[26] (along with *or*) – so we cannot use the Biber framework as a basis for assessing *but*'s salience. However, by treating *but* in the same way as the raw counts for factors in the Biber study – i.e. to work with scores normalised to counts per thousand, it is possible to identify contrasts between the data sets that are available to me. We will now enter another small detour, as an analysis of this problem provides a further useful exemplification of the way in which the combination of keyword analysis and Hoey's five questions can be used to find out interesting things about what might appear to be rather uninteresting words.

[26] In an illuminating comment on the way in which research models often develop, Biber reports (personal communication) that there was no specific reason for this omission: "BUT -- I don't remember what all went into my thinking there -- I notice that I also did not include OR, but again I don't know why. It would be interesting to know whether BUT patterns more like co-ordination or more like adverbial subordinators ??" (Personal communication from Doug Biber)

Corpus	Item	Count	Normalised Score	Total words
Romfict (LOB)	but	191	6.14	31,096
PP	but	66	0.59	111,114
BNC Spoken Core	but	6529	6.04	1,080,072
BNC Written Core	but	4324	4.12	1,050,344

Table 67 - but

51.2 The counts in *Table 67 - but* summarise the marked contrast between RomFict and PP, and a smaller but parallel difference between the counts for BNC Written and Spoken core. Because of the structure of the BNC Core data I currently hold, it is not practicable to separate out the different genres, so the data set counted here includes the full range in the corpus – from highly literate to highly conversational texts. *Table 68 - X2 test* indicates that the difference between RomFict and PP is statistically significant, while the difference between the counts for *but* in BNC Spoken and Written Core is not (a ✓ indicates that an item scores .025 or better and is unlikely to be the result of chance differences). Nevertheless there is a noticeable contrast between the two counts for these two corpora and a more detailed study of keyword lists for individual written genre in BNC, referenced against the full spoken set, could form the basis for a revealing study of spoken / written differences.

Probabilities									
0.010	**0.025**	**0.050**	***X2***	**E**	**(O-E)2/E**	**Romfict**	**(O-E)2/E**	**PP**	**BUT**
						31096.00		111114.00	Total words
						191.00		66.00	Count
×	✓	✓	**7.98**	2.25	6.76	6.14	1.21	0.59	Score (p.1000)
0.010	**0.025**	**0.050**	***X2***	**E**	**(O-E)2/E**	**BNC Spoken**	**(O-E)2/E**	**BNC Written**	
						1080072.00		1050344.00	Total words
						6529.00		4324.00	Count
×	×	×	**0.37**	5.08	0.18	6.04	0.18	4.12	Score (p. 1000)

Table 68 - X^2 test

51.3 Why *but* should be so much less common in PP than in Romfict or BNC Spoken is another question. It is distributed reasonably evenly across the 14 texts in the corpus (Table 69 - But: counts), although two texts (SG-CHER and

BC-MID) have high scores in relation to the number of words in the texts, and two contain no instances at all. Its presence does not, however, seem to be the result of a strong bias in any one text.

Text	count	count p1000	total
SQ-VLAD.FMT	18	1.56	11566
SQ-CHER.FMT	6	1.38	4343
BC-TERM.FMT	12	0.89	13543
BC-MID.FMT	6	0.72	8308
BC-ENV.FMT	7	0.53	13091
BC-SCI.FMT	5	0.48	10372
SQ-BULG.FMT	4	0.31	12729
BK-182.FMT	1	0.31	3260
BK-11.FMT	3	0.28	10865
BK-162.FMT	2	0.27	7278
BK-65.FMT	1	0.20	5014
BK-178.FMT	1	0.16	6313
BC-LAB.FMT		0.00	2861
BK-47.FMT		0.00	2738

Table 69 - But: counts

51.4 Looking more closely at *but*, and using the Hoey questions mentioned above, it is possible to see a marked contrast in the patterns of use of this word between PP and RomFict. Given the restricted data set being used, such a study cannot be conclusive. It is, however, suggestive of some of the reasons for the low incidence of *but* in these formal written texts when compared with literary "dialogue rich" production.

51.5 As I have said, the number of instances of *but* in our two text samples is not sufficiently large to provide the quantity of information required to say anything definitive about any differences that there may be between the ways in which the word used in the one set of texts or the other. However, the brief study reported here (based on the five Hoey questions) indicates that there are some interesting contrasts between the two corpora, and that further investigation with larger data sets of could be fruitful – particularly a contrastive study of spoken and written data. It may appear perverse to be

asking questions originally designed for an analysis of lexis about a grammatical item. However, the results below seem to justify this adaptation of the approach. These results are summarised in the following discussion; complete tables of all the data for *but* can be found in Appendix 4 - But analysis.

51.6 **Q1. What lexical patterns is the word part of?**

Q3. What structure(s) does it appear in? As we are already dealing with a non-lexical item, and will, in many instances, find ourselves considering the relations that it enters into with other non-lexical items, I have decided to bring together Hoey's question 1 and question 3. Any patterning that we identify in the use of *but* is likely to involve both collocational and colligational features, willy nilly. The relatively low frequencies of *but* and its role in discourse as a contrastive conjunction reduce the range of possible collocations it takes on. Most instances of *but* in the two text collections are in clause initial positions and this reduces the left context collocations and three word clusters in which it can appear. In PP there are only two such potential collocations *independent but complementary* and *limited but carefully targeted*, but there are too few of these instances to constitute any form of patterning. RomFict contains no such adjective + but + adjective patterns.

51.7 The form of patterning which is most important in both corpora involves two word clusters[27]. These combinations are more colligational than collocational, and there is a strong difference in the pattern of use in the two corpora. This contrast between the structural associations of *but* digraphs is given in the

[27] Clusters are words which are found repeatedly in each others' company. They represent a tighter relationship than collocates, more like groups or phrases (but I call them clusters because these terms already have uses in grammar) - Concord Help Text - Scott 1997

tables below. These summarise the major part of speech associations for *but + word* in the two corpora and demonstrate the clear contrast between the corpora.

PP	count	%
adverb / addition	13	19.70
pronoun	10	15.15
adjective	8	12.12
verb	6	9.09
time	3	4.55
verb modal	3	4.55
general adverb	1	1.52

Table 70 - PP But colligation

Romfict	count	%
pronoun	69	36.13
general adverb	13	6.81
verb	11	5.76
proper noun	10	5.24
question	10	5.24
verb nonfinite	4	2.09
verb modal	2	1.05
adjective	1	0.52

Table 71 - RomFict But colligation

51.8 The table below shows the two word clusters in PP and Romfict which contain *but* (WordSmith Cluster settings – words in cluster: 2 / minimum frequency: 3). The implications of this patterning will be commented on in the following paragraphs.

PP cluster	Freq.	Romfict cluster	Freq.
but also	12	but I	22
but the	5	but she	18
but we	5	but it	16
		but he	14
		but you	10
		but the	6

Table 72 - but (2 word clusters)

51.9 **Q2. Does the word regularly associate with particular other meanings?**

Apart from the obvious difference between the frequencies of occurrence of two word clusters in the two corpora, what stands out immediately here is the contrast between the company *but* keeps in the two data sets. In PP, it is most strongly linked in the digraph "but also" (and in most cases, with a preceding

"not only" – see Appendix 26: Not only … but.) Here the role of *but* in the unfolding of the discourse is primarily additive. Rather than signalling a simple contrast – a change in the direction of the argument – *but* is used to introduce a professionally warranted commentary or extension to the argument. Surprisingly, when the six *but the* pairs are investigated, it emerges that they too are found in a similar type of structure. A statement is made which either includes a broad generalisation or a statement of some form of problem, and this is immediately restricted by a comment introduced by *but*.

1	This is a pre-condition for economic transformation, and in all countries, it has been **essentially** spontaneous. **But** the degree and the speed of development of the SME sector, has depended on the extent to which people in each country have had business and other ……
2	It will be **difficult** to separate out the impact of this project from other economic changes and new initiatives, **but** the starting point is to have an agreed framework within which inputs and throughputs can be measured, progress assessed and changes proposed.
3	The **main** functions of the Employment Service are placing people into jobs and the payment of unemployment benefits, **but** the ESNRO has promoted numerous efforts and initiatives to combat unemployment, in various forms.
4	These projects are **largely** funded by multilateral and bilateral agencies, **but** the XXX XXXX also manages projects on a collaborative funding basis with Polish institutions. Field management is provided by a project uni….
5	There will, of course, be **local** variations **but** the principle will remain the same. The development of highly skilled trainers and consultants will follow the linear progression required to e.g. t……

51.10 This additive comment involving *but* has a particular local semantic prosody in the PP Corpus. Through such commenting clauses the writers are able to demonstrate their awareness of the various constraints and pitfalls which might lie in wait, and are also able to demonstrate their professional capacity to deal with these potential problems. Similarly, though it is part of a different pattern, *but we* is also used in suasion. This time the writer is using a polite but professional qualification of a preceding argument – *You the client may wish X, but we respectfully submit Y….*

spects of economic transformation, but we would also emphasise that th in an early stage of development. But we would expect to find some in an important part in this project but we think it premature to define s who wish to study each language, but we suggest that those staff who ly which areas each will focus on, but we believe that it is important

Table 73 - "but we"

51.11 In RomFict, the picture is very different. There is only one instance of "but also": "Grant obviously had not only been the notorious jewel thief **but** had **also** murdered Greta"), and none of the six instances of *but the* in RomFict have this additive quality. *but* in Romfict is used in its default contrastive role, usually signalling a closure:

> "Don't go on, Steve! Oh, don't think I don't like you, and I suppose I should say thank you for – for wanting to marry me. **But** it 's quite impossible – it always will be impossible!""
> A romance can do no harm to our publicity at the moment, **but** marriage must wait. Don't you agree?"

51.12 **Q4. Is there any correlation between the word's uses / meanings and the structures in which it participates?** This question has been answered to some extent in the earlier part of this discussion. *But* is a contrasting conjunction, and is restricted to a small number of roles in English – most commonly in clause co-ordination, and to some extent adjective or adverb co-ordination. The interesting result to arise from this investigation has been that one aspect of *but* – its capacity to add an *enhancing* contrast to an earlier statement has been used almost to the exclusion of others in PP, and that this use is not evidenced in RomFict. The suasive language of PP exploits one of the potentials of a very common word, the narrative purpose of RomFict draws on another.

51.13 **Q6. Is the word associated with (any positions in any) textual organisation?** As we have mentioned above, *but* tends to occupy an initial position in the second clause of a co-ordinated pair. Its use in sentence initial

position is still frowned upon in prescriptive grammars (see comment in Sinclair et al 1990: 376). As neither the PP nor RomFict corpora are held in parsed form or are marked up for clause boundaries, the only way of identifying the role of *but* in text organisation was through searching on punctuation. While this was not an ideal solution, it does indicate some contrasts (paragraph boundary information was not available for RomFict). The results of this survey are given in *Table 74 - "but" positions*. The counts for each of the categories were converted to percentages of the total instances of *but* in each corpus collection so as to provide a better basis for comparison.

	PP	Romfict
Sentence initial	9.09	22.51
Sentence initial in direct speech (indicated by ")	0.00	12.04
Clause punctuation (, ;)	50.00	52.36
Paragraph initial	00.00	*

Table 74 - "but" positions

* unaccountable for given the mark-up of the RomFict corpus

51.14 Of course, it has to be remembered that alternatives to *but* are available – notably *however*. When this word is considered in the two text collections, another kind of contrast emerges: there are many more instances of *however* in PP. This can be accounted for by the preference for *however* as a contrasting sentence conjunction in written communication.

HOWEVER	PP total 36	Romfict total 4*
Sentence initial	25	0
Sentence initial in direct speech (indicated by ")	0	0
Clause punctuation (, ;)	8	3
Paragraph initial	3	

Table 75 - However

*one instance: non-coordinating use

The PP / RomFict contrast is matched by the counts for the same items in BNC Core (*Table 76 - BNC Core – however/but*), where there are 37 times more instances of *however* in the 1 million word written set.

	however	but	% total
Written Set	632	4,360	0.4992
Spoken Set	71	6,529	0.66

Table 76 - BNC Core – however/but

51.15 **Conclusions** – Although this discussion of *but* has involved us in another small detour, it has been instructive from my point of view in that it has demonstrated:

- the usefulness of the negative keyword list in identifying potentially significant differences between small genre specific corpora
- the importance of looking at "boring" words (i.e. high frequency non-lexical items) as the combinations which they enter are often more important for the development of a particular style or set towards the reader than many of the more sexy words in texts.
- the usefulness of Hoey's five questions for the development of a rounded appreciation of how a word works across particular sets of texts.

51.16 After this sidetracking around *but*, we will now return to the main substance of this chapter and consider how Hoey's five questions can help us better understand the functions of the positive and negative keywords in the PP corpus.

52. What lexical patterns is the word part of?

52.1 Restrictions of space and time will limit the number of keywords in PP which we can analyse. I shall begin, therefore, by considering what Scott calls *key-keywords* (Scott M, 1997a:237) – that is keywords in a corpus which occur in all, or most, of the texts which make up the corpus you are studying. The table below contains the 7 most prominent key-keywords in PP. *Training* and *project* occur in all 14 texts, followed by the others in the sequence given. As it is in some senses the least obvious of the seven (which are all to do with what happens in projects), I shall choose to look at *experience*.

N	WORD	OF 14
1	TRAINING	14
2	PROJECT	14
3	MANAGEMENT	12
4	DEVELOPMENT	12
5	PROGRAMME	12
6	EXPERIENCE	11
7	IMPLEMENTATION	11

Table 77 - PP key-keywords

52.2 In considering the collocational patterning of the keywords in PP, we will use the following simple definition. "Collocates are the words which occur in the neighbourhood of your search word" (Scott 1996:help text). For practical purposes we will set the limits of our current enquiry to items within four words to the left or right of the search word (following Sinclair 1991:106, and focusing mainly on words occurring within two words to the left of the right). Apart from 'traditional' collocational information based on these criteria, we will use two other categories of collocation information (available within the WordSmith Tools suite):

- *clusters*

  definition: "Clusters are words which are found repeatedly in each others'

company. They represent a tighter relationship than collocates, more like groups or phrases." (Scott 1996:Help text)

- *patterns*

  definition: patterns "show the words adjacent to the search word, organised in terms of frequency within each column" (Scott 1996:Help text)

  The significance of these different views of collocational relationships will become clearer as we develop our account of *experience*.

53. Lexical patterns: collocates

53.1 A collocates table for *experience* gives us a number of different views of the company this word keeps. By sorting the data in different columns it is possible to get contrasting perspectives on the ways in which a word keeps company[28]. The table below shows collocating lexical and non-lexical items of *experience* which occur 10 or more times. The table has initially been sorted by the overall frequency of words in the table. Thus, *experience* is the most frequent word, followed by *of*, *in*, *the*, *and* etc.

WORD	TOTAL	LEFT	RIGHT	L2	L1	*	R1	R2
EXPERIENCE	**389**	0	0	0	0	389	0	0
OF	**185**	35	150	19	16	0	148	2
IN	**109**	10	99	10	0	0	94	5
THE	**87**	18	69	6	12	0	4	65
AND	**76**	53	23	15	38	0	20	3
HAS	**62**	57	5	33	24	0	1	4
HAVE	**27**	26	1	19	7	0	0	1
CONSIDERABLE	**25**	25	0	2	23	0	0	0
OUR	**22**	21	1	0	21	0	0	1
TRAINING	**21**	5	16	3	2	0	0	16
YEARS	**21**	21	0	4	17	0	0	0
TO	**20**	2	18	1	1	0	10	8
EXTENSIVE	**18**	18	0	3	15	0	0	0
RELEVANT	**17**	6	11	2	4	0	7	4

[28] The table also shows words (e.g. of, in) which will have to be treated when we come to consider colligates. In this part of the argument we will consider non-lexical words in so far as they contribute to the broader contexts / phrases in which the node word is found. In the section in which we deal specifically with colligates, these words will be treated along with others in the same part of speech category.

PROJECT	**15**	11	4	1	10	0	0	4
INTERNATIONAL	**15**	14	1	0	14	0	0	1
WITH	**14**	12	2	6	6	0	2	0
ELSEWHERE	**13**	0	13	0	0	0	12	1
THIS	**13**	9	4	3	6	0	1	3
AS	**13**	6	7	4	2	0	6	1
WIDE	**12**	12	0	1	11	0	0	0
PRACTICAL	**12**	11	1	0	11	0	0	1
WORKING	**11**	2	9	0	2	0	0	9
EU	**11**	7	4	0	7	0	0	4
HE	**10**	9	1	9	0	0	1	0
ON	**10**	9	1	5	4	0	0	1

Table 78 - experience collocates

53.2 This visualisation of collocation can be of mixed usefulness – especially as it includes so many non-lexical items. Some, such as *the* are distracting – it appears on the table because it occurs 65 times at R2 – i.e. where it is the determiner of another word and not directly associated with *experience*. *of* and *in*, however, will be interesting, particularly for the learner, as they provide immediate and relevant exemplification of the ways in which these two prepositional colligates of *experience* differ or are interchangeable. *And*, will also be revealing, as we have here a confirmation of the tendency towards non-phrasal co-ordination which emerged in the study based on the Biber framework. We will return to this in our consideration of colligational patterning. Outside these top two, the lexical collocations of *experience* are more predictably located in either the left or right contexts. Further sorting makes the picture clearer.

53.3 Sorting on column L1 shows the main categories of left collocate for *experience*. Of the 18 words that are left collocates. 15 are lexical items (although *has* and *have* may be delexicalised). Predictably, the majority of these lexical items are adjectives (6) CONSIDERABLE / EXTENSIVE / INTERNATIONAL / PRACTICAL / RELEVANT / WIDE, followed by nouns (5) EU / PROJECT / TRAINING / WORKING / YEARS.

WORD	TOTAL	LEFT	RIGHT	L2	L1	*	R1	R2
AND	76	53	23	15	**38**	0	20	3
HAS	62	57	5	33	**24**	0	1	4
CONSIDERABLE	25	25	0	2	**23**	0	0	0
OUR	22	21	1	0	**21**	0	0	1
YEARS	21	21	0	4	**17**	0	0	0
OF	185	35	150	19	**16**	0	148	2
EXTENSIVE	18	18	0	3	**15**	0	0	0
INTERNATIONAL	15	14	1	0	**14**	0	0	1
THE	87	18	69	6	**12**	0	4	65
WIDE	12	12	0	1	**11**	0	0	0
PRACTICAL	12	11	1	0	**11**	0	0	1
PROJECT	15	11	4	1	**10**	0	0	4
HAVE	27	26	1	19	**7**	0	0	1
EU	11	7	4	0	**7**	0	0	4
WITH	14	12	2	6	**6**	0	2	0
THIS	13	9	4	3	**6**	0	1	3
RELEVANT	17	6	11	2	**4**	0	7	4
ON	10	9	1	5	**4**	0	0	1
AS	13	6	7	4	**2**	0	6	1
TRAINING	21	5	16	3	**2**	0	0	16
WORKING	11	2	9	0	**2**	0	0	9
TO	20	2	18	1	**1**	0	10	8
IN	109	10	99	10	**0**	0	94	5
ELSEWHERE	13	0	13	0	**0**	0	12	1
HE	10	9	1	9	**0**	0	1	0
EXPERIENCE	389	0	0	0	**0**	389	0	0

Table 79 - experience: left sort

53.4 This view of the left collocates of a key-keyword in PP gives immediate access to a range of information that is of great importance for writers in L2. In a simple grid, a learner is now able to see how proposal writers make statements about their *experience*. Whether experience is *extensive*, *wide* or *practical*, it is something that the consultants or companies *have*. It is a property, part of their stock in trade. This grid also confirms the earlier observation we had made about the elevation of *international* to a strong collocate of *experience* (equating with *good*) in Project Proposals (Chapter 5: *Grammar and Style*), as well as confirming the importance of noun / noun modification in the corpus. It is pedagogically important to note that this type of pattern (e.g. *EU experience* / *project experience*) is much less common in learner writing than in expert texts (Granger S & C Tribble 1998), and a review of the collocates of keywords provides an excellent basis for the development of learner language awareness in this area.

53.5 The right sorted collocates offer information that is perhaps less obviously rich but which is still important.

WORD	TOTAL	LEFT	RIGHT	L2	L1	*	R1	R2
OF	**185**	35	150	19	16	0	**148**	2
IN	**109**	10	99	10	0	0	**94**	5
AND	**76**	53	23	15	38	0	**20**	3
ELSEWHERE	**13**	0	13	0	0	0	**12**	1
TO	**20**	2	18	1	1	0	**10**	8
RELEVANT	**17**	6	11	2	4	0	**7**	4
AS	**13**	6	7	4	2	0	**6**	1
THE	**87**	18	69	6	12	0	**4**	65
WITH	**14**	12	2	6	6	0	**2**	0
HAS	**62**	57	5	33	24	0	**1**	4
THIS	**13**	9	4	3	6	0	**1**	3
HE	**10**	9	1	9	0	0	**1**	0
CONSIDERABLE	**25**	25	0	2	23	0	**0**	0
OUR	**22**	21	1	0	21	0	**0**	1
YEARS	**21**	21	0	4	17	0	**0**	0
EXTENSIVE	**18**	18	0	3	15	0	**0**	0
INTERNATIONAL	**15**	14	1	0	14	0	**0**	1
WIDE	**12**	12	0	1	11	0	**0**	0
PRACTICAL	**12**	11	1	0	11	0	**0**	1
PROJECT	**15**	11	4	1	10	0	**0**	4
HAVE	**27**	26	1	19	7	0	**0**	1
EU	**11**	7	4	0	7	0	**0**	4
ON	**10**	9	1	5	4	0	**0**	1
TRAINING	**21**	5	16	3	2	0	**0**	16
WORKING	**11**	2	9	0	2	0	**0**	9
EXPERIENCE	**389**	0	0	0	0	389	**0**	0

Table 80 - experience: right sort

53.6 The point to comment on is that *experience* is not often used on its own. In 242 out of 389 instances (62.21%), *experience* is post-modified by a preposition (either *of* or *in*). What the nature of that experience is, is not immediately apparent from this two left / two right collocation grid and will have to be reviewed when we come to consider the clusters in which *experience* participates (although *training* and *working* are in fact members of three word clusters with *experience* and either *of* or *in*). Of the other immediate right collocates, only two are lexical items – *elsewhere* and *relevant*.

53.7 *Elsewhere* is interesting as it benefits from the same positive semantic prosody as *international*. One of the problems that the proposal writer has is to demonstrate that the experience their organisation has gained in delivering

other projects is applicable to the project for which they are now bidding. Although not wanting to pre-empt later discussion, it is relevant to note here how a concordance for *experience elsewhere* also reveals how the digraph is implicated in an important rhetorical ploy. In all but two of the instances in the PP corpus, *experience elsewhere* has either been placed in the first clause of the sentence by means of a fronted prepositional phrase, or is sentence initial by default as it is the grammatical subject of main clause – either as is, or with a first person plural pronoun determiner. *Experience elsewhere*, like *international*, is called upon by these writers as a means of sanctioning what would otherwise be an unsupported opinion. This is not to say that a subterfuge is being used, as it is often the case that you can only analyse one situation by reference to near analogues. It is interesting to note, however, how this keyword is used in the development of the argument.

Our input will be to draw on successful experience elsewhere and help local staf
operate within an open market economy. Experience elsewhere has shown us that t
dy Follow-Up Programme 3.52 Our experience elsewhere has demonstrated th
hem. These are discussed below. But our experience elsewhere in Russia, and in t
at are their skills and motivation. Our experience elsewhere is that starting wi
ion and CSME as well as with TACIS. Our experience elsewhere strongly reinforces
are responsible for signing contracts. Experience elsewhere suggests that an or
er, comment on the basis of substantial experience elsewhere that the region app
– private consultants. 2.13 Experience elsewhere would suggest that
ise four criteria that, on the basis of experience elsewhere, can be regarded as
y to be incomplete. On the basis of our experience elsewhere, we would also expe
the Board of Directors. On the basis of experience elsewhere, we know that it is

Table 81 - experience elsewhere

53.8 *experience relevant* is a more explicit way of connecting work done with work to do. Of interest for the learner is the fact that the structure in play here is *experience relevant to*…. . To what the experience is relevant cannot be seen in this representation of collocational information. However, another tool is available which does give access to a view of the wider collocations of the node word.

54. Lexical patterns: clusters

54.1 Cluster information gives us another view of collocation, providing insights into the frequently occurring combinations into which a particular word enters. This can be especially useful for the apprentice writer, as it offers access to the lexical / phrasal resources which experienced writers draw on (through intertextual awareness) when developing new texts. (Selzer G, 1993:173, De Cock S, et al 1998:67). For the purposes of this analysis we will look at 5, 4 and 3 word clusters associated with *experience*. The full results of this survey are provided in Appendix 12: Experience_Clusters: I will summarise main findings only here.

54.2 We noted in the preceding section that *of* and *in* appeared frequently in the immediate right context for experience but that the 2 + 2 collocation information gave little information on the lexical items with which *experience* collocated beyond that threshold. Clusters provides this information in a highly visible way. Below is a table which combines data for 3, 4 and 5 word clusters for *experience* + *of* or *in* (Table 82 - experience clusters). The table does not indicate absolute frequency as there is a significant number of duplicates – e.g. the three instances of the 5 word cluster *experience of assisting PMUs implement* subsume the three instances of *experience of assisting PMUs*, (although they do not account for one of the four instances of *experience of assisting*). However, the table does demonstrate the ways in which the analysis of clusters can be used to gain access to information on phrasal collocations. This information could be particularly valuable as it is not entirely restricted to the content areas which the texts in PP address (e.g. Poland, PMUs) but gives insights into patterns which could be used in a wide

range of analogous texts (e.g. *experience in the provision of*). These more extended collocations of *experience* are more clearly visible when non-lexical items are stripped out (see Table 83 - experience lexical collocates (clusters).

cluster	Freq.
experience in central and eastern	3
experience in Poland	6
experience in the provision	3
experience in the provision of	3
experience in the uk	3
experience of assisting	4
experience of assisting PMUs	3
experience of assisting PMUs implement	3
experience of designing	4
experience of designing programmes	3
experience of designing programmes for	3
experience of environmental education	3
experience of EU best practice	3
experience of managing	6
experience of project implementation	3
experience of tenderer and subcontractors	4
experience of the design	3
experience of the design and	3
experience of the development	3
experience of the PHARE	4
experience of the PHARE programme	4
experience of training	12
experience of training in	3
experience of work	5
experience of work in	4
experience of working	7
experience of working in	3
experience of working with	4
experience relevant to	7
experience relevant to this	5
experience relevant to this assignment	5
project experience in Poland	3

Table 82 - experience clusters

experience xx xx design
experience xx xx development
experience xx xx PHARE
experience xx xx provision
experience xx xx uk
experience xx assisting
experience xx central and eastern
experience xx designing
experience xx environmental education
experience xx EU best practice
experience xx managing
experience xx Poland
experience xx project implementation
experience xx tenderer and subcontractors
experience xx training
experience xx work
experience xx working
experience relevant xx
experience relevant xx xx assignment
project experience xx Poland

Table 83 - experience lexical collocates (clusters)

54.3 A "cluster" view of collocation is valuable as it gives an insight into the kinds of predictable structure which expert writers are able to draw on during text composition. We must also remember that we are at present only considering one key-keyword out of 333 positive keywords in PP, yet this keyword enters into a significant number of repeated relationships with other words (and keywords such as *considerable* and *relevant*). The value for writing instruction of this kind of analysis is significant as it provides learners with direct access to the kinds of language system knowledge mentioned in Chapter 2: *Teaching Writing*.

55. Lexical patterns: patterns

55.1 A pattern table gives access to other sorts of information. As Scott explains, when you are looking at a "patterns" list in WordSmith Tools, you see "the words adjacent to the search word, organised in terms of frequency within each column. That is, the top word in each column is the word most frequently found in that position. The second word is the second most frequent, etc." (Scott 1996: Patterns Help Text). *Table 84 - experience: patterns* shows the most frequent patterns for *experience* in PP. Columns L1 and R1 show the most frequently occurring words to the left and right of the centre – summarising in some senses what we have said earlier about immediate collocation. The whole matrix can also be seen as a summary of the combinatorial potential of *experience* in the context of a given genre – in this case PP. Reading from left to right, it is possible to generate a huge variety of three or four word combinations – and even to go up to seven word combinations such as "*their expertise and experience is in management …*" or "*he has considerable experience elsewhere in development*".

L3	L2	L1	centre	R1	R2	R3
and	has	and	experience	of	the	and
the	of	has	experience	in	training	in
has	have	considerable	experience	and	working	training
we	and	our	experience	elsewhere	to	with
over	in	years	experience	to	Poland	the
of	he	of	experience	relevant	that	PHARE
in	the	extensive	experience	as	management	of
consultant	with	international	experience	is	in	development
to	on	the	experience	the	work	this
their	expertise	wide	experience	over	has	has
as	as	practical	experience	with	expertise	to
he	years	project	experience	has	relevant	management
on	training	their	experience	work	project	project
with	this	have	experience	this	EU	that
training	development	EU	experience	he	we	work
have	extensive	this	experience	we	*consultant*	eu

Table 84 - experience: patterns

55.2 By way of contrast, the following table shows the most frequent patterns for *experience* in the SPORTS data set in a 250,000 word corpus of Independent Newspaper articles (MicroConcord Text Collections, OUP 1992).

L3	L2	L1	centre	R1	R2	R3
was	his	the	experience	of	the	and
the	the	vast	experience	for	in	by
of	wealth	racing	experience	the	his	will
his	another	and	experience	in	was	but
from	by	of	experience	to	we	the
with	is	his	experience	he	at	in
vast	had	good	experience	another	gained	has

Table 85 - Independent texts – experience: patterns

This extract demonstrates both the different set of relations (with lexical and non-lexical items) that *experience* enters into in this different genre, and indicates the relative paucity of non-unique instances in this 250,000 data set when it is compared with PP.

56. Lexical patterns: conclusion

56.1 By investigating the collocations of a single keyword from these three perspectives we have seen that it is possible to develop a detailed understanding of how the word operates in a particular genre. In the case of *experience* we have not only identified the specific lexical items with which it strongly collocates (a different set from its collocates in other genres or

domains), but have also begun to identify other patterns of association which we will deal with in following sections.

57. Does the word regularly associate with particular other meanings?

57.1 If Hoey's second question is asked of a word as it is used in general, we gain insights into what we will call its global lexical prosody. Hoey 1997a gives the example of a study of *consequence* (in its meaning of 'effect' as opposed to 'importance') in a large general corpus. Hoey identifies four semantic prosodies for *consequence*:

- the logic of underlying processes – 56% (inevitable, inexorable, likely, probable …)
- the badness of an outcome – 15% (dire, appalling, regrettable …)
- the seriousness of an outcome 11% (important, decisive …)
- the expectedness or otherwise of an outcome – 9% (unintended, odd …) (Hoey M 1997a:3)

A similar study by Stubbs (Stubbs M, 1995:247) identified a predominantly negative semantic prosody for the verb CAUSE (e.g. cause an *accident, cancer, death, pain* etc.)

57.2 While accepting the usefulness of this notion, Hoey argues that it should be extended so that it not only covers broad categories such as 'unpleasant' or 'positive' but also includes more specific prosodies such as e.g. "occupation". Hoey justifies this by drawing attention to a phrase such as *train as a* which not only occurs with common collocates such as *teacher, nurse,* or *lawyer,* but

is also found in rare combinations e.g. *sathin*, *boxing second*, *kamikaze pilot* (Hoey 1997a:2). These latter would never be thrown up as obvious collocates of *train as a*, but they are, nevertheless, strongly implicated in a semantic prosody: PROFESSION.

57.3 For the lexicographer or student of lexis these insights are important. For our present purpose, however, although a knowledge of a word's *global* semantic prosody will be important, it may be more useful in developing an understanding of our difficult text to find out about the *local* semantic prosody of the word under scrutiny. I gave the example of the positive semantic prosody taken on by *international* in PPs in Chapter 5: *Grammar and Style*, and proposed that words in certain genres may establish *local* semantic prosodies which only occur in these genres, or analogues of these genres.

57.4 The question I will ask in this section is, therefore: "Does the word regularly associate with other particular meanings *in this context*?" I am not assuming that all keywords in a text will have specific local semantic prosodies, but I am proposing that this is an aspect of language use worth considering as it will constitute important local knowledge for writers in a specific genre. What I have found interesting in the case of *experience* is that there do appear to be identifiable differences between the meaning with which *experience* is associated in PP and its meaning in a general population of texts. This was not something that I had predicted, but the findings I present below make a strong case for this being so.

57.5 The local semantic prosody of *experience* in PP results from the predominant associations it takes on in this environment – and these associations are common to all the proposals in the PP corpus as *experience* is a key-keyword.

Table 79 - experience: left sort and *Table 80 - experience: right sort* provide clear examples of the collocates of *experience* in PP. The typical contexts of use are:

Preceded by	#	%	Preceded by	#	%
and	38	9.77	EU	7	1.80
has	24	6.17	have	7	1.80
considerable	23	5.91	their	7	1.80
Our	21	5.40	This	6	1.54
years	17	4.37	with	6	1.54
-	16	4.11	direct	5	1.29
of	16	4.11	substantial	5	1.29
.	15	3.86	)	4	1.03
extensive	15	3.86	broad	4	1.03
International	14	3.60	on	4	1.03
the	12	3.08	relevant	4	1.03
practical	11	2.83	consultancy	3	0.77
wide	11	2.83	depth	3	0.77
Project	10	2.57	His	3	0.77
-	7	1.80	professional	3	0.77

Table 86 - PP experience left context (82.53% of all instances)

Followed by	#	%	Followed by	#	%
of	148	38.05	is	5	1.29
in	94	24.16	from	4	1.03
and	20	5.14	*consultant name*	3	0.77
elsewhere	12	3.08	both	2	0.51
to	10	2.57	for	2	0.51
.	9	2.31	includes	2	0.51
,	7	1.80	over	2	0.51
relevant	7	1.80	which	2	0.51
as	6	1.54	with	2	0.51

Table 87 - PP experience right context (86.6% of all instances)

57.6 The overall picture of *experience* in the context of PP is that it is:

- frequently linked to another noun with *and*
- frequently associated with the verb *have*
- frequently qualified by a noun or adjective (considerable, extensive, international, practical, relevant, wide, EU, project, training, working, years) which emphasises the superior quality of the experience
- frequently followed by further specification of the kinds of *experience* in question (postmodifying prepositional phrase introduced by *of* or *in* – 62.2% of all right context words).

57.7 An interpretation of *experience* in the context of PPs can be summarised by a COBUILD dictionary style definition (my apologies to professional lexicographers!!):

DEFINITION: Experience is a form of professional capital which can be used to warrant opinions or recommendations and establish the authority of one consulting or management agency over and above that of others.
EXAMPLE: *company name* has gathered extensive knowledge and experience in transferring and adapting …
…The wider knowledge and experience of management training…
…Our input will be to draw on successful experience elsewhere and help local staff… … These examples of project experience demonstrate not only a strong …
…DHV has accumulated the knowledge and experience to assist institutions…

57.8 Such a definition stands in contrast (quality of defining style apart) with those provided by the COBUILD dictionary itself:

> **Experience** is knowledge or skill in a particular job which you have gained because you have worked at the job for a long time.
> EXAMPLE: I had no military experience...
> ...in my experience as a teacher...
> ...experience of working with children...
> He was senior to me in experience...
>
> **Experience** is the state or process of feeling something or being affected by it.
> EXAMPLE: The experience of colour is wholly subjective...
> ...the experience of fear.
>
> **Experience** is all the events, knowledge, and feelings that make up an individual's life or the character of a society.
> EXAMPLE: Everyone learns best from his own experience...
> ...speaking from personal experience.
>
> An **experience** is something that happens to you or something that you do, especially something important that affects you.
> EXAMPLE: The funeral was a painful experience...
> ...my later experiences in the village.
> (Sinclair et al, 1987)

57.9 While COBUILD Definition 1 does in some senses include the PP meaning of the word, it has not been designed to accommodate the specific local semantic prosody which *experience* gains in the PP environment. This is not a criticism of the dictionary; rather, it is a comment on the way in which a particular environment (co-texts, readership) colours the meaning of words. The spoken subset in BNC Core offers 100 instances of *experience* which do conform[29], more or less, to the senses offered in the COBUILD definitions. Examples are given in the table below.

Definition 1. office experience work experience skills experience	**Definition 3.** my experience is in your experience is it something that you talk… personal experience I speak from experience Madam Chairman firsthand experience
Definition 2. *no instances*	**Definition 4.** depressing experience good experience terrifying experience horrifying experience I had an experience with a bike it was quite an experience What an experience One told me his experience, on the phone from Waco prison

Table 88 - BNC Spoken Corpus – Experience: semantic prosodies other than "professional"

The spoken corpus also provides examples of the sense in which I feel *experience* is used in PP.

```
at wherever possible, the expertise and experience of the Board will be
 idea of building on past knowledge and experience. There 's
e is so much fund raising knowledge and experience amongst our volunte
ts so I think Deutsch Aerospace had any experience in that at all did
o come up with these proposals have any experience in recruitment.
and er they have picked up considerable experience from that but, you
```

Table 89 - experience: BNC spoken corpus data

57.10 Significantly, these 'professional experience' uses occur in three particular environments:

[29] A full concordance listing is in Appendix: Experience_BNC_Spoken

- non-phrasal co-ordination – already commented on as significant in the formal suasive wording of PPs (and paired with *knowledge* – a strong collocate of *experience* in PPs)

- as the object of lexical verb *to have* + following preposition *in* – again, a feature of *experience* in PPs

- qualified by *considerable* – once again, a strong collocation in PPs

The contrast between *experience* in PPs and in Spoken BNC Core is now clear: while Spoken BNC Core contains most meanings offered by the COBUILD lexicographers, PP contains no instances of COBUILD Meanings 2, 3 and 4, and the two isolated instances of Meaning 1 are exceptions which prove the rule. In example 1, "work experience" is part of the consultant's earlier profile, but he has since taken on a much broader set of *professional* interests. In example 2 "work experience" is what the professional provides for other people…

> "His education, and earlier **work experience**, was an industrial chemist, but since 1988 his main interest has been in SME development.
>
> … selection; implications in training, **work experience** and placement; sexual harassment … "(BNC Spoken Corpus data)

57.11 What we see in operation in PP is a local semantic prosody which has been specific to this genre or to PPs and other analogous genres. *Experience* has not been given a new, technical meaning in PPs. Rather, a local semantic prosody (implicit in the COBUILD 1 definition), which may be unique to PPs – though intuitively I do not consider this to be the case – has been similarly

exploited by writers in three different organisations as part of the suasive rhetoric of the proposals.

58. What structure(s) does it appear in?

58.1 In paragraph 48.2 we made the connection between this question and the term *colligation*, defining this as "the grammatical company a word keeps". Our earlier consideration of *experience* has already introduced many of its main colligates in PPs. These are summarised for PPs in *Table 90 - BNC/PP experience: left colligates*, in which, by way of contrast, the same data is also provided for colligations of *experience* in BNC Spoken Core. This corpus has again been chosen as a comparator as there are insufficient instances of *experience* in the RomFict collection we have used in earlier analysis to provide a basis for useful comparison.

58.2 A complete account of the immediate colligations of *experience* in these corpora is provided in the Appendices (Excel Spreadsheet 3: Experience colligations (XL Experience_colligation and other data.xls)). The two tables below summarise information for the top 5 immediate colligates in each text collection..

PP top 5 left context		**BNC top 5 left context**	
30.88	adjective	23.00	adjective
13.12	noun	22.00	noun
9.77	coordinating "and"	20.00	phatic
7.97	verb have	13.00	coordinating "and"
6.68	determiner	4.00	verb have
73.42	percentage of total	87	percentage of total

Table 90 - BNC/PP experience: left colligates (all counts are percents)

PP top 5 right context		**BNC top 5 right context**	
38.05	of	18.00	preposition
29.3	preposition	15.00	full stop
5.14	coordinating "and"	13.00	of
3.34	adverb	11.00	clause punctuation
3.1	adjective	6.00	coordinating "and"
83.93	percentage of total	68	percentage of total

Table 91 - BNC/PP experience: right colligates (all counts are percents)

58.3 Taking left colligations, there is a striking similarity between the incidence of the most frequent immediate colligates. *Experience* is premodified by either an adjective or noun in both corpora to almost exactly the same extent: PP = 44% / BNC Spoken Core = 45% – though the balance of adjectives is greater in PP. Likewise, coordinating *and* and verb *have* are both in the top five. The only major point of contrast between the two corpora arises directly from their different modes of production – the category *phatic* subsuming the various "mms", "hmms" and "ums" that arise in the BNC Spoken Core transcripts. Immediate right colligates show a greater contrast with the striking predominance of *of* + *prepositions* in PP (67.35% of all instances / 31% in BNC Spoken Core), reflecting this tendency in PP as a whole.

58.4 To look in more detail at the colligational patterning in PP it was necessary to exploit the POS coding which had been added to the corpus. Although WordSmith makes it possible to sort on individual codes, it does not permit automatic counting of patterns. To do this it has been necessary to transfer the output from the concordancer to a spreadsheet where such computations can be made. Data is then available in a form which gives access to three colligates to the left and right of the centre word.

L3		L2		L1	Centre	R1		R2		R3
[NN1]	of	[IO]	their	[APPGE]	experience	[NN1]	in	[II]	Central	[JJ]
[.]	From	[II]	our	[APPGE]	experience	[NN1]	in	[II]	Nizhny	[NP1]
[VV0]	31	[MC]	Our	[APPGE]	Experience	[NN1]	in	[II]	Assisting	[NP1]
[NP1]	;	[;]	Our	[APPGE]	experience	[NN1]	in	[II]	Hungary	[NP1]
[TO]	bear	[VVI]	their	[APPGE]	experience	[NN1]	from	[II]	elsewhere	[RL]
[RR]	,	[,]	their	[APPGE]	experience	[NN1]	in	[II]	transferring	[VVG]

Table 92 - Colligational information

58.5 The procedure developed to obtain the data summarised in the tables below is reported in Appendix15 : Extracting colligation information from POS coded text. Full data is given on the disk which accompanies this thesis (File Name:

XL Experience_colligation and other data.xls). Initial conclusions about the colligational relations of *experience* in PPs are summarised in the table below: *Table 93 - PP vs BNC Colligation*[30] which gives information on Left 1 + centre + Right 1 colligations. When interpreting this table it should be remembered that R1 (NN1) represents the POS tag for *experience*. L1 is the position of the POS tag for the word immediately to the left of *experience*, and R2 the position of the POS tag for the word immediately to the right of the search word.

PP L1 R1 R2	totals	% of 390.00	BNC Spoken L1 R1 R2	totals	% of 100.00	BNC Written L1 R1 R2	totals	% of 124.00
[JJ] [NN1] [II]	46	11.83	[JJ] [NN1] [.]	6	6.00	[AT] [NN1] [IO]	6	4.84
[JJ] [NN1] [IO]	42	10.80	[AT] [NN1] [IO]	4	4.00	[JJ] [NN1] [IO]	6	4.84
[VHZ] [NN1] [IO]	20	5.14	[NN1] [NN1] [.]	4	4.00	[JJ] [NN1] [,]	5	4.03
[CC] [NN1] [IO]	16	4.11	[APPGE] [NN1] [,]	3	3.00	[JJ] [NN1] [.]	5	4.03
[-] [NN1] [IO]	15	3.86	[APPGE] [NN1] [VBZ]	3	3.00	[JJ] [NN1] [II]	5	4.03
[NN1] [NN1] [II]	12	3.08	[APPGE] [NN1] [CC]	2	2.00	[CC] [NN1] [.]	4	3.23
[APPGE] [NN1] [II]	11	2.83	[APPGE] [NN1] [FU]	2	2.00	[JJ] [NN1] [CC]	4	3.23
[AT] [NN1] [IO]	10	2.57	[APPGE] [NN1] [II]	2	2.00	[JJ] [NN1] [VHZ]	4	3.23
[NNT2] [NN1] [II]	10	2.57	[AT] [NN1] [CC]	2	2.00	[II] [NN1] [,]	3	2.42
[.] [NN1] [IO]	7	1.80	[AT] [NN1] [CST]	2	2.00	[AT] [NN1] [CC]	2	1.61
percentage of total		48.59	percentage of total		30	percentage of total		35.49

Table 93 - PP vs BNC Colligation

58.6 This table gives evidence for the following colligational relations for *experience*:

PP	BNC Spoken / Written
1. top ten colligates in PP = 48.59% of total 2. there is a greater concentration of instances of colligation in a smaller number of types than in the comparator corpora	*top ten Spoken = 30%* top ten Written – 35.49%
3. as could be predicted from earlier comments all of the top 10 colligates in PP have a following preposition (II) or "of" (IO).	*only 2 out of Spoken top ten have a following preposition (II) or "of" (IO)* 3 out of Written top ten have a following preposition (II) or "of" (IO)
4. the pattern Adjective + *experience* + preposition / of accounts for 22.63% of all L1+R1+R2 colligates in PP – *experiences* are frequently premodified and further specified by *of* or prepositional phrase	*no instances in Spoken top ten* 4.84% of L1+R1+R2 colligates in BNC Written
5. Verb "has" + experience + of = 5.14% of instances in PP	*no occurrences in top ten in Spoken or Written*

[30] If the reader is not familiar with the codes used in CLAWS POS mark-up they may wish to refer to APPENDIX - CLAWS7 TAGLIST

Table 94 - Experience: colligates

58.7 Other three-word patterns can be identified by sorting and counting L2 + L1 + R1 and R1 + R2 + R3 – see *Table 95 - PP Colligation (Left/Right).* None of these combinations has the same degree of patterning as L1 + R1 + R2; the highest incidence (10% of instances) being found in the right colligation *experience* + of + present participle (e.g. experience of managing…), although this is worth noting as there is some evidence to suggest that it is a structure which is not well known or used by foreign language learners[31]. Other larger combinations are relatively rare (Table 96 - PP Colligations Left + Right patterns), although again it is worth noting that there are 13 instances of [JJ] [NN1] [IO] [VVG] (i.e. … considerable experience of managing..) and 7 instances of [VHZ] [JJ] [NN1] [IO] [VVG] (i.e. … has considerable experience of working …). It would seem that the (possibly prefabricated) chunks of text which writers of PPs work with can be quite large.

L2 L1 R1	totals	390.00
[VHZ] [JJ] [NN1]	34	8.72
[NN1] [CC] [NN1]	22	5.64
[VH0] [JJ] [NN1]	19	4.87
[MC] [NNT2] [NN1]	15	3.85
[NN1] [IO] [NN1]	12	3.08
[.] [-] [NN1]	11	2.82
[JJ] [NN1] [NN1]	10	2.56
[NN2] [CC] [NN1]	10	2.56
[PPHS1] [VHZ] [NN1]	10	2.56
[CC] [JJ] [NN1]	9	2.31

R1 R2 R3	totals	390.00
[NN1] [IO] [VVG]	39	10.00
[NN1] [II] [AT]	33	8.46
[NN1] [IO] [NN1]	28	7.18
[NN1] [IO] [AT]	27	6.92
[NN1] [IO] [JJ]	24	6.15
[NN1] [II] [VVG]	16	4.10
[NN1] [II] [JJ]	15	3.85
[NN1] [II] [NN1]	14	3.59
[NN1] [II] [NP1]	11	2.82
[NN1] [IO] [NN]	8	2.05

Table 95 - PP Colligation (Left/Right)

L1 R1 R2 R3	totals	390.00
[JJ] [NN1] [IO] [VVG]	13	3.33
[JJ] [NN1] [II] [AT]	12	3.08
[JJ] [NN1] [II] [VVG]	12	3.08
[JJ] [NN1] [IO] [NN1]	10	2.56
[JJ] [NN1] [IO] [AT]	9	2.31
[JJ] [NN1] [II] [JJ]	8	2.05

L2 L1 R1 R2 R3	totals	390.00
[VHZ] [JJ] [NN1] [IO] [VVG]	7	1.79
[MC] [NNT2] [NN1] [II] [AT]	6	1.54
[VH0] [JJ] [NN1] [IO] [NN1]	6	1.54
[VHZ] [JJ] [NN1] [II] [AT]	5	1.28
[VHZ] [JJ] [NN1] [II] [VVG]	5	1.28
[VHZ] [JJ] [NN1] [JJ] [II]	5	1.28

[31] Of 183 instances of experience in a 400,000 word sample from the Longman Corpus of Learners' English there were only 8 instances of experience of *ing and four of these were direct quotations from an essay with the title "Discuss your experience of learning a foreign language"!

[JJ] [NN1] [IO] [JJ]	7	1.79
[-] [NN1] [IO] [VVG]	6	1.54
[NNT2] [NN1] [II] [AT]	6	1.54
[VHZ] [NN1] [IO] [VVG]	6	1.54

[.] [-] [NN1] [IO] [JJ]	4	1.03
[.] [-] [NN1] [IO] [VVG]	4	1.03
[VHZ] [JJ] [NN1] [IO] [AT]	4	1.03
[DD1] [JJ] [NN1] [II] [AT]	3	0.77

Table 96 - PP Colligations Left + Right patterns

59. Is there any correlation between the word's uses / meanings and the structures in which it participates?

59.1 In explanation of the possible implications of the relationship between collocation, colligation and meaning, Hoey proposes what he calls the "drinking problem"[32] hypothesis. Although the medium in which I am working prevents me from replicating Hoey's graphic demonstration of this hypothesis at the time when it was first proposed at the Łódź *Practical Applications of Language Corpora* Conference in 1997 – tipping a polystyrene beaker of water onto the middle of the forehead rather than into the mouth! – I am able to offer Hoey's (less striking but no less clear) summary:

> "- Where it can be shown that a common sense of a word favours common colligations, then the rare sense of the word will avoid these colligations.
>
> - Where two senses of a word are approximately as common (or as rare) as each other then both will avoid the colligational patterns of the other.
>
> - Where either a) or b) do not apply, the effect will be humour, ambiguity (momentary or permanent) or a new meaning combining the two senses." (Hoey M, 1997a:6)

59.2 We have already seen that BNC Spoken Core contains three of the senses of *experience* offered in the COBUILD Dictionary (paragraph 57.8) and I have argued above that the sense of "*experience* as professional warrant" is neither fully covered by the dictionary definitions nor frequently occurring in

[32] Hoey was referring here to a joke in the film "Airplane" – the joke depended on the potential ambiguity in phrases such as "drinking problem" – a problem with alcohol? a problem with getting liquid into my mouth?

comparator corpora. A review of *experience* in BNC Written Core (143 instances) reveals a similar pattern (see Appendix 11: Experience – BNC Written.doc). The only instances where there may have been strongly similar colligations in BNC Written and PP were for the [JJ] [NN1] [IO] (adjective + experience + of) pattern. When the actual contexts which generated this result are seen, there only proves to be one likely candidate – example 6 below. The others are cancelled by the existence of contradictory elements in the extended collocational contexts – e.g. "little" in 1., "my" in 2., "No" in 3., "single" in 4., and "My" in 5.

1	y by users with little previous experience of computers. People who were
2	losest friend. Because my first experience of love was here, I have neve
3	iv Intra-party discord No long experience of membership of a constituen
4	as stopped dead, and the single experience of one coin being spun one ha
5	div1 THE CLOUDS GATHER My first experience of service with HM forces was
6	ide opportunities for firsthand experience of work in the health service

Table 97 - "experience of" in BNC Written

59.3 In other instances, the dominant colligational patterns for *experience* in PP – adjective + *experience* + preposition, adjective + *experience* + of, verb "has" + adjective + *experience* – are absent, as is the specific sense in which the word is used in PP. A comparison of the normalised counts (per thousand) for PP and the BNC Core sets (Table 98 - experience: counts) confirms the relatively high incidence of *experience* in PP – although this difference in frequency of occurrence would not be sufficient on its own to account for the different colligational and collocational relations that *experience* takes on in the PP Corpus. It would seem – as is predicted by the drinking problem hypothesis – that a different meaning for a word requires different colligations and collocations.

	BNC Written	BNC Spoken	PP
Total	1,080,072.00	1,050,593.00	114,490.00
Experience Count	143.00	100.00	389.00

Normalised Count	0.13	0.10	3.40

Table 98 - experience: counts

60. Is the word associated with (any positions in any) textual organisation?

60.1 Hoey (1997a:10) reported that he found the word *ago* is twice as likely to occur in paragraph initial position, and 13% of all sentence initial instances of *ago* were text initial in the (largish) corpus he had to hand. Hoey (1997b) also reports an experiment in which he demonstrates the role that lexical clues play in enabling subjects to identify paragraph boundaries. In his 1997 study he showed that when certain word classes (e.g. names of people) are sentence initial, they also have a strong tendency to be paragraph initial:

> "What appears to be the case is that the wording at the beginning of the sentence dictates whether we perceive it to be a topic sentence. In other words, topic sentences are the product of micro-wording choices, they do not dictate such choices." (Hoey M, 1997b:163)

A study of lexis seems to have something to teach us about texts.

60.2 I initially considered two major positions that *experience* can occupy in PP: sentence initial, and paragraph / sentence initial. My test for sentence initial was that the word (or noun phrase where *experience* was preceded by a determiner) occurred immediately after a preceding full stop (*.) or a paragraph start code (<para> in PP Corpus – a paragraph was defined as any span of text bounded by double hard returns: ¶¶). The first results indicated that there are 85 instances in PP where *experience* is sentence initial. Closer examination shows that this figure is misleading as there is a significant number of textual units in PP which are neither paragraphs not sentences. I propose to use the categories *headings* and *bullet points* to account for these categories.

60.3 Using these categories, *experience* can be found in PP in the text unit initial positions listed below. (figures for BNC Core Written and the OUP / Independent Newspaper corpus are given for comparison)

Corpus	**# words**	**instances of *experience***	**initial**	**% instances**	***% text initial***
PP – Paragraphs / Sentences	114,490	389	85	21%	
Sentence			*29*	*7.4%*	*34% of text initial*
Paragraph			*7*	*1.7%*	*8% of text initial*
Bullet Point			*15*	*3.9%*	*18% of text initial*
Heading			*48*	*12.4%*	*56% of text initial*
BNC Core Written	1,080,072	142	5	3.5%	
OUP / Independent Newspaper	1,087,858	128	9	7.1%	

Table 99 - PP sentence initial

60.4 The PP figure of 21% for *experience* as text unit initial is strikingly high, although looking at the sentence initial category on its own, it is important to note that sentence initial *experience* is only 7.4% of total instances – a similar percentage to sentence initial *experience* (7.1% of 128) in the 1 million word OUP Independent text collection. The high incidence of text unit initial *experience* is, clearly, largely accounted for by its use in headings and bullet points (74% of text initial *experience*). This demonstrates a conjunction of two noteworthy aspects of the PP corpus. The first has already been commented on – the prominence of certain key words. The second is the pronounced use of a hierarchic structure with numbered headings and subordinate bullet points. This is a form which is not explicitly demanded by the Terms of Reference themselves (see Appendix 33: PHARE Contract - TOR.doc), although it is clearly encouraged by the way (e.g. *Table 100 - Document structure*) in which the Terms of Reference themselves are structured:

WORK PROGRAMME

a2.2 The overall objective of this project is to help develop a successful and sustainable RDA. The proposed approach is shown schematically in figure a2.1. The following paragraphs describe briefly the main elements of each of the tasks shown in figure a2.1.

Phase 1: initial review

Objective: to agree with the client the main training requirements and potential projects for further development and implementation in phase two, based on a realistic assessment of the area's needs and resource availability, and drawing on best practice from elsewhere.

Task 1: review current activities and plans

a2.3 The first main task will involve a review of all current activities and plans of the RDA. This will need to be undertaken in the light of the following factors, for which background information will be collected as part of this task:

- the expectations and constraints of CCC
- regional economic and physical development needs and priorities
- financial resources

Table 100 - Document structure

60.5 These text organisers are referred to as textual-mapping devices by Bhatia who comments on their rapid development in legal English:

> "Some fifty years ago, these textual-mapping devices were almost non-existent, but, in present-day British legislation, they are quite common." (Bhatia V 1993:142)

… going on to comment on the way in which their use was much rarer in other Commonwealth legislatures. Here is an example of the way in which generic changes introduced in one discourse community are not necessarily immediately incorporated by analogous communities. Their use in PP, and the intertextual resonance they have for readers (PPs are more like *these* texts, less like *those*), will be discussed in the final chapter in this section,

61. Conclusion

61.1 We began this chapter with a discussion of some of the pedagogic limitations of the Biber framework for analysis, arguing that learners not only need information about the stylistic and grammatical features of difficult texts, but that it is also helpful for them to develop an understanding of the lexical patterning of a target genre. This was followed by two short detours. The first considered the value of keyword analysis to show differences between genres. The second showed how a combination of keyword analysis and Hoey's "five questions" provided a practical way of finding things out about the important words in a text. These two detours were then followed by an extensive account of one of these words: *experience*.

61.2 The results we have obtained so far can be summarised as follows:

- It is possible to identify important words (keywords) in a text or set of genre exemplars. Importance can be ascribed on the basis of significant presence or significant absence.

- Software tools allow you to study hundreds of examples of these words and to identify their collocations, colligations, semantic prosodies and roles in text development.

- The results of such analysis give the learner access to the phrasal collocations of these keywords, along with an understanding of their roles in discourse development.

61.3 The degree of analysis to which *experience* was subjected could be considered a cruel and unusual punishment, and would certainly not be required in order

to develop a pedagogically sufficient understanding of which words are important in a difficult text and how these words behave. The analysis undertaken above does, however, offer a range of possible techniques which a teacher or learner could make use according to their particular needs. We will consider how such an approach to the teaching of this sort of text could be implemented in practical teaching contexts in a later chapter.

SECTION TWO – AN ACCOUNT OF THE TEXTS

Chapter 7: Organisation

> "You should be aware that PHARE tender documents are highly standardised and the form of the tender is very prescriptive. Thus, 'salient common features' of the text are highly probable." (correspondence from supplier of corpus data May 1996)

62. Introduction

62.1 When I made my initial request for data to consultancy groups interested in winning the management of PHARE Projects, I met with two main kinds of reaction. The first was what I had expected – the documents I wanted to study had been developed after a significant investment in time and money, and could not be released because they were commercially confidential. Another group of correspondents had a very different response. They were a little surprised that I was interested in the texts they wrote – the comment given at the beginning of this chapter being fairly typical. Although writers in these organisations also recognised that a large amount of effort and cost was involved in the development of project proposals, they felt that – qua texts – there was not much difference between the proposals themselves, and that I would not find much of interest in their language. What really mattered in their opinions was the CVs of the consultants and the bottom lines in the financial proposals. This was what won the contract.

62.2 I have argued in the chapter "Writing Difficult Texts" that one way of accounting for difficulty in writing a text is in terms of a writer's unfamiliarity with a given genre – and that this was the kind of difficulty on which I would focus. In that chapter, I also said that as a result of the work undertaken in this thesis, I hope to be able to offer "pedagogic pathways into unfamiliar texts so

that that which was difficult because it was unfamiliar becomes less strange, and thereby easier". From such a perspective, if there are predictable regularities across a particular genre, this does not constitute a disadvantage. Rather, it makes the genre easier to approach, both for teachers and for apprentice writers – especially those working in their first or a second language. In the chapters "Grammar and style" and "Lexical dimensions" I have already noted several non-intuitive regularities which have potential pedagogic value. In this chapter I will review the extent to which there are other regularities above the clause complex and at whole text levels.

62.3 Bhatia (1993) suggests a seven step approach to genre analysis. This provides a useful summary of much of what has been done up to and including this chapter:

"1. Placing the given genre-text in a situational context
2. Surveying existing literature
3. Refining the situational/contextual analysis
4. Selecting the corpus
5. Studying the institutional context
6. Levels of linguistic analysis
    analysis of lexico-grammatical features
    analysing the text-pattern or textualisation
    structural interpretation of the text-genre
7. Specialist information in genre analysis"
(Bhatia VJ 1993:22-34)

62.4 Steps 1 - 4 were largely accounted for in "Writing difficult texts" and "Approaching the data", and we addressed a number of the key issues in step 6 in "Grammar and style" and "Lexical dimensions". The major steps that remain are:

- a more detailed consideration of the *institutional context*
- the structural interpretation of the genre

- the collection of specialist information.

  Specialist information and comments on the institutional context will be presented in Chapter 8. In this present chapter I shall pay attention to aspects of the structural interpretation of the genre, focusing on the organisation of Project Proposals.

62.5 As I have already mentioned, practitioners do not appear to feel that there is much of interest to discover in an analysis of the organisation of PPs. At first sight this is, perhaps, true. So long as you read the TOR and the ITB carefully, the overall structure of the proposal can be treated as given. I have found, however, that an analysis of the texts in the corpus can reveal information and insights of considerable potential value to the learner. First, it shows the extent to which the TOR are not only a source of guidance with regard to the structure of the PP, but also a very valuable linguistic resource. Secondly, and more importantly, the analysis indicates that the writers of the PPs in the corpus share remarkably similar approaches to the organisation of their arguments at paragraph and section level.

62.6 This has a bearing on comments that I have made in earlier chapters in this thesis and offers good reasons for investing the time required to come to a fuller understanding of the organisation of the genre in which you have an interest. Thus, in Issue 2 at the end of Chapter 2: *Teaching Writing*, I said that students need multiple examples of a genre in order to have a reasonable chance of making appropriate generalisations about a difficult text they are learning to write; and in Chapter 3: *Approaching the data*, I commented that it would be important for learners to review all of the sections of a new text if

they were going to give themselves the best chance of success in writing into a new genre. Although it will not help learners collect examples of the target genre, the kind of analysis that I present in this chapter does offer a way of working with them once they have been collected!

62.7 In undertaking this analysis of the organisation of PPs I have adopted two approaches, each offering quite different perspectives on what is going on in these texts. PPs can be considered to be reader-driven texts *par excellence*; their highly structured hierarchic heading systems largely determined by the requirements of the agency that issued the TOR and ITB. The first analysis, therefore, relates to the macro structure of PPs, focusing on the immediately apparent organisational information that they all share – section headings. My interest here is to see if it really is the case that the TOR have such a strongly determining role in deciding the organisation of PPs.

62.8 The second analysis was chosen to complement this first approach. My overall conclusion regarding macro structure is that although there is much that we can say about the *language* of section headings and the relationship between the TOR and the resultant PP, we cannot build a model of PP organisation in general. Each proposal is very strongly linked to the TOR to which it is responding. In this sense, the correspondent quoted above is correct: "PHARE tender documents are highly standardised and the form of the tender is very prescriptive" – by implication, if you want an explanation of a PPs macro organisation, go back to the TOR. What I felt I needed in order to complement the first analysis was, therefore, something that would allow me to look at units above the clause, and which would allow me to see if there were any significant patternings in PPs as discourses that were independent of

the TOR. Notions of minimal discourses developed by Winter and Hoey (Hoey M 1983, Winter EO 1976, 1977, 1982, Hoey M & E Winter 1986) offered such an analytic framework, and it has proved to be very productive.

63. Macro structures

63.1 All of the texts in the PP Corpus have explicit organisational features. As a minimum, these take the form of numbered main headings (usually conforming with ITB specifications) and numbered paragraphs. Some consultancy groups have opted for a more elaborate hierarchy of sections headings and subsection headings, using either automatic or manual systems for number maintenance. Others have used relatively complex systems of apparent heading and subheading, but these have been signalled typographically rather than numerically, and often, in fact, prove to be inconsistent across the individual document. In order to ensure that I was making meaningful comparisons between texts in the corpus, it was necessary, therefore, to edit the data to ensure consistent heading and subheading numbering across the corpus. When the original texts were available as word processor files, "sections" were identified on the basis of specific MS Word *heading number* styles or on the basis of internally referenced typographic conventions (hierarchies of font size, emboldening and italicising). When texts were scanned, internal paragraph and section numbers had to be identified on the basis of typographic conventions alone. In all instances, the original documents have been reprocessed and section heading levels have been made consistent so that a new "table of contents" can be generated with an appropriate level of delicacy for the particular text. The most complex section hierarchies had four levels. The simplest had two. When I talk of

"levels" and "sections" below, it is, therefore, on the basis of these explicitly signalled text organisation features.

63.2 A first way of seeing if there is any overall structural similarity between texts in the PP Corpus is to consider the explicit headings that the writers ascribe to sections of the proposals. The first level of analysis of this issue can be purely quantitative - how many labelled sections do the texts have? The second level of analysis will be concerned with the wordings used in the headings themselves. In the same way that Swales' comments …

> "The genre names inherited and produced by discourse communities and imported by others constitute valuable ethnographic communication, but typically need further validation." (Swales J 1990:58)

… the names of the headings and subheadings in PPs also give insights into the concerns and rhetorical purposes of this group of writers.

63.3 Three consultancy groups are represented in the PP Corpus – CGA, CGB, CGC – and they have contributed fourteen texts in all – CGA (5), CGB (6), CGC (3). Each contains a number of distinct "declared" sections – i.e. explicit section headings which have been indicated as such either typographically or numerically. Table 101 lists the number of declared sections in the 14 texts in the PP Corpus, and also gives information on the number of words in the complete text, along with a normalised count for declared sections per thousand words in order to provide a more useful basis for comparison than the raw counts.

consultancy group + proposal identifier	section count	total words	normalised section count (per 1000 words)
CGA LAB	29	2,861	10.14
CGB BK182	20	2,873	6.96
CGB BK65	20	3,413	5.86
CGC BULG	68	12,879	5.28
CGA ENV	55	13,263	4.15
CGC CHEREPOV	17	4,345	3.91
CGA MIDEP	27	8,597	3.14
CGB BK47	16	5,171	3.09
CGB BK-178	16	6,458	2.48
CGC VLAD	28	11,703	2.39
CGB BK-162	15	7,434	2.02
CGA SCIT	20	10,510	1.90
CGB BK-11	13	10,867	1.20
CGA TERM	12	13,824	0.87

Table 101 - PP Corpus - level counts

63.4 This analysis reveals a pronounced range in the use of section headings across the PP corpus, both within consultancy groups and between individual proposals, irrespective of the relative length of the particular documents. Consultancy Group A (CGA) contributed one text with the highest level of section elaboration (10.14 sections per 1000 words) as well as the lowest (0.87 sections per 1000 words). Similarly Consultancy Group B (CGB) provided texts with the second highest (6.96 sections per 1000 words) and second lowest levels of elaboration (1.20 sections per 1000 words). Why writers have elected to use section headings in such different ways is not evident from the texts themselves.

63.5 We have already said that the specific wording of the sections themselves warrants further consideration as there is a clear increase in diversity as one moves down the hierarchy. A complete summary of the section headings used in the corpus is given in Appendix 28: Organisation - Section Counts. When this is analysed it reveals that there is a significant degree of common ground in level 1 headings, and, predictably, much greater diversity at lower levels in the hierarchic text structure. Table 102 shows the counts for section headings at each of the four text levels identified in PP Corpus texts. Table 103 shows

the increasing diversity in section labels at lower hierarchic levels. Section Heading Level 1 has the lowest type:token ratio, while Section Heading Levels 3 and 4 have the greatest diversity, with no duplication of types evidenced.

section heading levels	section heading count
Level 1	62
Level 2	151
Level 3	98
Level 4	45
Grand total	356

Table 102 - Heading level counts

section headings	tokens	types
Level 1 Count	62	40
	Type / token =	**0.645**
Level 2 Count	151	123
	Type / token =	**0.814**
Level 3 Count	98	98
	Type / token =	**1**
Level 4 Count	45	45
	Type / token =	**1**

Table 103 - Sections type:token

63.6 When Heading Level 1 headings are looked at in more detail, the way in which they reflect the language and organisation used in the Invitation to Tender and the Terms of Reference issued by PHARE PMUs becomes apparent. For example, the text in Table 104 is part of a PHARE Invitation to Bid (ITB) and has been marked up (intuitively) to show the terms that are commonly found in such documents (the full text can be found in Appendix 31: PHARE Contract - ITB). Table 105 shows the different Heading level 1 section headings in the PP Corpus. Reading the marked up ITB in conjunction with Table 105 and the other heading levels exemplified in Appendix 31: PHARE Contract - ITB, provides an insight into how the terminology and superficial organisation of the PP texts are indeed constrained as a result of the "very prescriptive" nature of the form of the tender.

compulsory elements | optional / interpretive elements | optional / interpretive elements

3. CONTENT OF TENDERS

The tender submitted by the Tenderer shall fully comply with the requirements set out in the tender dossier and comprise:

3.1 A **technical proposal** consisting of:

3.1.1 **Statements and Terms of Reference** (future Annex A to the Contract):

3.1.2 A **statement of intention** to provide the services in conformity with this tender dossier by the Tenderer or his duly authorised agent.

3.1.3 The **Terms of Reference** for the services as they appear in the tender dossier, initialled on each page by the Tenderer.

3.1.4 A copy of the first page of the General Conditions for Service Contracts financed from PHARE Funds, signed by the Tenderer.

3.1.5 A **statement** concerning the sub-contracting envisaged for parts of the services, if sub-contracting is envisaged, signed by the Tenderer.

3.1.6 A statement concerning the bank account to which payments may be made.

3.1.7 A signature by the Tenderer or his duly authorised agent

3.2 **Organisation and Method** (future Annex B to the Contract): This part shall cover in particular:

a) **Organisation of the project**, **method of work** and **experts** proposed by the Tenderer as well as proposed **human resources**, **sub-contractors** and **physical means** to be provided by the Tenderer for the **project implementation**. This shall include a description in **narrative and chart form** of the consultant's distinctive **approach** to the requirements outlined in ANNEX A. The **criteria of evaluation** should take into consideration the categories which form the basis of the contractor's **inception report**.

b) A precise indication must be given concerning the **total amount of man-day / man-months / man-years** proposed for each **expert**, and whether these experts are EC or local experts. The proposal must be coherent with the Terms of Reference. The tenderer should indicate that the candidates proposed and the quantity of **services** offered; the number and the **job titles** of the staff concerned will be provided, with a description of the foundation and specific tasks assigned to the nominated **team leader** and **individual experts**, a **bar chart** indicating the **professionals**, showing periods on the duty station and in the home county, and estimates of total numbers of man-months required broken down by individual staff. The **financial evaluation** will consist in comparing the global price of comparable offers. With regard to human resources, the Tenderer should note that the Terms of Reference belong into one of the following categories, depending on the nature of the project:

- **Fixed Term Expert Assignment** (for instance for typical **technical assistance** oriented projects): In this case **manning requirements** in quantity is clearly given in the Terms of Reference and Tenderers must provide exactly the requested manning. The **technical evaluation** will concentrate on the candidates proposed by the Tenderer. The financial evaluation will consist in comparing the global prices (total prices minus reimbursables).

- Other composite services, where the Contracting Authority does not fix the inputs needed in advance and the Tenderers themselves must evaluate (and justify) the optimum manning needed. The technical evaluation will take into consideration both the candidates proposed and the quantity of services offered. in this case, the number and job titles of the staff concerned will be provided, with a description of the function and specific tasks assigned to the nominated team leader and individual expert, a bar chart indicating the professionals, showing periods to be in duty station and in the home country' and estimates of total numbers of man/months required broken down by individual staff. The financial evaluation will consist in comparing the unit prices (total prices minus reimbursables, divided by the number of offered working man-months of EC experts).

c) **Logistics Plan**: A description of the logistics planning foreseen for the whole **programme implementation**. Particular attention should be paid to **transport arrangements**, **freight forwarding terms**, and **INCOTERMS** in accordance with PHARE procedures. **Route maps** for overland travelling should be included.

d) **Risk Analysis**: On the basis of Annex A's requirements, the contractor should provide a risk analysis outlining the probability of success in the **time period** and highlighting **key factors** which have a bearing on the efficient and economic implementation of the project.

e) **Time Schedule** :The Tenderer must submit a **workplan** with envisaged specific reference to mobilisation of the team, **submission of reports** and documents, **specific meetings**, and **draft time** to methods of accelerating the **procurement** and **delivery** of

goods without associated cost escalating factors.

f) **Remarks**, **comments** and **suggestions** which the Tenderer may consider it advisable to raise.

g) A **List of Staff**; including **Curriculum Vitae** (future Annex C to the Contract) following the standard model enclosed in the tender dossier should be included.

h) Furthermore, the following should be included:

- A signed **commitment** from each of the experts to accept the work proposed by the Tenderer, if the Tenderer is awarded the contract,

- A document showing the **firm's experience** in the relevant field, and other information such as the **firm's structure** and size or the firm's **headquarter ability of back-stopping** the on-site operations (the same applies to the proposed sub-contractor(s), if there is any).

Table 104 - extract from PHARE ITB

Total	TITLES - HEADING LEVEL 1	Total	TITLES - HEADING LEVEL 1
7	ORGANISATION AND METHOD	1	GENERAL APPROACH
5	LIST OF STAFF	1	GENERAL CONDITIONS FOR SERVICE CONTRACTS
4	STATEMENT AND TERMS OF REFERENCE	1	INPUTS
3	INTRODUCTION	1	KEY ISSUES AND APPROACH
2	INTRODUCTION AND SUMMARY	1	LIST OF STAFF + CVS
2	KEY ISSUES	1	ORGANISATION OF THE PROJECT
2	STATEMENTS AND TERMS OF REFERENCE	1	PROJECT MANAGEMENT
2	TECHNICAL RESPONSE	1	PROJECT MANAGEMENT AND STRUCTURE OF TEAM
2	TERMS OF REFERENCE	1	PROPOSED APPROACH
2	WORK PROGRAMME	1	PROPOSED PROJECT TEAM
2	WORKPLAN	1	PROPOSED RESPONSE
1	ADVISERS	1	QUALIFICATION OF TERMS OF REFERENCE AND SCOPE OF WORK
1	ANTICIPATED OUTCOMES	1	RATIONALE
1	BACKGROUND AND EXPERIENCE	1	STAFF AND EXPERIENCE OF TEAM
1	COMMENTS ON THE TERMS OF REFERENCE	1	STATEMENT OF INTENTION
1	COMMITMENTS FROM EXPERTS	1	STATEMENT OF UNDERTAKING
1	CURRICULA VITAE	1	SUMMARY
1	CURRICULA VITAE OF PROFESSIONAL STAFF	1	TECHNICAL APPROACH AND WORKPLAN
1	EXPERIENCE OF TENDERER AND SUBCONTRACTORS	1	TECHNICAL PROPOSAL
1	EXPERTISE OF [Consultancy Group] 1 & [Consultancy Group] 2	1	WORK PROGRAMME RESOURCES AND SCHEDULE

Table 105 - Level 1 headings

63.7 The apparent match between the language of the TOR sample and observed language use in section headings was checked more objectively by creating a list from the words and phrases which had been underlined in the extract from the ITB (this list contains 51 items), and then counting the words which occur in both this list and the full listing of section headings. Details of this exercise can be found in Appendix 28: Organisation: Section Counts. The final result obtained shows that over 52% of the words and phrases informally identified in the extract from a single ITB were also to be found in list of Section

Headings collected from the whole PP Corpus. Such an approach proved helpful as I was not able to apply an automatic procedure to identify this kind of phenomenon. In particular, the Scott style key-word list for an individual ITB revealed no patterning connected with section headings. The terms used in headings gain their significance through their position in the text, not as a result of high frequency. This finding has a bearing on the issue of automatic text analysis as it indicates that it is important not only to consider the frequency of words, but also to consider *where* words are used. Hoey has already demonstrated how "boring" words can become interesting as a result of their position in a text – e.g. the word "sixty" is sentence initial position in 71% of the instances in which it appears in a large newspaper corpus, and is text initial in 14% of these instances (Hoey M 1997:21). By the same token, it also appears that it can be instructive to account for the provenance of the language used in obvious features like headings and sub-headings. There can be more to these features than meets the eye.

63.8 Two conclusions can be drawn from these initial analyses, and both have implications for pedagogy. The first is that the ITB and TOR constitute an important resource for proposal writers. This may appear to be a truism for expert writers, but is a fact that can easily be overlooked by apprentices. A close *reading* of the demands set by the ITB and TOR will usually provide clear guidance for *writers* about the overall structure the proposal should be given, and the expectations and motivations of potential readers[33]. The other conclusion is that ITBs and TORs are also valuable linguistic resources for writers approaching PPs for the first time. In particular, their terminology

[33] For confirmation of this observation, see Chapter 8.

gives a clear indication of the expectations of the PMU or other organisation which has issued the tender documents. Thus, they should be read carefully to ensure that a) concepts have been understood, and b) terminology can be adopted where appropriate. For example, in the extract from the ITB quoted above, 'backstopping' is an interesting neologism which was echoed in a level 2 section heading: *Backstopping, support and management*.

63.9 Other issues arise for pedagogy when lower heading levels are considered. One is that the proposals present considerable diversity with regard to the number of levels of section headings that are used. This leads one to ask *why* the writers have opted for high or low explicitness in the signalling of text organisation. The second point is that it is often the case that relationships between levels in the hierarchy are less than transparent. Again, what do writers hope to achieve through their use of numbering schemes and section headings, and to what extent are they aware of the impact of their use of these schemes on the overall reader-friendliness of their texts? The interviews that are reported in Chapter 8: *Writing Project Proposals* give some answers to these questions – particularly regarding the importance of the ITB as a resource for writers, and to a lesser extent, the way in which the different approaches that consultancy groups adopt to planning and developing a proposal determine the extent to which they go beyond the levels of organisation specified in the ITB. Further work on this aspect of PPs and similarly tightly specified genres could, however, be of interest.

64. Moves

64.1 If an analysis of explicit section headings gives us an indication of the predictable macro structures of PPs, are there other ways of coming to a clearer understanding of how these texts work as discourses[34] rather than as formal text structures? One possible starting point is the idea of the *move* as developed in Sinclair and Coulthard (1975). Drawing on the Hallidayan categories of "structure, system, rank, level, delicacy, realization, marked, unmarked" (Sinclair J & M Coulthard 1975:24), they proposed a five level system for the analysis of classroom discourse (I. Lesson; II. Transaction; III. Exchange; IV. Move; V. Act). There were five classes of move in Sinclair and Coulthard's analysis of classroom language: "*Framing* and *Focusing* moves realise boundary exchanges and *Opening*, *Answering* and *Follow-up* moves realise teaching exchanges." (Sinclair J & M Coulthard 1975:44). Although these categories were developed with a particular kind of interaction in mind, Sinclair and Coulthard suggested that some moves, e.g. Framing, were probably typical of all spoken discourse (ibid:44).

64.2 The notion of the move has since been applied to the analysis of written texts by e.g. Swales (1981) who proposed a typical four move structure for introductions to research article in the human, social and physical sciences:

Move 1:	**Establishing the research field**
Move 2:	**Summarising previous research**
Move 3:	**Preparing for present research**
Move 4:	**Introducing the present research**

[34] I use *discourse* here in Widdowson's sense: "Discourse is a communicative process by means of interaction. Its situational outcome is a change in a state of affairs: information is conveyed, intentions made clear. Its linguistic product is *text*." (Widdowson HG 1984:100). I find it clearer than those definitions offered by, for example, Fairclough 1989 or Clark & Ivanić 1997, where it has been given a "catch-all" status (Clarke R and R Ivanić 1997:14). Although it is narrowly cognitive, if one accepts that all texts are mediated in social contexts of one form or another, it has the benefit of letting one focus on how the text works as communication, rather than on how texts are produced.

64.3 Similarly, Bhatia discusses the "cognitive move structures" of Acts of Parliament (Bhatia VJ 1993:32), and Dudley-Evans discusses "move cycles" in academic articles, raising the important questions:

> "…. first, how do we make decisions about the classification of moves; and, second, how can we be confident of the validity of the moves and the move cycles that are posited?" (Dudley-Evans 1994:226)

For Dudley-Evans, practical answers to this question can be found in approaches such as that taken by Crookes G (1986)[35], or through recourse to direct discussion with informants with a specialist background in the field (Dudley-Evans T 1994:227).

64.4 When analysing classroom language or acts of parliament, a large part of the analyst's task is necessarily concerned with identifying stretches of text which have analogous communicative purposes and then establishing a classificatory system for the range of moves which the context requires. This kind of move analysis has proved to be less valuable for PPs than it might be for texts which have not been so explicitly labelled. On the basis of the 213 Level 1 and Level 2 headings that were discussed in paragraphs 63 ff. above, I have been able to identify six potential classes of *content element* in PP (for details see Appendix 27: Organisation – content elements). These are:

content element	level
Administration	15
Context	12
Expertise of the Firm	14
Obligatory element	71
Project Team and Consultants	24
Optional element	77
Grand Total	213

Table 106 - content elements

[35] A group of informants were asked to apply a Swales 4 Move model to the same article. Crookes was able to obtain agreement at above 0.6, thereby demonstrating the general usefulness of the Swales model (reported in Dudley-Evans T 1994:227)

64.5 I would not, however, be willing to call these *moves* in the sense proposed by Sinclair and Coulthard or Swales. They certainly do not have the same kind of predictive force as e.g. *Methods* / *Results* in laboratory reports (Dudley-Evans T 1985). Rather, they appear to have little or no predictability, but result from the interaction between the experience of earlier proposals prepared by the team of writers in the bidding organisation and the Terms of Reference (TOR) issued by the funding organisation. Thus, the content elements that I have marked Administration, Context, and Obligatory element relate directly to the requirements of the TOR which define the overall structure of the proposal. Optional elements, on the other hand, appear to be specific to a proposal – or even to the house style of the consultancy group involved in the bid (See Chapter 8 for a discussion of this aspect of text organisation).

64.6 While it is possible to attempt a categorisation of some of these content elements into discrete moves – e.g. AIM, METHODS and TEXT MAP within Obligatory Content Elements (see Appendix 27: Organisation – content elements).

label	move
aim	8
methods	52
text map	11
total	71

Table 107 - moves in obligatory content elements?

… such a set offers little explanatory value, and would offer no more assistance to learners than would a close reading of the TOR they were responding to. While this is not to claim that the idea of moves is of no

value[36], the analysis I have attempted of PPs indicates that the notion of *moves* is less useful if the text in question is highly specified by its potential readership.

65. Minimum Discourses

> "It is not enough to identify problems. Our teams are accustomed to finding solutions to the toughest of problems". (quotation from Proposal BK178 in the PP Corpus)

65.1 If an attempt to analyse moves has not been particularly fruitful, an alternative analysis of the organisation of PPs has, however, proved useful. This is derived from the work of Winter (1977, 1982), Hoey (1983), Hoey M & E Winter (1986) on minimum discourse patterns – notably **Situation**, **Problem**, **Response**, **Evaluation** (SPRE) (Hoey 1983:34). Although in his 1983 study Hoey (ibid:31) stresses that his choice of this particular pattern for close study should not be taken as privileging it above other common patterns (e.g. general ➔ particular, reason ➔ result), SPRE does appear to offer particularly rich insights into the structure of PPs. It has proved possible to identify a sequence of SPRE cycles in use across a number of randomly sampled passages from the Technical Component of PPs and this has offered a useful insight into the way in which PP writers organise their argument.

65.2 An individual example from CGA ENV is given in the following paragraphs. This proposal, written by Consultancy Group A, offers an example of the way in which a series of SPRE moves provides coherence to text which might otherwise be little more than a to-do list organised around a set of pre-

[36] Continuing work on academic discourse (e.g. Thompson P 1997, 1998) demonstrates that, although problematic, move analysis can have powerful explanatory value, especially in settings where writers are given little or no formal guidance regarding the requirements they must fulfil when writing into a particular genre).

determined section headings. A four level Table of Contents for CGA ENV is given in Appendix 5: CGA ENV – Table of contents. This confirms both the way in which the ITB / TOR drive the overall organisation and how the terminology of the section headings is driven by TOR language. In the first instance, we will focus on two content elements under (III) General Approach: *Goal* and *Strategy*.

65.3 In what is beginning to be an established tradition (Hoey M 1983) I tested my own analysis of this text by offering it in jumbled form to seven informants (all either teachers or administrators working for the British Council in Colombo, Sri Lanka; and all with educated native-speaker command of English). The text of this exercise is given below in Figure 18, the original task and tabulated results are in Appendix 43: Text sequencing activity)

Jumbled sequence

Strategy

It is acknowledged that the strengthening of Environmental Education is an essential element of reaching sustainable development in Poland. Teaching school-age children to care for the environment is an investment in the future.

It is our understanding that this project forms part of a programme of public awareness and curriculum development, building on existing educational activities and helping to create an institutional framework for Environmental Education. We highlight below the factors which we consider most important to maximise the impact and benefit of the technical assistance programme.

This project takes place in the context of a period of great change in Poland with the emergence of a modem advanced country taking its place in the European community of nations. The environmental challenges facing the country are complex with industrial and economic interests in the past being given precedence over ecological interests. A higher priority is now being given to environmental issues.

We acknowledge that changes are taking place rapidly in Poland and that development may alter the picture of Environmental Education, Our plan and our consultants are highly flexible and can respond to developments as they happen.

Figure 18 - Text sequencing exercise

65.4 Results from my seven respondents are shown below. 4 out of 7 of the respondents were able to reconstruct the passage exactly (57%). More interestingly, 86% were able to identify the opening and closing Situation /

Evaluation sections, with the Problem statement causing greatest confusion – although there was no consensus regarding the likely candidates for Problem.

seq	response 1	response 2	response 3	response 4	response 5	response 6	response 7
one	1	3	3	3	3	3	3
two	3	4	2	1	1	1	1
three	2	1	1	2	2	2	2
four	4	2	4	4	4	4	4

original	% original
1	86%
2	57%
3	71%
4	86%

65.5 By doing this very small study I was not attempting to add anything to the argument around minimal discourse patterns; that case has been well made elsewhere. What I was interested in was to check if my intuition about the relevance of this pattern in this particular context was reasonable – and it seems to have been. The original sequence of this extract is given in Figure 19 below. It will be noted that the Problem / Solution sequence were, in fact, in the same numbered paragraph in the original text. It is certain that the re-ordering would have been a great deal easier to do had the jumbled text been offered as a three paragraph problem.

Strategy

14. This project takes place in the context of a period of great change in Poland with the emergence of a modem advanced country taking its place in the European community of nations. The environmental challenges facing the country are complex with industrial and economic interests in the past being given precedence over ecological interests.

15. A higher priority is now being given to environmental issues. It is acknowledged that the strengthening of Environmental Education is an essential element of reaching sustainable development in Poland. Teaching school-age children to care for the environment is an investment in the future. It is our understanding that this project forms part of a programme of public awareness and curriculum development, building on existing educational activities and helping to create an institutional framework for Environmental Education. We highlight below the factors which we consider most important to maximise the impact and benefit of the technical assistance programme.

16. We acknowledge that changes are taking place rapidly in Poland and that development may alter the picture of Environmental Education, Our plan and our consultants are highly flexible and can respond to developments as they happen.

Figure 19 - ENV: Text sample

65.6 The extract exemplifies a rhetorical strategy that appears to be common to many of the texts in the PP corpus. SITUATION and PROBLEM are given elements, drawn directly from the TOR. As a result, in finite clauses in S/P cycles the grammatical subject / actor is associated with neutral, impersonal subject noun phrases:

SITUATION:	the project / environmental challenges / a higher priority
PROBLEM:	it / the strengthening of environmental education / teaching school age children

By contrast, responsibility in the RESPONSE / EVALUATION cycle lies with the consultancy group, and the grammatical subject / actor has, accordingly, a very different reference:

RESPONSE:	it (our understanding) / this project / we / we
EVALUATION:	we / changes / development / our plan / our consultants / they (developments)

Identifying this local phenomenon has led me to make a more detailed study of first person pronouns in the PP corpus, a study which has shed light both on the extent to which SPRE is spread across the proposals in the corpus, and the ways in which first person plural pronouns are strongly implicated in

Response / Evaluation. Results of this study are reported in paragraphs 65.7 to 65.14.

65.7 My earlier analysis of PP had led me to conclude that personal pronouns were a relatively low frequency element in PPs and were unlikely to play a major role in the structure of the texts. What I found was that first person *plural* pronouns actually have a relatively high frequency in PPs – a conclusion that runs counter to my earlier understanding of the genre. This apparent contradiction arises from a limitation in Biber 1988 – viz. that there is no differentiation between first person singular and first person plural pronouns. All first person pronouns are lumped together as a single class. What becomes clear when you look at the difference between uses of we/our/us and I/my/me in PP in comparison with a general corpus such as LOB, is that there are striking and important contrasts between the proposals in the PP corpus and other texts (Table 108).

	PP	normalised	LOB	normalised
we/our/us	767	6.699275	4942	4.015275
I/my/me	2	0.017469	10950	8.896653
first-person pronouns	***769***	***6.716744***	***15892***	***12.91193***
total words	114,490		1,230,800	

Table 108 - first person pronouns in PP and LOB

In PPs the number of first person singular pronouns is negligible (and in fact is the result of the inclusion of the text of a letter of intent in one of the PP texts) – especially when compared with LOB. However, normalised counts (per thousand) for first person plural pronouns shows that the count is in fact *higher* in PP than it is across LOB. This contrast is, however, masked, if only the combined counts for personal pronouns are taken. In this case the normalised count for PP comes out *lower* than in LOB.

65.8 This contrast is maintained if one looks at first person pronouns in PP and their nearest analogues in LOB+. Again, there is an initial impression that PP has fewer first person instances than everything apart from Academic Prose (Table 109):

Linguistic feature	PP	Press Rep	Press Rev	Religion	OffDoc	Ac Prose	Let Prof
first-person pronouns	6.7	7.50	7.50	16.60	10.00	5.70	40.90

Table 109 - first person pronouns (formal writing)

Although data for a split singular/plural count are not available for LOB+, I would predict that the reality of the matter will be that the PP will prove to have a similar count to Academic Prose, but that the gap between e.g. Official Documents or Religion would be reduced if singular and plural counts were separated.

65.9 What appears to be the case in PPs is that in a significant number of instances (e.g. 28% of the instances of "our" in the PP Corpus[37]) it is typically associated with moments when the writer is presenting either:

- a SOLUTION e.g. :

> "… We propose to contact other publishers if WSIP are not able to meet **our** delivery deadlines …"
> "… **Our** approach to the technical assistance will therefore be to build up and support local capacity …"

- or an EVALUATION in which the consultancy group's previous experience is called on to justify their analysis e.g. :

[37] in the other 60% of instances it is typically associated with the capacities or personnel associated with the firm

> "… In particular, our partnership can bring the following benefits to the programme …"
> "… Throughout, our aim will be to maintain quality and relevance of the work in an environment which is rapidly changing …"
> "… These clarify our understanding of the terms used and offer suggestions which we consider will maximise the effectiveness of the inputs …"

65.10 The right collocates for the 61 instances of "our" in these contexts given in Table 110 show how this particular pronoun is strongly associated with proposed actions and purposes or earlier experience.

experience	21
approach	20
general (approach)	4
understanding	4
view	3
aim	2
intention	2
initial (aim)	1
input	1
knowledge	1
objective	1
use	1

Table 110 - right collocates "our" in the PP Corpus

65.11 Similarly, the right collocates for "we" (in this case from the ENV proposal alone) indicate a similar set of rhetorical organisers. On the one hand *our experience* might suggest a solution. On the other, *we* might *propose* a solution.

propose	28	feel	1
have	8	highlight	1
will	7	intend	1
consider	5	list	1
understand	3	offer	1
acknowledge	1	provide	1
are	1	shall	1
believe	1	would	1
expect	1		

Table 111 - "We" in ENV

65.12 There is a good reason for the predominance of SPRE patterns in PP. The essence of proposal writing is having to respond to a set of givens. Proposal writers are neither inventing a project, nor presenting their independent analysis. Rather, they are responding to the analysis of the situation which is

presented by the TOR. Although they are invited to *comment* on this analysis, there is a limit to the extent to which it is profitable to challenge the analysis presented in the original project specification. If you are bidding to manage a project, you are obliged to accept the fundamental rationale for the proposed reform.

65.13 The way in which SPRE and associated patterns are used can be seen in many sections of the texts in PP. Two further examples are given below. The first shows a Problem / Response / Evaluation cycle where there is once again a clear shift from the neutral impersonal grammatical subjects / actors in the Problem element, to first person plural grammatical subjects / actors in Response and Evaluation.

CGC (CHEREPOV)
[PROBLEM] Task 5: design planning processes and procedures a2.11 This is an essential stage in development of the planning and delivery capabilities of the RDA. **[RESPONSE]** We will work with the RDA staff to design appropriate planning processes and procedures. The processes will cover project identification, involvement of key actors in the region, packaging to attract resources, and monitoring and review. **[EVALUATION]** Our input will be to draw on successful experience elsewhere and help local staff to develop processes appropriate to local circumstances and the types of projects to be undertaken.

Table 112 - CGC Cherepov

65.14 The same pattern is shown in a more extended example. Once again the Problem is stated through the language of the Terms of Reference and the Response and Evaluation are signalled by the use of first person plural pronouns.

BK162
AIMS AND OBJECTIVES

The wider objective of the Programme is:

[PROBLEM] "the strengthening of the strategic approach to social policy by way of maximising coordination and upgrading the information and project implementation functions within the Ministry of Labour and Social Welfare."

Within this, the immediate objectives of the programme are:

- the strengthening of the in-house capacity of the recently created Policy Coordination Unit (PCU) to provide policy advice within the MLSW;
- to provide assistance to the recently created Information Technology Centre in the development of integrated information systems within the MLSW;
- to assist in the development of the implementation capacity of the PMU;
- to provide assistance in the development of Occupational Safety and Health (OSH) Policy;
- to assist in the design and implementation of appropriate social policy responses to the consequences of economic restructuring.

1.2.3 METHODOLOGY

General Approach

[RESPONSE] Our general approach to this project will have the following characteristics:

- flexible and appropriate response to Bulgarian needs;

  The structuring of this contract is designed to allow the use of short-term expertise at short notice. The Team Leader will work closely with colleagues in the MLSW, and in particular with the Director of the PMU, to identify the short-term TA needs. This identification will carried out within the context of the medium-term strategy developed during the inception phase of this Programme.

- working in partnership with our Bulgarian colleagues;

The purpose of this contract is to assist the MLSW and the PMU to find Bulgarian solutions to Bulgarian problems. While it is important to share EU and other countries' experience and best practice with our Bulgarian colleagues, it is at the same time essential that this experience be translated efficiently so that it is relevant to the Bulgarian situation.
[EVALUATION] Our experience is that it is only by working in close partnership with our local counterparts that such a translation can take place.

Table 113 - BK162

66. Conclusion

66.1 This brief discussion of the organisation of project proposals closes this central section. In the first chapter (*What are we looking for?*) I outlined some of the problems which were met in applying a corpus based approach to the study of a specific genre. In the second (*Grammar and Style*) we saw how the results of a corpus based analysis using the framework developed in Biber 1988 could offer valuable insights into the linguistic construction of a particular genre, particularly along the interpersonal dimension, and how such an approach had potential value in writing pedagogy. In *Lexical dimensions*

an analysis of Keywords (Scott 1996) was used to provide an account of the wordings chosen by PP writers.

66.2 In this last chapter, *Organisation*, we have accounted for two aspects of the organisation of PPs which are likely to have pedagogic value in the development of an argument in PPs – explicit section headings and the use of a minimal discourse pattern (SPRE). Clearly, neither of these observations on the structure of PPs has been intended as a definitive statement about the structure of "difficult texts". The purpose of the chapter has been to demonstrate the pedagogic potential of seeing what kinds of structure exist in genres in which one is interested. If an assumption is being made, it is that generically related texts will have some kind of structure and that the structure of one member of a genre is likely to be related to the structure of another. In PPs the common elements are:

- the organising language of the TOR/ITB and its impact on the organisation of the proposal itself, and

- the minimal discourse pattern which underpins the arguments in the proposal and which is related to the conditions in which PPs are produced

The patterns and organisational features will be different in other genres, but they will exist and they are something that it will be useful for the writer to identify as part of the preparation for writing. In the final chapters of the thesis we will see to what extent the observations that have been made in this chapter and earlier chapters in Section Two accord with the understanding of writers in the organisations which provided the texts in the PP corpus, and what relevance they have for writing pedagogy.

SECTION 3 – TALKING TO WRITERS

Chapter 8: Writing Project Proposals

"My overall responsibility was to win the project."
(Project proposal writer in conversation 12 May 1998)

67. Introduction

67.1 In Chapter 3 (*Approaching the Data: dealing with genre*), I said that this thesis would have three areas of focus: pedagogic, corpus and case study. Issues in pedagogy were introduced in Chapter 1 (*Writing Difficult Texts*), and Chapter 2 (*Teaching Writing*) and will be returned to in Chapter 9 (*Helping learners write difficult texts*); corpus issues have been addressed in the four chapters of Section 3. In this chapter I will describe the approach taken in the development of the case studies, and will report findings that have arisen from them.

67.2 In conducting these interviews, I have become more aware of the value of mediating insights gained from corpus analysis with additional, more ethnomethodologically oriented data. The research into writing reviewed in Chapter 2 contains many studies of the composition processes of individual writers, and there is a growing literature which is able to demonstrate the value of a corpus linguistic approach to the study of specific written genres (Scott MR 1998, McKenna B 1997, Thompson P 1998, Tribble C 1998), or the study of writing processes from the perspective of the organisation in which the text is written (Odell L & Goswami D 1985, Spilka R 1993, Gunnarson B 1997). However, I am not aware of studies which combine a corpus analysis of a specific genre with a review of writing practices and preferences in the organisations which have provided the original corpus data. The results

reported in this present chapter indicate the potential of such a combined approach, and further applications of the approach should yield significant insights in future studies.

67.3 Although the case studies reported in this chapter are small-scale – three interviews with writers from two of the organisations which contributed to the PP corpus – they have proved useful. Specifically, they have achieved their main purpose of enhancing our understanding of the ways in which contextual knowledge (knowledge of audience, of other texts etc.) is drawn on when writers are making selections from the meaning potential of English during the composition of texts which are part of the Genre Specific Potential of a specific communicative context. Thus, the greatest immediate value of the interviews has been that they provide insights into:

- the writing process adopted in successful tendering organisations
- the linguistic and text organising preferences of experienced writers within these organisations
- the reasons why expert writers of PPs have such preferences.

67.4 The interviews proved to have an additional value as they have allowed me to test conclusions reached in the earlier corpus focused chapters. These were:

- There is a close relationship between the language of the documents provided in the Invitation to Bid and the final text of the technical proposal. This relationship goes beyond a simple re-statement of required section headings, as it includes a conscious re-working and incorporation of themes and arguments developed by the writers of the TOR

- Writers of PPs draw on a knowledge base of preferred keywords when writing proposals. The lexical set used is not necessarily dependent on the immediate purpose and focus of the proposal, but is, rather, related to the desire of the writer to present the professional standing of their organisation in a favourable light. In the context of PPs (and, I would predict, in other clearly delineated genres) this tendency to draw on a set of preferred lexis also produces local semantic prosodies as part of the genre specific potential.

67.5 In the next section I will describe the organisations which participated in the survey and the individuals who took part in interviews. I will then go on to outline the instruments I have used in the case study and results that have been obtained.

68. The organisations

68.1 Members of staff from two organisations (Organisation A, Organisation B) agreed to participate in this stage of the research project: one from Organisation A, and two from Organisation B.

68.2 Organisation A is a large UK based organisation with a world-wide network of offices and over 50 years' experience in aid and development work, and the capacity to bid for the management of both large and small human resource development projects. The interview with Interviewee 1 took place in the organisation's UK headquarters. Although the organisation employs a large number of staff, and can, therefore, draw on a very rich human resource, it also suffers from frequent losses of continuity in its project teams as a result of the continual movement of staff to and from the UK. While, once again, this has some advantages in terms of the network of overseas contacts the

organisation is able to call on, it can have a serious impact on its capacity to respond quickly and effectively to bid proposals. With senior staff sometimes recently arrived in post, and the ever-present risk that key staff will be moved or promoted to other parts of the organisation, it can be difficult to bring appropriate writing teams together at short notice.

68.3 Two staff members in Organisation B agreed to be interviewed. Organisation B has a very different history and structure. It was established 10 years ago by three founding partners, and is still managed by this trio. Specialising in economic and management consultancy it has developed a strong reputation in social reconstruction. It has a small headquarters building in the UK and two overseas offices, and currently employs 50 specialist consultants. Being a small organisation it has the capacity to respond rapidly to the changing needs of its clients, but it also lacks the large-scale network required to manage extensive projects independently, so often enters into strategic alliances with other suppliers in order to win and manage contracts. This is an advantage in so far as it allows the company a high degree of flexibility in its operations, but also has a down side as it often means that members of the core team can become seriously over-stretched!

69. The interviewees

69.1 Interviewee 1 has worked in Organisation A for over 15 years. She is a British national with a long experience of working in Eastern Europe. Although she has this long experience of project management, she has only recently begun to be directly involved in proposal writing. She is the author of one of the

texts in the PP Corpus, and was initially interviewed during the preparatory stage of this research project.

69.2 Interviewee 2 has worked in Organisation B for around eight years. Her first language is German, but she is fully effective as a writer of English and has been responsible for the preparation of a number of successful Technical Proposals. During her time with Organisation B she has worked as a consultant on economic and social development projects in Britain and central Europe, and is now based in the headquarters office with a management role in a number of UK projects.

69.3 Interviewee 3 has worked in Organisation B since it was established. At the moment, he is more actively involved in long term consultancies in Eastern and central Europe than Interviewee 2, spending part of his year in field activities, and the rest of his time in the UK headquarters. He is currently responsible for project pursuit within the region and has been involved in the development of a series of technical proposals during the last three years.

69.4 I was fortunate to be able to gain access to these three interviewees as they represent a useful cross-section of proposal writers. Interviewee A has only recently begun to take responsibility for proposal writing, and was much more tentative in many of her conclusions than the other two interviewees. Additionally, the consultancy group which employs her stands in sharp contrast to the company which employs the other two interviewees. Being both large and highly bureaucratic (at least at headquarters level) it imposes very different kinds of constraint on the proposal writer from those imposed by Organisation B. Not only do Interviewees B and C work in this smaller, more flexible organisation, but they have had contrasting experiences of

proposal writing. Interviewee B has worked for several different organisations and tended, in discussion, to view proposal writing from a "best practice" perspective. Interviewee C, on the other hand, has worked in the same organisation through most of his post-university career, and tended to be slightly more cavalier – or, at least, pragmatic – in his view of the process of proposal development. All this has meant that although the picture of proposal writing gained from the interviews is by no means exhaustive, it is more extensive than might initially be indicated from the numbers involved.

70. Interview structure

70.1 The interview was planned to last for one hour (the pre-agreed time) and was based on two activities: a structured interview and a text analysis task.

70.2 The interview (see Appendix 19: Interviews - Questionnaire) had two main sections. In the first section, the writing context was discussed through two sets of questions. The first set included three questions, all relating to the ways in which the writing of the Technical Proposal is managed in the organisation:

- What happens in your organisation before project proposal writing begins? (focus on project identification / team formation …)

- What happens in your organisation at the beginning of proposal writing? (focus on team management / task allocation / analysis of TOR / identification of consultants …)

- What happens in your organisation during the process of proposal writing? (focus on developing drafts / peer review …)

70.3 Questions in the second set enabled a more detailed discussion of the choices that were made in the development of text, focusing on:

- the use of the Terms of Reference (TOR), and the Invitation to Bid (ITB) in the development of the proposal
- the overall organisation of the technical proposal (TP)
- the reasons for the naming of key sections in the TP.

70.4 Section 1 of the interview was designed to elicit information about the writing culture which exists within the tendering organisation and the roles of proposal managers. One of the main findings of the corpus study components of this research project has been the strong influence of the language of the TORs on the organising language of the Technical Proposals (section headings and sub-headings). I wanted to use the opportunity for discussion with experienced writers to investigate the extent to which writers are aware of the influence of the TORs and to clarify their reasons for drawing on them in the development of the PPs.

70.5 Section 2 was based on a text analysis task in which the interviewee was asked to read a short extract from a TP (not written by a member of the consultancy group) and to mark those words or passages in it which were:

- *different* from the way in which they would have worded such a passage or
- *similar* to their way of writing.

70.6 This task was designed to provide interviewees with an opportunity to provide specific exemplifications for the opinions they had offered in Section 1.

Asking interviewees to undertake a detailed analysis of a short span of text proved to be a useful technique as it elicited information which could not have been obtained from questioning alone. Although findings from this section reveal a significant lack of homogeneity and approach within and between consultancy groups, they also showed an overall similarity of approach – despite differences in interviewee experience and writing context. These results will be discussed in the following sections.

71. Interview results: Question Set 1 – Starting

71.1 ***What happens in your organisation before project proposal writing begins?***

In the context of the PHARE programme, both organisations had similar approaches to project pursuit, using a combination of web-based research and active 'getting alongside' approaches to the government departments and individuals who might become involved in future EU projects. The reason for this preparatory work is to ensure that the organisation is (a) aware of what will be coming up in its field, and is (b) able to maximise its chances of being shortlisted as a potential supplier once it has submitted a formal expression of interest.

71.2 Deciding whether or not to express an interest is in itself a complex process. The criteria which are common to both organisations are:

- Will the contract provide a satisfactory return on invested time? PHARE, for instance, became significantly less interesting for British organisations during 1997/98 as a result of the very strong pound sterling.

- Is the organisation's bid likely to be favourably viewed by the national government? In some circumstances there can be a significant preference for

suppliers from one country rather than another, and it may not be worth your while starting a tender.

- Does the organisation have appropriate potential partners? PHARE projects typically involve two or three EU countries, and having the right partners can be critical to the success of a bid.

- Does the organisation have the right staff available? There is a general assumption amongst writers with an interest in this kind of bidding process that the CVs of the experts are of paramount importance – i.e. if you do not have these right, you might as well not bother![38]

71.3 ***What happens in your organisation at the beginning of proposal writing?***

Once the organisation has received the Invitation to Bid (ITB) or Invitation to Tender, one of the first issues it has to face is whether or not the proposal is to be written by one person or a team. The theoretical position seems to be that:

> …once you get the Invitation to Tender, something called a proposal team is put together which includes the proposal manager – who's the person who will write the bulk of it. But it will also have a proposal director who's the person who will make some sort of quality control – reads it hopefully before it goes out! And then other members who might be in a position to write things for it and who will also later be in the team to do the work. (Interviewee 2)

This however, is not always possible or appropriate, and the second interviewee from Organisation B reported that:

[38] This assumption does not fit comfortably with the considerable effort that organisations put into proposal writing. The fact of the matter appears to be that while "names" are important, their importance is in itself a textual construction. While the expertise of individuals is not in doubt, names only become "names" as a result of the accumulated weight of other documents (e.g. consultancy reports, project evaluations), and the way in which earlier experience is represented through the arguments, histories and connections presented in the technical proposal itself.

> In theory, a proposal is written by several people. In practice […] it's very often the same person - because it's quicker - one person thinking through the approach and assembling the bits is better than several (Interviewee 3)

71.4 Such a view supports conclusions reached by Kleimann in her 1993 study of the relationship between workplace culture and review processes (Kleimann S 1993). Organisation B is a small operation with good internal communication, but also puts its staff under quite considerable pressure in terms of the scale and range of tasks they have to undertake. In such circumstances the specific workplace culture encourages the occasional sacrifice of best practice in favour of getting the job done as quickly as possible.

71.5 If an organisation does decide to opt for a team approach, all interviewees agreed that there is a strong need for single point management, and that, at the end of the day, one person makes a better job of pulling all of the text components together than many.

> … my responsibility was to produce a document that was coherent, that had no erratic leaps, that had an internal intellectual logic to it so that one section flowed into another without any gaps. And that a document which had different authors contributing could be seen as a single whole. (Interviewee 1)

71.6 Interviewees also acknowledged that prior experience of writing proposals for the agency which issued the ITB is a key qualification for the proposal manager – while recognising that this too can produce its problems.

> We see that [having written this kind of proposal before] is very important, because you have to be experienced in reading the TOR and understanding the TOR and then putting across the message of how you want to do it - so there is an element of having done it before which helps you do it again. But there's a danger probably that you become too stale - we have personally in the company one or two people whose proposals I find are just deadly boring because I just feel that they are so stereotypical and mechanical (and some of those that we haven't won recently!) Now I'm not sure why, but sometimes, there comes a point where we just churn things out and that's not good either, so I think you need to bring a certain freshness to it too […] (Interviewee 2)

71.7 One way to avoid this sort of problem is to ensure that the proposal draws on the ideas of more than one person. Interviewee 2 described the use of 'brainstorming' sessions as a way of ensuring freshness:

> … these tend to be very useful given that we've got such a range of people working here and it's proved to be very helpful if one can have an hour, an hour and a half, together to spell out some key points. There's always a kind of "think-piece" bit in those proposals, where you demonstrate that you've understood the TOR, that you're competent in the field and that you can add value to the TOR at this stage (Interviewee 2)

71.8 ***What happens in your organisation during the process of proposal writing?***

As a further strategy to avoid the staleness that can come from over-familiarity, the bidding organisations interviewed here also invest time and resources in ensuring that the proposal draws on expert opinion, and that this is closely matched to the needs and expectations of the potential recipient country.

> … I suppose you're parallel tracking - building up the team at the same time as you're building up the approach. You're also gathering information on how to bid - which may involve a visit to the country and will probably involve speaking on the phone to people, maybe gathering general country information […] Typically if you're dealing with outside people, you're dealing with them because they have some specialised knowledge either of the country or the subject, so that is feeding into what you're writing as well. (Interviewee 3)

71.9 In such a context, the writer's task is relatively clear. This understanding is neatly summed up by Interviewee 1:

> My overall responsibility was to win the project. Winning the project in our proposal meant demonstrating that we knew what we were doing, that we'd taken account of all aspects of the Terms of Reference, and that we were able to illustrate our competence, the experience that was relevant to the implementation of this particular project, and our ability to provide a flexible response to the terms and conditions of this particular project …

The corpus study confirmed the general impact across all of the texts in the PP corpus of this need simultaneously to demonstrate professional competence and institutional responsiveness to the needs of the client. Professional competence was signalled through devices such as the suasive language of the proposals (e.g. the high incidence of attributive adjectives and phrasal coordination) and the exploitation of local semantic prosodies. Professional flexibility was often signalled through the echoing of the language of the TORs which accompanied the invitation to bid. (See sections 72 and 73 below for a fuller discussion).

71.10 However, while being aware of this over-riding responsibility, all interviewees stressed the time pressure that they are under when writing:

> "One of the criticisms directed at us regularly is that we don't have time to check the final product with all the partners, so they don't get a final say in the shape of the thing that's submitted. But life's a bit like that, you know, you don't get enough time" (Interviewee 1)

They also commented, that although their organisations have produced guidelines for proposal development, they are frequently ignored in the pressure of getting the job done:

> "We do have what is called a proposal writing handbook which a colleague of mine put together which maps out the procedure how one should go about it, but to be honest we don't always follow that – which is unfortunate because it is very sensible, the way he has structured it." (Interviewee 2)

72. Interview results: Question Set 2 - Writing the technical proposal

72.1 The second set of questions in the interview focused on the interaction between the proposal writer and the documentation provided in the ITB. All those involved in the interviews stressed how important this is to effective proposal writing:

> "the Terms of Reference are central, absolutely central to writing a proposal " (Interviewee 1)

– a view which confirmed the findings reported in the Section 3 chapters *Lexical Dimensions* and *Organisation* where the relation between the lexis and heading labels of PPs and the language of PHARE ITBs and TORs was shown. Although the documentation which informs the proposals is not the only source of the lexicon of PPs, it is certainly a major resource (whether it is consciously or unconsciously drawn on). Of specific interest in the study reported below is the fact that the PP writers interviewed were strongly aware of the importance of the TORs, and went to considerable trouble to ensure that their proposals demonstrated that they valued the contribution which the writers of the TORs had made.

72.2 How (if at all) do you make use of the documents provided in the TOR / ITB for a potential contract? A common perception of the TOR was given by Interviewee 1. She commented:

> "… you do like to echo the words that are in the TOR or even phrases, because it's a bit like active listening on paper, isn't it? That you're reflecting back to someone something that they thought they were very clever to devise, and people always like and can relate to something that has an element of themselves in it."

In such a context, the proposal is thus not only a demonstration of the organisation's competence, but also of its responsiveness and sensitivity to the local and technical issues raised by the authors of the TOR. Mirroring the language used by the TORs' authors is not simply flattery, it is a conscious strategy designed to show that the potential service supplier is willing to ensure that their views harmonise with those of the agency which will select the service supplier.

72.3 Interviewee 2 offered a similar view when she said: "I try to reflect the TOR because I feel they've obviously done their work and they want that to be seen …" Although she goes on to add: "… the TOR normally reflect the work programme, not the issues. We normally have an issues section and a work programme, and the issues section is where you're allowed to put a bit of intellectual input into it", drawing attention to the more intellectual style she reports as being typical of Organisation B, and the need to strike a balance between originality and deference to the controller of the budget. Striking this balance may lead to tensions during the development of proposal, but it is at such points that the professional judgement of the writer becomes critical.

72.4 It is possible that the TOR in PHARE projects are more important in the development of a PP than is the case with bids to other organisations. Interviewee 3 commented:

> "…in a UK job in economic development – or in some EC jobs – you'd typically be able to speak to the person who wrote them [the TOR] and try to understand what was behind them. That's not usually the case in PHARE – partly because they've usually recruited some other consultant to write them …"

Not having recourse to discussion with the original authors of the TOR, and also knowing that those evaluating the proposal will often be as dependent on the ITB documents as the bidding organisation, the TOR take on a central importance for those writing the proposal. This places such proposals into a very specific category, as their intertextual dependence appears to be much greater than that of proposals submitted to other organisations.

72.5 The two remaining question areas in the second section of the interview were:

- How do you (and your colleagues) decide on the main section divisions of the Technical Proposal?

- How do you (and your colleagues) decide on the different section headings of the TP?

72.6 Discussion under these heads brought out the extremely close relationship between the overall structure of the PP and the language and organisation of the TOR. Interviewee 1 illustrated this when she commented:

> "The [section heading] labels for PHARE and TACIS projects are almost given - so you have to make sure that you follow the regulations there, that you're giving what they ask you for and no more, or not too much more…"

There is, however, a contrast between this view and that of Interviewee 2 from Organisation B. In her opinion, while the work-programme component of the proposal (the section which responds directly to the tasks implied by the TOR)

may draw very closely on the language of the ITB, the proposal should also contain an "Issues" section which is written in a much more open way:

> "Well, what we usually do is say in all our proposals: 'In a regional development programme of this kind in the late 1990s, in this kind of environment, the following issues are relevant …'. And that's where the brainstorming session is good because that's what provides the intellectual input into the project. And then I would normally have bullet points and say, 'Partnership models' for instance, or 'Sustainability for development' or 'International networking' – or you name it! Say the six key issues which we'd identified in the brainstorm…" (Interviewee 2)

72.7 This comment throws light on the ITB/PP relationship mentioned in paragraph 67.4 – a question which was also raised in an earlier chapter:

> "This leads one to ask *why* the writers have opted for high or low explicitness in the signalling of text organisation. […] What do writers hope to achieve through their use of numbering schemes and section headings, and to what extent are they aware of the impact of their use of these schemes on the overall reader-friendliness of their texts?" (*Organisation*)

Apart from the normal sign-posting that is offered by such devices, it appears that writers have two (possibly contradictory) purposes in their use of section headings:

- to demonstrate their *conformity* to the requirements of the TOR
- to demonstrate the *originality* of their approach.

Thus, the main headings used in the majority of the proposals tend to reflect the requirements of the TOR (see the list of Level 1 status headings in Table 114 below):

Organisation and method
List of staff
Statement and terms of reference
Introduction
Introduction and summary
Key issues
Statements and terms of reference
Technical response
Terms of reference
Work programme
Workplan

Table 114 - Primary headings

Secondary and tertiary headings, by contrast, more frequently reflect the writers' attempts to add value in the proposal and usually move away from the language of the ITB. The diversity of lexis deployed is much greater and the themes addressed are more specific (see Appendix 28: *Organisation - Section Counts* and example in Table 115 below)

Adding value
Approach to delivery stage
Approach to inception stage
Building trust
Business advice (including some local consultants training)
Company expertise
Consortium organisation
Contracting and financial services
Current environmental education in Poland
Curriculum development, primary education; curriculum development, vocational education
Delays to the start of the project
Development phase
Dissemination phase
Eastern European experience
Improving the image of entrepreneurial activity

Table 115 - Tertiary headings

73. Text Analysis Task results

73.1 During the text analysis task, interviewees were asked to mark up a text extract from a PP (not written by their organisation) and to maintain a protocol commentary. The protocol was tape recorded but has not been transcribed.

The original text extracts were marked with coloured highlighters – red (*not* like I write) and green (*like* I write). A word-processed version of these results can be found in Appendices 20, 21 and 22. Red highlighting has been replaced by ~~**bold / strikethrough**~~ text. Green highlighting has been replaced by **shaded text.** The passage used in the activity is quoted in full below. It comes from an early section in the Technical Proposal and was selected because it makes reasonable sense in isolation from the rest of the text.

Our general approach to this project will have the following characteristics: - flexible and appropriate response to xxxxx needs; The structuring of this contract is designed to allow the use of short-term expertise at short notice. The Team Leader will work closely with colleagues in the MLSW, and in particular with the Director of the PMU, to identify the short-term TA needs. This identification will be carried out within the context of the medium-term strategy developed during the inception phase of this Programme. - working in partnership with our xxxxx colleagues; The purpose of this contract is to assist the MLSW and the PMU to find xxxxx solutions to xxxxx problems. While it is important to share EU and other countries' experience and best practice with our xxxxx colleagues, it is at the same time essential that this experience be translated efficiently so that it is relevant to the xxxxx situation. Our experience is that it is only by working in close partnership with our local counterparts that such a translation can take place. - location of project operations within xxxxx; It is our practice to develop adequate administrative structures for the project operations in-country. Our intention is that our long-term experts will live in apartments rather than hotels and we shall employ sufficient xxxxx staff from our own budget to ensure the efficient administration of this important project. In addition to their important technical and administrative contribution, our xxxxx employees will provide our foreign advisers with an essential xxxxx perspective and sensitivity to local needs. (258 words)

Table 116 - Text analysis passage

73.2 There was general approval from all interviewees for the use of headings and bullet points – the consensus being that bullet points make information more easily accessible and allow the writer to highlight the strengths of their argument. The only interviewee who objected to the wording of the bullet points themselves was interviewee 1. Her feeling was that a rhetorical opportunity had been missed as a result of the author's failure to use grammatical parallel structures across the three bulleted lines (i.e. *responding*

flexibly and appropriately … / *working* in partnership … / *locating* project operations …).

73.3 Interviewee 1 proved to have very different responses from the other two. First, she was the only one to approve of the use of *translated efficiently* and *such a translation* in the paragraph under *working in partnership* …this metaphorical use being strongly disapproved of as potentially confusing by the two other interviewees.

> I would avoid the word "translation" – simply because "translation" means "language translation", and translating from one situation to another is a confusing use of language for people reading it … (Interviewee 3)

A probable explanation for this stance is that the interviewees from Organisation B appeared to be more aware of the issue of writing proposals to be read by speakers of languages other than English than the interviewee from Organisation A:

> "I think most of us are aware now that one should try and keep the English a bit more straight-forward in the PHARE proposal in that most of the people reading it either in the recipient country or in Brussels will not, however good their English is, it will not be their first language. Some of us write in a complicated way left to ourselves, and it's best to perhaps review what we've written to make sure it comes across relatively straight-forwardly." (Interviewee 3)

73.4 Secondly, Interviewee 1 commented positively on the repetition of *our* throughout the passage (11 instances) – the others finding it inappropriately repetitious, or more dangerously, overly tub-thumping.

> "There's a lot of 'our' in here … our, our …I would not use that so much, I would try to give an impression that we know what we are talking about and that we've done this kind of work before, but this is overplaying it a bit …I feel it's patronising and I wouldn't like to use it." (Interviewee 2)

73.5 Away from questions of style, Interviewee 1 was also the only one who did not comment on the inclusion of low-level administrative information in the passage – e.g.

> "Our intention is that our long-term experts will live in apartments rather than hotels and we shall employ sufficient xxxxx staff from our own budget to ensure the efficient administration of this important project."

This was seen as being entirely inappropriate by the other interviewees, their position being that the reader had no interest in such matters at this point in the document (and that these kinds of detail were usually covered by PHARE administrative procedures and required no special attention from the service provider).

73.6 Where all interviewees were in agreement was in their approval of a set of words and phrases – referred to as "motherhood words"[39] by one interviewee. "Motherhood" – a particularly apt bit of jargon – can be defined as:

> "Apple-pie, great, things that are incontrovertibly good things, and therefore you put in anyway …Because everyone expects them they're only interesting if you can say something that says how you'd do them or give some example of how they would work …" (Interviewee 3)

73.7 The list in the first column of Table 117 - "motherhood vs keywords" shows the approved (motherhood) words and phrases which the interviewees marked up in the 258 word text extract used in the final interview task. The point of interest here is that these motherhood terms map remarkably closely onto the key-words discussed in *Lexical dimensions* (see paragraphs 5.5 ff.). Thus 18 of the 31 words or phrases identified by the interviewees also proved to be

[39] Or – as plain "motherhood": as in "*this one's a real motherhood.*" (reported by Lou Burnard during the TALC 1998 conference)

keywords or keyphrases (clusters set at 2) in the PP corpus (columns 2 and 3 in Table 117).

1	2	3
"motherhood" terms	**PP/BNC Keywords**	**PP/BNC Keywords (2)**
administration	ADMINISTRATION	
advisers	ADVISERS	
appropriate (Attr. Adj. 1)	APPROPRIATE	
experience	EXPERIENCE	
expertise	EXPERTISE	
foreign (Attr. Adj. 2)	FOREIGN	
inception	INCEPTION	
local (Attr. Adj. 3)	LOCAL	
long-term (Attr. Adj. 4)	LONG-TERM	
partnership	PARTNERSHIP	
phase	PHASE	
PMU	PMU	
PMUs	PMUS	
Programme	PROGRAMME	
project	PROJECT	
short-term (Attr. Adj. 5)	SHORT-TERM	
technical (Attr. Adj. 6)	TECHNICAL	
best practice (Attr. Adj. 7)		BEST PRACTICE
contribution	x	
efficient administration	x	
efficiently	x	
flexible (Attr. Adj. 8)	x	
foreign advisers (Attr. Adj. 9)	x	
important (Attr. Adj. 10)	x	
important project (Attr. Adj. 11)	x	
inception phase	x	
local counterparts (Attr. Adj. 12)	x	
perspective	x	
sensitivity	x	
TA	x	
technical contribution (Attr. Adj. 13)	x	

Table 117 - "motherhood vs keywords"

73.8 Not only is there this closeness of fit in terms of the words that the interviewees have approved in this very short passage, they also included a large number of examples (13/31) of one of the features strongly identified with PPs as a genre (attributive adjective). The fact that unprompted expert informants felt the need to mark this word class for approval provides further support for their significance in the construction of allowable contributions to the genre.

74. Conclusion

74.1 These brief case studies have provided an opportunity to test conclusions reached in earlier chapters in this study, and have also demonstrated the value of discussion with expert practitioners when it comes to understanding the nature of expert performance (Bazerman C 1994, Tribble C 1997a). The writers in this study have confirmed the importance of context knowledge (of co-texts and social context) in text formation, and have given clear examples of the kinds of professional consensus on allowable language use that develop in the process of genre consolidation.

74.2 The studies have also shown the potential value for pedagogic practice of such simple interviews. By focusing on what expert writers in a genre *do* when they are writing, and asking interviewees to reflect on the language that they prefer to use, it is possible to develop accounts of how different communities of writers go about their work. This information can provide the basis for practical and effective learning programmes in which apprentice writers are not required to imitate the results of best practice – the texts, but are, rather, asked to reflect on why certain writing practices have developed and what their value might be to them, in their own development as writers.

SECTION FOUR – IMPLICATIONS FOR PEDAGOGY

Chapter 9: Helping learners write difficult texts

> "… what we call our knowledge is of its nature provisional, and permanently so." (Magee B 1973:27)

75. Introduction

75.1 The research project reported in this thesis has had the overall aim of identifying ways of helping apprentice writers to learn how to write unfamiliar factual genres. In the first chapter, I described what I hoped to achieve during the project in terms of a series of outputs that would contribute to this overall aim. The outputs were:

Output 1	a statement on the nature of *writing*, *difficult* and *texts*
Output 2	a summary of key issues in current approaches to the teaching of writing
Output 3	an account of possible applications of empirical linguistics in the development of writing instruction materials
Output 4	an account of the ways in which expert writers view a) the texts that have been the subject of the corpus analysis, and b) the writing processes required for the production of these texts
Output 5	a report of findings and recommendations for possible pedagogic applications.

75.2 Each of the first four outputs has been elaborated and reported in earlier chapters. This final chapter completes Output 5 – major findings related to specific areas of research have already been presented in each of the earlier sections. In Section One: *Setting things up*, I established the scope of the thesis and set out a working definition of the "difficult texts" in the title. I also reviewed current issues in teaching writing, and presented some ideas on how the tools and resources used by writing teachers could be expanded. During this discussion, I stated my understanding of how the notion of genre could be of practical relevance to those who want to write difficult texts, and some of

the problems which teachers and students have to face when the notion of genre is misapplied. In Section Two: *An account of the data*, I reported the results of a detailed analysis of examples of a clearly defined genre: i.e. project proposals submitted to the EU PHARE programme and other development agencies. This section made a major contribution to the development of Output 3:

> " an account of possible applications of empirical linguistics in the development of writing instruction materials"

as it demonstrated the potential of the frameworks for language description outlined in Biber D 1988 and Scott M 1996 and 1997b for practical language teaching. Section Three: *Talking to writers* is also important as it shows the extent to which the findings of the corpus based study were confirmed by staff in organisations which write project proposals, and demonstrates the value of referring to expert informants as a means of checking conclusions based on corpus data.

75.3 In this final Section, I will make some suggestions for how insights gained in earlier chapters can be used by teachers and students who are concerned with the writing of difficult texts. I have always intended that the research reported here should be considered as an applied linguistics project, agreeing with Widdowson when he comments:

> "I think it is the responsibility of applied linguistics to demonstrate whether or not linguistics can provide insights of use to the language teacher and to investigate other sources of insight in the search for relevant models." (Widdowson HG 1984:8)

As someone who has taught writing for many years I feel that the job of writing teachers is particularly challenging – especially in the way we find ourselves having to be:

> " a motivator, an interpreter of the task, a designer of meaningful tasks, an organizer, a resource, a support person, an evaluator, and a reader for information." (Grabe W & R Kaplan 1996:254)

I hope, therefore, that this thesis will make a contribution to the work of those who are involved in this aspect of language education. The extent to which I have found "insights of use" to those who take on the complex roles of the writing teacher will be a measure of the extent to which I have fulfilled the major purpose of this thesis – helping learners write difficult texts.

76. Frameworks for teaching writing

76.1 At the end of Chapter Two I asked five questions in response to a series of issues that had been raised about aspects of the teaching of writing. These questions are given below (in a sequence different from the original):

> **Question:** What constitutes a theoretically adequate pedagogic framework for teachers who want to help learners to write difficult texts (as we have defined them here)?
>
> **Question:** What practical guidance can be given to help teachers develop appropriate methodologies for the writing courses they offer their students?
>
> **Question:** What practical means are there to help teachers decide whether or not the texts they are dealing with are exemplars of an identifiable genre?
>
> **Question:** What kinds of examples do teachers and learners need when approaching the problem of writing into a new genre?
>
> **Question:** Is there any help for teachers who cannot get hold of examples of the kinds of texts their students need to write?

76.2 I will use these questions as a way of organising this final chapter. In so doing, I want to offer a framework for teaching writing which goes some way to meeting the criteria for integrating a genre approach into writing pedagogy which I also outlined in Chapter 2 – i.e. an approach which:

- provides clear guidelines for the selection of appropriate pedagogic examples
- gives a practical basis for analysing these examples
- enables students to develop useful hypotheses about language use in genres that are important to them, so that they will be better able to write texts which will achieve their various purposes

The discussion in this chapter will also be informed by my understanding of the four essential sets of knowledge that writers need when approaching a writing task – i.e. *content knowledge, writing process knowledge, context knowledge,* and *language system knowledge*[40].

76.3 The practical suggestions for helping learners to write difficult texts that I make here will, therefore, be framed by the five questions I set out above. The suggestions themselves can be tested by asking what potential they have for usefully extending learners' knowledge in relation to the difficult tasks they wish to write.

77. Question 1

> "What constitutes a theoretically adequate pedagogic framework for teachers who want to help learners to write difficult texts (as we have defined them here)?"

[40] Quoted in Chapter 1: *Writing difficult texts*

77.1 An underlying assumption of this thesis has been that the notion of genre has sufficient explanatory force for it to be used as a starting point for writing pedagogy. Genre offers a basis for the selection and analysis of examples of the kinds of texts learners want to write, and for developing an understanding of the communicative context in which those texts are written. Hasan has commented:

> "Genre bears a logical relation to CC [Communicative Context], being its verbal expression. If CC is a class of situation type, then genre is language doing the job appropriate to that class of social happenings." (Halliday MAK & R Hasan 1985:108)

If this view is accepted, a major part of the job of the writing teacher can be seen as providing learners with opportunities to develop their understanding of the communicative context for which the genre is the verbal expression, and then to develop an analysis of how language is used as a means of ensuring that the requirements of that communicative context are met. In so doing, the teacher is helping students to gain knowledge of the conditions in which the texts they want to write are produced, and the constraints that writers of such texts have to contend with. In addition, they are also ensuring that learners develop the linguistic resources they need in order to write texts which will be allowable contributions in the chosen genre.

77.2 The first stage in this process, the development of a fuller understanding of communicative context and co-texts (*context knowledge*) can be achieved with the help of a model such as that proposed by Berkenkotter and Huckin (1995). They have suggested a theoretical framework for the consideration of genre from a sociocognitive perspective, and the five principles they propose, can usefully inform a syllabus for writing. These are:

> "*Dynamism*. Genres are dynamic rhetorical forms that are developed from actors' responses to recurrent situations and that serve to stabilize experience and give it coherence and meaning. Genres change over time in response to their users' sociocognitive needs.
>
> *Situatedness*. Our knowledge of genres is derived from and embedded in our participation in the communicative activities of daily and professional life. As such, genre knowledge is a form of "situated cognition" that continues to develop as we participate in the activities of the ambient culture.
>
> *Form and content*. Genre knowledge embraces both form and content, including a sense of what content is appropriate to a particular purpose in a particular situation at a particular point in time.
>
> *Duality of structure*. As we draw on genre rules to engage in professional activities we *constitute* social structures (in professional, institutional and organizational structures) and simultaneously *reproduce* these structures.
>
> *Community ownership*. Genre conventions signal a discourse community's norms, epistemology, ideology, and social ontology."
> (Berkenkotter C & TN Huckin 1995:4)

77.3 Taking as a practical example the needs of learners who want to write project proposals, a syllabus component focussing on *context knowledge* could involve two major steps. The first would give learners opportunities to come to an understanding of the constraints and expectations that are commonly encountered in the contexts in which they will be writing (the *Situatedness* of the genre). For PPs this would involve developing an understanding of e.g. EU institutions and tendering practices, and the sources and kinds of information that are available to organisations submitting proposals. It would also include a consideration of peer review procedures, house style guides, and the roles of the proposal manager and of specialist contributors (either from within the organisation, or sub-contractors brought in to help develop a specific bid). In this study of Project Proposals, we have seen the importance

of the ITB and TOR and the efforts which Consultancy Groups go to to follow the evolution of policy in the organisations to which they make bids. In other settings, this component may involve a process of institutional induction (e.g. the kinds of pre-sessional course commonly provided for overseas students coming to study in the UK), or it may require course participants already working as practitioners to reflect and report on their own professional experience.

77.4 Step two in this syllabus component would require a review of the expectations that potential readers have of the texts in question (Smart G 1993, Freedman A, Adam C & G Smart 1994). Such a review would oblige learners to consider the *Duality of structure* and the *Community ownership* of the genre, but it would not imply that learners should accept what they observe as immutable and hegemonic (the kinds of imposition which Clark R & R Ivanić 1997 argue against). A major value of this kind of analysis is that it allows students to develop an understanding of the constraints which writers face within a specific communicative context and enables them to make their own decisions about how they manage their relationship with other owners of the genre.

77.5 During this review process learners would also analyse examples of PPs in order to assess the extent to which their assumptions about these texts match the reality of what professional writers actually write. This would satisfy the principle of *Dynamism*, in the same way that e.g. a course in writing for publication in the social sciences would need to take into account recent changes in the nature of allowable contributions to the genres owned by specific discourse communities (Bloor M & T Bloor 1991, Edwards H 1997).

77.6 If our hypothetical course in proposal writing is seen as a set of modules, alongside the teaching units which enable learners to come to an understanding of the sociocognitive dimensions of PPs, there would also be a need for modules in which learners would have opportunities to identify the *Forms and content* that are appropriate to the genre. In other words, in order to ensure that students who want to write difficult texts have the *language knowledge* that they require, the syllabus must offer appropriate text resources and text analysis tools, along with the means for making meaningful comparative statements about the texts in question. This is the second main element of the syllabus specification, and leads us usefully to the second of our questions.

77.7 If genre is used in these ways, it offers a practical basis for the elaboration of writing programmes. Huckin and Berkenkotter's sociocognitive model gives teachers and materials writers a way of addressing the contextual dimension of a genre, and, as we shall see in the next section, corpus linguistics offers practical tools for dealing with the elaboration of language system knowledge.

78. Question 2

> "What practical guidance can be given to help teachers develop appropriate methodologies for the writing courses they offer their students?"

78.1 In Chapter 2, I gave examples of different teaching materials currently used in writing pedagogy, some of which were strongly form-focussed, while others paid more attention to cognitive and social dimensions of writing. During this discussion, I concluded that it was inappropriate to attempt to impose a single methodology for teaching writing – a view that harmonises with current

opinion in ELT where there is general agreement that teachers need to match teaching methods to local conditions (Bamforth R 1993, Brookes A & Grundy P 1990). Using genre as a starting point for writing instruction does, however, have implications for what happens in classrooms (Christie F 1993, Gee S 1997), and practical suggestions can be made for the management and focus of writing instruction if genre is adopted as a theoretical framework for the writing syllabus.

78.2 One such account has been proposed by Callaghan, Knapp and Noble in their report of work done at secondary level in the "Language and Social Power" project in Sydney, Australia (Callaghan M, Knapp P & G Noble 1993:179-202). The diagram given in Figure 20 summarises key features of the approach taken:

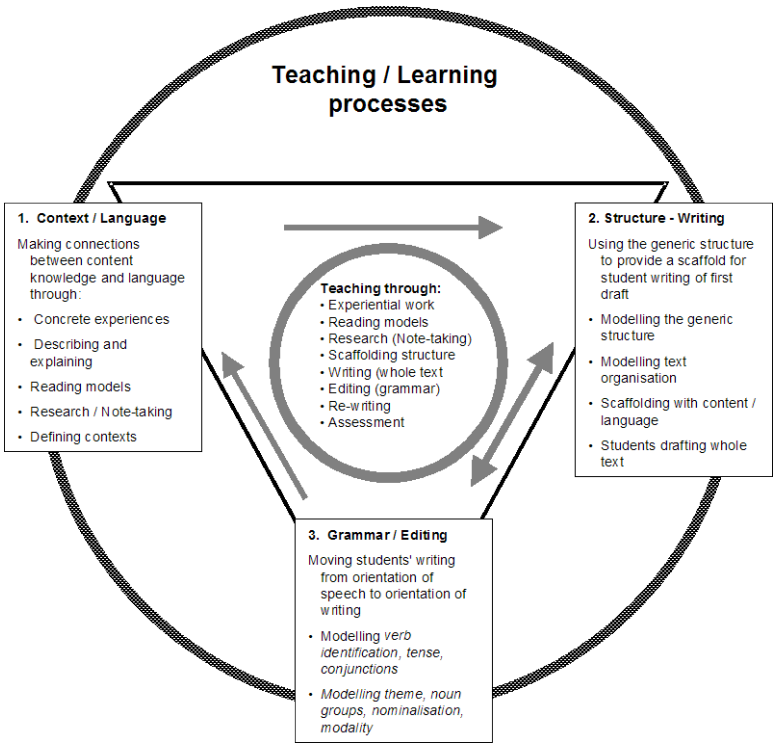


Figure 20 - Callaghan M, Knapp P & G Noble 1993:196

78.3 The three major elements in the teaching / learning processes summarised in this diagram offer a workable basis for the development of the kind of factual

writing programme we have been discussing. Often beginning from an analysis of examples of a genre and the social context in which they have been written (modelling), students go on to develop an understanding of the content and context relevant to the genre. They are then able to use examples of the genre as a starting point for their own writing (scaffolding) before developing their own texts, and, at a later point, to use these examples as a resource for the improvement of their own independently developed texts. Because the model assumes the possibility and value of recursion, the three sets of learning processes presented in the model should not be seen as being in a fixed sequence. Activities in the learning processes can be used at any stage in a learning cycle. In the model, teachers have expertise as analysts and informants and it is part of their job to make this expertise available to the learners. The learners are also recognised as having the capacity and right to criticise and subvert the models that are presented during teaching / learning cycles if they wish to.

78.4 If teachers decide to use genre as a starting point for helping learners to write difficult texts, they are likely to use the kinds of teaching / learning cycles indicated by Callaghan, Knapp and Noble. In such a teaching / learning setting, the corpus linguistic tools and techniques demonstrated in the four chapters of Section Two: *An account of the texts* have considerable potential to offer ways of organising the learning experience. Stylistic analysis drawing on Biber 1988 can give learners essential insights into the structure of the interpersonal dimension of the genre. Keyword analysis can enable learners to build an understanding of the lexis patterning of allowable contributions.

78.5 Although such applications of corpus linguistic approaches to text and genre analysis can be extremely useful to the development of language system knowledge, it is import that they be seen as complementing rather than replacing other, more familiar, classroom activities. As we have seen in Chapter 2: *Teaching Writing*, the approaches that are proposed here are not a panacea. They are, rather, practical suggestions for ways in which teachers can extend their normal classroom practices by drawing on new technologies to deal with very old problems. I will discuss some of the ways in which corpus linguistic tools can be incorporated in writing pedagogy in the next section: *Working with corpus data.*

79. Working with corpus data

79.1 One of the main implications of earlier chapters has been that it is possible to use electronically held corpora to help teachers and learners develop an appreciation of aspects of language use in particular settings. In this present writing, I do not propose to offer detailed examples of how concordance data can be integrated into classroom learning or teaching materials. I have done this elsewhere (Tribble C 1991, 1997b, 1997c; Tribble C & G Jones 1997). Moreover, a range of corpus based ELT task types and classroom activities now exists and has begun to be discussed more widely (Botley S, Glass G, McEnery T & A Wilson (eds.) 1996; Flowerdew J 1993; Flowerdew J 1996; Granger S & C Tribble 1998; Johns T & P King (eds.) 1991; Melia J & Lewandoska B [eds.] 1997; Scott M 1997b), and Tim Johns at Birmingham University continues to demonstrate how valuable the combination of a creative and imaginative teacher and an appropriate corpus of texts can be to

the language learner (Johns T 1991,1994,1997, http://sun1.bham.ac.uk/johnstf/homepage.htm).

79.2 What I propose to offer here, based on this earlier experience, and on the work done in this thesis, are some principles for using corpora in writing instruction. I shall consider some of the implications of adopting such principles by looking at a practical example: how to teach learners to write project proposals.

79.3 The principles for using corpora in writing pedagogy are:

Principle 1 Work comparatively. The salient features of a specific genre will not become clear unless they can be seen against a background of other genres, or other populations of texts. This requirement is built in to Scott's Keyword techniques (Scott M 1996), and is the basis both of early work in stylistics e.g. Leech and Short (1981), and of recent studies by Stubbs (1996). The LOB and Brown corpora or the BNC (Sampler, full version, or genre specific subsets) constitute invaluable resources as yardsticks against which your own text can be measured[41]. The results of Biber 1998 (and the results of this study whatever its limitations) offer a very different kind of resource. They make it possible to reference not only the lexical dimensions of the texts under consideration, but also to profile these texts in terms of grammatical use.

Principle 2 Work systematically. First, if your data is not well organised and consistent, you cannot expect consistent, reliable results – rubbish in, rubbish out! Secondly, as we saw in Chapter 3: *What are we looking for*, if you do not ask the right questions (in the right way) you may get the answers

that you *want*, but these may not be the *right* answers.[42] Hoey's *Five questions* offers a good example of the kind of framework needed for analysing words or phrases which have been identified as being of interest in a particular genre. We saw in Chapter 6: *Lexical dimensions* how combining such a check-list with keyword analysis makes it possible both to identify the saliently prominent words and phrases in a corpus, and to come to a clearer understanding of how this lexis is used by writers to achieve the rhetorical effects that they require.

Principle 3 Work with whole texts wherever possible. This is said not only because of the kinds of problems which can arise if studies are based on text fragments (Stubbs M 1996:4), but also because we are concerned here with teaching writing, not grammar or lexis. If you do not have access to the whole text, (ideally in its original form with all its other associated features e.g. figures, diagrams, typography – even typeface), you can only ever have a partial understanding of the kind of impact the text might have on a reader, and of the choices the writer has made during the production of the text. In Chapter 7: *Organisation* we saw how features such as section headings could be used in developing an understanding of the interaction between Terms of Reference and Project Proposals. If it had not been possible to identify those features in the corpus, a valuable level of analysis would have been lost.

Principle 4 Don't ignore frequent words. The discussion of *of* in Chapter 6: *Lexical dimensions* and recent work by Hoey (1997a) shows that a

---

[41] The first of these (LOB / Brown) is currently available internationally from ICAME (http://www.hd.uib.no/icame.html); the second should be available internationally by 1999 from the BNC consortium (http://info.ox.ac.uk/bnc/)

[42] See Widdowson H G "The Critical Analysis of Text" *Applied Linguistics* (Forthcoming) for a criticism of Stubbs 1994 for this kind of failure.

systematic consideration of frequent or apparently uninteresting words can reveal aspects of the text which might not be shown by a study of lower frequency items. Similarly, the striking contrast between the use of definite article "the" in the Romantic Fiction data set as compared to PPs, and its implication in the specificity of Proposal language would not have become apparent without keyword analysis. Very high frequency words such as definite articles are frequently excluded from corpus studies that are driven by lexicographic or grammatical description agendas (Sinclair J 1987, 1991). In the analysis of genres, they should be seen as potential sources of essential information.

Principle 5 Don't assume that the corpus says it all. As we have seen in Chapter 8: *Writing Project Proposals*, there can be considerable benefit in talking about the texts you have been studying with the people who wrote them. One of the benefits of this kind of discussion is that it ensures you are not imposing your own perspective on the data. The other is that you will often find that new insights become available for results you have obtained (e.g. *motherhoods* in PPs).

79.4 What of our practical example of project proposal writing? We have already seen in Chapter 3: *Approaching the data*, and Chapter 4: *What are we looking for*, that a great deal of work can be involved in obtaining examples of a genre and then putting it into an appropriate form for analysis. However, assuming that a consistently formatted and coded corpus of project proposals in machine readable form is available, (Principles 1 and 2) and that teachers / learners have access to a PC with WordSmith, and a reference wordlist based on a

corpus such as LOB or Brown, or the BNC Sampler, a useful procedure which would conform to the principles I have presented above might go as follows.

Step 1 Review the overall organisation of the genre exemplars under analysis – this should be done with access to both the paper version and an electronic version where possible. Identify the ways in which text organisers and non-verbal information are used and try to decide the motivation for this use. Where possible, review any support documentation which was required in order for the texts to be written (e.g. TOR / ITB). If it is felt to be useful, make separate lists of the language used in explicitly marked section headings (Principles 2 and 3).

During this review, ask if it is possible to identify differences or similarities between the ways in which texts in the research corpus are organised when compared with other text genres, and what motivates these differences (Principle 1).

Step 2 Analyse the examples in the research corpus to identify any discourse patterns they may share (e.g. Reason / Result, General to Particular, SPRE). How are arguments in paragraphs and sections organised? Are there any significant patterns in subject/theme (Principles 2 and 4)? In what way does the research genre differ from other genres (Principle 1)?

Step 3 Make a frequency sorted wordlist and a keyword list for the research corpus. It can be helpful to do this for single words and two word clusters (Principles 2 and 4). Review how *positive* and *negative* keywords are used in the texts. Use Hoey's five questions as a starting point for this investigation. Reference the investigation against the use of the word classes

in specific text dimensions in Biber 1988 (Principle 1). This review will make it possible to identify:

- key words and phrases associated with the genre (this offers *content knowledge* and *context knowledge* as these words and phrases will give clear indications of the elements required in allowable contributions to the genre, and *language system knowledge* as the keywords and phrases associated with them indicate the extent to which a text is e.g. nominally or verbally oriented, the kinds of subordination that are used, how noun phrases are structured etc.)

  major wordclasses that are used in the genre and the impact of the use of these wordclasses on the interpersonal dimension of the text (this will relate to particularly to Biber's Factor 1 "Involved versus Informational Production" and Factor 5 "Abstract versus Non-abstract Information". This also extends *context* knowledge and *language system* knowledge and prepares learners for effective modelling from the texts in the corpus. (Principles 1, 2 and 4)

  Step 4 Check conclusions that arise from this analysis with expert informants where-ever possible. It can be very helpful to get the agreement for a "return visit" to organisations that provide text samples. Once learners start to write independently, another way of checking conclusions against expert opinion is to use classroom techniques such as *reformulation*[43] (Allwright RL, M-P Woodley & JM Allwright 1988; Tribble C 1997a) (Principle 5).

79.5 Once such a review has been completed (either by teachers or materials writers preparing to help students, or by students investigating a genre as part

of a learning process) the research corpus becomes a means for ensuring appropriate independent writing at the end of a learning process. At the end of such a review, teachers or materials writers will be positioned to develop appropriate, motivating and stimulating learning materials which match to the needs of your students. Learners will be enabled to approach the writing task with a fuller understanding of the task and the results that they are aiming to achieve.

80. Questions 3 and 4

80.1 Discussion in the earlier parts of this chapter has assumed that the texts which have been the subject of analysis have been members of the same genre. What I have not done as yet, is to offer guidance as to how teachers and learners can feel reasonably sure that this is the case for the corpus resources they themselves assemble. The third and fourth question of the five that I set out at the beginning of the chapter address this problem:

> "What kinds of examples do teachers and learners need when approaching the problem of writing into a new genre?"
>
> "What practical means are there to help teachers decide whether or not the texts they are dealing with are exemplars of an identifiable genre?"

80.2 I have treated these two questions together, as, clearly, they are both concerned with the same issue – the need for appropriate models when genre approaches to writing instruction constitute part of a teaching / learning strategy. Flowerdew refers to this issue when he discusses the ways in which the modelling of texts is used by writers approaching new genres:

---

[43] A technique in writing instruction in which a student draft is sympathetically reformulated by an expert informant and the two versions of the text are used as learning / teaching materials

> "… this skill of seeking out instances of genre-dependent language use in English and incorporating them in one's own writing or speaking is not limited to foreign languages. Many native speakers make use of others' writing or speech to model their own work in their native language, where the genre is an unfamiliar one." (Flowerdew J 1993:313)

80.3 A short, unhelpfully circular response to the first question: "What kinds of examples do students and teachers need?" would be: "Examples of the genre". This is unhelpful because of the ambiguity which is still associated with the term *genre* itself. A more useful, though longer, answer is that if teachers and students are interested in working with a genre approach, they need to make sure that:

- the texts they are aspiring to write are part of the GSP of a clearly delineated communicative context (as we shall see below, communicative purpose is the most useful test of genre equivalence), and that

- the examples they use for analysis or modelling are truly analogous to one another.

Without such a narrow definition, genre becomes uncomfortably woolly as a category – we have seen in Chapter 2: *Teaching Writing* (in the discussion of brochures) the kinds of confusion which writing teachers can face if they do not ensure that the genre they are teaching *is* a genre, and that the texts they present to learners are members of that genre.

80.4 Having set out the criteria which should inform the selection of examples, I would want to stress that (in conformity with Principles 1 and 3) learners will need the opportunity to work with examples of whole texts, and, as we have seen in Chapter 2, they will need many rather than single examples.

80.5 The answer to Question 2 also implies the response to Question 3. The "practical means" to help teachers decide whether or not text samples are members of the same genre is to be found in the communicative purpose of the texts at issue. As an illustration I give below an example of an unambiguously delineated genre: RECIPE:

A	**Agnés Sorel** – Stuff the omelette with minced mushrooms tossed in butter and cohered with a thin chicken purée, lay some roundels of tongue on the top, surrounded with a thread of thickened gravy (Saulnier 1914, p66)
B	***Piselli alla Francese*** *Shell enough young garden peas to fill a two-pint jug. The younger and sweeter the better.* - *Stew, over a low flame, 2 chopped lettuces, 1 chopped onion, salt and pepper in 10 gms of butter for about 10 minutes.* - *Add the peas along with a generous bunch of parsley and thyme and two cups of good stock.* - *Cook over a high flame stirring continuously until the liquid has reduced and the peas are cooked.* *This will provide a main vegetable dish for five or six people. Our cook, Maria, gave it to us for the first time in the early 1950s - she had adapted it from a French recipe she learned during the war. Our only justification for putting it in a collection of Italian recipes is that we have eaten it so regularly in that part of the world.* (Ross J & M Waterfield 1973:121)
C	Put your tunny steaks in hot olive oil to brown lightly on each side. Remove and put aside to keep warm. Now put into your olive oil 2 shallots, chopped; 2 carrots, sliced; a stick of celery, finely chopped, and 4 tomatoes, skinned and chopped. Add a sprig of thyme and cook for 15 minutes. Now put your tunny steaks back into the braising pot, and add wine to cover them. Put the lid on your pan and cook for 1 hour in a slow oven [170^{o}C / 325^{o}F, gas 3]. (Lassalle G 1976:177)

Table 118 - Recipes (Tribble C 1997:54-55)

80.6 Each of these texts has a clear main communicative purpose – to enable the reader to produce a certain number of servings of a specific dish to a specific standard. Although recipes can be used for other purposes (entertainment, historic research etc.) there is usually one main reason for writing them down – to help people cook the dishes they want to serve others, or eat themselves.

While they all share this communicative purpose, the recipes do not share structure, syntax, lexis or style, and some can be seen as more central to a genre than others (e.g. recipe A is probably less central than recipe B). In terms of communicative purpose they are analogues; in terms of surface features they differ in many respects.

80.7 The practical means whereby teachers can assess whether or not they are dealing with texts that are members of the same genre lies, therefore, in a test for communicative purpose. This position harmonises with Hasan (as stated above), and with Swales, when he says:

> "Communicative purpose is both a privileged criterion and one which operates to keep the scope of a genre as here conceived narrowly focused on comparable rhetorical action." (Swales 1990:58)

… and offers a reasonable test for classroom teachers and researchers alike.

80.8 The usefulness of such a test can be seen when we consider the problems posed to teachers in foreign language learning or in mother tongue instruction by "letters" or "essays" where the texts as traditionally taught are usually accounted for on the basis of form rather than purpose. A form-driven view of texts risks leading to the mechanical imitation of text-book models. A view driven by communicative purpose starts with the *context* of writing and considers form in the light of the needs of writers and readers. In the latter material the sort of modelling proposed by Flowerdew becomes possible. It is much harder to achieve this with materials that are driven by form. This problem can be seen clearly in the "handbooks for commercial correspondence" that are widely available in the EFL and native speaker market (e.g. Ashley A 1992). In such materials, although some teaching units

are organised around themes such as: *Enquiries* or *Complaints and adjustments*, others focus on broad contexts such as "Electronic correspondence" or "In-company communication"; and, all of the teaching units are restricted by the fact that only single examples are offered for imitation.

80.9 These recipe examples also demonstrate the value of Principles 1 and 3 for working with corpus data. Working comparatively allows learners to see how the genre stands in relation to other genres (e.g. TECHNICAL INSTRUCTIONS, PRODUCT INFORMATION) and to understand the social motivation for the differences between one genre and another. Having access to many examples allows learners to appreciate the range of allowable contributions to the genre and to identify marginal and central instances. When, as is often the case with teaching materials such as the commercial correspondence handbooks referred to above, this is not the case, learners have no means for reaching this kind of conclusion.

80.10 In finding answers to the first four questions posed at the beginning of this chapter, we have, in many senses, been dealing with ideal cases where teachers and learners have access to the data they need and the resources necessary for its analysis. What of those teachers who work in less than ideal circumstances?

81. Question 5

> "Is there any help for teachers who cannot get hold of examples of the kinds of texts their students need to write?"

81.1 However persuasively I may (or may not) have made a case for using genre and corpus analysis as way of helping learners to write difficult texts, the last

of my five questions still requires an answer. All too often, it will prove to be extremely difficult for teachers to find examples or analogues of the kinds of texts that their students need to learn to write. Interestingly from my perspective, this problem is not restricted to learners with an interest in recondite disciplines or professional activities, it is a major problem for most learners and most teachers. This problem exists for several reasons - e.g.

- In many cases learners in educational or training settings are asked to write texts which are only written in those settings – the prime example of this being the still ubiquitous "essay", a genre which has no life outside the classroom.

- In academic settings it is usually difficult, or impossible, to have access to one of the key academic genres – examination responses. Although this is for good administrative and ethical reasons, it nevertheless makes it extremely difficult for learners to gain an appreciation of how to go about writing into these genres.[44]

81.2 Various solutions have been proposed to this kind of problem. Granger (1998) recommends the use of corpora of essays written by native speakers (NS) of British or American English as a way of giving learners exemplar texts for modelling ESSAY. I have argued against this position (Granger S & C Tribble 1998) – not because of any problem of shared communicative purpose, but because of the highly problematic nature of unedited NS production – and have suggested that the essays currently found in CD ROM

encyclopaedias offer an interesting set of analogues to various genres of writing commonly required of learners. (Tribble C 1997c). Johns at Birmingham University uses a corpus of New Scientist texts as the basis for a remedial grammar course for postgraduate students at that institution (Johns T 1997 & Tim Johns' home page).

81.3 The main point to make here is that when it is not possible to collect examples of precise analogues of the genres that learners need to write, teachers can still use genres that stand in a reasonably close relation to the target genre (e.g. encyclopaedia essays or newspaper editorials as relatives of student essays). Such data will not allow the highly specific modelling which the PP corpus could make possible for writers of analogous proposals. However, so long as learners understand that there are restrictions on the conclusions they draw from this corpus data, such a corpus will still give students infinitely more scope to develop language system knowledge than other approaches. Thus, in the case of the New Scientist collection used by Johns, Myers has clearly demonstrated that the communicative purpose of scientific journalism is markedly different from that of academic journals, and requires strongly contrasting instantiations of equivalent knowledge (Myers 1994) – such articles are not analogues of journal articles or MA dissertations. Johns has shown, however, that such texts can be useful in extending learners' language system knowledge, even though the texts are not members of the genres that learners need to write. While they cannot be used as a basis for modelling

---

[44] A good example of this problem occurred in my own experience teaching LLM students in the University of London. The legal "problem" essay that students have to write is a major genre in the LLM examination, yet it is never published nor publicly circulated. It is produced by candidates, consumed by examiners and then destroyed by the university after a brief period in the archives of Senate House.. The examiners know what a "good" answer looks like – but candidates can only find out about the genre

major academic genres such as examination essays, laboratory reports, or MA dissertations, because they have face validity with John's learners at Birmingham, they do constitute a valuable starting point for the development of language system knowledge. (Johns T 1997)

82. Some conclusions

82.1 So where have we arrived? We certainly do not have a recipe book for writing project proposals – or any other kind of difficult text. That has never been my intention. What I hope we do have is, first, a clearer sense of the ways in which teachers and learners can draw on the notion of genre – that is, to paraphrase Hasan, "language doing the job appropriate to a specific class of social happenings" – as a way of framing a pedagogy for writing. Secondly, I hope that I have demonstrated the potential of certain techniques in empirical linguistics (notably corpus analysis) for application to language education.

82.2 If teachers (and especially text book writers) who read this thesis are persuaded that there might be some value in using some of the approaches proposed here, it could have an interesting long term impact on the kinds of materials that are published for teaching writing. At the very least I look forward to text books for teaching writing which offer assistance to learners in writing real-world genres, and which give learners the chance to develop appropriate knowledge of the context, language system and writing processes required to write into those genres. At best, I look forward to text books which offer all of this, and are accompanied by a CD ROM containing a corpus of appropriate text examples and appropriate text analysis software.

---

at second hand, usually through a very short "examination preparation" course offered by one of the

82.3 As a final note, as with most research projects, I have probably raised more questions than I have resolved. Some of the potential research areas which have arisen during earlier discussion, and which I feel would merit further consideration are:

- Does a sudden obligation to develop a capacity to write into new genres cause changes in social relations? Administrators and academics in the former communist countries of central and eastern Europe have not only had to learn new ways of working, they have also had to learn how to write new kinds of text (in their mother tongue and in English). To what extent has the need to write into these new genres changed the ways in which people think about the problems they are addressing, and change the way they work with one another?

- Studying "Negative Keywords" proved to be much more fruitful than I had expected – especially when combined with Hoey's five questions (e.g. the discussion of *of* in Chapter 5: *Lexical dimensions*). There appears to be considerable potential for further study of both positive and negative keyword contrasts between different written genres or different modes of production (spoken / written). It would be interesting to see the extent to which these techniques would permit a classification of texts which was equivalent to Biber 1988, but which did not require a marked up corpus.

- The interviews that are reported in Chapter 8: *Writing Project Proposals* give important insights into the ways in which texts such as Invitations to Bid (ITB) interact with the texts which are written in response to such invitations.

---

tutors who is also an examiner.

It would be interesting to investigate the importance of such texts as a resource for proposal writers, and to get a better understanding of the ways in which the contrasting approaches that consultancy groups adopt to planning and developing a proposal determine the way in which they follow or flout the text specifications made in the original invitation.

82.4 Enough. As I have noted in the epigraph to this chapter, what we call knowledge is provisional. I hope that in writing this particular difficult text I have made a contribution to our understanding of some of the issues that cluster around writing, and teaching writing. I certainly do not expect it to be the last word.

BIBLIOGRAPHY, TABLES & FIGURES

BIBLIOGRAPHY

Allwright RL, M-P Woodley & JM Allwright (1988) "Investigating Reformulation as a practical strategy for the teaching of academic writing" *Applied Linguistics* 9/3:122-128

Ashley A (1992) *A handbook of commercial correspondence* Oxford University Press Oxford

Aston, Guy (1996) "The British National Corpus as a Language Learner Resource" in Botley S, Glass T, McEnery T & A Wilson (eds.) *Proceedings of Teaching and Language Corpora 1996* UCREL Technical Papers Volume 9:178-191 UCREL, Lancaster University Lancaster

Baker M, G Francis & E Tognini-Bonelli (eds.) (1993) *Text and Technology: In honour of John Sinclair* John Benjamins Publishing Company Philadelphia

Bakhtin MM (1986) "The problem of speech genres" in Emerson C & M Holquist (eds.) *Speech Genres and other late essays*. University of Texas Press Austin

Bamforth R (1993) "Process vs. Genre: Anatomy of a false dichotomy" *Prospect: a journal of Australian TESOL* 8/1-289-99

Bazerman C (1994) *Constructing Experience* Southern Illinois University Press Carbondale and Edwardsville Il

Benson JD & WS Greaves (1980) "Field of discourse: theory and application" *Applied Linguistics* 1:45-55

Bereiter C & M Scardamalia (1987) *The psychology of written communication* L Erlbaum Hillsdale NJ

Bereiter C & M Scardamalia (1993) "Rethinking genre from a sociocognitive perspective" *Written Communication* 10:475-509

Berkenkotter C & TN Huckin (1993) "Rethinking genre from a sociocultural perspective" *Written Communication* 10/4: 475-509

Berkenkotter C & TN Huckin (1995) *Genre Knowledge in Disciplinary Communication* Lawrence Erlbaum Associates Hillsdale NY

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

Biber D (1988) *Variations across Speech and Writing* Cambridge University Press Cambridge

Biber D (1989a) "A typology of English texts" *Linguistics* 27:3-43

Biber D (1989b) "Styles of stance in English: lexical and grammatical markings of evidentiality and affect" (Special edition: *The pragmatics of affect* – edited by Elinor Ochs) *Text* 9: 93-124

Biber D (1990) "Methodological issues regarding corpus based analyses of linguistic variation" *Literary and Linguistic Computing* 5 /4: 257-269

Biber D & E Finegan (1989) "Drift and the evolution of English style: a history of three genres" *Language* 65: 487-517

Biber D, S Conrad & R Reppen (1994) "Corpus-based approaches to issues in Applied Linguistics" *Applied Linguistics* 15/2:169-188

Bizzell P (1987) "Language and literacy" in Enos T (ed.) *A sourcebook for basic writing teachers* :125-137 Random House New York

Black A, Carling F & G Leech (1993) *Statistical Grammars of the English Language* Rodopi Amsterdam

Bloor M & T Bloor (1991) "Cultural expectations and socio-pragmatic failure in academic writing" in Adams P, Heaton B & P Howarth (eds.) *Socio-cultural issues in English for academic purposes* Macmillan London

Botley S, Glass G, McEnery T & A Wilson (eds.) (1996) *Proceedings of Teaching and Language Corpora 1996* UCREL University of Lancaster Lancaster

Bowers C A (1987) *Elements of a post-liberal theory of education* Teachers College Press New York

Brandt D (1986) "Text and context: how writers come to mean" in Barbara Couture (ed.) *Functional approaches to writing: research perspectives* Francis Pinter London

Britton J (1983) "Shaping at the point of utterance" in Freedman A, Pringle I & J Yalden (eds.) *Learning to write: first language, second language* :13-19 Longman London

Brooks A & P Grundy (1990) *Writing for Study Purposes* Cambridge University Press Cambridge

Burnard L (1995) *The British National Corpus Users Reference Guide (SGML version)* Oxford University Computing Services Oxford

Callaghan M; Knapp P & G Noble (1993) "Genre in practice" in Cope B and Kalantzis M (eds.) *The powers of literacy: a genre approach to teaching writing* The Falmer Press London

Campanelli P & JM Channell (1994) Training: an exploration of the word and the concept with an analysis of the implications for survey design Employment Department London

Censor (1982 – first published c1880) *Don't* Pryor publications London

Clark R & R Ivanić (1997) *The politics of writing* Routledge London & New York

Christie F (1993) "Curriculum genres: planning for effective teaching" in Cope B & M Kalantzis (eds.) *The powers of literacy: a genre approach to teaching writing* The Falmer Press London & Washington

Coe RM (1987) "An apology for form: Or, who took the form out of process?" *College English* 49:13-28

Cook G & Seidlhofer B (1995) *Principle and Practice in Applied Linguistics: studies in honour of HG Widdowson* OUP Oxford

Cooper M (1986) "The ecology of English" *College English* 48:364-375

Cooper M (1989) "Why are we talking about discourse communities? Or, foundationalism rears its ugly head once more" in Cooper M & M Holzman (eds.) *Writing as social practice* :202-220 Boynton . Cook Portsmouth NH

Cope B & M Kalantzis (1993) The powers of literacy: a genre approach to teaching writing The Falmer Press London

Coulthard M (ed.) (1994) *Advances in Written Text Analysis* Routledge London

Couture B (1986) "Bridging epistemologies and methodologies: research in written language function" in Couture B (ed.) *Functional approaches to writing: research perspectives* : 1-10 Frances Pinter London

Crookes G (1986) "Towards a validated analysis of scientific text structure" *Applied Linguistics* 7/1:57-70

De Beaugrande R & W Dressler (1981) *Introduction to text linguistics* Longman London

De Cock S (1998) "An automatic approach to the phrasicon of EFL learners" in Granger S (ed.) *Working with Learner Language* Longman Harlow

Doherty M, L Knapp & S Swift (1987) *Write for Business* Longman Harlow

Dudley-Evans T (1985) *Writing Laboratory Reports* Nelson London & Sydney

Dudley-Evans T (1994) "Genre analysis: an approach to text analysis for ESP" in Coulthard M (ed.) *Advances in written discourse* :219-228 Routledge London & New York

Dunning T (1993) "Accurate methods for the statistics of surprise and coincidence" *Computational Linguistics* 19/1:61-74

Edwards H (1997) "Subject-specific EAP: issues in developing a syllabus" in Langley G (ed.) *BALEAP Professional Interest Meetings Report 1995-1997* BALEAP London

Ellsworth E (1989) "Why doesn't this feel empowering? Working through the repressive myths of critical pedagogy" *Harvard Education Review* 59; 3: 297-324

Emig J (1983) *The web of meaning* Boynton/Cook Upper Montclair NJ

Faigley L (1986) "Competing theories of process: a critique and a proposal" *College English* 48/6:527-542

Fairclough N (1989) *Language and Power* Longman Harlow

Fairclough N (1992) *Discourse and social change* Polity Press Cambridge

Fielding N (1993) "Qualitative Interviewing" *Researching social life* Sage Publications London

Firth JR (1957) "A synopsis of linguistic theory, 1930-1955" reprinted in Palmer FR (ed.) *Selected Papers of JR Firth, 1952-59* Longman London

Fish S (1980) *Is there a text in this class?* Harvard University Press Harvard, Mass.

Flower L (1994) *The construction of negotiated meaning: a social cognitive theory of writing* Southern Illinois University Press Carbondale and Edwardsville Il

Flowerdew J (1993) "An educational or process approach to the teaching of professional genres" *ELTJ* 47/4305-316 Oxford University Press Oxford

Flowerdew J (1996) "Concordancing in Language Learning" in Martha Pennington (ed.) *The Power of CALL*: 97-113 Athelstan Houston Tx

Foggart J (1993) *Real Writing* Hodder and Stoughton London

Foucault M (1972) *The archaeology of knowledge* Harper & Row New York

Fox G (1993) "A comparison of 'policespeak' and 'normal speak': a preliminary study" in Sinclair J, M Hoey & G Fox (eds.) *Techniques of description* Routledge London

Freedman A & Medway P (1994) "New Views of Genre and their implications for Education" in Freedman A & Medway P (eds.) *Learning and teaching genre* Heinemann Boynton / Cook Portsmouth NH

Freedman A; Adam C & G Smart (1994) "Wearing suits to class: simulating genres and simulations as genres" *Written Communication* 11/2:193-226

Fulcher G (ed) (1997) *Writing in the English language classroom* Prentice Hall Europe ELT Hemel Hempstead

Garside R, Leech G & G Sampson (1987) *The computational analysis of English* Longman London

Gee S (1997) "Teaching writing: a genre based approach" in Fulcher G (ed.) *Writing in the language classroom* Prentice Hall ELT Hemel Hempstead

Geertz C (1983) *Local Knowledge: further essays in interpretive anthropology* Basic Books New York

Giroux H (1988) *Schooling and the struggle for public life: critical pedagogy in the modern age* University of Minnesota Press Minneapolis

Grabe W & R Kaplan (1996) *Theory & practice of writing* Longman London & New York

Graddol D (1997) *The Future of English* The British Council London

Graddol D & S Goodman (eds.) (1996) *Redesigning English: new texts, new identities* Routledge London

Granger S (ed.) (1998) *Working with Learner Language* Longman Harlow

Granger S & C Tribble (1998) "Exploiting Learner Corpus Data in the Classroom: form-focused instruction and data-driven learning" in Granger S (ed.) *Working with Learner Language* Longman Harlow

Graves D (1983) *Writing: teachers and children at work* Heinneman Educational Books Portsmouth NH

Graves D (1984) *A researcher learns to write* Heinemann London & Portsmouth NH

Gunnarsson B (1997) "The writing process from a sociolinguistic viewpoint" *Written Communication* 14/2, April 1997:139-188

Halliday MAK (1959) *The language of the Chinese 'Secret History of the Mongols'* Publication 17 of the Philological Society Blackwell Oxford

Halliday MAK (1973) *Explorations in the Functions of Language* Edward Arnold London

Halliday MAK (1975) *Learning how to mean: explorations in the development of language* Edward Arnold London

Halliday MAK (1978) *Language as social semiotic* Edward Arnold London

Halliday MAK (1989) *Spoken and written language* Oxford University Press Oxford

Halliday MAK & R Hasan (1976) *Cohesion in English* Longman London

Halliday MAK & R Hasan (1985) *Language context and text: aspects of language in a social-semiotic perspective* Oxford University Press Oxford

Hamp-Lyons L & B Heasley (1987) *Study Writing* Cambridge University Press Cambridge

Hayes J & L Flower (1983) "Uncovering cognitive processes in writing: an introduction to protocol analysis" in S Rosenburg (ed.) *Advances in applied psycholinguistics: Vol 2 Reading, writing and language learning* Cambridge University Press Cambridge

Hedge T (1988) *Writing* OUP Oxford

Hoey M (1983) *On the surface of discourse* George Allen and Unwin London

Hoey M (1997a) "From concordance to text structure: new uses for computer corpora" to appear in Melia J & Lewandoska B (eds.) *Proceedings of PALC 97* Łodz University Press Łodz, Poland

Hoey M (1997b) "The interaction of textual and lexical factors in the identification of paragraph boundaries" in Thiele W & R Mechthild (eds.) *Grammar and text in synchrony and diachrony: in honour of Gottfried Graustein* Vervuert Iberoamericana Leipzig

Hoey M & E Winter (1986) "Clause relations and the writer's communicative purpose" in Couture B (ed.) *Functional Approaches to Writing* : 120-141 Frances Pinter London

Holroyd M (1971) *Lytton Strachey: a literary biography* Penguin Books London

Hopkins A & C Tribble (1989) *Outlines* Longman London

Jacobs SE (1989) "From syntax to genre: Writers' use of definite constructions in fifteen editorials" *Written Communication* 6/4: 528-554

Johns AM (1990) "L1 compositions theories: implications for developing theories of L2 composition" in Kroll B (ed.) *Second language writing: research insights from the classroom* :24-36 Cambridge University Press Cambridge

Johns T (1988) "Whence and Whither Classroom Concordancing?" in T Bongaerts et al (eds.) *Computer Applications in Language Learning* Foris Dordrecht

Johns T (1991) "Should you be persuaded: Two examples of Data-Driven Learning Materials" in Johns T & P King (eds.) *Classroom Concordancing – Special edition: English Language Research Journal* 4:1-16 Birmingham University Birmingham

Johns T (1994) "From printout to handout: grammar and vocabulary learning in the context of data-driven learning" in Odlin T (ed.) *Approaches to Pedagogic Grammar* :293-313 Cambridge University Press

Johns T (1997) "Kibbitzing one-to-ones" in Langley GV (ed.) *BALEAP professional interest meetings report 1995-1997* British Association of Lecturers in English for Academic Purposes London

Johns T & P King (eds.) (1991) *Classroom Concordancing – Special edition: English Language Research Journal* 4 Birmingham University Birmingham

Jones L & R Alexander (1989) *International Business English* Cambridge University Press Cambridge

Jordan RR (1992) *Academic Writing Course* Nelson London

Kleimann S (1993) "The reciprocal relationship of workplace culture and review" in Spilka R (ed) *Writing in the workplace: new research perspectives* Southern Illinois Press Carbondale and Edwardsville Il

Kress G (1989) *Linguistic processes in sociocultural practice* OUP Oxford

Kress G (1993) *Learning to write* Routledge London

Kuhn TS (1970) *The structure of scientific revolutions (2nd edition)* University of Chicago Press Chicago

Lamott A (1994) *Bird by bird: some instructions on writing and life* Doubleday New York

Langan J (1993) *College writing skills with readings* McGraw-Hill New York

Lassalle G (1976) *The adventurous fish cook* Macmillan London

Leech G & M Short (1981) *Style in Fiction* Longman Harlow

Louw B (1993) "The diagnostic potential of semantic prosodies" in Baker M, Francis G & E Tognini-Bonelli (1993) *Text and technology: in honour of John Sinclair* :157-176 John Benjamins Philadelphia/Amsterdam

Magee B (1973) *Popper* Fontana Paperbacks London

Martin JR (1985) "Process and text: two aspects of human semiosis" in Benson JD & WS Greaves (eds.) *Systemic perspectives on discourse, Vol. 1* Ablex Northwood NJ

Martin JR (1989) *Factual Writing: exploring and challenging social reality* Oxford University Press Oxford

Maybin J (1994) "Teaching writing: process or genre" in Brindley S (ed.) *Teaching English* The Open University / Routledge

Mayer J (1990) *Uncommon Sense* Heinemann Portsmouth NH

Melia J & Lewandoska B (eds.) (1997) *Proceedings of PALC 97* Łodz University Press Łodz, Poland

McArthur T (1986) *The written word* Oxford University Press Oxford

McEnery T & A Wilson (1996) *Corpus Linguistics* Edinburgh University Press Edinburgh

McKenna B (1997) "How engineers write: an empirical study of engineering report writing" *Applied Linguistics* 18/2:189-211

Miller C (1984) "Genre as social action" *Quarterly Journal of Speech* 70: 151-167

Millington H (1995) *Windows Grep 1.6* Huw Millington East Grinstead UK

Murison-Bowie S (1996) "Linguistic corpora and language teaching" *Annual Review of Applied Linguistics* 16:182-199

Murray D M (1982) *Learning by teaching: selected articles on writing and teaching* Boynton/Cook Montclair NJ

Myers G (1990) *Writing Biology: texts in the social construction of scientific knowledge* University of Wisconsin Press Madison

Myers G (1994) "Narratives of Science and Nature" in Coulthard M (ed.) *Advances in written text analysis* Routledge London

O'Dell F (1996) *CAE writing skills* Cambridge University Press Cambridge

Odell L & Goswami D (eds.) (1985) *Writing in non-academic contexts* Guildford New York

Oostdijk N (1988) "A corpus linguistic approach to linguistic variation" *Literary and Linguistic Computing* 3:12-25

Oppenheim AN (1992) *Questionnaire design, interviewing and attitude measurement* Pinter Publishers London

Owen R (1992) *BBC Business English* BBC English London

Magee B (1973) *Popper* Fontana Paperbacks London

Quirk R, Greenbaum S, Leech G & J Svartvik (1985) *A comprehensive grammar of the English language* Longman London

Raimes A (1993) "Out of the woods: Emerging traditions in the teaching of writing" in Silberste S (ed.) *State of the Art TESOL Essays* TESOL

Reither J (1985) "Writing and knowing: towards redefining the writing process" *College English* 47:620-628

Richards J, Platt J & H Weber (1985) *Longman Dictionary of Applied Linguistics* Longman London

Ross J & M Waterfield (1973) *Leaves from my Tuscan notebook* Penguin Books Harmondsworth

Samraj BR (1989) "Exploring current issues in genre theory" *Word* 40 1/2: 189-200

Saulnier L (1914) *Le Répertoire de La Cuisine* Leon Jaeggi & Sons Staines, UK

Scott M (1996) *WordSmith Tools* Oxford University Press Oxford

Scott M (1997a) *WordSmith Tools Manual* OUP Oxford

Scott M (1997b) "PC Analysis of key words - and key key words" *System* 25/1 : pp 1-13

Scott M (1997c) "The right word in the right place: Key word associates in two languages" *Arbeiten aus Anglistik und Amerikanistik* 22/2: 239-252 Tubingen

Scott M (1998) "Focusing on the text and its key words" in Burnard L (ed.) *Teaching and Language Corpora 1998* Unpublished proceedings of the 1998 TALC conference, Oxford

Selzer G (1993) "Intertextuality and the writing process" in Spilka R (ed.) *Writing in the workplace: new research perspectives* Southern Illinois University Press Carbondale and Edwardsville Il

Sinclair J (ed.) (1987) *Looking Up* Collins ELT London & Glasgow

Sinclair J (1991) *Corpus, Concordance, Collocation* OUP Oxford

Sinclair J et al. (1987) *Collins COBUILD English Language Dictionary* Collins London

Sinclair J et al (1990) *Collins COBUILD English Grammar* Collins ELT London

Sinclair J & M Coulthard (1975) *Towards an analysis of discourse* Oxford University Press Oxford

Smart G (1993) "Genre as invention: a central bank's response to its executives' expectations as readers" in Spilka R (ed.) *Writing in the workplace* Southern Illinois University Press Carbondale and Edwardsville Il

Spack R & C Sadow (1983) "Student-teacher working journals in ESL freshman composition" *TESOL Quarterly* 17/4:575-593

Spilka R (1993) *Writing in the workplace: new research perspectives* Southern Illinois Press Carbondale and Edwardsville Il

Stevens V (1995) "Concordancing with Language Learners: Why? When? What?" *CAELL Journal* 6/2:2-10

Stubbs M (1987) "An educational theory of (written) language" in Bloor T and J Norrish (eds.) *Papers from the Annual Meeting of the British Association for Applied Linguistics* CILT / BAAL London

Stubbs M (1994) "Grammar, Text, and Ideology: Computer-assisted Methods in the linguistics of Representation" *Applied Linguistics* 15/2:201-223

Stubbs M (1995) "Corpus evidence for norms of lexical collocation" in Cook G & Seidlhofer G (eds.) *Principle and Practice in Applied Linguistics: studies in honour of HG Widdowson* OUP Oxford

Stubbs M (1996) *Text and Corpus Analysis* Blackwell Oxford

Summers D & M Rundell (1990) *Longman Dictionary of Contemporary English* (2nd Edition) Longman Harlow

Swales J (1981) *Aspects of Articles Introductions* Aston ESP Research Reports No. 1 University of Aston Birmingham

Swales J (1990) *Genre Analysis* Cambridge University Press Cambridge

Swales J & CB Feak (1994) *Academic writing for graduate students* University of Michigan Press Ann Arbor, Michigan

Swales JM & P Rogers (1995) "Discourse and the projection of corporate culture: the Mission Statement" *Discourse and society* 6/2: 223-242

Thompson P (1997) "Current issues in EAP writing instruction" in Langley G (ed.) *BALEAP Professional Interest Meetings: Report 1995-1997* British Association of Lecturers in English for Academic Purposes

Thompson P (1998) "Examining PhD theses in a corpus" in Burnard L (ed.) *Teaching and language corpora* Unpublished proceedings of the 1998 TALC conference, Oxford

Tribble C (1985) *Style and Purpose: or why things get written in different ways* MA dissertation, Institute of Education, London University London

Tribble C (1989) *Word for Word* Longman London

Tribble C (1991) "Some uses of electronic text in English for academic purposes" in J Milton & K Tong (eds.) *Text Analysis in Computer Assisted Language Learning* The Hong Kong University of Science and Technology Hong Kong

Tribble C (1997a) *Writing* Oxford University Press Oxford

Tribble C (1997b) "Corpora, Concordances and ELT" in Boswood T (ed.) *New Ways in using Computers in ESL* TESOL

Tribble C (1997c) "Improvising corpora for ELT: quick-and-dirty ways of developing corpora for language teaching" in Melia J & B Lewandowska-Tomaszczyk (ed) *PALC '97 Proceedings* Łódź University Press Łódź

Tribble C (1998) "Genres, keywords, teaching: towards a pedagogic account of the language of project proposals" in Burnard L (ed) *Teaching and Language Corpora 98* (unpublished proceedings of the 1998 TALC conference, Oxford)

Tribble C & G Jones (1990) *Concordances in the Classroom* Longman Harlow

Tribble C & G Jones (1997) *Concordances in the Classroom: A Resource Book for Teachers* Athelstan Houston Tx

Turk C & J Kirkman (1989) *Effective Writing, improving scientific, technical and business communication* E & FN Spon (Chapman and Hall) London

van Leeuven T (1993) "Genre and field in critical discourse study" *Discourse and society* 4/2: 193-223

White R & V Arndt (1991) *Process Writing* Longman Harlow

White R & D McGovern (1994) *Writing* Prentice Hall International Hemel Hempstead

Widdowson HG (1984) *Explorations in Applied Linguistics 2* Oxford University Press Oxford

Widdowson H G (Forthcoming) "The Critical Analysis of Text" *Applied Linguistics*

Williams R (1976) *Keywords* Fontana London

Wilson A & T McEnery (eds) (1994) *Corpora in language education and research: A selection of papers from TALC94* UCREL Lancaster University Lancaster

Winter EO (1976) "Fundamentals of information structure" mimeo: Hatfield Polytechnic Hatfield

Winter EO (1977) "A clause relational approach to English texts: a study of some predictive lexical items in written discourse" *Instructional Science* 6/1: 1-92

Winter EO (1982) *Towards a contextual grammar of English* Allen & Unwin London

Witte S (1992) "Context; text and intertext: Towards a constructionist semiotic of writing" *Written Communication* 9: 237-308

Wynne M (1996) *A post-editor's guide to CLAWS7 tagging* UCREL University of Lancaster Lancaster

Zamel V (1983) "The composing processes of advanced ESL students: 6 case studies" *TESOL Quarterly* 17/2165-187

LIST OF TABLES

Table 1 - What writers need to know....10
Table 2 - Business English materials (Contents)....37
Table 3 - Business English Materials (Owen R, 1992:59)....39
Table 4 - Letter of rejection (Corpus of administrative correspondence – personally held)....42
Table 5 - pleased / unfortunately....45
Table 6 - EAP Programme....47
Table 7 - Brochure counts....64
Table 8 - Proposals: break-down by agency....74
Table 9 - Proposals: location....75
Table 10 - Biber 1988: Linguistic Features....95
Table 11 - Neighbouring Genres....107
Table 12 - Counts (First Algorithms)....110
Table 13 - Major discrepancies....111
Table 14 - Neighbouring texts....112
Table 15 – Discrepancies....113
Table 16 - Neighboring texts (by Dimension)....121
Table 17 - Ranked scores....125
Table 18 - high frequency items....126
Table 19 - low frequency items....126
Table 20 - top 34 out of 852 attributive adjectives....129
Table 21 - bottom 10 out of 852 attributive adjectives....129
Table 22 - Attributive Adjectives....130
Table 23 - Considerable....130
Table 24 - Wide....131
Table 25 - Key....131
Table 26 - Colours....131
Table 27 - International....132
Table 28 - Top PP classifying adjectives....133
Table 29 - International....134
Table 30 - international (BNC)....135
Table 31 - Nominalisations: punctuation....137
Table 32 - Nominalisations: verbs following....138
Table 33 -Top 20 nominalisations....138
Table 34 - coordination: adverb....140
Table 35 - coordination: verb....141
Table 36 - Doubling effect....143
Table 37 - Verb preceding noun coordination....145
Table 38 - predictive "will"....146
Table 39 - this + predictive "will"....147
Table 40 - words in Theme....148
Table 41 - will + verb....148
Table 42 - adverbs in PP and Romfict....149
Table 43 - Adverbs in PP....150
Table 44 - Adverbs in RomFict....151
Table 45 - verb:noun....151
Table 46 - male/female pronouns....152
Table 47 - their....152

Table 48 - private verbs – RomFict 153
Table 49 - PP private verbs 154
Table 50 - vagueness 154
Table 51 - integrity 155
Table 52 - linguistic features 156
Table 53 - "PP / Guardian Wordlists" 160
Table 54 - PP/BNC Keywords 168
Table 55 - PP/Guardian Keywords 169
Table 56 - BNC / Guardian Keywords – PP Frequency 170
Table 57 - PP/BNC: negative keywords 172
Table 58 - PP/Guardian: negative keywords 172
Table 59 - PP/BNC/Guardian Negative Keywords 173
Table 60 - RomFict: positive keywords 174
Table 61 - RomFict: negative keywords 175
Table 62 - Romfict: frequency 176
Table 63 - Positive keywords (word class) 182
Table 64 - Positive keywords 182
Table 65 - Negative KW: word class 185
Table 66 - "that" in the PP Corpus 186
Table 67 - but 187
Table 68 - X^2 test 187
Table 69 - But: counts 188
Table 70 - PP But colligation 190
Table 71 - RomFict But colligation 190
Table 72 - but (2 word clusters) 190
Table 73 - "but we" 192
Table 74 - "but" positions 193
Table 75 - However 193
Table 76 - BNC Core – however/but 194
Table 77 - PP key-keywords 195
Table 78 - experience collocates 197
Table 79 - experience: left sort 198
Table 80 - experience: right sort 199
Table 81 - experience elsewhere 200
Table 82 - experience clusters 202
Table 83 - experience lexical collocates (clusters) 202
Table 84 - experience: patterns 204
Table 85 - Independent texts – experience: patterns 204
Table 86 - PP experience left context (82.53% of all instances) 207
Table 87 - PP experience right context (86.6% of all instances) 207
Table 88 - BNC Spoken Corpus – Experience: semantic prosodies other than "professional" 209
Table 89 - experience: BNC spoken corpus data 209
Table 90 - BNC/PP experience: left colligates (all counts are percents) 211
Table 91 - BNC/PP experience: right colligates (all counts are percents) 211
Table 92 - Colligational information 212
Table 93 - PP vs BNC Colligation 213
Table 94 - Experience: colligates 214
Table 95 - PP Colligation (Left/Right) 214
Table 96 - PP Colligations Left + Right patterns 215

Table 97 - "experience of" in BNC Written....216
Table 98 - experience: counts....217
Table 99 - PP sentence initial....218
Table 100 - Document structure....219
Table 101 - PP Corpus - level counts....228
Table 102 - Heading level counts....229
Table 103 - Sections type:token....229
Table 104 - extract from PHARE ITB....231
Table 105 - Level 1 headings....231
Table 106 - content elements....235
Table 107 - moves in obligatory content elements?....236
Table 108 - first person pronouns in PP and LOB....241
Table 109 - first person pronouns (formal writing)....242
Table 110 - right collocates "our" in the PP Corpus....243
Table 111 - "We" in ENV....243
Table 112 - CGC Cherepov....244
Table 113 - BK162....245
Table 114 - Primary headings....263
Table 115 - Tertiary headings....263
Table 116 - Text analysis passage....264
Table 117 - "motherhood vs keywords"....267
Table 118 - Recipes (Tribble C 1997:54-55)....287

LIST OF FIGURES

Figure 1 Clark and Ivanić 1997:11....17
Figure 2 - Flowers & Hayes writing process model (Grabe W & R Kaplan 1996:92)....25
Figure 3 - An alternative representation of the process of writing as a social practice (Clark & Ivanić 1997:98)....26
Figure 4 - Writing as communicative language use (Grabe W & R Kaplan 1996:226)....27
Figure 5 - Swales 1990:84....31
Figure 6 - Doherty Knapp & Swift 1987:13-15....41
Figure 7 - Hamp-Lyons L & B Heasley 1987:100....50
Figure 8 - Dudley-Evans T 1985:5-8....51
Figure 9 - White R & D McGovern 1994:22-4....54
Figure 10 - Jordan RR 1992:49-50....56
Figure 11 - O'Dell F 1996:87....63
Figure 12 - Involved vs. Informed: first attempt....104
Figure 13 - Narrative vs. Non-narrative: first attempt....104
Figure 14 - Explicit vs. Situation Dependent: first attempt....105
Figure 15 - Overt expression of persuasion: first attempt....105
Figure 16 - Abstract vs. non-abstract: first attempt....106
Figure 17 - On-line informational elaboration: first attempt....106
Figure 18 - Text sequencing exercise....238
Figure 19 - ENV: Text sample....240
Figure 20 - Callaghan M, Knapp P & G Noble 1993:196....277

LIST OF APPENDICES

1. Attributive Adjectives_Excel formulae.doc
2. Brochure (BNC search results).doc
3. Brochure texts.doc
4. But analysis.doc
5. CGA ENV – Table of contents.doc
6. CLAWS7 TAGLIST .doc
7. Environmental Education TOR.doc
8. Experience Immediate left collocates.doc
9. Experience_ sentence initial.doc
10. Experience_BNC_Spoken.doc
11. Experience_BNC_Written.doc
12. Experience_Clusters.doc
13. Experience_left and right sort.doc
14. Experience_sentence initial.doc
15. Extracting colligation information from POS coded text.doc
16. Interview - Text analysis (Respondent 1).doc
17. Interview - Text analysis (Respondent 2).doc
18. Interview - Text analysis (Respondent 3).doc
19. Interviews – Questionnaire.doc
20. Interviews – Respondent 1 .doc
21. Interviews – Respondent 2 .doc
22. Interviews – Respondent 3.doc
23. Keywords_PP_BNC.doc
24. Keywords_PP_GDN.doc
25. Nominalisations .doc
26. Not only_but.doc
27. Organisation – content elements.doc
28. Organisation – Section Counts.doc
29. Organisation – summary of proposal sections.doc
30. 'our' in PPs.doc
31. PHARE Contract - ITB.doc
32. PHARE Contract – Technical Proposal.doc
33. PHARE Contract - TOR.doc
34. Phrasal coordination.Adjective.doc
35. Phrasal coordination.Verb.doc
36. PP Themes.doc
37. Predictive modal will.doc
38. Preparing the PP Corpus.doc
39. Private verbs.doc
40. Search Algorithms_final set.doc
41. Search Algorithms_initial set.doc
42. Tendering Instructions (PHARE Project).doc
43. Text sequencing activity.doc

Printed reports from the following spreadsheets are included in the Appendices.

1. Corpus Results Biber Factor scores (XL Corpus Results Biber Factor scores.xls)
2. Corpus Results Original Counts (XL Corpus Results Original Counts.xls)
3. Experience colligations (XL Experience_colligation and other data.xls)