The University of Reading

A pedagogically-motivated corpus-based examination of PhD theses: macrostructure, citation practices and uses of modal verbs

Submitted for the degree of PhD

School of Linguistics and Applied Language Studies

Paul Thompson

September 2001

Dedicated to the memory of my father

Abstract

This thesis is a pedagogically-motivated study of the genre of 'PhD thesis'. Variation in the macrostructure of such texts is investigated in a corpus of sixteen theses, half of which represent work in an applied science (Agricultural Botany) while the other half belong to the social science domain (Agricultural and Food Economics). The degree of variation between theses in the two departments, and also between theses in a single department are further examined through two empirical studies, one which looks at the different uses of citations, while the other describes the range of functions that the core modal auxiliaries play in the texts. The purpose of these studies is to identify what choices are available to thesis writers in different disciplinary contexts.

The contexts in which the theses were produced is explored through a set of interviews with the PhD supervisors, which indicates that the two sets of theses perform essentially different functions: the Agricultural Botany theses are fundamentally reports, while the Agricultural and Food Economics theses are primarily discursive.

Four rhetorical macrostructures for the theses are described, two for the Agricultural Botany, and two for the Agricultural and Food Economics theses, and the latter are shown to be less conventionalized. The two disciplinary communities are shown to be distinct, in their practices and their discourse.

The analyses of the citation and modal verb data indicate that PhD theses in these two departments vary substantially, in their use of the research literature, and in the ways that writers position themselves in relation to their texts, their findings and to their audience.

Pedagogical implications of the research are discussed, and the value of 'genre' as a text category is discussed.

Acknowledgements

My first thanks go to my supervisors Clare Furneaux and Pauline Robinson, who have given me timely support and guidance through the last five years. They have always been ready to listen, to challenge my ideas, suggest new ways forward, and have offered masses of positive encouragement. Thanks to this I have been able to develop both professionally and intellectually, and explore possibilities I never dreamt of before I started out. Many thanks also to past and present colleagues at the Centre for Applied Language Studies, and latterly within the School of Linguistics and Applied Language Studies, who have always supported me, and have been a pleasure to work with.

I would also like to express my gratitude to all the members of staff, both academic and administrative, in the Departments of Agricultural Botany and Agricultural and Food Economics, who have been kind enough to offer me their time and have been always immensely helpful. Thanks also to the doctoral graduates who allowed me to use their theses. I hope they don't feel that I have misrepresented them in any way!

Last but not least, my thanks to my family: my wife, Naoko, and daughter, Emily, for their patience, during those dull weekends when I ignored them and sat for hours in front of the computer. And to my mother, for all her help through the years.

Table of Contents

List of Appendices

List of Tables

List of Figures

Chapter 1: Introduction

1.1 The rhetorical situation

In the introduction to her PhD thesis, a genre-analytic comparison of research article introductions in Spanish and English, Sally Burgess wrote of the task facing her in writing her own introductory chapter:

> One cannot spend several years examining how others go about 'creating a space' for their research, to employ Swales' (1990) terminology, without facing the same challenge in a state of abnormally heightened awareness of the nature of the task and the strategies available' (Burgess 1997: 1)

In writing my own PhD thesis, I find myself in a similar situation, yet on a broader scale. For the last five years, I have collected, read, reformatted, and then tagged, analyzed, given presentations and written about, PhD theses, in a range of subject areas. Through these experiences, I have learned much about the ways that PhD theses can be organized, the multiplicity of tasks that are performed in the creation of such large, substantial texts, and the thousands of decisions that have to be made about how best to stage, and how best to frame one's arguments.

Now it is my turn to write a thesis. This increased knowledge of the range of options available to the doctoral student writer may perhaps be an advantage, but at the same time it makes me fully aware of the enormity of the task that lies before me.

The doctoral candidate's range of options derives from knowledge of the language system, knowledge of the subject area, and the knowledge of the rhetorical conventions of one's discipline, as learned from writing in the various genres, and from the reading of those, and other, genres of text in one's disciplinary communities, and from the process of drafting, reviewing and revising one's writing in the light of comments from supervisors, teachers, peers, and other audiences. The range of options that is available to me are perhaps greater than is usual, as a result of the months of study of rhetorical practices in different disciplines, the results of which are reported and discussed in the present text.

The rhetorical situation prompts a number of questions: for example, in the introductory chapter, should I start off by indicating how my study relates to real-world concerns, or should I place my work within a context of theoretical dispute? Should I summarize what is known about the subject of my enquiry in general, or should I leave that until the second chapter? Should I introduce my research questions, state the objectives of my study, and provide an overview of the whole thesis?

In addition to these organizational questions, there are also stylistic issues to address. For example, should I aim for what has been termed 'author-evacuated' prose (Geertz 1983) – writing in which the author's presence in the text is de-focussed by the avoidance of the use of the first person singular pronoun – or can I make occasional use of the personal voice? How much use is acceptable? Many writing guidelines for academic writers[1] counsel against the use of the first person singular pronoun, arguing that academic writing should project an image of objectivity. And yet we can find countless examples of successful incorporation of 'I' into academic writing, such as in the thesis quoted above (Burgess 1997), and in many articles in recent issues of respected journals. Inspection of a recent issue of the *Applied Linguistics* journal, for example, reveals that only one article (Littlemore 2001) out of six does not use the first person pronoun to introduce the personal voice at all. Within the corpus of theses that are studied in the following chapters there are examples of science writers using a personal tone in certain contexts, and also, in the interviews with supervisors reported in Chapter 4, we will find one science supervisor who recommends students to use 'I' for certain purposes.

The injunction to avoid the use of 'I' is based on perceived convention, and convention plays a major role in determining what is acceptable and what is not. The writer's options for what to say in a given context, and how to say it, are constrained by conventions. At the same time, the writer can choose whether to follow or to ignore the

[1] For example, online guidelines for academic writers at Canterbury Christ Church College, UK, advise: 'In academic writing try to avoid words such as: "personally", "actually", "to be honest", "frankly" etc. Omitting such words and expressions will make your writing more objective. You should also avoid writing in personal terms such as: "I/my/me", "you/your", "we/our/us". Instead you should consider writing in the 'passive voice'. (http://www.cant.ac.uk/sssu/guides/guide21.htm [Viewed 18/9/01])'.

conventions, or even to subvert them. Malcolm Ashmore's PhD thesis in the sociology of knowledge, which was latterly published as 'The Reflexive Thesis' (Ashmore 1989) is a classic example of a text that deliberately (and effectively) subverts conventions to achieve its own rhetorical ends. The opening chapter, for example, is written in the form of a transcript of a lecture in which the writer is the lecturer, and the inquisitive student at the back of the classroom, who bombards the lecturer with questions is his advisor, and an expert in the field, Michael Mulkay. The decision not to follow a convention must, though, be guided by a judgement on the possible effects of this, particularly in the case of assessed academic writing where the relationship between writer and reader is an unequal one – two or more of the readers, after all, have the power to pass or fail the text. Ultimately the choices made will depend upon the rhetorical *purpose* of the writer. The writer's purposes, and the linguistic forms and rhetorical organization chosen to accomplish these purposes, are a major theme of this present thesis.

One feature of PhD theses that Swales (1990) suggests is highly characteristic of such texts is the relatively high use of metadiscourse (language about language – 'the complex of devices used by authors to comment on the texts they write and keep readers informed of the rationale of composition' (1990: 24)), in order to guide the reader through the text and the development of the argument. The forms of metadiscourse that are most characteristic are *preview* and *review* statements (as they are termed by Crismore and Farnsworth 1990), where the writer informs the reader of what is to appear later in the text, or summarizes what appeared in an earlier point of the text. One reason for the prevalence of preview and review statements in a thesis is that theses are usually very long texts (in comparison to, say, a research article), but there may be other reasons, such as that proposed by one of the supervisors interviewed (see Chapter 4): the readership for the thesis includes the examiners who are perhaps not as expert in the subject of the research as the writer is, and the writer should make the structure of the argument and the findings of the research as explicit as possible, to help the examiners, rather than to perplex them. The motivation for the use of metadiscourse in this case can be seen as highly strategic, and we see once again that an explanation for a particular feature of the genre can be given in terms of rhetorical purpose.

1.2 Theme of the thesis

This study is an examination of language and rhetorical organization in PhD theses in two departments at the University of Reading. The uses of the **core modal verbs** and of **citations** are quantified and then analysed in the complete texts, and also in component parts, of sixteen PhD theses, in order to explore the extent of variation in use of these features both within a single text and between texts, for a better understanding of the nature of the texts that are called PhD theses. The analysis of these features is primarily **functional**, in the sense that the aim of the analysis is to identify the rhetorical functions that these features help to perform.

The idea for this study grew from a question that came to mind when I was looking at texts included in the Reading Academic Text corpus. This corpus was established at the Centre for Applied Language Studies, at the University of Reading, in 1995, as a resource for pedagogically-motivated research into the rhetorical and linguistic features of academic writing (Carne 1996). The corpus contains both research papers and PhD theses. In the design phase, it had been decided that the corpus should contain both exemplars of "expert" academic writer performances, in the shape of research articles written by academic staff at the university, and also exemplars of successful student performance. In the documentation for the corpus, the theses were classified simply as belonging to the genre, 'PhD thesis'.

On closer inspection, however, those texts classified as PhD theses seemed to be more diverse than I expected, and the question arose, 'To what extent can we say that PhD theses belong to a genre?' Academic naming conventions lead to the grouping of these texts under the heading 'PhD thesis' but what similarities and what differences exist between texts that are so classified? Walter Nash (1990: 9), in his introduction to an edited collection of studies on academic writing, poses the question, "Is 'academic writing' a genre, as, in a broad sense, 'journalism' is a genre?" Just as the status of 'academic writing' as a genre can be questioned (as indeed can journalism), so too can question marks be placed over the ascription of theses, written in different disciplinary areas, into a single category: 'genre'. PhD theses perform a social function: they are texts that provide evidence to examiners of a candidate's suitability for the award of a doctorate. A number of criteria may be laid down by university examination councils,

and there will also be received notions of what constitutes an acceptable submission, but can we assume that because these texts are categorised under the same label that they actually have much in common, beyond their social function?

Within EAP and applied linguistics, the most commonly accepted conceptualization of *genre* is that the member texts of this category perform social actions (Miller 1984/1994) and they share a set of communicative purposes (Swales 1990). If genre is understood to be social action, and the actions have meanings within particular communities, then we would expect a genre to be either particular to a community, or for a genre to develop differently within different communities, which would lead to a plethora of forms of the genre. Furthermore, while it can be seen that theses share a common social function, it is not clear as yet whether they can be said to share a set of communicative purposes. If they possess markedly different sets of communicative purposes, then this would suggest that the category of 'genre' needs further refinement, or it should be admitted that the genre embraces a wide, or even disparate, range of texts. If the exponents are so disparate, what value does the category have for description?

A further problem encountered was that of the terminology in which discussion of writing in academic settings is carried out. It is common to talk about academic or disciplinary communities, about disciplinary discourses, and about students serving an apprenticeship in these communities (see, for example, Spack 1988; Berkenkotter, Huckin et al. 1991; Dudley-Evans 1991; Currie 1993; Belcher 1994; Berkenkotter and Huckin 1995). Closer contact, however, with individuals working in two different departments within one university, and with the texts that are produced in these departments, has led me to question these terms. To what extent is it meaningful to talk about an academic community, and is there such an entity as a disciplinary community? What is a community, and what is a discipline? When we see, for example, that a particular thesis was written and submitted to the Department of Agricultural Botany, at the University of Reading (an institutional grouping), does this tell us anything about what discipline the thesis is representative of, and what community, or communities, it was addressing? Does the successful submission of a thesis mean that the writer has now been accepted into a community, and, if so, which one, or ones? When we talk of apprenticeship, and of initiation, what are the students serving an apprenticeship for,

and what are they being initiated into? Does the metaphor of apprenticeship define or does it constrain our understanding of these processes?

1.3 Motivation for the study

The primary motivation for the study is pedagogic: the aim is to achieve a richer understanding of the nature of the texts that are produced as PhD theses, in order to improve the substance and quality of tuition for non-native speaker doctoral students who are preparing to write a thesis. The content of the thesis applies most immediately to the subdivision of English for Specific Purposes (ESP) that is known as English for Academic Purposes (EAP) (Robinson 1991).

One of the characteristics of ESP work is the importance placed upon meeting the specific needs of the learners (Dudley-Evans and St John 1998: 4). This emphasis on the needs of the students has led to ESP/EAP being perceived as following a pragmatist ideology, and this pragmatism has been the subject of attack by critical theorists such as Benesch (1993) and Pennycook (1997), who claim that EAP adopts a conservative stance towards the dominant academic structures. Allison (1996), however, defends the pragmatist approach by arguing that EAP is not necessarily conformist: EAP has the potential to present learners with the means to make their own choices, rather than merely accept the choices of others. The present thesis is written from the standpoint of a pragmatist, who endorses Allison's position, and who believes in the importance of identifying what choices are available to writers in different contexts, or, when those choices cannot easily be identified, to know what questions to ask.

The pragmatism of EAP can be seen in the importance placed on the needs analysis. For short courses, such as intensive courses in English for Japanese and Korean police in advance of the 2002 World Cup, or courses in radio communications for flight control staff, the content of the syllabus is planned following a needs analysis, which aims to identify exactly what language (and other) skills the students require. In the case of mainstream EAP courses, this often is based on received notions of what is required. Pre-sessional courses offer writing tuition based on notions of what constitutes 'academic writing', both at the level of register and of genre. A formal academic register is understood to be objective, with a tendency to use the passive voice, omission of the personal pronoun, and so on (see, for example, Jordan

1992:101), and the genres of academic writing that are practised are typically those of the essay, and the research project report. The reasons for working at this level of generality are often practical: the students in the class may come from a variety of backgrounds, and have various levels of proficiency in English; they may also be planning to enter courses in different departments and possibly also at different levels. On top of this, they may need to pass exams such as the IELTS exams in order to satisfy the requirements of the department that they have applied to, and these exams have a washback effect on the language studied.

At higher levels of proficiency an EAP course may target a different level of generality, and aim to prepare students to write the kinds of texts that they will be expected to write in their own departments. This approach is highly pragmatic – the students need to learn skills for particular purposes. The major problem with this approach however is that it is often not clear *what* or *how* the students are expected to write, in their departments. The EAP teacher is not an expert in each of the subjects that her students take, and needs to draw on external sources of information about what the writing expectations are. The teacher can look at textbooks for guidance, or can conduct research by asking members of faculty, or by asking students to bring examples of the kinds of texts they are asked to write, by interviewing other students, or by trying to find statements of what is required by the department. In many cases however such information is difficult to access, and the data collected can be difficult to analyze. One lecturer's mental construct of what constitutes a good essay is very different from another (Creme and Lea 1997; Lea and Street 1998).

The problem is perhaps most pronounced in the case of PhD theses. When doctoral students come to an EAP unit for in-sessional help with their writing, the EAP instructor is in a difficult position. The instructor is unlikely to have written a thesis and therefore has little knowledge to draw on. It is often not easy to gain access to representative texts, and PhD theses are such long texts that it is difficult to know where to begin one's analysis, let alone to know which texts will act as good exemplars of the kind of thesis that any given student is going to write. If the plan is to have students look at representative texts and analyze them to develop 'genre knowledge' (Bhatia 1999), it is usually easier to ask students to bring examples of the journal

articles that they read for analysis in class, rather than such large and possibly impenetrable texts as PhD theses.

One solution for the EAP instructor, as suggested above, is to use an approach that draws on the wealth of research into writing in academic settings, such as the work that has been done on the research article (for example, Swales 1981; Gunawardena 1989; Swales 1990; Brett 1994; Burgess 1997; Holmes 1997; Thetela 1997; Gledhill 2000). Two widely-used textbooks that are based on the research done in this area are Swales and Feak (1995) and Weissberg and Buker (1990). This approach takes as its prototypical form the experimental research article which conventionally consists of the following sections: introduction, methods and materials, results, discussion and conclusion. This is referred to as the **IMRD**, or IMRAD model. Variations of the model are recognized: in some disciplines, the Results and Discussion are merged, in other disciplines there is often no Conclusion section, and in some disciplines the Methods section can appear at the end of the paper.

There are complications with this approach, however, depending on the students in the class: students from some disciplines, such as History or Anthropology, complain that they do not need to write in the IMRD format. Even where there are similarly named rhetorical sections, they may be quite different in composition, as Holmes (1997) has demonstrated in his move analysis of discussion sections in research articles in three social science disciplines (History, Politics and Sociology). On top of these problems, a question that must be addressed before making too great a use of research articles as exemplars to analyse and learn from is: Are the texts that students write for their PhD theses similar to, or different from research articles in their discipline? And, further, do the research articles that members of the discipline write use the conventional IMRD format or are there types of article other than the experimental research report? Do writers also need to learn to write case reports, survey reports, government reports, state-of the art review papers, technical articles or position papers? It may well be that such papers also are written in a discipline and influence the type of text that is written as a thesis in that subject area.

An alternative solution for the EAP teacher is to turn to the ‘guidelines’ literature on the PhD thesis. Paltridge (forthcoming) reviews nine published guides written for the

master's or PhD student. He finds that the majority of advice given relates to the processes of conducting research work, and that advice on the writing of the thesis or dissertation is limited. Phillips and Pugh (1994), for example, contains an eleven page chapter on 'The form of a thesis' which has three pages actually devoted to discussion of the form, and in which the following headings are suggested: 'introduction', 'literature survey', 'methods', 'results', discussion', and 'conclusions'. This clearly reflects an IMRD structure for the reporting of research, and is a very brief treatment of the question of structure. Of the nine books reviewed by Paltridge, Evans (1995) was exceptional in that the book contains extensive discussion of the organisation of a thesis, with nine chapters out of fourteen on this topic.

The EAP teacher can also turn to the research literature on the PhD thesis. Thompson (1999), Paltridge (forthcoming) and Ridley (2000) have written about the different organizational structures of PhD theses, based chiefly on readings of the Tables of Contents of theses written in different disciplines. There are a number of case studies of individual students developing their theses and learning to adjust to the expectations of their supervisors (James 1984; Bloor and Bloor 1991; Dudley-Evans 1991), and Dong (1998) presents a set of self-reports by non-native speaker students on their experiences of learning the conventions of dissertation writing. Hewings (1993) compares the conclusions of dissertations with those of research articles. More recently, Bunton (1998) has analysed the rhetorical moves within the Introduction and Conclusion chapters of PhD theses, and identified a different set of moves compared to those used in research articles. To date, however, the literature on dissertations and theses is relatively sparse, as Dudley-Evans (1999) observes, and there is a need for much more research. One area in which the research literature is notably lacking is in close analysis of the language and the rhetorical organization of PhD theses as complete texts. Secondly, though the work of Paltridge, Thompson and Ridley points to the different rhetorical organization structures of theses across disciplines, there is no research to date, as far as I am aware, into specific disciplinary differences of the language and organization in PhD theses. The present thesis aims to fill this gap.

1.4 The contribution of the present study

As stated in the previous section, there has been relatively little genre-analytic research into the PhD thesis, and what applied linguistic research there has been has tended to

focus on specific sections of the thesis (introductions, or conclusion sections, for example). Hanania and Akhtar (1985) looked at verb tense usage across the sections of Master's dissertations following an IMRD format, but there is no comparable study of PhD theses. Bunton (1999) studied uses of metatext in 13 PhD theses written by non-native speaker students in Hong Kong, with comparisons of density of use by chapter and between two broad disciplinary areas, sciences and humanities. Bunton's study is, however, more a study of register, as he does not attempt to tie use of metatext to rhetorical purpose, nor does he relate usage to specific rhetorical sections. As yet, there is no study of complete texts which compares theses written in a scientific disciplinary area with those in a social science, and relates these uses to communicative purposes.

This thesis attempts to deal with theses as complete texts, in order to add to our knowledge of how such large texts are organized, and to see how writers use particular linguistic features in particular ways to achieve their rhetorical purposes. One aim of the study is to assess the degrees of variation in macrostructure, and also in uses of the linguistic features that are analysed, both between the two departments, but also within the departments. Theses are large texts and this may be one reason why so little research has been done on theses as complete texts. The approach taken in this study is to utilize computer technology in order to make the texts more tractable to analysis. The texts are tagged on computer to identify particular features, and then the instances are quantified and analysed with the use of concordancing software and Excel spreadsheets. The patterning of language around particular features is also explored through the use of concordancing software. Use of the technology makes possible the processing of data on a large scale that would have been difficult to achieve before.

At the same time, it should be noted that the study attempts to work on a large scale, taking a broad horizontal sweep by quantifying particular features over large amounts of text. The quantity of data makes it impossible to analyze the data in any great depth. At one extreme, one can take a small amount of data and examine this, to continue the spatial metaphor, vertically – in depth. At the other extreme, one can take a broad view, as with Biber's (1988) studies of register variation in speech and writing, and estimate the relative frequencies of linguistic features in different kinds of texts. This thesis attempts to work at both extremes, and to mix the quantitative with the qualitative, the

horizontal with the vertical, the textual with the contextual, in what Yunick (1997:329) describes as a ‘dynamic tension’.

Apart from the immediate application of findings about theses to pedagogical practice, the thesis also makes a contribution to the debate over how texts should be categorized for particular purposes. The naming of particular texts as PhD theses has a social origin, to denote an extended written text that has been submitted for university examination purposes – but does this naming have value for the classification of texts for other purposes? Is this naming useful for pedagogical purposes, or for ‘descriptive’ purposes? Should the texts written by students in one department be considered comparable to the texts written by students in other departments? In the work by Douglas Biber and his colleagues, for example, in the compilation of the Longman Grammar of Spoken and Written English (Biber, Johansson et al. 1999), the use of language in academic writing is compared with that in journalistic texts. How valid are such comparisons, and how useful are these in descriptions of language? Such comparisons have some value at the level of broad descriptions of language register, and thus may be useful in introductory courses for academic language teaching, but it is not certain that these descriptions will be of similar value to writers at more advanced levels. The value of a particular naming and classification will depend upon the purposes to which the information is to be put, and this thesis interrogates the notion of ‘genre’ as a useful category for pedagogically-oriented research into academic writing products and practices.

At a more specific level, this thesis, in examining the frequency and functions of certain features of writing in PhD theses, proposes frameworks for the analysis of citations and of modal verbs that are suited to computer-based studies of academic text. The framework for the analysis of citations is useful in determining what citation forms are used in different rhetorical sections of a text, and which citation types are more commonly used in a particular research approach. The categories are easy to apply and can be used with students in classroom work that analyses not only the forms but also the functions of citation. Similarly, the framework for the functional categorization of modal verbs that is described at the end of Chapter 7 provides a practical model for the treatment of modal verb use in an advanced academic writing course, particularly at Master’s or PhD level. Discussion of these points will be made in Chapter 8.

Finally, this thesis provides information on the rhetorical organisation of PhD theses in different disciplinary areas that is based not only on information drawn from chapter headings but also from analysis of language in different sections of a number of theses. These models of organization may prove useful in academic writing courses and they should also provide starting points for developing frameworks for further applied linguistic investigation of PhD theses.

1.5 Overview of thesis

Chapter 2 presents a discussion of terminology; the terms 'discipline', 'genre', 'discourse community' and 'communicative purpose' are examined in turn and problematized. I stress the importance of viewing these concepts as plural and dynamic. The review of approaches to genre analysis points to the necessity of acquiring information about the contexts in which the texts to be studied were produced, which, in turn, argues for the combination of qualitative methods of research with the quantitative. The chapter continues with a review of the applied linguistic research literature on disciplinary discourses and concludes with a brief review of contrastive rhetoric research on academic writing, to illustrate some of the problems that international students have with writing in English for academic purposes.

In Chapter 3, I propose a framework for the conceptualization of PhD thesis writing within different layers of context. Intertextuality and modality are identified as features of the texts that can productively be examined. On the basis of this discussion, I select citations and modal verbs as the subjects for analysis in the study.

Chapter 4 describes the principles and the procedures for the development of the corpus of theses that was established for this study. The assessment of the representativeness of the corpus and of the names used for classification raise questions about the terms in which academic writing is discussed.

In Chapter 5 the contexts of production of the theses are investigated through a set of interviews with supervisors of the theses. Following this, questions regarding the nature of a discipline, and of the PhD thesis are discussed further, with particular reference to the two departments studied, and with reference to the literature. Finally, a working

model of the patterns of rhetorical organization is developed, on the basis of the information gathered in the interviews, and of relevant studies. This model will provide a framework in which the analyses of the following chapters can be placed, and will be adapted in the light of those analyses.

Chapters 6 and 7 present the two quantitative empirical studies, of the uses of citations and of modal verbs. In Chapter 6, a new framework for the categorisation of citation types is developed, and in Chapter 7 a functional account of the various uses of the core modal verbs is proposed. The variation of practice and of rhetorical purpose between writers in the two departments and also between writers within the same department is the focus of these two chapters, and from these analyses a more detailed picture of variation in language use and of rhetorical organization in PhD theses in different disciplines emerges.

Chapter 8 considers the pedagogical and research implications of the findings of the analytical sections, and draws the thesis to a conclusion with an overall evaluation and a set of recommendations for further research.

Chapter 2 Review of the literature

2.1 Overview of the chapter

In this chapter, the research literature on aspects of writing in the disciplines, on genre and the analysis of genre, is reviewed. The review is not comprehensive, as it selects the aspects that are of either direct or possible relevance to the study. Firstly, terms and concepts are discussed, and relevant work in these areas is surveyed. The key terms examined are 'discipline', 'genre', 'discourse community' and 'communicative purpose'. Research into disciplinary discourses, and approaches to genre analysis are then reviewed.

The final part of the chapter reviews recent work in contrastive ESP rhetoric, which indicates that learning to write in the modes conventional to British academia may well be an extremely difficult task for many international students. The purpose of this section is to create a context for discussion, in Chapter 8, of the pedagogical implications of the findings of the study.

2.2 Terms and concepts

As reported in Chapter 1, Nash (1990:9) questions whether 'academic writing' can be called a 'genre'. The term 'academic writing' is used to describe a wide range of different types of text, ranging from undergraduate essays produced under timed examination conditions, to laboratory reports, and, further, to dissertations and theses. The products are highly diverse, and the resemblances are in some cases almost impossible to find ('like chalk and cheese', ibid: 8). Undergraduate essays in different fields of enquiry, and at different stages of an undergraduate course, can take a multitude of forms and are assessed according to widely differing criteria; Crème and Lea (1997: 29-32) report, from their own research, some criteria for what might constitute a good essay in different disciplines:

- using the personal to illustrate points (Social Anthropology)
- correct use of technical language and terminology (Biology)
- information processing skills (Law)
- location in historical context (History)

- use of good argument, premise, reasoning and conclusion (Politics)
- use of primary texts (English)
- translation of the highly academic for the lay person's understanding (Management Science)

Rather than treat academic writing as a unitary concept, a more productive line of enquiry is to look at the different kinds of writing expected of students in different disciplines. To do this, however, we need to be able to define what a discipline is, or, rather, what our present understanding of 'discipline' is to be.

2.2.1 Discipline

On the surface it may seem a simple task to define a discipline. According to the OED Online[2], a discipline is 'a branch of instruction or education; a department of learning or knowledge; a science or art in its educational aspect'. It is reasonably easy to discern the content and perhaps also the boundaries of core, established subjects such as Physics and Economics. Problems arise however with the less 'core' members of disciplinary groupings. If we extrapolate from the second part of this definition ('a department of learning or knowledge'), to interpret 'department' in an institutional sense, we could ask whether institutional groupings represent disciplinary divisions. Is the fact that there is a department of Agricultural and Food Economics at the University of Reading sufficient grounds to identify this as a discipline? The answer to this is 'No', as institutional decisions over which members of staff should be grouped together are not always driven by taxonomic concerns (based on the classification of the branches of knowledge), but may also be driven by marketing decisions, by political interests, or by administrative concerns. An institutional grouping, therefore, does not of itself constitute a discipline.

Becher (1989) is a major study of the cultures and epistemologies of different disciplines. In his discussion of what constitutes a discipline, he admits the difficulty of defining exactly the notion of a discipline, and where the boundaries lie between one discipline and another. One approach that can be taken is to specify a number of requirements that groupings must satisfy before they can claim disciplinary status. King

[2] http://dictionary.oed.com/entrance.dtl

and Brownell (1966, cited in Becher) propose the following criteria: the grouping should have a network of communication, a tradition, a particular set of values and beliefs, a domain, a mode of enquiry, and a conceptual structure. These requirements appear too restricting, however; they project an image of homogeneity, and a static representation, which is unsatisfactory. When a discipline encounters a paradigm shift (Kuhn's 1962 term), for example, there will be a period of conflict within the discipline when the values and beliefs are challenged. In a changing world of concepts of knowledge, too, as Geertz (1983) has argued, there is a blurring of genres and of boundaries between domains. The claim that a discipline should have a single, central mode of enquiry, and a particular set of values and beliefs, no longer seems to be applicable (if it ever was).

A more useful (and useable) set of criteria is one that stipulates that a discipline should have an organizational structure, an international community, professional organizations and specialist journals (Becher, op cit). In the case of applied linguistics, then, the fact that there are several national associations of applied linguistics, and that there are also international journals and international conferences of applied linguistics (AAAL, AILA, for example), would qualify the subject area as a discipline. There may not be agreement within the discipline over what the unifying features of the discipline are, and there may be no central mode of enquiry, but, I would argue, there is sufficient reason to identify applied linguistics as a discipline.

This approach to the definition of a discipline implies that the discipline is not merely the branch of knowledge, in the abstract, but is also the people who have specialised expertise in the forms of knowledge and the employment of the techniques, and the modes of thinking, of that department of knowledge. The discipline is thus also realised through the activities, the culture and the artefacts of the members of the discipline.

Despite the difficulties inherent in trying to define whether a grouping constitutes a discipline or not, Becher claims that those groupings that are seen to be disciplines often follow characteristic practices. They share certain attitudes, activities and cognitive styles, as well as congregating around a domain of knowledge and structuring that domain in certain ways. In an earlier study (Becher 1987), he looked at linguistic features of disciplinary discourse. The terminology of the subject, he claims, is of no

particular interest, as this merely reflects a difference in subject content. It is the other words, the words by which members of the group position themselves and express the values and assumptions of their community, that characterise the grouping. He gives as an example the language of appraisal used in different disciplines: sociologists use epithets such as ‘well-argued’, ‘powerful’ and ‘persuasive’ to praise, whereas ‘anecdotal’ and ‘contentious’ are used for disapprobation; physicists, on the other hand, use ‘elegant’ and ‘economical’ to praise, ‘sloppy’ to criticize; historians like work that is ‘original’ or ‘scholarly’ but not ‘jargon-ridden’ or ‘tendentious’ (Becher 1987: 264).

There is a danger with a term such as ‘discipline’ that it is conceived of as describing a unitary entity. Becher (1989) stresses that disciplines are *plural* entities and his purpose in his study is to find models for mapping the territories that these different tribes (the metaphor he employs in the title to the book) inhabit, and to find distinctive characteristics in their cultures and their world-views. The approach he takes is to ask members of institutional groupings how they perceive the activities and structures of their subject area. For a conceptual framework, he draws on the work of Biglan (1973) who places disciplinary activity along three axes: pure–applied; hard–soft; life–non-life. Taking the first two axes, we can see that a particular discipline, such as physics, might be placed in the hard-pure sector, and then we would try to characterize the disciplines that are in this area. Becher proposes that hard-pure disciplines view research in terms of accretion of knowledge, whereas hard-applied subjects employ a heuristic, ‘trial-and-error’ approach to the acquisition of knowledge.

Placing a discipline on these axes does not seem a wholly satisfactory solution to the problem of defining disciplines, however. In a broad sense, Economics might be classified as a hard, pure subject, but it must be remembered that there is diversity within the discipline. Micro-economics, for example, can employ ethnographic methods in its research, Econometrics involves use of statistics and mathematical methods, while New Institutional Economics draws on approaches in the political sciences and in constitutional law. While this does not refute the possibility that there is a central core of knowledge that members of the discipline are likely to share in common, it does create problems for outsiders who try to understand the activities of the discipline. It is not valid to assume that all work performed in Economics research belongs in the hard-pure domain. Our understanding of a discipline, therefore, must

accept that there is diversity within the grouping, and that the distinction between hard and soft, pure and applied, does not exclude the possibility of activities within the discipline that may have more in common with the work and world-view of other disciplines than with the 'parent' discipline. The perception that a discipline is 'hard' and 'pure', for example, identifies this disciplinary grouping as centred around activities that are 'hard' and 'pure', but does not exclude the possibility that some areas of the discipline do not fit easily into that categorisation.

2.2.1.1 A working definition of 'discipline'

For the purposes of the present study, based chiefly on the ideas of Becher (1989), I will use the following working definition of 'discipline':

> A discipline is a branch of education or instruction. There is a core body of knowledge and set of methodologies in this field of study, that are distinctive, but the field of study will also contain a diversity of approaches and concepts, some of which have a more or less contested status within the discipline. In its human and institutional manifestations, the discipline has an organizational structure, an international community, professional organizations and specialist journals.

This definition is deliberately loose. In calling a discipline a *branch* of instruction, we employ a tree metaphor. If we view Economics as a discipline, that suggests that Applied Economics, Agricultural Economics, Econometrics, and so on, must be viewed as somehow at a lower level than the 'parent' discipline, or, if we are to continue the tree metaphor, the main branch discipline. The question then becomes: At what point does the offshoot become a branch in itself? There are international journals of Applied Economics, Agricultural Economics, and Econometrics, and there are also international communities for each. Whether or not economists would view these subject areas as disciplines is another question, and we will need to ask expert informants for their views on this, while recognizing that there may well be dispute amongst the informants too. The problem for the applied linguist, as an outsider to these groupings, is to know at what level to work. Is it more useful to compare the discourses of main branch disciplines, or is it more productive to work at a level of greater diversity? This must surely depend upon our research purposes .

2.2.2 Genre

Genre, too, is an elusive concept. As Kay (1994) observes, when trying to grasp the meaning of the term through a reading of the literature, one is soon confounded by the plethora of gradations of genre, of higher and lower order: macro-genres, micro-genres; complex genres, minimal genres; discourse genres, text genres, rhetorical genres; supra-genres, sub-genres. Where do genres begin and where do they end? A pragmatic solution to this problem is offered by Henderson and Dudley-Evans (1990:9): 'An economics textbook could either be considered as a genre in its own right or as a sub-genre depending on whether we are interested in comparing one economics textbook with another or whether we are concerned with comparing textbooks, in various disciplines with a prototypical textbook.' In other words, as in the case of 'discipline', the definition of a genre depends upon our purpose.

Genre is a term used to describe a group, or class, of texts, or interactions. Our chief interest in the present study is with written text, so we will refer in the following pages mainly to *texts*, and to *text genres*. The criteria by which texts are ascribed to a genre are not always clear, and this can lead to confusion over what degree of regularity we should expect to find in the format, rhetorical organization and register of language in exponents of a genre. Texts that are produced for transactional purposes, such as invoices, are likely to be highly conventionalized because the emphasis is on the speedy transfer of information within a commercial context, while texts that are produced for artistic purposes (such as novels, or poems) may be relatively divergent because these texts aim at originality (although it should also be pointed out that many art forms are highly formalistic).

The concept of 'genre' has had considerable influence on ESP/EAP research since it was first used in an ESP context by Tarone *et al* (1981/1988) and Swales (1981). As Swales (2001) observes:

> … a focus on genre redrew the map of academic discourse by replacing rhetorical modes such as exposition, or registral labels such as scientific language with text-types such as research article, term paper, final examinations, MA thesis and conference abstracts … (Swales 2001:47)

In the following sections, we will look at how genre has been conceptualized in the research literature, and in particular the key concepts of 'discourse community' and 'communicative purpose'.

2.2.2.1 Swales' definition of genre

Swales' (1990) definition of genre is the most influential within the field of English for Specific Purposes. Genre is, he proposes:

> A class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale of the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choices of content and style. Communicative purpose is both a privileged criterion and one that 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. If all high probability expectations are realized, the exemplar will be viewed as prototypical by the parent discourse community. The genre names inherited and produced by discourse communities and imported by others constitute valuable ethnographic communication, but typically need further validation. (Swales 1990: 58).

This is a much-cited definition and does not require extensive discussion here. The concepts and propositions that I wish to draw attention to, and problematize, are those of the *discourse community*, and the postulate that *communicative purpose* is a privileged criterion in the identification of a genre. In the following sections I will examine each of these in turn, before returning to the discussion of genre.

2.2.3 Discourse community

'Discourse community' is a concept which has achieved considerable currency in the research literature on genre analysis. It is a problematic concept, however, as it is not clear exactly what the community (or communities?) consists of, and where the boundaries lie.

Swales (1990) proposed the following six defining characteristics of a discourse community:

1. A discourse community has a broadly agreed set of common public goals.

2. A discourse community has mechanisms of intercommunication among its members.
3. A discourse community uses its participatory mechanisms primarily to provide information and feedback.
4. A discourse community utilizes and hence possesses one or more genres in the communicative furtherance of its aims.
5. In addition to owning genres, a discourse community has acquired some specialist lexis.
6. A discourse community has a threshold level of members with a suitable degree of relevant content and discoursal expertise. (Swales 1990:24-27)

As an example of a discourse community, Swales described the activities and the features of an association that he belonged to, the Hong Kong Study Circle, a group that aims to foster interest in the postage stamps of Hong Kong. The example works because it is a special interest group, with a clearly defined membership, but it is more difficult, as Swales himself has admitted (Swales 1998:21), to apply the criteria to other, less clearly defined groupings. Despite the fact that this definition of a discourse community is a much-cited one, and has often been used in studies of discourse practices in academic settings, it is better adapted to the description of clubs and societies than to less formalized groupings.

The concept of a discourse community which is built around ‘membership’, ‘a broadly agreed set of common public goals’ and ‘specialist lexis’ is not easily applied to the university community as a whole, for several reasons. Firstly, it is not certain that all students, especially those at undergraduate level, can be said to aspire to participation in an *academic* community; their motivation may be more instrumental than integrative. Many students take degrees in order to further their careers in industry, for example, and have no intention of remaining in an academic community. The advantages of taking a course in such cases may be to receive higher training, or merely to get a qualification that will enhance chances for promotion. Others may study for a doctorate for intellectual stimulation and to pursue their interests in a particular subject. In this case, their desire may be to engage *with*, but not necessarily to engage *in*, a community. There are different degrees of engagement.

Secondly, there are increasingly sharp divisions in orientation between disciplines (Barnett 1994, for example, contains a set of papers by British educationists addressing the crisis of fragmentation in academia) and it is difficult to see the various disparate groupings that constitute 'academia' as forming a single community, unless community is held to be a highly diffuse grouping. Thirdly, the nature of the community is uncertain, too. Is it uniform, or diverse? Does it promote conformity through strict rules or does it allow for difference? The tendency to regard community as a force for conformity has been challenged by those who argue for a plural concept in which diversity is respected (for example, Rafoth 1990; Clark 1994).

The concept of a discourse community may apply more readily to a disciplinary community, a grouping of people that have an area of study and activity in common. Ramanathan and Kaplan (2000:176-180) elaborate a convincing illustration of how Swales' six characteristics can be seen to apply in the case of the TESOL community, and yet there is an element of messiness in the way that they equate TESOL with the interests of the AAAL (American Association of Applied Linguistics) community. The TESOL group may be one of the interest groups in American Applied Linguistics, but it is not the only one. As seen in the previous section, it is difficult to identify the boundaries of a discipline.

Another question to be posed is: what does a community consist of? Miller (1994) distinguishes between two metaphorical kinds of community: one, a *taxonomical* community, in which the members have similarities which the classifier attributes to them (eg, UK passport holders); the second is a *relational* collective, in which members have a real relation to each other, such as in a school, where there are teachers, administrative staff and pupils. Miller sees the former as demographic, the latter as material. She proposes a third type – the 'virtual community': the community as invoked, as presupposed in rhetorical discourse. The virtual community, it is argued, is something that is internal to the genre, a construct. Once again, this community should be seen to be plural.

Myers (1989) distinguishes two levels of audience for an article: one, *endophoric* and the other, *exophoric*. The former is the inner circle of researchers in the same specialist area, who understand the concepts and the techniques and whose sensibilities as peers,

but also as competitors, need to be considered; the latter is the wider audience. In taking the expectations and requirements of an audience into consideration, a writer is likely to have different audiences in mind during the process of writing, and may therefore construct not a single audience, but a plurality of audiences in the text.

Communities create discourse, and discourse creates communities (Devitt 1993). This idea derives from the structuration theory of Giddens (1984), a theory that has been highly influential in recent formulations of genre theory. Giddens posits that individuals both constitute and instantiate structures. This offers a view of genre as both constraint upon the individual and also as a means for expression of the individual. Similarly, genres embody the conventions of communities, but also in a sense can be said to constitute the communities.

Swales (1993) redrew his ideas of 'discourse community' in the light both of Miller's article,[3] and of Giddens's structuration theory. In his revised definition, Swales envisaged 'discourse community' as existing only through people's engagement in the conventions of the community. This dynamic view is more satisfactory, as it allows for the change over time that has been described in diachronic studies, such as those by Bazerman (1984), Dudley-Evans and Henderson (1990) and Bhatia (1993).

More recently, however, Swales (1998:22) has reviewed his 1993 statements about discourse community and has criticized these views as reductionist, on the grounds that they imply that discourse communities only exist through texts. Swales now perceives two types of discourse community:

- a Place Discourse Community (PDC) which is a group of people who regularly work together, have developed a set of genres for regulation of the roles that each has to play within the community, and has a set of traditions and a sense of its own history;
- the other type, a Focus Discourse Community (FDC) is a grouping of people that are joined by a shared focus of interest (this could be a disciplinary community, or a professional association).

[3] This was originally presented as a paper at the 'Re-thinking Genre' conference in Ottawa, 1992

While this distinction allows the ethnographer the means to conceive of communities as physical entities again, for the text analyst it is useful to see the discourse community on two levels: the *physical* and the *textual* . People participate in groups, and develop an identity within the grouping, both through social interactions and through sharing in the community culture. This is the physical aspect of the community, and can be understood to be similar to the PDC. Because this is a *discourse* community, a community that shares the same discourse world, it can be said also that conversely the community culture is expressed through the texts, or genres, which constitute the community, and that the community is recreated through its texts. The same holds true of the FDC, which usually has less of a physical existence (at meetings such as seminars, conferences), and which also communicates through various genres. To understand the community we must study its texts. At the same time, though, we cannot understand the texts without knowledge of context, which returns us to study of the physical/cultural. The two approaches should be seen therefore as complementary, rather than exclusive, and will involve the analyst in what Johnson (2001) has termed 'complementary cyclical research'.

Bex (1996) has talked of different types of community that he belongs to, and he productively describes community as consisting of a series of circles of differing radii, and exerting differing degrees of attraction. In terms of the university environment in which he works, he acknowledges close allegiance to the ring of colleagues that he is in immediate contact with, while admitting to little sense of identity with members of other departments, unless an outside threat, such as cuts in funding, unites all in a common cause. The concept of discourse community that emerges, therefore, is of a plurality of circles of relationship, of differing strengths of attraction.

2.2.3.1 Working definition of 'discourse community'

The term 'discourse community' is defined here strictly for the purposes of this study, drawing on the definitions proposed by Miller (1994) and Swales (1998). It is taken to be a loose term used to describe either the actual physical community that shares the same disciplinary or research interests, or the audience within a text.

- In the former case, the members of the community range from the people that work in the same institutional grouping (to which the writer will have varying

degrees of closeness of interest, and allegiance), to people in other locations who have similar research interests or concerns.

- The audience within the text is a virtual community, as constructed through the text. The construction of this community is constrained by genre expectations within the physical community, and the writer must construct this virtual community to the satisfaction of representatives of the expert membership of the disciplinary community – in the case of the PhD thesis, the examiners and the supervisor.

The two parts of the physical community can be equated with Swales's 'Place Discourse Community' (the people that work in the same institutional grouping) and 'Focus Discourse Community' (people in other locations with similar research interests or concerns).

2.2.4 Communicative purpose

'A genre comprises a class of communicative events, the members of which share some set of communicative purposes.' (Swales 1990: 58). A clear justification for this criterion is given in Swales (ibid) and Bex (1996): two texts (one, for example, a critical article and the other, a parody of a critical article, such as the famous 1996 spoof article by Allan Sokal that was published in the journal *Social Text*), cannot be distinguished by presentation or language alone, but they can be distinguished by their communicative purposes, which will normally be recognized by expert members of the community (though not by the unwitting editors of *Social Text*, who failed to spot the parody!). The research article introduces, justifies and reports research, whereas the parody imitates the research article to achieve some comic effect. In addition, the research article has a more prescribed set of conventional forms, while the nature of the parody genre is that it inhabits the forms of other genres, like a hermit crab. Following this, Swales argues that letters, in general, do not form a genre, as they may have similarity in form, but they do not have a communicative purpose in common. Letters of complaint, on the other hand, form a genre as they have the purpose of complaining in common.

However, Askerhave and Swales (2001) have recently challenged the notion that communicative purpose can be granted a privileged status in the identification of text genres. They problematize the notion of communicative purpose by arguing that

communicative purposes in any given communicative event are often complex, and it is dangerously reductive to simplify purpose to a single aim. Furthermore, they argue that communicative purpose in many texts is not easily determined, although the text type itself may be recognizable. They therefore propose that communicative purpose may be provisionally ascribed in the early stages of analyzing a genre, but that the purposes of the texts must be identified through 'extensive text-in-context enquiry' (Askerhave and Swales 2001: 209), after prolonged fieldwork. In the case of the PhD thesis, this implies that we might ascribe particular purposes to the thesis in the preliminary stages, but these will have to be tested through analysis of the texts and the contexts in which the texts were produced.

2.2.4.1 Perspective on communicative purpose for this study

In the present study, it will be assumed that communicative purpose is not a privileged criterion. Instead, it is the accepted social function of the text (in the case of the thesis, a text that provides evidence to examiners of a candidate's suitability for the award of a doctorate) that determines the genre, and the identification of communicative purpose is a secondary stage that may inform our understanding of why there is variation within the genre.

2.3 Genre and genre analysis

2.3.1 Genre: text, context and structure

Paltridge (1997) dates the use of genre as a classificatory term back to the *Poetics*, in which Aristotle divided literary works into generic groups, such as 'poetry' and 'drama'. The fragments of the *Poetics* that remain suggest that Aristotle aimed to establish exclusive categories for the different types of poetry, and also to define each category by reference to the finest example of that category, on formal and thematic criteria. During the twentieth century, a number of researchers in other disciplinary fields have individually adopted a *functional*, rather than a formal, approach to the study of genre, arguing that texts cannot be understood in isolation from the contexts of their production. Among these was the anthropologist, Bronislaw Malinowski; he contended that the folktale, as a genre, is more than simply a text and that the meaning of the text cannot be grasped without reference to the meaning it holds for the particular discourse community to whom it was addressed (Malinowski 1960).

From Malinowski come two terms that are widely used in discussion of genre theory, particularly in the approach taken towards genre by the Australian systemic functional school: the 'context of situation' and the 'context of culture'. The term 'context of situation' is used in Malinowski (1946/1923) to refer to the cultural and affective elements of context that spoken language is situated in; 'context of culture' appears in Malinowski (1935). It is not immediately clear from a reading of the texts how Malinowski himself differentiated between the two terms, but in systemic functional studies, the former can be taken to refer to the immediate physical, institution-bound contexts in which texts are created, while the latter refers to the wider sociocultural contexts (Eggins 1994). The terms provide valuable means for distinguishing between two layers of context, and I will use them in the framework that I develop in Chapter 3.

2.3.2 Genre analysis and register analysis

Lee (2001) discusses the terms 'genre' and 'register' in relation to corpus text categories. He argues that the two terms describe different points of view on text: register is concerned with linguistic patterns that are commonly associated with situations, and genre is used when we talk about texts belonging to culturally recognisable text categories. Texts in a genre are thus instantiations of a register (or registers) and may 'vary with the times, with fashion, and with ideological movements in society' (ibid: 11). The genre labels tend to remain the same, however, as they are descriptors of socially constituted categories of text. He cites Couture (1986:82):

> While registers impose explicitness constraints at the level of vocabulary and syntax, genres impose additional explicitness constraints at the discourse level … Both literary critics and rhetoricians traditionally associate genre with a complete, unified textual structure. Unlike register, genre can only be realized in completed texts or texts that can be projected as complete, for a genre does more than specify kinds of codes extant in a group of related texts; it specifies conditions for beginning, continuing, and ending a text.

In this view, then, when we talk about genres, we talk about whole texts, and therefore about text structures. This accords with Swales' assertion that 'genres have beginnings, middles and ends of various kinds' which consequently 'leads to an analysis of discourse structure' (Swales 1990: 41).

Swales (ibid) has observed the movement in EAP research away from register analysis to genre analysis, or, as Yunick (1997) puts it, away from correlational analyses towards 'thick descriptions'[4]. One of the main proponents of an approach to texts that includes ethnographic research is Vijay Bhatia, who claimed that register analyses 'reveal very little about the true nature of genres and about the way social purposes are accomplished in and through them in settings in which they are used' (Bhatia 1993:18). Yunick (op cit) however posits that the two approaches should not be placed in exclusive opposition but that researchers should view the opposition as a dynamic, productive tension. The tensions are between the search for general properties and the investigation of situated practices, between generalization and contextualization, and between quantitative and qualitative analyses. A potential danger in following a predominantly ethnographic approach, for example, is that the researcher relies too heavily on insider definitions of particular contexts and of what language features are most salient:

> Quantitative, correlational work serves two functions alongside ethnographic work to identify not only phenomena general to many genres, but also significant pattern of meaning making that might not emerge from ethnography alone (Yunick 1997:326)

Yunick suggests that a broad quantitative approach which identifies recurrent language patterning across texts within a genre and also between genres can indicate to what extent certain phenomena are genre-specific and to what extent they are more widespread. This quantitative approach can be combined with qualitative methods of data analysis to produce a rich description of situated textual practices.

2.3.3 Three 'schools' of genre analysis

Hyon (1996) distinguishes three major schools of genre analysis within applied linguistics:

- an **ESP** school,
- a North American school of **New Rhetoric**, and

[4] The term 'thick description' was coined by the philosopher Gilbert Ryle to characterize a description of an event that takes into account the contextual factors that influence individual participants' behaviour. The term was adopted by Geertz (1973) to refer to a particular approach to anthropological research work.

- an **Australian** school, strongly influenced by Hallidayan systemic-functional linguistics.

In Hyon's account, the ESP school views genre as text types that are defined by formal properties and communicative purpose, in social contexts; the New Rhetoric School focuses on social purposes and actions, and uses ethnographic methods to explore the situational context; the Australian school applies Hallidayan analytical frameworks to texts, and centres attention on linguistic features of texts. Where the Australian and ESP schools tend to focus on linguistic and rhetorical move analyses, according to Hyon, the North American school is more concerned with the roles of genres in institutions. Yunick (op cit), building on this distinction, observes that while the Australian genre theorists draw strongly on the Hallidayan school of systemic-functional linguistics, and the North American New Rhetoric movement is multi-disciplinary, employing anthropological, social and literary/rhetoric theories, the approach of writers such as Swales, Bhatia and Dudley-Evans in ESP studies has tended to be eclectic, synthesizing a variety of models rather than privileging a single linguistic model.

For the purposes of this research, I am concerned with genre in an ESP/EAP context, and thus identify more strongly with the perspective of this school. It should be recognized, however, that there has been an increasing degree of exchange, or influence, between the schools. Researchers in the North American field of rhetoric studies, for example, are beginning to address questions concerning ideology and education (Freedman and Medway 1994), influenced by the Australian school, and have used linguistic analysis to explore rhetoric (for example, Vande Kopple 1985; Crismore and Farnsworth 1990; MacDonald 1992). ESP/EAP, in turn, has been greatly influenced by North American research. Swales (1993) acknowledges the influence of Miller (1984/1994), whose conception of genre as social action helped to move attention away from the creation of taxonomies, both in ESP and in Australian studies. Bazerman's studies of laboratory practice, in conjunction with the studies conducted by sociologists of scientific knowledge (for example, Latour and Woolgar 1979; Knorr-Cetina 1981) introduced a social constructionist perspective, which perceives knowledge as socially manufactured. The effect of these studies was to focus attention on the contexts in which texts are produced, as well as the texts themselves.

Within the Australian school, two approaches to the location of texts within social practices can be detected (Cope and Kalantzis 1993): Martin arrives at social contexts through study of the text[5] , while Kress takes the opposite approach, first examining the wider social context and the meanings that are attached to the construction of the texts, before beginning the linguistic analysis (e.g., Kress 1993). As Swales (1993) remarks, however, in the Australian systemic linguistics school, reference is often made to the 'context of situation' and 'context of culture', and yet little discussion of these factors is made in the actual analysis of text.

A highly influential schema for locating texts within social practices in the ESP/EAP tradition is that of Bhatia (1993). He works through the following steps:

- placing the text-genre in a situational context
- surveying the literature
- refining the contextual analysis
- selecting an appropriate corpus
- studying the institutional context
- selecting the levels of analysis (lexico-grammatical, move structure)
- checking results with specialist informants

His approach, like that of Kress, works from the social context to the text, in order to situate the texts within the contexts of situation and culture, and provide explanatory force to the analyses. This procedure, in an adapted form, informs the methodology used in the present study.

2.3.4 Genre as conventionalized social action

A fundamental element of the current conceptualization of genre is that it is about social action; genres are 'how things get done, when language is used to accomplish them' (Martin 1985:250). A genre, following Miller (1984/1994), develops from the repeated performance of similar communicative events. As the events are repeated, conventions become established. This conventionalization serves at least two purposes:

- to regulate social interaction, and

[5] Martin (1993) attempts generalizations about the research practices and world views of historians and scientists based on his study of grammatical metaphor in history and science textbooks.

- to simplify the communicative event by setting up *expectations* of how the event will proceed, and by providing the actors with ready-made forms.

In the case of activities such as the advancement of new knowledge claims within the physical sciences the conventionalization of the communication of these claims assists in the speed at which knowledge can be manufactured, as Bazerman (1988) has observed.

'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.' (Berkenkotter and Huckin 1993:4) The conventions of a genre are not fixed, but change over time, and they also have varying degrees of rigidity. Bhatia's (1992) study of Indian and British law texts revealed that, while the latter showed some signs of evolution, the former was rigidly conventionalized and had changed little over the decades. On the other hand, Dudley-Evans and Henderson's (1990) historical study of Economics research articles over the last 90 years, and Bazerman's (1984) study of spectroscopy articles, also over a period of 90 years, revealed that these genres have altered considerably over the years as the readership, the research practices, and the uses of the knowledge changed.

Genres are thus based in the societies that develop them and need to be learned by individuals who aspire to be accepted into the society, or at least to the extent that they can engage with the community. Furthermore, for different genres there are different degrees of tolerance of variation within the genre. For the novice, learning a new genre involves learning what it is acceptable to say, and which rhetorical options it is possible to consider in the circumstances. Freadman (1994) uses the metaphor of the rules of etiquette to describe the rules of a genre, implying that the rules are not immutable, and that they depend on social norms. The norms derive from social interactions at a range of levels – the general culture (for example, the scientific community at large), the local culture (the university department, and at another level, the laboratory, in the case of a science discipline), and the interplay between different levels (see the discussion of 'discourse community' above).

An alternative view of genre conventions is that they can be seen to perform a gatekeeping role. If writers fail to follow the conventions satisfactorily, they will not be

admitted through the gate, as it were. Berkenkotter *et al*'s (1991) case study, following a student writer through the first year and a half of a doctoral programme illustrates the importance of learning to write in a disciplinary voice, using the conventions of that discipline, such as the use of citations to create a rich intertextuality. Texts that fail to meet the genre expectations of the grade awarders will be unsuccessful. The term 'gatekeeping' itself is interesting and open to different interpretations – expert members of the community (or at least of the institution) have the power to decide who will be able to pass through the gates, but a relevant question to ask is: pass through the gates to get where? One answer might be to say that the successful candidate is then admitted into the academy; this is a metaphor which has strong currency, judging by the 'apprenticeship into the academic community' rhetoric. A different interpretation of 'gatekeeping' is to see it in terms of allowing students to move from one stage of their life, to another. Rather than talk in terms of 'initiation into an academic community', therefore, it is preferable to include as part of one's investigation of an academic genre a questioning of the different functions that it can perform within the community, and also beyond.

2.3.5 Interrelatedness of form and content

One of the principles underlying the genre framework of Berkenkotter and Huckin (1995) is that genre knowledge 'embraces both form and content, including what content is appropriate to a particular purpose in a particular situation at a particular point in time.' (Berkenkotter and Huckin 1995:13) Form is an important feature in ESP/EAP analysis of texts produced within the genre, as these forms indicate something of the conventions of the genre, and of the 'beginnings, middles and ends'.

From one point of view, it is difficult to see how content can distinguish genre. The argument would be: each letter of complaint must presumably be dealing with a different cause for complaint, and it is difficult to imagine any generic feature pertaining to content. Coe (1987, 1994), however, argues powerfully against the separation of content from form. He rejects the view that form is a fixed mould for content to be poured into, arguing instead that form fits the content and the rhetorical purpose, and that there can be no meaning without form – 'information is made by putting data *in formation*, by forming' (Coe 1987:16).

In the case of pedagogy, this implies that students should be encouraged to view genres as potential forms, but not as models for writing; the analysis of exemplars of a genre should be considered in terms of the rhetorical choices that a writer had to make and the forms of language and organization that were available to help the writer to achieve the purpose. As can be seen from the discussion in Chapter 1, this is a view that is consonant with my own position. The point that is important in the present discussion, however, is that form and purpose should be seen to be inseparable, when viewing text from a genre perspective, if we are to accept that genre is a classificatory term used to describe groups of texts that have a broad set of purposes in common. If the rhetorical organization of texts in a given genre is found to vary noticeably from text to text, this may be attributable to the variation in content and in communicative purposes for each text.

2.3.6 Community ownership

In Swales' original (1990) definition of 'discourse community' (see above), he wrote of discourse communities possessing 'one or more genres in the communicative furtherance of its aims' (1990:26). In Swales (1998:207) he discussed further the question of which comes first, the genre or the discourse community, and left the question in the balance, claiming that in some cases the genres predate the communities and in others the communities develop new genres.

This conceptualization of discourse communities having proprietary rights, of ownership, however, seems to be misplaced, as it would seem more logical to say that discourse communities typically recast pre-existing genres to suit their own purposes, rather than that they develop their own genres.

2.3.7 Genre: working definition

Our interest here lies with written, not spoken, text and so we will confine ourselves to the consideration of what genre means in relation to written text. We will perceive 'genre' to be concerned with texts that are complete rhetorical entities, and we will expect genre analysis to reveal something about the beginnings, middles and ends of texts. 'Genre' is a socially constructed concept to describe a set of texts that are perceived to perform similar functions. Texts belonging to a genre are conventionalized, to differing degrees, in terms of sequencing, of layout, of

phraseology, and there are expectations of, and constraints on, the structure and linguistic expression of such texts. These expectations can vary from one disciplinary community to another. The forms that the texts take can also vary, depending on the range and diversity of purposes that exponents of the genre are asked to serve.

The analysis of genre aims:

- to identify the regularities of form, of rhetorical organization and of linguistic features within the genre
- to relate these regularities of form, of rhetorical organization and of linguistic features to communicative purpose
- to establish which features are obligatory and which are optional, within given discourse communities
- to understand why and how the genre has developed into its present form, and what functions the genre plays within the community

Having reviewed the literature on genre, and having formed working definitions of 'discipline', 'discourse community', and 'genre' for the purposes of this research, we turn now to a view of what research has shown about variation in writing across the disciplines.

2.4 Research in writing across the disciplines

Becher (1987), as noted above, investigated the terms of appraisal used in different disciplines and proposed that the ways in which writers in a discipline position themselves, and express their values, are highly characteristic. From an applied linguistic perspective, another point of departure in research into writing across the disciplines is to contrast the rhetorical organisation of discourse. Dudley-Evans (1994), for example, examined lectures and dissertations in the fields of Highway Engineering and Plant Biology, and found that the former tend towards a situation-problem organization, in which solutions are evaluated in terms of their applicability to the specific problem, with relatively little reference to previous research; in contrast, the latter work around the IMRD (Introduction, Materials and Methods, Results, Discussion) framework that is conventional in the natural sciences, and discussion of the results places the results within the wider context of research within the field.

2.4.1 Move analyses

In the ESP/EAP tradition, the dominant approach to genre analysis has been the description and analysis of moves and steps. Dudley-Evans and St John (1998) explain the terms thus:

> A 'move' is a unit that relates both to the writer's purpose and to the content that s/he wishes to communicate. A 'step' is a lower level text unit than the move that provides a detailed perspective on the options open to the writer in setting out the moves. (Dudley-Evans and St John 1998:89)

Since the publication of Swales's move analyses of Introductions to Research Articles (Swales 1981; 1990) which were carried out chiefly on scientific articles, there have been a number of comparable studies of moves in Introductions in other subject areas, and also of the moves in other sections of the Research Article.

Brett (1994) and Holmes (1997) have looked at, respectively, Results, and Discussion sections in articles in the social sciences, comparing the moves they identified in their corpora with the move patterns identified by Hopkins and Dudley-Evans (1988) in their study of research articles and dissertations. Holmes examined move structures in History, Political Science and Sociology research articles and concluded that, while there was a broad similarity in the kinds of moves made, both within the corpus of articles he assembled, and also the articles that Hopkins and Dudley-Evans analyzed, there were noticeable disciplinary variations, such as in the tendency for the Political Science writers to cycle moves, while the History writers did not. Holmes makes the suggestion that social science research articles, unlike those in natural sciences, tend towards complexity and elaboration at the beginning of the articles rather than at the end.

Al-Ali and Holme (1999) analyzed research articles in a range of different subject areas and found considerable variation in move structures, depending on the discipline, and on the arrangement of the sections. In chemical engineering articles, for example, the papers ended with Results and no attempt was made to evaluate the results. Bloor (1999) employs Swales' (1990) distinction between fast and slow Methods sections to investigate variation in practice across disciplines. Fast texts are those that assume that their readers have expert knowledge of the subject, and of the methods used in the field,

and they do not include examples or present justifications for their choice of methods; consequently, fast texts are condensed and relatively brief. Slow texts, on the other hand, are explicit about procedures, offer examples and justify the choices made, and are thus longer texts. In her study, Bloor's fast text came from Materials Science, while the slowest text came from Applied Cognitive Psychology.

It is worth observing at this juncture that there is not often discussion in these studies about what exactly is meant by 'research article', nor to what degree the research articles in social science subjects can be treated as comparable to natural science research articles. As Holmes (ibid) notes, most of the research that has been done on 'the' research article has concentrated on texts in the natural sciences, and these articles tend to be reports of experimental research. When Holmes selected articles for his study, he used the criterion that the articles should be reports of 'original research'. Original research can take many forms, however, depending on the form of research that is undertaken. Bazerman (1995) distinguishes between experimental, interpretative and reconstructive approaches to academic enquiry. History would fit into the latter category, and it can be argued that there is little justification for comparing the rhetorical organization of history articles with those of chemical engineering as the writing tasks involved are so different. Are these different genres, or is there a genre of the 'research article' that has different prototypical forms?

Let us consider once again the quotation at the beginning of 2.2.2 above:

> An economics textbook could either be considered as a genre in its own right or as a sub-genre depending on whether we are interested in comparing one economics textbook with another or whether we are concerned with comparing textbooks, in various disciplines with a prototypical textbook. (Dudley-Evans and Henderson 1990:9)

The implication of the final part of this statement is that there *is* a prototypical textbook, and that the prototypical textbook is a singular entity. Again, here is a sentence from Swales's definition of genre: 'If all high probability expectations are realized, the exemplar will be viewed as prototypical by the parent discourse community.' (Swales 1990: 58). A possibility that has not been addressed, it would appear, is that there might be more than one prototype. There may be more than a single prototypical pattern for a particular category of text, especially, one might

hypothesize, in the case of highly complex, less common texts such as the PhD thesis. An alternative interpretation, following the logic of the Swalesian definition of genre, would be to say that if there are different prototypes, then these are different genres.

2.4.2 Grammatical features

There have also been a number of linguistic studies that have examined grammatical features of writing in different disciplines. Martin (1991) analyzed nominalization in science and history textbooks and suggested the following differences: the reasoning of science tends to be concrete, whereas that of history tends to the abstract; science *analyzes*, history *synthesizes*; in science, reasoning is expressed *between* sentences, in history *within* sentences. MacDonald (1992) explored differing attitudes towards knowledge and communication within the disciplines of psychology, history and literature through analysis of the grammatical subjects of sentences in articles in those fields. She found that psychology articles tend to foreground research methods and justifications for these methods, and to emphasize generalizable phenomena, while the literature articles focussed more directly on the phenomena with little epistemic justification. The history articles tended to occupy the middle ground.

2.4.3 Hedging

The importance of hedging in academic writing has been the subject of much research in recent years. The main researcher in this area has been Hyland who has published widely (Hyland 1994; 1996a; 1996b; 1999). Hyland argues that it is necessary to examine the use of language as rhetorically motivated within specific contexts, and that modal expressions can only be properly understood when they are viewed within their context of utterance. He then claims that in science research article writing the use of reader-oriented hedges (by this he means hedges that acknowledge the reader's right to make judgements and to engage in a dialogue) is not always motivated by a genuine sense of deference but by the norms of the scientific community – if writers are to be accepted by the community, they must act in a deferential way. The motivation here is to conform, then, but the valuable point that Hyland makes, is that the writer follows the norms of a given discourse community and that hedging is therefore culturally based.

This would imply that hedging in different disciplinary cultures can play a different role. Hyland's conclusions are drawn from his analysis of a corpus of molecular biology research articles. An indication that epistemic modality can be used in different ways is given in Simpson (1990) who analyses the uses of modality in literary-critical discourse. Simpson argues that F.R. Leavis, in his major work of literary criticism, *The Great Tradition*, uses modality in a highly rhetorical fashion, qualifying propositions that are not controversial, while strongly asserting propositions that would normally be open to contestation. The strength of assertion made in highly contentious claims makes them difficult to challenge without taking a similarly strong counter position (Hoey (1984) has noted similar rhetorical strategies at play in the writings of Chomsky).

Bloor and Bloor (1993) examine the modification of propositions in a corpus of 11 research articles from the *Economic Journal*. They take Myers' (1989) concept of hedging as a politeness strategy and evaluate the uses of such modification in the economics texts. Where scientists are perceived to hedge propositions that articulate new knowledge claims, the Economics writers sampled are found to modify *field-central* claims (claims that relate to their models or methods), but not to hedge *substantive* claims (claims about the state of the world). Bloor and Bloor (ibid) conclude that in Economics it is the development of new models that is central, rather than the production of new knowledge. In a separate study, Dudley-Evans (1993) analyzed the language used in a debate over Milton Friedman's views that was carried out in a series of public articles by eminent economists. Although the genre is different from that of the research article, and the writers are leading members of the discipline rather than student writers, it is interesting to note Dudley-Evans' comment that Friedman uses hedging as a defensive strategy, to try to lessen possible criticism of his shifting position. This suggests that hedging may play different roles in the discourses of different disciplines.

2.4.4 Disciplinary thinking

A study of differing rhetorical approaches, that points to the way that writers are 'encultured' (Joliffe and Brier 1988) into their disciplines, is Hansen (1988) who observed two researchers, one, a sociologist following a positivist paradigm of research, and the other, an anthropologist conducting an ethnographic study. Both of

the researchers were investigating black communities in the United States, but there were marked differences in the resulting texts, in the ways that the writers structured their arguments. One of Hansen's conclusions is that a novice in a discipline needs primarily to learn to *think* in the way that is 'sanctioned' by the disciplinary community (ibid: 207), and to use the forms of argumentation espoused by the disciplinary community. Williamson (1988) investigated the different functions writing fulfils within three different disciplines in the processes of learning to think in the 'sanctioned' way: students in biology were trained to write in the 'disciplinary voice' (ibid 128), chiefly reporting what they have done and observed, while students in the English field were developing personal voices through writing; as for English majors, the training of sociology students placed emphasis on the development of critical thought through writing, but also aimed to move the writing from a personal voice to a 'sociological' one.

In a study of three texts from different fields (Molecular Biology, Sociology and Literary Criticism), Bazerman (1988:46) concluded that in the texts:

> the three statements of knowledge are different things. In mediating reality, literature, audience, and self, each text seems to be making a move in a different game … the words arise out of the activity, procedures, and relationships within the community.

This implies that we cannot study the texts that are produced within different disciplinary groupings without looking also at the activities, procedures and relationships within the groupings.

Hyland (2000) is a book length treatment of disciplinary discourses, which is based on the premise that academic writing is a form of social interaction. He looks at disciplinary differences in the writing of research articles, book reviews, abstracts, textbooks and the 'scientific letter', and seeks to explain these differences through the cultures and epistemologies of the disciplines. In his analysis of variation between disciplines, Hyland focuses on citation practices, demonstrating that 'hard' and 'soft' disciplines differ in the ways that they cite the work of others, both in density of citation and in the forms of citation. In Chapter 6 I will refer to this work in detail, and compare the results of my own study of citation practices in PhD thesis writing with the results of Hyland.

2.4.5 Summary

In conclusion, then, research into writing across the disciplines has identified differences in the rhetorical structures used; in grammatical features such as thematisation and nominalization, that reveal differences in the ways that the disciplines view research and knowledge; in reasons for hedging; in acceptability of projection of a personal voice. These differences arise from the varied epistemological bases and conventions that underlie social interactions in the disciplines.

2.5 Contrastive studies of academic writing

Although I do not intend to make contrastive studies of the theses in my corpus with comparable thesis corpora in other languages, I include a short review here of what Mauranen (1993) has termed 'contrastive ESP rhetoric' studies, in order to illustrate some of the problems that international students face in learning the discourses of another academic culture and tradition. The evidence given here will be referred to again in the discussion of the pedagogical implications of the present study in Chapter 8.

Trzeciak (1996) has reviewed the problems that non-native speaker students have with academic writing in English. He considers particular problems that NNS students face under the headings of: macro-discoursal patterns; coherence and style; degrees of commitment and detachment; use and attribution of source material. These headings are used for the sub-sections that follow.

2.5.1 Macro-discoursal patterns

In terms of macro-discoursal patterns, Trzeciak takes the conclusion as an example of how different cultures hold different ideas of what the purpose of a particular rhetorical section should be. To support this observation, he cites Régent (1992: 73) who notes that French medicine texts hardly ever finish with recommendations, while it is quite common for English texts to do so. Trzeciak also considers the research on cultural differences in patterns of organization, from Kaplan's (1966) patterns of organization (which actually refers to paragraph structure rather than text development, though many later researchers have failed to note this, as Trzeciak adroitly points out) to Clyne's (1987) observation of the commonness of digressions in German academic

writing in some fields, and Kachru's (1988) demonstration of a spiral pattern of organisation in Indian English expository writing. However, the assumption that first language patterns will necessarily assert themselves over second language forms is not justified, as Trzeciak argues; rather it is likely that the two patterns will be in conflict, during the period of development of an interlanguage.

Burgess (1997) performed move analyses of Introduction sections on texts written by English and Spanish academics, with some of the latter writing in Spanish and some in English. The interesting conclusion to the study was that while there were differences in the move structures of the different groups of writers in her sample, Burgess proposed that the greatest influence on the move structures was the audiences that were being addressed.

Cmejrkova (1996) also performed move analyses, using Swales' (1990) CARS (Create a Research Space) model, of the Introduction sections of linguistics articles, in this case, articles written by Czech academics and she found that while the writers used Moves 1 (Establishing a territory) and 2 (Establishing a niche) regularly, they did not use Move 3 (Occupying the niche). Move 3 involves either outlining the purpose of the research or announcing it, then announcing principal findings and describing the structure of the article. The reason that her informants gave for not performing Move 3 was that they did not feel comfortable explaining what the rest of the paper was about, as they believed the article should be written more like a detective story. Cmejrkova identifies this as an Eastern European approach to writing. Another feature of Czech academic writing that Cmejrkova discusses is that a reader-responsible prose is the preferred option, a subject discussed in the next section.

2.5.2 Coherence and style

Hinds (1987) makes the useful distinction between 'reader-responsible' and 'writer-responsible' writing. In the former case, it is the responsibility of the reader to make sense of the text while in the latter it is the job of the writer to make the meaning of the text explicit to the reader. English writing is taken to be an example of 'writer-responsible' prose, while a language such as Japanese prefers ambiguity and vagueness. The different orientations can be detected not only through differences in rhetorical organization but also in degrees of use of metatext. Three studies which have examined

the use of metatextual features across languages are Crismore *et al* (1993) and Mauranen (1993), who both contrasted metatext in Finnish and Anglo-American writing, with Mauranen looking at the writing of economics research articles, and Valero-Garcés (1996) who repeated Mauranen's research design with Spanish and Anglo-American texts. The studies found that Anglo-American writers use more metatext, and are more concerned with guiding the reader through the text than writers of the other two nationalities. This is assumed to indicate a more writer-responsible prose and a generally more explicit textual rhetoric. Mauranen argues in her discussion of these findings that the differences are due to contrasting notions of politeness: in 'reader-responsible' cultures, it may be considered patronising to be too explicit and leave no space for the reader to make their own inferences.

Several studies have focussed on the use of connectors by L2 writers, in contrastive analysis (for example, Ventola and Mauranen 1991; Field and Yip 1992; Milton and Tsang 1993; Granger and Tyson 1996). Two studies, Granger and Tyson (1996) and Milton and Tsang (1993), examined connector usage in large corpora of L2 student writing (the International Corpus of Learner English, and the Learner's Corpus, respectively) and found that, while students did not overuse connectors in general, they did overuse or misuse certain items, and this was often influenced by L1 transfer. Common connectors in French, for example, were translated into a rough English equivalent and used without an understanding of the pragmatics of the connector.

2.5.3 Degrees of commitment and detachment

Groom (2000) has examined the problems that international students may have with indicating propositional responsibility and the ownership of ideas in their academic writing. Appropriate citation of sources, and the clear signalling of the writer's position towards the texts cited, are difficult both for novice writers (see Campbell 1990; Borg 2000) and for L2 writers. As Groom shows, a writer may fail to attribute a proposition clearly to an author and thus be held to be responsible for a proposition that the writer actually does not agree with. Alternatively, L2 writers may be too dependent on source texts and fail to achieve a distinctive voice in their own text, as noted in 2.5.4 below.

Hedging has been reported by practitioners to be a major problem for L2 writers, although there is little research into the problems that specific cultures have. Bloor and

Bloor (1991) suggest that Czech and German students are taught to write directly but there is no empirical evidence provided. Myers (1989) has explained hedging in terms of politeness and it may well be that notions of politeness, as in the case of metatext noted above, vary from culture to culture, which could lead to the wrong strategies being applied.

Looking at uses of modal verbs in a parallel corpus of English and Swedish, Aijmer (1999) found that certain modals in each language had wider ranges of sense than the apparent equivalent in the other language; the modal *kan* in Swedish, for example, has both epistemic and root modality whereas the English *can* rarely conveys an epistemic sense, except in negative and interrogative constructions. In this case, first language influence may lead to errors in usage of modal verbs.

The adoption of a depersonalized voice can also create difficulties for students from certain cultures. Miyahara (1986) analyzed the speeches that American and Japanese university students write for public speaking contests, and found that the Japanese students exemplify their points by reference to personal experience, as a rhetorical strategy for involving the audience in a shared viewpoint ('to establish a bond of feeling and harmony', ibid: 91), while the American students choose to support their ideas with facts and figures. The personal experience or anecdote, therefore may be included by a Japanese student even when writing in English, as support for an idea, where the Anglo-American preference is for objective evidence.

2.5.4 Use and attribution of source material

As for use of source material, there may be differences of concepts of originality and of ownership of ideas. In Chinese writing, it is customary to respect the ideas of past authorities, and the reproduction of their ideas is expected – in the Anglo-Saxon world, where the individual expression of ideas is valued, this is sometimes taken to constitute plagiarism (Bloch and Chi 1995). In both Chinese and in French it is sometimes considered unnecessary to give full bibliographic details for all citations (Régent 1992; Bloch and Chi 1995). Thompson and Tribble (2001) found that NNS writers tend to overuse particular types of citation, and to have limited understanding of the rhetorical purposes of different types of citation format.

2.5.5 Summary of ESP contrastive rhetoric research

Trzeciak's four headings of areas of difficulty have been used to indicate some of the problems that non-native speaker students face in writing English for Academic Purposes. The review is by no means complete, as there is a substantial literature on contrastive rhetoric. What Trzeciak rightly points out is that the perception of cultural differences should not be used to deprecate the other culture, to perceive it as deficient in some respect, but to illuminate our understanding of why students write in the ways that they do. This point will be taken up again in the discussion of pedagogical implications in Chapter 8.

2.6 Summary of chapter and conclusion

In this chapter, the relevant literature on genre, and disciplinary discourse has been reviewed. For the purpose of the present study, working definitions of 'discipline' and 'discourse community' have been proposed. Studies of disciplinary discourses have shown that students in different disciplines write in different ways and also have to learn to think in discipline-specific modes. The contrastive ESP rhetoric literature has identified a range of problems that students from other cultural backgrounds may face, such as a reluctance to be explicit, difficulties with citation, and a different understanding of the function of a particular rhetorical section, such as a research article introduction or a conclusion.

The main argument developed through this review was that the concepts of discipline, genre, discourse community and communicative purpose are complex and plural. Disciplines are groupings that contain conflict and diversity, while genre does not imply a single fixed prototypical model. Discourse communities exist at different levels of coherence and attraction, in both physical and in virtual forms, and communicative purpose in genre exemplars is complex not unitary. Furthermore, the perception that form is bound to content and to communicative purpose allows us to see that a genre can vary in its forms according to the content to be expressed and the rhetorical aims of the writer, within the bounds of the expectations and rules of the community.

If we are to hold that a genre must have a single prototypical form, then we must expect to be able to define a prototypical PhD thesis. If, however, we believe that genres are

groupings of texts that are socially perceived to perform a similar function, and moreover that these texts will vary in form according to communicative purpose and content, then we can maintain a concept of genre that allows for a variety of forms.

Chapter 3 Framework

3.1 Introduction

In this chapter, a theoretical framework for visualizing the rhetorical context surrounding the thesis-as-text is developed. The relations between the classical rhetorical entities of writer, text and reader are discussed, from a social constructionist perspective, and a diagrammatic representation of the relations is presented.

From the discussion three features of the thesis-as-text emerge as points for study: intertextuality, as expressed through reference to other texts, verbal processes in reporting verbs, and the uses of modal verbs to modulate relations between text, reader, writer, and the worlds of phenomena and concepts.

In the final section of the chapter, the research questions are stated.

3.2 Development of a framework

The basis of the framework contains three parts: a writer, a text, and an audience. In the case of the PhD thesis, we are concerned here with the writer of the thesis, the thesis itself, and the readership for the thesis. The writer is a physical entity, yet at the same time is a textual entity, as constructed within the text, both by the writer in the process of writing, and by the reader in the process of interpretation; the reader too has both physical presence, in the form of the supervisor(s) and the examiners, and possibly research colleagues, but the reader also is a constructed entity both within the mind of the writer (the imagined audience) and also within the text (cf the virtual community, discussed in Chapter 2). The text itself has both a concrete existence, in terms of paper and print, and also a constructed, dynamic existence within the minds of those who read it (see Figure 3.1 overleaf).

Figure 3.1 The writer-text-reader relationship as both physical and virtual entities

3.2.1 The writer

The writer is both agent and subject, in that the writer has individual meanings to communicate and purposes to achieve, and yet must follow genre conventions to do this. The conventions are, in a sense, the codification of audience expectations within the research community (Berkenkotter, Huckin *et al*. 1991), and so should be seen to be more than merely the conventions of academic prose. The diagram shows the relation between writer and genre conventions as a dynamic cycle, between the social and the individual, equivalent to Giddens' theory of structuration: the individual is both subject to the genre and instantiates the genre. To what degree the relation is one of equality, however, is a point for debate. In the case of Malcolm Ashmore, the subverter of conventions in his thesis 'The Reflexive Thesis', there was a productive exchange between convention and invention. Such theses, however, are rare, and it may be that the force of convention on the thesis writer can be heavy, and restricting. All the same, our interest here is not with the process of writing, but with how powerful the conventions of the genre are, for the writer within a given area of research.

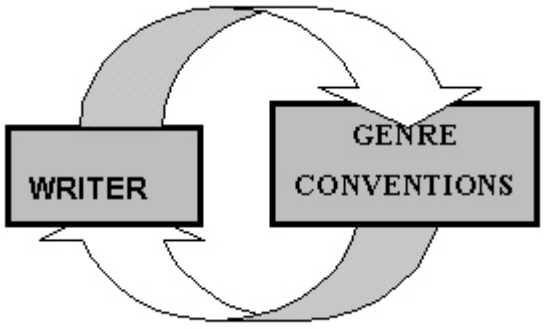

Figure 3.2 The writer-genre relationship as symbiotic

Expanding the picture further, the writer brings to the process of writing a range of actions, data and ideas that will form the content of the text. The actions are those of the writer-as-researcher, and of other researchers; the data is either the outcome of the

writer's research or may be the outcome of others' work; and the ideas are those that underpin the research and also emerge from it, coming from engagement with the ideas of others and from analysis of the data. These are represented verbally through different processes: the *reporting* of actions, the *describing* and *analysing* of data, the *discussing* and *evaluating* of ideas.

Furthermore, the job of the writer-as-researcher is to add to knowledge (to make an original contribution to knowledge), and the writer must indicate what is known in order to make it clear how the research has added to the bank of knowledge. In one sense, what is known is a form of data (something that is given), but it must also be reconstructed textually. In another sense, though, there are epistemological foundations to the research practices within a given discipline that are not always made explicit, and which can be said to underlie the text. These underpin the means by which research is conducted, the terms in which it is reported, and are at the level of the context of culture.

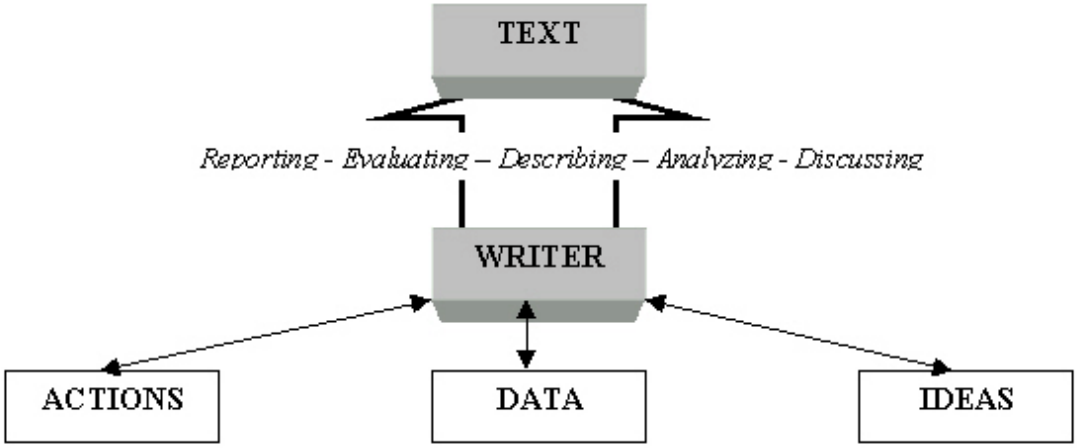

Figure 3.3 Verb processes

No text exists in and of itself. Following Bakhtin (1986) we can say that any text is involved in dialogue with other texts. Fairclough (1992) talks of two forms of intertextuality in a text: *manifest* intertextuality, which is realized through explicit references to other texts, as in the case of citation, and *constitutive* intertextuality which is implicit in the echoes of other texts conveyed through shared patterns, wordings, themes, and concepts. The writer creates a new text out of previous experiences of texts, from the various textual schemata that are formed from readings, and writings, of

other texts. In the case of a PhD thesis, this may involve a form of initiation, as Berkenkotter *et al* (1991) put it, into the research community that they engage with, through the literature and frameworks of the community. Berkenkotter *et al* (op cit) tracked a doctoral student through the first year and a half of his doctoral and observed student's gradual mastery of the appropriate linguistic conventions and rhetorical forms for communicating within the research community, through reading and writing, and interacting with faculty and peers. The later writings that the student produced are heavily intertextual, in that they are thick with references to the literature of the subject (manifest intertextuality) and with the concepts and terminology of the literature (constitutive intertextuality). Making appropriate references (appropriate to the conventions of the research community) to the work of others, and to report the actions, ideas and findings of others is an essential facet of the thesis.

At the same time, however, it should be borne in mind that a writer can belong to more than one community (Hudson 1980), and that the writer may draw on experiences of texts from other communities. For example, a PhD student may have experience of writing reports for corporate audiences, and this experience can also influence the writing of the thesis, although it could also be the cause of conflict between the writer and the supervisor, if the supervisor deems the influence of the non-academic genres to be unacceptable.

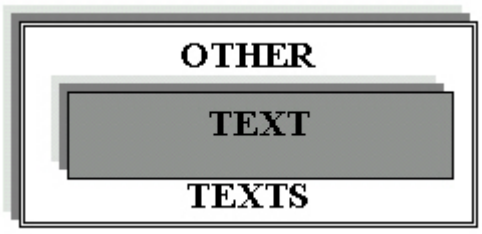

Figure 3.4 Intertextuality

3.2.2 The reader(s)

No text exists in isolation (Bazerman 1994 writes of the systems of genres), and texts as lengthy as theses are usually not written in isolation either. They will have been drafted several times, and various versions will have been read by the supervisor and possibly also by research colleagues and friends. The feedback from these readers,

especially that of the supervisor, will contribute to the form of the final text. In the final stages, the writer will also be asked to make revisions on the recommendations of the examiners.

Figure 3.5 The collaborative nature of text production

Drawing on schema theory, we could say that the schemata that writers employ when creating new texts are fashioned from previous texts, and previous experiences of both reading and writing texts. In the case of the PhD thesis, we might ask what schemata a writer has available for the writing of a new text genre. It is likely that the writer has to develop new schemata, or adapt old schemata, and this process may involve a considerable amount of negotiation with the supervisor and other members of the research community. Through writing and interacting with members of the community, the writer develops.

3.2.3 The text

As mentioned above, the text can be considered to have both a physical and a virtual entity. The physical entity is the book that is produced; according to the University of Reading 'Rules for the submission of theses for higher degrees', this must 'be completely typewritten or completely printed on one side only of A4 paper', and the final version must be hardbound. What concerns us here, however, is the latter: the text that is created through the interaction of verbal and visual representation with human interpretation. Narrowing our focus further, though we admit the importance of the visual aspect of the text (layout, figures and tables), these do not form a part of the present study, as the focus is on orthographic language. Texts are understood here not to be merely neutral representations, but to be mediated constructs that are attempts to

construe the world, the writer, the reader, and other texts in certain ways, according to the writer's intentions, both conscious and unconscious. In choosing to express an idea in a particular way, the writer is constructing the world, and a persona, and in addressing an imagined audience, is constructing a readership through the text. Ivanic (1998), for example, has argued that self-representation in writing is a construct, and Charles (2000) has demonstrated how thesis writers construct appropriately academic identities within texts through the use of expressions such as 'It is prudent to …' or 'It is reasonable to …' (which indicates that the writer is a reasonable person), or 'It is straightforward to …' (which establishes the writer's competence as a researcher). Charles's approach could be extended to analyses of how the writer then constructs the reader (and also manipulates the reader), the world and other texts. It could also be argued that the writer creates the text within the text itself, through the use of metadiscourse. Intratextual references, for example, construct the text as an entity, helping the reader to build an image of the text (the metaphor is intentional) – its structure, its status, its purposes.

An important aspect of the mediation of the relationship between these different textual entities and the writer is in the use of modality, or, more broadly, the use of modal verbs. Through modality, the writer can hedge propositions, for example. Myers (1989) has argued, convincingly, that hedging in academic writing is a politeness strategy, allowing the readers the opportunity to take up different positions on issues. In that sense, hedging both constructs the reader as an astute expert in the subject, and it also indicates that the writer is sensitive and considerate of the reader. The modal verbs, while not only the only conveyors of epistemic uncertainty, play a major role in hedging.

Modality can also construe a proposition about the 'real-world' as accepted fact, or as claim. It can indicate whether the likelihood of a phenomenon occurring is certain, or merely possible, according either to evidence that the writer has, or to received ideas (the knowledge of the discipline, or of society at large). Legitimacy of claim can be invoked by the use of a modal verb, as can the writer's intention to recommend a course of action, to external actors. Modality is thus an important factor in the relationship between the textual and intertextual entities.

3.2.4 The wider context

The full set of relationships is represented diagrammatically in Figure 3.6 below.

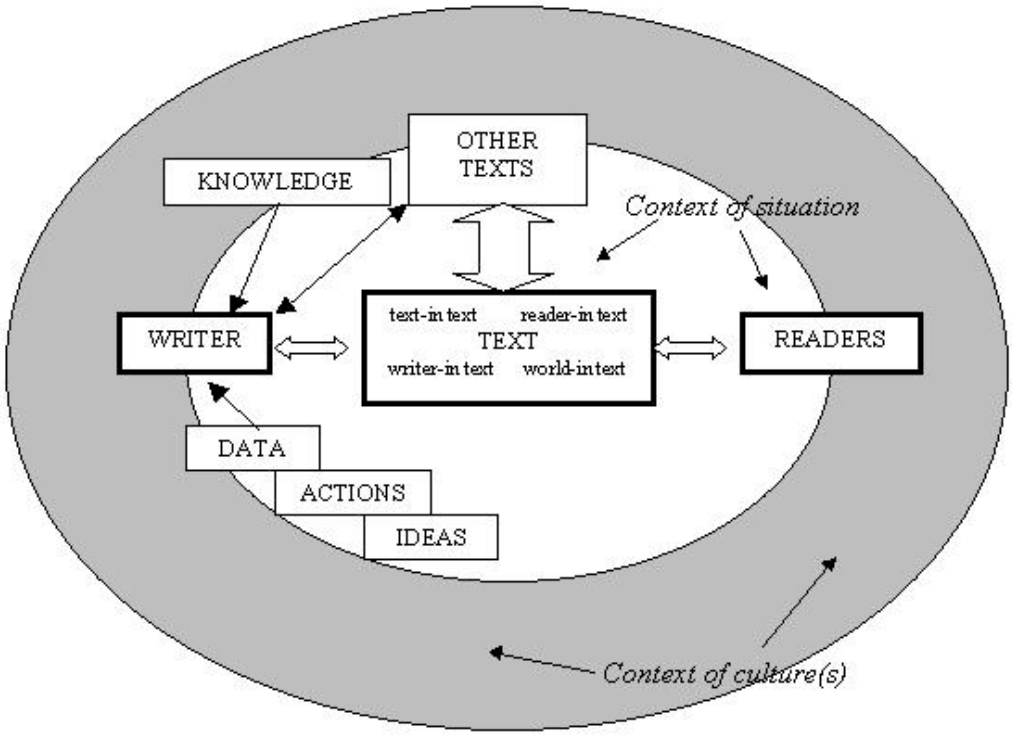

Figure 3.6 Rhetorical framework for thesis writing

The wider context is the context of culture. An 's' has been appended to the word 'culture' to convey the idea that there is no single culture, with a clear boundary; the context of culture is a loose term used to describe the implicit values, knowledge structures, and codes of practice followed by a grouping. There may be different cultures involved in this interaction between writers, readers and texts, as an individual can belong to several cultures at once, and at the same time it needs to be remembered that culture and community are both concepts that should admit of diversity.

At a more immediate level, the main actors meet within a context of situation – a local context. The situation might be that the writer and the supervisor are colleagues in the same department, or that the writer is working in one country, separated from an academic environment, and visits the supervisor for short periods every few months. Clearly the opportunities for the former to be initiated into the ways of the research community are greater, but this will depend on motivation. The student in the latter case may be studying towards a doctorate for instrumental purposes, seeing that a PhD

will increase the opportunities for promotion within the institution that he or she is employed at.

The writer has a range of actions, of ideas and of data to write about and also is informed by knowledge and by other texts. These texts in turn report, discuss and evaluate their actions, ideas and data, and they must be reconstructed (or reconstrued) within the new text. The interaction between the text in question and other texts is a powerful and necessary one; as Berkenkotter *et al* (ibid) have shown, to be accepted, one must learn to write and cite in the expected ways of the community.

Within the text itself are constructions of the writer (the writer's persona), the reader (in interactional language), the literature and the world. These are constructed in a variety of ways, at different levels, but the construction of these will be subject to some degree of conventionalisation. The exact degree of conventionalisation is a point of interest. As Bhatia (1993) has shown, the language of one law genre can be strictly codified and allow little variation while the language of another is far more open to change. At a local level, it should prove possible to look for instances of patterned language in the theses which indicate the preferred ways of putting things in that research community, and this is a feature that can be looked for in the use of the modal verbs.

An element of conventionalisation in genre that can be examined is that of rhetorical structure. The IMRD structure for experimental research articles is a case in point – the conventional form means that readers know clearly where they will find the information that they are looking for. If they are looking for the data, they look at the Results section; if they want to know what methodology was used, they look either at the abstract for a summary, or they turn to the Methods section. In the case of the PhD thesis, is there a simple conventional format used by writers in department, or are the forms more Protean?

3.3 Research questions

As stated in Chapter 1, the aim of this study is to achieve a richer understanding of the nature of the texts that are produced as PhD theses, in order to improve the substance and quality of tuition for non-native speaker doctoral students who are preparing to

write a thesis. I have argued, in Chapter 2, that texts cannot be understood in isolation and that it is necessary to understand something of the contexts in which the theses were produced. Ideally the writers of the theses would be interviewed, but that is not practical as the majority of the writers have left the university, and their memory of the thesis may also be unreliable, several years after the event. The alternative, followed in the present study, is to interview the supervisors of the theses, and to find out what their estimation of the thesis was, what they think is the conventional structure of a thesis in their area, and what the purpose of a thesis is, amongst other questions. The interviews will be described in Chapter 5.

In order to have a framework for discussing the rhetorical purposes of any thesis in the corpus, it is necessary to have an idea of the macrostructure, particularly if we are to look at the use of particular linguistic features across whole texts. Biber and Finegan (1994) have shown how the uses of linguistic features change from one section of a Medical IMRD research article to another, and this information is useful if we have a simple conventional structure such as IMRD. If, however, the structure is less conventionalised, the information on dispersion of linguistic features will be of little value without some account of what the rhetorical purposes of each section are. It is necessary therefore to examine the macro-structure first.

On the other hand, it may not be possible to understand the purposes of each section without closer examination of the text. The analysis of language use across the whole length of the thesis may illuminate our understanding of the larger structure of the text, and so, after studying the uses of particular features, we should return to a consideration of the wider picture.

The linguistic features we have chosen to examine are those of citation practices, and the uses of modal verbs, in theses written in two departments, those of Agricultural Botany, and Agricultural and Food Economics, in the University of Reading. It is intended that the study of citation practices will inform our understanding of how writers construct their theses upon the texts of others, while the examination of uses of the modal verbs should reveal how and where writers modify their propositions, how they construct a persona within the text, and how they represent knowledge about the world.

The present study is exploratory in orientation, not hypothetico-deductive, and thus hypotheses are not stated, and subsequently tested. The exploration of the texts is guided by the following research questions:

1. Is there variation in rhetorical organization, at the macro-level, between theses within each department, and between theses in the two different departments?
2. If there is variation in rhetorical organization at the macro-level, what factors contribute to the variation?
3. Is there variation between theses in the two disciplinary areas, and between theses in the same disciplinary area, in the types of citations used? If there is, what do these differences reveal about the purposes of the writers, the ways that they structure their texts, and the ways that the writers position themselves in relation to other texts?
4. Within and between the two disciplinary areas, is there variation, in quantity and rhetorical purpose, in the uses of specific language items that are acknowledged to be of central importance in academic writing: modal verbs?
5. How do PhD theses compare to research articles in the uses of modal verbs and in the uses of citations?

Chapter 4 The Corpus

4.1 Introduction

This chapter explains the corpus that was used for the main part of this study. Firstly the principles underlying the design of a corpus are discussed in general terms, so that the particularities of the present corpus can be understood in context. Following this, the process by which the corpus of theses was built up, standardized and tagged is explained, and finally general statistics on the corpus are presented.

4.2 Principles underlying corpus design

A corpus-based approach to genre analysis promises to make the task of examining theses far more tractable than it would be if only hard copy versions of the texts were available. Concordancing software programs make it possible to search for individual lexical items in large quantities of text at high speed, to assemble comprehensive lists of concordances, and also to detect patterning across large stretches of text. The software can first locate all instances of a particular lexical item, and then show the instances within a single line, or more, of text. It is thus possible to observe how particular features of language are exploited throughout a large text.

It is essential, however, that the corpus on which analysis is conducted is well designed and documented. Among the criteria by which a corpus is designed are the following:

- Purpose
- Size
- Representativeness
- Balance
- Contextual information

(summarized from Kennedy, 1998; McEnery, 1996)

The purposes for which the corpus are to be used will greatly influence decisions about how sophisticated the corpus needs to be, as well as what data needs to be included. If, for example, the corpus is to be used for linguistic analysis which will attempt to make generalizations about features of a language as used by a particular speech community, then the corpus must contain samples of the language performance of members of

speech community and must contain a sufficient quantity of data for valid generalizations to be made. If a corpus is being set up for the purposes of syntactic, morphological and phonological analysis of the language of a single child, then the corpus should contain morphosyntactic and phonological tagging, and probably information on contextual features as well; on the other hand, if the purpose of the compiler is to provide a resource simply for quantification of word forms in the child's speech, then an untagged corpus might be adequate. While the compilers of large scale corpora, which are made available to the research community at large, have to concern themselves with the preservation of as much of the original data as possible, with comprehensive tagging and documentation (as for example in the creation of the British National Corpus – for a detailed account, see Aston and Burnard, 1998), the individual researcher, with a particular set of questions to pursue, and with limited resources, can work with a simpler set of resources. Tribble (1997), for example, describes a set of procedures for what he describes as 'quick and dirty' means of building text corpora for the classroom teacher. While the methods he describes are quicker and dirtier than are required for the present research, the point to be made is that a corpus need be no more complex, in its design, and tagging, than is required for the purposes for which it is created.

4.3 Creation of the corpus

4.3.1 The Reading Academic Text corpus

As stated above (Chapter 1), the present study is as an investigation of the Reading Academic Text (RAT) corpus. The RAT corpus was established in 1995 at the Centre for Applied Language Studies, the University of Reading, as a resource for pedagogically-motivated research into the rhetorical and linguistic features of academic writing (Carne 1996). The original corpus was a mixture of academic papers written by practising academics, and theses that had been written by successful doctoral students in the Faculty of Agriculture at the University of Reading. The choice of the Faculty of Agriculture was motivated by the fact that the university has a large intake of international students that study for a doctorate in agriculture-related departments, and it was felt that research into the writing practices and products of students in that faculty would contribute most to improved pedagogical practice at the Centre.

4.3.2 The corpus of theses: RABET

4.3.2.1 Size of new corpus

The original corpus contained, as of October 1996 (when the present study began), eight PhD theses and a set of 20 research articles and other papers written by academics at the University. The eight PhD theses were all written by native speaker graduates from the Faculty of Agriculture at the university, who had graduated from 1989 onwards. The Faculty of Agriculture was at that time made up of 11 departments, and the eight theses had been submitted from within three of them: Agriculture (4 theses), Agricultural Botany (3 theses), and Agricultural and Food Economics (1 thesis).

For the purposes of the present study, a special sub-corpus had to be developed, and this was named the 'RABET' corpus, standing for 'Reading Agricultural Botany and Economics Theses'. The RABET corpus consists of two sections: one for Agricultural Botany, and the other for Agricultural and Food Economics theses. The number of theses in each section had to be increased to a more representative number, while still remaining manageable, and practicable. PhD theses are long texts, running up to 100,000 words in length, and they are also difficult to obtain as there are usually not many theses written in any one department, at a single university. One principle underlying the design of the original RAT corpus was that all the writers should be 'native speaker' writers of English. This concept had not been clearly defined, but in this study 'native speaker' has been taken to mean that the person had been educated from primary school onwards, in a schooling system in which English is the medium of instruction and is also the first language, thus allowing the inclusion of first generation immigrants who had lived in the UK or North America, for example, from early childhood. However, the number of 'native speaker' doctoral students in a department over a period of a decade is low, and this meant that it would be difficult to gather many theses. With these constraints in mind, the figure of 8 theses for each of the two departments was established as the target.

4.3.2.2 Representativeness and balance

Representativeness, as stated in 4.2 above, is a major issue in corpus design, and the question of what exactly it can be said that a corpus of 16 theses, written by native speakers, and submitted to two departments, within a single university in the UK,

represents is an important one. The two departments are, to some extent, unusual – there is only one Department of Agricultural Botany in the UK (there are several courses in Agricultural Botany offered within other departments), and the only other Agricultural Economics departments in the UK are at Wye and Newcastle-upon-Tyne. While there are several international journals of Agricultural Economics (the *American Journal of Agricultural Economics*, the *Australian Journal of Agricultural Economics*, the *European Review of Agricultural Economics* and the *Journal of Agricultural Economics*) there is no single dedicated journal for Agricultural Botany. Both of these departments, as it will be shown in Chapter 5, are interdisciplinary subjects, employing a range of approaches to solve problems. Within the Department of Agricultural and Food Economics, there are researchers who specialize in Economics, Marketing and Management, and there also further specializations in food marketing, development economics, the economics of research. In the Department of Agricultural Botany, there are laboratory based approaches, and there are field study approaches; the one works mainly at the level of molecules and cells, while the other works more at the level of crop management and crop ecology. Some research mixes these approaches, and other research (not represented in the corpus) takes a modelling approach, using sophisticated computer programs to simulate complex plant-soil chemical relations.

The theses in the RABET corpus cannot be said to be representative of all the theses submitted to each of these two departments, let alone of any particular disciplines, as the restrictions on access to theses made it impossible to control the variables. It should also be noted that there is no assurance that all the theses are of the same quality, as PhD theses are not given grades. It is therefore not possible to determine in advance whether a thesis is considered a 'good' thesis or a bare pass. Furthermore, it only became apparent that there was so much diversity of research types within each department after a long period of investigation into each department. It can be said, however, that the texts are *indicative* of the diversity of text types within each department, and that this also indicates the diversity of exemplars of PhD theses possible within comparable departments (departments which specialize in applied research).

4.3.2.3 Collection of texts

Firstly, the names of apparently native speaker graduates who had received their doctorates from these two departments from 1989 on were collected from searches of the university library catalogue. Then, after permission had been obtained from the heads of department, the departmental secretaries were asked to supply contact details, and to confirm that the people identified were in fact native speakers; from this, one further graduate, whose name is not typically 'British', was identified and added to the list.

Letters explaining the purpose of the research and asking for electronic copies of the thesis were then sent to the graduates, with a stamped and addressed postcard enclosed. Those who were willing to send copies returned the postcard, and were then sent an agreement form allowing reproduction of the thesis for research purposes (to be signed, and returned), a set of floppy disks (preformatted) and a stamped addressed envelope.

In all, a further six Agricultural Botany theses were collected and a further seven Agricultural and Food Economics theses, making a total of 9 Agricultural Botany theses and 8 Agricultural and Food Economics theses. Only eight of the Agricultural Botany theses were used in the study: one was excluded because the supervisor, when asked to agree to be interviewed (see Chapter 5), said that the thesis in question was not at all representative of work in the Department of Agricultural Botany. The writer of the thesis had been assigned to the supervisor, not because the research had direct relevance to agricultural concerns, but because the research was a major work of taxonomy, and taxonomy is one of the supervisor's areas of expertise. Details of the theses in the RABET corpus can be found in Appendix A.

4.3.2.4 Categorization problems

The argument presented by the supervisor raised several questions, which relate to the central theme of this study. Firstly, who decides whether any one thesis is typical of work in a discipline? If we employ prototype theory (Rosch 1975) as proposed by Swales (1990) to say that certain members of a category are more prototypical than others, at what stage do we say that certain exemplars are too far from the centre to be included in the category at all? Also, who decides what the prototype of the class

‘Agricultural Botany thesis’ is? There are also questions about the categorization: is Agricultural Botany a discipline, or are we simply talking about a department? If the latter, then should it not be possible to say that any thesis that has been written by a doctoral student registered within that department is therefore a member of the class ‘Agricultural Botany thesis’? On the other hand, if the term is being used to refer to a disciplinary grouping, it should also be possible to say that a thesis written by someone within a different department, the Department of Agriculture, for example, could be an exemplar of the class ‘Agricultural Botany thesis’ because it is research typical of that discipline. These are complex questions, which challenge the terms in which discussion of academic discourse is conducted, and they will be explored further in the coming chapters. In Chapter 5, the views of supervisors as to whether or not their subject is a discipline are reported and compared, and in the following chapters the similarities and differences between theses in each department are examined, and related to issues of research purpose and paradigm.

4.3.3 Naming and formatting of files

Theses were given identification numbers as they were added to the corpus. The fourth Agricultural Botany thesis, for example, is TAB-004, where the letters TAB stand for ‘Thesis Agricultural Botany’ and the number identifies this as the fourth thesis in that category. In this thesis, the various texts will be referred to henceforth either in the form ‘TAB-004’, or in the form ‘tab4’.

The theses were all, bar one, received in an electronic format. Files then had to be transferred to a hard drive, and, in several cases, converted from Macintosh, AmiPro or WordPerfect formats to Word for Windows format. The one thesis which was not received on disk was scanned and converted to Word files, using the FineReader® OCR programme. The files for each of the theses were checked against the hard copy version of the thesis, held in the library, to check that the contents of the files were the final version, by comparing every tenth page of the printed text against the files. In two cases, it was discovered that the text was different. The files that had been sent on floppy disk were from earlier versions of the thesis, probably the submitted version, prior to final changes made after the viva. In these two cases, the complete texts were checked and the electronic files were updated to match the printed versions.

The original RAT corpus had divided theses into the files for the Contents, Abstract, Acknowledgements and for each individual chapter, and this division was followed with the new additions to the corpus. However, the original corpus had not distinguished between the different departments that the theses emanated from. The distinction between disciplines is of importance in this study, and therefore the naming system had to be revised. Each file was given a name based on the simple thesis names described above, so that the files for the fourth Agricultural Botany thesis, for example, are as follows: tab4abs (the Abstract); tab4ack (Acknowledgements); tab4c1 (Chapter 1), tab4c2 (Chapter 2); tab4c3, tab4c4 and so on.

4.3.4 Tagging

For the purposes of analysis using concordance software, files eventually have to be converted to text format. In the conversion process, much information is liable to be lost, as raw text files do not hold information on formatting, they cannot contain graphics and they are restricted to the characters in the ASCII (American Standard Code for Information Interchange) character set. To preserve information, therefore, certain features of the text must be coded before conversion, and other information preserved in file header sections. The original RAT corpus had used COCOA tags for its header tags and the tags that had been inserted to indicate certain text structure details (section headings, etc), and the location of non-verbal information such as tables, diagrams and plates. COCOA tags are of the form: <G PhD thesis> where the letter G indicates the category of 'genre' and the following content identifies the member of that category (in this case, 'PhD thesis').

COCOA tags had been chosen because they were required by the concordancing programme (Micro OCP) that was used at the Centre in 1996, but COCOA tags are now scarcely used, and other, more flexible and powerful concordancing programmes, such as WordSmith Tools (Scott 1996), which do not depend on COCOA tagging, have since been developed. It was therefore decided to change the coding to fit in with a form of Standard Generalized Markup Language (SGML).

SGML is the international standard for defining markup languages that was first defined by International Standards in the document *ISO 8879:1986 Information Processing – Text and Office Systems – Standard Generalized Markup Language*. The

first version of the British National Corpus, for example, used a form of SGML mark-up, and the second version has been converted to another form of SGML, XML. XML allows the scripter to create as many tags as required, provided that the meaning of each tag is defined in a separate statement of the rules, or syntax, of the document, that is called the Document Type Definition (DTD).

The problem with DTDs, however, is that they can be cumbersome and they require special programs which can interpret a DTD. At the time of the development of the corpus, such programs were not common. What was required for the corpus was a set of text files that could be analyzed using the WordSmith Tools concordancing package, and which could also be easily read for an understanding of the text. The solution chosen was to prepare the texts as HTML files so that the files could all be read in standard web browsers such as Netscape or Internet Explorer, and these files could then be stripped of all tags (using the NoteTab Pro ® text editor, which performs this action at a single key command) to produce text files. Each text in the corpus, thus, is available in two formats: HTML and text.

4.3.5 Preparation of the texts

In text files, only characters within the ASCII set are available, and it is therefore necessary to code characters that are outside this set. Where possible, these characters were coded using the SGML conventions for special characters, as recommended by the World Wide Web Consortium (a complete listing of character entity references, as they are termed by the W3C can be found at www.w3.org/TR/REC-html40-971218/sgml/entities.html). For example, the pound sign (£) does not exist in the ASCII set, and so is coded as £ (where the opening ampersand and the final semi-colon identify the string as a character entity reference), or by the equivalent decimal code, £.

The focus of this study was on the language used within the body of the text, and not on the visual representation of data or ideas, and therefore the texts had to be stripped of all figures, tables, and plates, and these were replaced in the files by non-standard (in other words, not used in standard corpus mark-up schemes, such as the Text Encoding Initiative guidelines) tags: -FIGURE-, -TABLE-, -PLATE-. It is possible to create tags within SGML that are conformant to the TEI guidelines, within XML conventions,

provided that they are defined in the DTD for the document. In this case, the tags would be: <figure>, <table>, <plate>. This option was rejected however for the reason that tags are hidden when the text is shown in a browser, and it was considered important that the tags should be visible when viewed in a browser so that it would be possible to see where tables, figures or plates appeared in the original text.

The use of mathematical equations and expressions causes further problems for the corpus compiler, as the mathematical symbols also cannot be represented in ASCII. There is a further difficulty in deciding what the relationship is between mathematical and verbal forms of meaning representation. At a simple quantitative level, is the expression 10^5 equivalent to 'ten to the power of five' (six words), or could it be represented lexically in an alternative form? Further, the mathematician would argue that the latter formulation is economical and allows propositions to be conveyed more elegantly than through the use of words, and thus it would seem illogical to try to convert such symbolic means of representation into words. Consequently, all mathematical equations and expressions, as well as chemical reactions and DNA sequence maps, were replaced by: -EQUATION-, -EXPRESSION-, -REACTION-, -SEQUENCE-. Examples of these are shown in Table 4.1.

Tag	Example of what is replaced
-EQUATION-	$x_i(w,y) = y/\alpha_i$
-EXPRESSION-	a_{ij}
-REACTION-	$CH_2 + NH_2 \rightarrow CH_2 + NHOH$
-SEQUENCE-	M_1, QTL A, M_2, M_3, QTL B, M_4

Table 4.1 The feature tags used in the corpus

In addition to the feature tags, and the use of character entity references, HTML tags were used to indicate the use of bold or italic letters, the status of text as heading, list or body, the beginning of each paragraph, and the uses of quotation from other texts. The tags used are summarized in Table 4.2.

Tag	Use	Tag	Use
…	The text that is enclosed is a heading of the highest level (ie, chapter heading)	<p>	Beginning of a paragraph
…	The text that is enclosed is a heading of the second highest level (ie, section heading)	<blockquote> … </blockquote>	The enclosed text forms a block quotation
…	The text that is enclosed is a heading of the third highest level (ie, sub-section heading)	<qt> … </qt>	The enclosed text is quoted text
…	The text that is enclosed is a heading of the fourth highest level (ie, sub-sub-section heading)	…	Bold text
… …	Enclosed text is a list (unordered or ordered)	…	Italicicized text
…	Enclosed text is a list item	<u> … </u>	Underlined text

Table 4.2 The HTML tags used in the corpus

These tags allow the analyst to identify whether particular instances of a word or string of words appear in the body of the text, in a list, in quoted text (and therefore not in the words of the author) or in a heading to a section. The use of bold or italic formatting was removed from headings, but preserved elsewhere, as this might show evidence of author emphasis, in addition to the conventions of italicizing particular words in different disciplines. The paragraph tag used at the beginning of each paragraph had the extra benefit of making it easy to count how many paragraphs there were in a particular file, by quantifying the use of <p> tags.

The following extract from TAB-009 shows what the tagged text looks like:

```
<h3>4.1.4 Objectives of this study</h3>
<p>This experiment had two aims;
<ol>
<li>to study the seed production of natural populations of weeds found in a winter sown wheat crop following application of low rates of fluroxypyr for effects on size, number and weight of seeds.</li>

<li>to assess the progeny of these seeds for maternal inheritance of effects attributable to herbicide use.</li>
</ol>

<h2>4.2 Materials and methods</h2>
<p>A three year rotation experiment was designed to investigate the interaction between crop species, crop cultivar, sowing date and nitrogen fertiliser on weed incidence in three crops. The layout is described in detail elsewhere in this thesis (Chapter 2). In the first year a number of plots were used for a split-split-plot trial to investigate maternal inheritance effects of reduced herbicide doses on weed seeds and the subsequent progeny.
```

The <h3> tags show the enclosed text is a heading, at the third level (that is, that it is a sub-section; a section heading is the second level and the chapter heading is the first level). The <ol> tags surround the list text, and designate this as an ordered (by default, numbered) list, and the individual tags surround each list item. This completes the sub-section. The section heading follows with <h2> tagging, and this is followed by a paragraph block which is introduced by the <p> tag. In an HTML browser, this would be rendered as follows:

4.1.4 Objectives of this study

This experiment had two aims;

1. to study the seed production of natural populations of weeds found in a winter sown wheat crop following application of low rates of fluroxypyr for effects on size, number and weight of seeds.
2. to assess the progeny of these seeds for maternal inheritance of effects attributable to herbicide use.

4.2 Materials and methods

A three year rotation experiment was designed to investigate the interaction between crop species, crop cultivar, sowing date and nitrogen fertiliser on weed incidence in three crops. The layout is described in detail elsewhere in this thesis (Chapter 2). In the first year a number of plots were used for a split-split-plot trial to investigate maternal inheritance effects of reduced herbicide doses on weed seeds and the subsequent progeny.

Finally, each file was given a header with background information, as in the following example taken from tab3c7, in order to convey some contextual details:

```
<HEAD>
  <META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=iso-8859-1">
  <META NAME="Corpus" CONTENT="Reading Academic Text corpus">
  <META NAME="Genre" CONTENT="Thesis">
```

```
  <META NAME="Institution" CONTENT="Reading University">
  <META NAME="Department" CONTENT="Agricultural Botany">
  <META NAME="Author" CONTENT="Andrea Grundy">
  <META NAME="Title" CONTENT="The implications of extensification for crop weed interactions in cereals">
  <META NAME="Year" CONTENT="1993">
  <META NAME="Paragraphs" CONTENT="12">
  <META NAME="Thesiswords" CONTENT="31390">
  <META NAME="Docwords" CONTENT="2528">
  <TITLE>TAB-003: Chapter 7</TITLE>
</HEAD>
```

The category names are all self-explanatory, other than the first Meta Name line which identifies the character set used in the HTML page – this information is required by HTML browsers, such as Netscape Navigator and Internet Explorer.

The foregoing section describes the procedures for creating the HTML files. The final stage is to convert these files into text, by stripping out all the tags. Once this has been completed, the word counts for each file can be calculated. This information can then be added into the Header for the file as shown above, where the meta name 'Thesiswords' shows the total number of words in the thesis (counting the chapters only), and 'Docwords' shows the number of words in the body of the text for that particular document, or chapter (in other words, chapter 7 of TAB-003 contains 2528 words).

4.3.6 Word counts

The overall statistics for word counts are shown in Table 4.3 overleaf. The second column in the table shows the total number of words (bear in mind that this excludes all tables, figures, and so on) for each thesis. What is most remarkable about the figures in this column is that the Agricultural Botany theses are on average half the length of the Agricultural and Food Economics theses. At a total of 248,917 words for eight theses, the average Agricultural Botany thesis is approximately 31,110, while the average Agricultural and Food Economics thesis is 63,470. As the university guidelines for the submission of theses in the mid-nineties stated that a thesis should be between 60,000 and 80,000 words in length, the low figures for the Agricultural Botany theses are surprising.

	total	*c1*	*c2*	*c3*	*c4*	*c5*	*c6*	*c7*	*c8*	*c9*	*c10*
tab1	***18452***	3338	4683	5907	4524						
tab2	***38992***	3253	2380	9106	5647	6286	7596	4724			
tab3	***31390***	1832	6400	1097	8404	5686	5443	2528			
tab4	***28072***	5690	4048	4105	1853	2129	2249	2666	3678	1654	
tab5	***38738***	7949	989	1966	4197	4269	3710	3941	7199	4518	
tab7	***14774***	2646	2801	1191	2704	3202	2230				
tab8	***35936***	1047	3526	3186	8107	6755	3674	5146	4495		
tab9	***42563***	6755	19372	4690	5297	3480	2969				
	248,917										
tae1	***75691***	3929	2318	16740	7983	14917	4029	19220	6555		
tae2	***39336***	1901	3706	4005	9686	4470	7590	3784	4194		
tae3	***68962***	2135	6911	18135	5053	6338	3058	7490	6176	11039	2627
tae4	***55219***	8126	7605	22512	6029	7091	3856				
tae5	***96644***	1890	7452	23067	16745	22364	20142	4984			
tae6	***52666***	1851	16150	2291	8118	2467	8494	3119	3531	1572	5073
tae7	***41917***	2480	8186	8676	7735	7008	4475	3357			
tae8	***77349***	6938	5259	5913	11583	9905	9271	7313	8348	6914	5905
	507,784										

Table 4.3 Word counts for each thesis, file by file. The first column shows the name of the thesis (tab1 represents the first Agricultural Botany thesis – TAB-001), and the names in the first row indicate the number of the chapters (c1 is Chapter 1, and so on). The figures in the shaded cells are the sum of word counts for all the theses in the section.

In the case of corpora composed of samples of texts, it is easy to maintain comparability of sections because most of them contain a roughly equal number of words. In the present corpus, composed of whole texts, there is the problem, as can be seen in Table 4.3, that each chapter is of a different length, in number of words, with the shortest being 910 and the longest 23,067 (which is longer than the complete texts of either TAB-001 or TAB-007!). As the chapters are of unequal length, frequency counts in this thesis appear mainly as aggregates (per 1,000 words). The frequency of occurrence of a particular search term, such as the modal *may*, for example, is adjusted for each chapter, so that if it appears 10 times in Chapter 1 of TAB-001, this would amount to 10 occurrences in 3,338 words, which is equivalent to 3 occurrences per 1,000 words. This aggregated figure can then be compared with the aggregated figure for other chapters. While this method is not ideal, as it does not reflect how many instances there were in which the writer could have chosen to use a modal verb – the ratio of use of modal *may* to finite verb clauses would be a better measure – in an unparsed corpus the simple measure of aggregated occurrence per 1,000 words is the best option.

4.3.6 Secondary corpus: ResArt

For purposes of comparison, a small corpus of research articles (named hereafter as the ResArt Corpus) was established, to act as a reference corpus. Hyland (1999) contains studies of citation practices that are similar to those undertaken in Chapter 6. His study was based on a corpus of 80 research articles, ten each from 8 different disciplines. The figures that he derived from his analyses showed clear differences in citation practices between disciplines (as will be reported and discussed at greater length in Chapter 6). The disciplines represented were: Mechanical Engineering, Electronic Engineering, Physics, Biology, Marketing, Applied Linguistics, Sociology and Philosophy. His results made interesting points for comparison with my own data, but I wanted to have a further point of comparison, with research articles related to the areas that were addressed in the theses in my corpus.

The research articles were chosen by searching through the concordances of tagged citations (see Chapter 6 for details) and picking out all the references that appeared more than twice in a given thesis. These references were then checked against the bibliographies (on electronic file, if available, or by reading the hard copy on the library otherwise). If the reference was to a book, it was removed from the list, and this meant that it was not possible to find enough research articles for two of the theses (TAE-001, and TAE-006), as the majority of references were to books. In the case of the Agricultural Botany theses relating to crop science and ecology, it was necessary to include a few papers from conference proceedings, which it may legitimately be claimed are not from the same genre as the research articles, and this point will be discussed in Chapters 6 and 7, when the results of comparative analysis are presented. Another constraint on choice of texts was whether or not the journal in question was held in the university library. In the end it was possible to locate 42 articles, 22 representing Agricultural Botany and 20 representing Agricultural Economics. Bibliographic details of the articles included can be found in Appendix C.

These articles were photocopied and scanned, then prepared for inclusion in the corpus following the procedures used for the theses, and described above. The file names for each article identifies which of the theses it relates to, and the name of the first-named author. ‘tab1boll’, for example, identifies the paper as being referred to in TAB-001,

and the following four letters (boll) refer to the authors of the article: Boller & Kende. The Agricultural Botany portion of the corpus contains 68,528 words, and the Agricultural and Food Economics part contains 90,245 words; as is the case in the RABET corpus, the Agricultural and Food Economics texts are much longer.

The rationale behind the choice of research articles in this manner was that it would indicate influential articles for each of the two research areas: influential in content, firstly, but possibly also influential in style and organization. As argued above, the PhD writer has few exemplars of theses to draw on when writing a thesis, and it is likely that the research article, while remaining a different genre, written for a different audience, for different purposes, will still influence the writer's performance of this peculiar genre. By constructing this reference corpus I intended to be able to explore this possibility.

4.4 Summary of chapter

In this chapter, the principles underlying the design of the corpus, and procedures by which the corpus was assembled and prepared were described. Problems with the categorization of the texts were briefly discussed in 4.3.2.4, and the possible inadequacy of the information contained in the descriptors 'genre' and 'department' were noted.

The corpus, named the RABET corpus, consists of sixteen theses, eight each from the two departments of Agricultural Botany and Agricultural and Food Economics, and these theses are divided into files by chapter. The corpus files are held in both HTML and text files. The word counts for the two parts of the corpus show that Agricultural Botany theses are, on average, half the length of Agricultural and Food Economics theses.

A second corpus (the ResArt corpus) was created for purposes of comparison. This corpus consists of papers that were referred to more than twice in the theses in the RABET corpus. The papers are a mixture of articles from journals, and papers from conference proceedings.

Having established the procedures by which the corpus was created and developed, we now move on, in Chapter 5, to an exploration of the contexts in which these texts were produced, through a set of interviews with the supervisors.

Chapter 5 Exploring the contexts of writing: Interviews with supervisors

5.1 Introduction

In this chapter, the contexts of situation and of culture within the two departments are explored through a set of interviews with the supervisors of the theses in the RABET corpus. The purpose of these interviews was to set up a framework in which the textual analyses (Chapters 6 and 7) could be placed, and which would contribute to the understanding of how language is used by the various writers to achieve their communicative purposes. One of the main aims of the interviews was to establish what the expectations of the supervisors were and to what extent the format of the thesis is considered to be conventionalized.

The format of the thesis is important for the textual analysis of the theses, as it allows us to make comparisons between broad rhetorical sections. If each of the theses has a methods section, for example, comparisons can then be made between the methods sections of the Agricultural Botany theses and the Agricultural and Food Economics theses. Following the account of the interviews, I discuss the macro-organization of the theses in the RABET corpus, in the light of the interview data, of recent studies of thesis macro-structure (Ridley 2000; Paltridge forthcoming; Bunton 1998) and also my own initial readings of the theses in the corpus.

5.2 The interviews

Eight PhD supervisors in the two departments were interviewed about their views of doctoral research and the thesis: four from the Department of Agricultural Botany, and four from the Department of Agricultural and Food Economics. In the Agricultural Botany section of the corpus, there are eight theses, written for four different supervisors:

- TAB-001 and TAB-007 for one supervisor;
- TAB-002, TAB-005 and TAB-008 for another;
- TAB-003 and TAB-009
- TAB-004

In the case of the eight Agricultural and Food Economics theses, there were six different supervisors and I was able to interview four of them. One supervisor had left the university, and was not available for interview. The remaining five supervisors were all contacted but one declined to give an interview, and thus the number dropped to four. Three of the four supervisors that I interviewed had supervised one thesis, and one had supervised two of the theses in the corpus.

It would have been ideal to have interviewed the writers too, but this proved impractical. Only one of the Agricultural Botany thesis writers and three of the Agricultural and Food Economics writers are still at the university, and this was considered too small a number, and did not provide a fair balance. In addition, it was felt that, as the writers had written their theses between five and ten years before, they would have possibly unreliable memories of the experience. While it is acknowledged that the views of the writers themselves are of considerable interest for the purposes of the present study, the constraints of time, access and reliability made this impossible.

5.2.1 Background: the departments

The Departments of Agricultural Botany and Agricultural and Food Economics are two distinct departments with their own separate facilities within the Faculty of Agriculture and Food (as it was known at the beginning of the nineties; its name has since changed to that of 'Faculty of Life Sciences'). The Department of Agricultural Botany is within that Faculty, yet, since 1986, has also been a part of the School of Plant Sciences, along with the Departments of Botany, and of Horticulture (Giles 2000). It offers courses in crop protection, plant ecology, agronomy, plant bio-technology, genetics, crop physiology, while Agricultural Economics deals with agricultural and food economics, development economics, marketing, and management. They are both applied subjects, one falling within the life sciences domain, the other within the social sciences.

5.2.1.1 Agricultural Botany

The Department of Agricultural Botany was founded in 1909 and John Percival, who is credited with establishing Agricultural Botany as an academic discipline by writing the first textbook on the subject in 1900 (Bunting in Giles 2000: 70), held the first chair in Agricultural Botany. In 1925 the department offered the first honours course in

Agricultural Botany. It can thus be seen to be a department with a relatively long history.

The Department of Agricultural Botany is divided, physically, having its laboratory based research conducted within one building (the Plant Sciences building) and its field studies (crop ecology, and agronomy) unit based in a set of buildings in another area of the campus. The doctoral students attached to a supervisor will work in the same building as the supervisor, unless they are part-time students working outside the university. Doctoral students tend to work in teams, with a professor and other research staff, on a particular project, and there is a generally relaxed, collegial feeling to the department. The Department has weekly research seminars, shared with the other parts of the School of Plant Sciences, which most members of staff and postgraduate students attend and at which invited speakers or doctoral students present their research. I attended several of these seminars over the course of the last five years, to learn more about the kinds of research that are conducted within the department, as I also did for the weekly Agricultural and Food Economics seminars. This observational research work is not reported on explicitly in this thesis, as it was not conducted in a methodical manner, but some of the insights gained from the experience inform the analyses that are made in the following chapters.

There is a first-year doctoral student programme, which includes sessions on writing up research reports, and on the presentation of data in reports. Many of the doctoral students are sponsored and they are required to complete their doctorate within 3 to 4 years. They are encouraged to write up their research for publication in journals (the list of authors to include the names of their professor and the other members of the research team) during the second or third year, and also to give presentations at conferences. The doctoral student programme is conducted within the School of Plant Sciences and it is this institutional grouping that has the closest sense of community for members of the department.

5.2.1.2 Agricultural and Food Economics

The Department of Agricultural and Food Economics at the time that the theses were researched and written was located in a set of buildings adjacent to the Earley Gate entrance to the campus. Since its establishment in 1923, the Department has had links

with the Ministry of Agriculture (in its different manifestations), conducting government-funded research projects and farm surveys. Increasingly in recent years this has changed to a wider European-oriented focus, especially as the department has diversified to include more work in food economics and marketing. Until the year 1995, the Department was known as the Department of Agricultural Economics and Management.

Doctoral students in the department are given office space, usually shared with one or more students with similar research interests. There are weekly research seminars, one which is a closed seminar only for members of the department and the other an open seminar at which invited speakers (academics, consultants and civil servants) make presentations. The first-year doctoral programme is run in conjunction with the Department of Economics, and this suggests that the strongest sense of affiliation for students in the department is with the discipline of Economics. However, the increasing number of researchers involved with marketing and management issues (represented within the Agricultural and Food Economics section of the RABET corpus by TAE-001 and TAE-006) is creating a greater diversity of interests and affiliation. In the 1990s, several of the doctoral students in the department had teaching responsibilities (unlike the Agricultural Botany students who were involved in full-time research either at the university or in industrial research laboratories), and also were involved in either research or consultancy projects. During the period of time in which they were preparing their theses, they may have written reports, for sponsors, journal articles, or magazine articles.

5.2.1.3 Research orientations of the departments

In terms of Biglan's (1973) axes of hard-soft and applied-pure, both these departments conduct work that is clearly towards the applied end of the axis, and the research is towards the hard end of the other axis. However, as argued in Chapter 2, this distinction is a loose one, and it should be recognized that the research work conducted by different academics within the department may cover a considerable range. Methodologies within Agricultural and Food Economics, for example, can range from the purely mathematical to the use of focus groups, questionnaires and participatory rural appraisal. Methodologies within Agricultural Botany do not include such a wide

range between quantitative and qualitative orientations, but there are differences between work at the molecular level and that at the ecological or systems level.

5.2.2 Questions

In the interviews, I was interested in exploring the following broad research questions:

- Are theses in each department highly conventionalized in their overall organization ?
- How do theses differ from books or articles?
- What purposes does doctoral research serve within each department? What purpose does it serve for the department, for the student, and for the outside world?
- What appears to be the central communicative purpose of a thesis in each department?

These questions address issues of conventionalization (with the possibility of determining possible reasons for the strength, or laxity, of conventions), of purpose and of the community/communities that a thesis can be seen to be addressing, and which the authors may seek to enter. The aim of the second question was to establish to what degree the thesis is seen by expert members of the disciplinary community (as represented by the supervisor) as distinct from an article or a book, as this could help to establish to what degree the thesis is seen as a distinct genre, within that community.

Before the interview, each supervisor received a set of open-ended questions and a copy of the Table of Contents pages for the relevant thesis/theses. The questions reported on in this chapter are as follows:

1. What makes a good PhD thesis, in your view?
2. What are the purposes of PhD research, in your opinion?
3. What was the background of the student and the research that was done? How long did the writing of the thesis take? What kind of job did the student take up after completion of the doctorate?
4. What is the essential nature of a thesis in your field of study? To what extent is a thesis different from a book, or a journal article?

5. What kind of audience do you think that a thesis writer should ideally have in mind when writing?
6. Is there a conventional way of organizing/structuring a PhD thesis in your field of study?

Each interview lasted between 30 and 60 minutes, and can best be described as semi-structured. Where discussion led to new questions, the conversation was allowed to develop in new directions. The interviews were recorded, on the condition that the tapes would not be made public. As a result, it is not possible to include the transcripts for the interviews in this thesis. The interview data is summarised below, firstly for Agricultural Botany and secondly for Agricultural Economics (this sequence is followed throughout the thesis), with the numbers referring back to the question as shown above.

5.3 Synopsis of interviews: Agricultural Botany (4 supervisors)

5.3.1 Q1 – What makes a good PhD thesis, in your view?

The supervisors were first asked to say what they thought constituted a good thesis. As with most questions, there was no unanimity in the range of answers provided, but the supervisors did all refer to the importance placed upon clarity of expression. Two of the four supervisors stressed that they felt that a good thesis was one that addressed current concerns, which was, they said, characteristic of an applied science. That the other two supervisors did not support this idea is not an indication that they did not share the opinion, but that the discussion did not lead them to express a similar opinion.

Two of the supervisors, again, said that a thesis should have a reasonable, logical storyline: a set of questions is posed at the beginning and the ensuing chapters relate the story of how the researcher attempted to answer these questions. The text should be ordered by logical steps (a question posed and explored, followed by the posing of another question that arose from the consideration of the first, and so on) rather than reporting the chronological sequence of the experiments. Myers (1990) has termed this the ‘narrative of science’, an ordering of events that follows the logic of scientific enquiry, rather than the order of time or of nature. One of the supervisors stated that it

would actually be disadvantageous to report the work in chronological sequence as this would make the logic of the enquiry less clear.

5.3.2 Q2 – What are the purposes of PhD research, in your opinion?

All four Agricultural Botany supervisors stressed that the PhD process provides a *training*:

1. in research techniques and the management of a major research project, leading to an ability to work independently
2. in logical enquiry and the effective communication of one's work through writing. One supervisor asserted that science was of no value unless it was made public, and that students therefore had to learn to write up their research effectively.

5.3.3 Q3 – What was the background of the student and the research that was done? How long did the writing of the thesis take? What kind of job did the student take up after completion of the doctorate?

Of the 8 doctoral students in Agricultural Botany, six had been employed by the department as Research Assistants on large-scale research projects, with predefined objectives. The situation is typical of scientific postgraduate research programmes as described by O'Brien (1995: 5): doctoral students contribute to a cumulative development of knowledge, in a predominantly collaborative enterprise. It was interesting to note that one supervisor praised a particular thesis for its success in generating further research, and attracting further funding, by establishing new paths for exploration. Scientific research is expensive, and an ambitious department must be able to attract ample funding for its research projects.

The remaining two students were already employed in industry and they funded their own studies, with the support of their companies. They also chose the subject of their doctoral research themselves.

After completion of the doctorate, which took between three and five years, five students went on to become researchers in commercial laboratories, one continued on to post-doctoral research, one became a researcher in a government-funded laboratory,

and the last remained at the university, becoming a lecturer. One supervisor confirmed that the majority of doctoral students go on to posts in research and development in industry.

The writing-up phase took on average six to nine months, according to the supervisors, although one writer (TAB-007) wrote up in six weeks, a feat facilitated by the fact that the writer had already written, and had published, two research articles based on his research.

5.3.4 Q4 – What is the essential nature of a thesis in your field of study? To what extent is a thesis different from a book, or a journal article?

According to the supervisors, the fundamental nature of a thesis in Agricultural Botany is that of a *report*. The emphasis is on the work that has been done in the laboratory or field, and the information gained should be presented clearly and economically. One supervisor even claimed that in many parts, the tables and graphs are the focus, and that the text is ancillary, but this view was not shared by all supervisors. As another supervisor put it, the thesis is 'the lens through which the research work is seen'.

One of the two students who were working at the time of the doctorate, according to his supervisor, took the concept of report to an extreme. As the student in question was used to writing simple factual reports for his company, with short, direct recommendations at the end, he had to be pushed to include further discussion of different possibilities. This suggests two points:

- though the Agricultural Botany thesis is basically a report, it must also include sufficient discussion of implications and alternative possibilities;
- the final product of the thesis may be a curious mixture of varied conventions and influences regarding acceptable content, organization and style, as suggested in Chapter 3 above

In this case, the audience that the student was accustomed to addressing had different expectations from the audience that the supervisor intended the writer to address. Despite the advice of the supervisor, however, the thesis in question remains essentially a report of what the researcher did; quantitative analysis of the thesis, as will be

described in Chapter 7, reveals that there is scarcely any use of the modal verbs except ‘may’, with only two occurrences of ‘would’, one of ‘will’ and none of ‘should’, which indicates that the predominant style of the text is simple reporting statements. As will be seen in Chapter 7, this writer makes much less use of the modals than any of the other Agricultural Botany thesis writers. Ultimately the success of the thesis depends on the quality of the research conducted or the argument presented, rather than on the strict adherence to a set of conventions, and this thesis is a case in point, showing a negotiation between writer and supervisor that leads to a compromise.

As for the differences between a thesis, a book and an article, the supervisors said that a thesis should place the writer’s own work within the context of other people’s work, whereas a book has a broader base, dealing with other people’s work equally. The article reports one’s work with a clear focus, using only the salient parts. The thesis, on the other hand, can include failed experimental work as this can tell the story of the project, and there is little restriction on length. The comments mainly related to the focus and content of the different types of text. Only one supervisor referred to the different audiences, saying that while a thesis is guaranteed at least three readers, a book has to attract readers, and an article has to satisfy the reviewers.

5.3.5 Q5 – What kind of audience do you think that a thesis writer should ideally have in mind when writing?

When asked about the audience for the thesis, all four supervisors stated that the thesis should be addressed to fellow researchers who are interested in the broad research area, on a level of parity (neither speaking up nor down to the reader) and they felt that the thesis should definitely not be written for the supervisor, nor the examiner.

5.3.6 Q6 – Is there a conventional way of organizing/structuring a PhD thesis in your field of study?

While one supervisor met the question, ‘Is there a conventional way of organising/structuring a PhD thesis in your field of study?’ with a conclusive ‘Yes!’, another supervisor said that the conventions were growing weaker. The difference of opinion became more marked in later discussion (though the two were interviewed separately) where the former argued strongly against the use of personal pronouns in

thesis writing, on the basis that the experiments are the focus not the researcher, while the latter related that he encouraged students to use first person pronouns where this helped them to communicate their ideas.

All four supervisors described the conventional format of writing in their discipline as the IMRD structure, with the most common form in a thesis being a cyclical, complex pattern. The two patterns of organization that they described are shown in Figure 5.1.

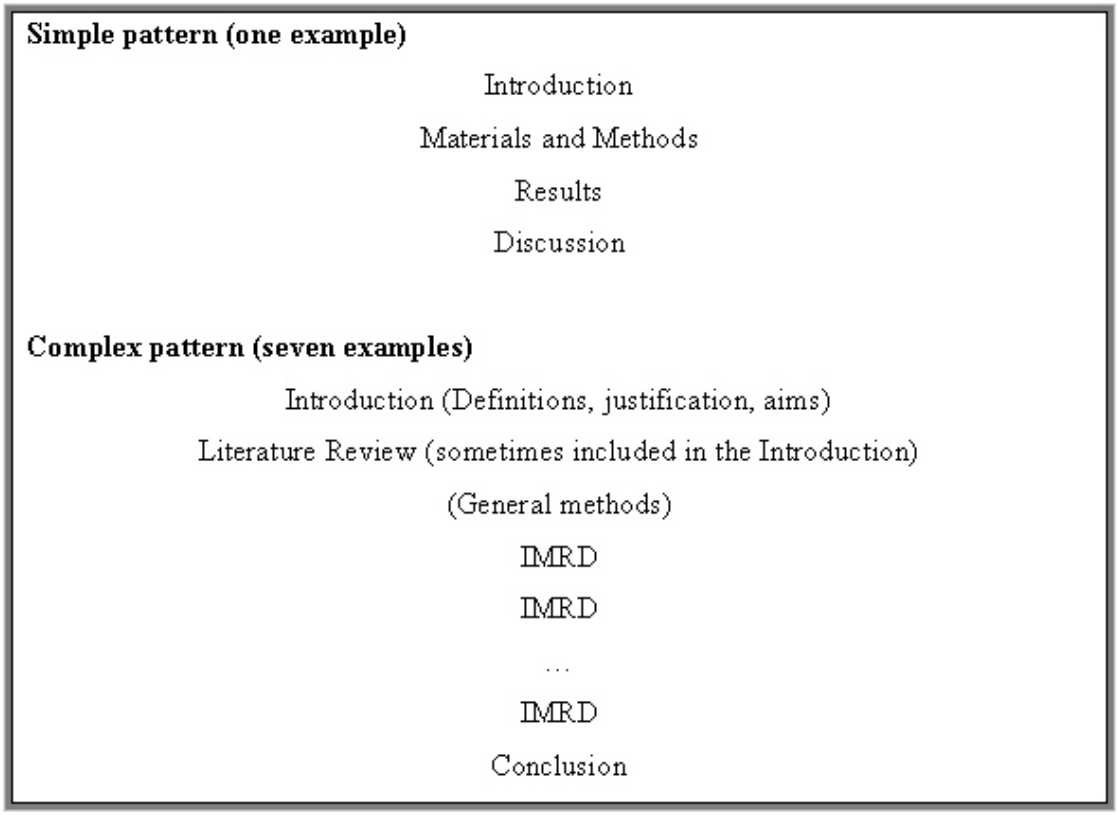

Figure 5.1 Frameworks for a thesis in Agricultural Botany. The information in brackets after the name of the pattern indicates how many examples of the pattern are found in the eight Agricultural Botany theses in the RABET corpus

These patterns of organization were evident in the theses in my corpus. The simple pattern was only used by one of the writers (the text is one of the shortest and is also the oldest – 1990), while the complex pattern was used by all others. In the complex pattern, the description of 'General Materials and Methods' is optional, depending on whether a core set of materials and methods was used for most of the experiments reported; if not, the procedures followed are given in each chapter. The IMRD chapter pattern is recursive (indicated by the three dots in the figure), with each section obligatory. A chapter reports a set of experiments, and in some cases, more than one set

of experiments so there is more than one IMRD cycle within the chapter. An extra section may be added after the discussion section of a chapter, to summarise the main points of the chapter, usually in a bullet list.

One supervisor mentioned that two of his colleagues specialized in computer modelling of plant-soil interactions, and that the report of this kind of research would probably lead to a quite different macrostructure, as the bulk of the thesis would involve development of the model. As there were no examples of such research in the corpus, no further discussion is made here, but it should be noted that there is the possibility that alternative macrostructures can be used in Agricultural Botany theses.

5.4 Synopsis of interviews: Agricultural Economics (4 Supervisors)

5.4.1 Q1 – What makes a good PhD thesis, in your view?

When asked 'What make a good thesis?', the Agricultural Economics supervisors made reference to 'coherence', 'concepts', 'comprehensibility' and 'plausibility'. A major difference from Agricultural Botany was that the Agricultural Economics thesis was described as fundamentally an *argument*, and a good thesis is one that is 'well defended'. A thesis is discursive, and it presents novel perspectives on field-central 'issues', a word that I found was often used in the Agricultural and Food Economics departmental seminars too.

Comments were also made about style: a good thesis should be 'stimulating and readable' and 'it should not be obscure'. An interesting observation, although not necessarily of direct use to the EAP teacher, was that 'the structure of the thesis shows that the person has understood the subject', and that a good thesis was therefore one that had a clear and appropriate structure. There was no mention of structure in the case of the Agricultural Botany theses, which one might surmise is due to the conventionalization of rhetorical organization in those theses, which means that it is the ordering of the information that is of importance rather than the overall structure of each chapter, which is usually a given.

5.4.2 Q2 – What are the purposes of PhD research, in your opinion?

When asked about the purpose of PhD research in Agricultural Economics, two of the supervisors judged the process to be a training of the individual's critical faculties, and another said that it was training to undertake a major piece of research. The emphasis on training is evident in both departments, particularly in terms of a training to conduct research on one's own. However, one Agricultural Economics supervisor (now retired) spoke of the writing of a thesis as 'a ritual dance', and as 'a hurdle to be surmounted rather than a useful training'. He criticized the insistence of Research Councils on a four-year completion deadline for doctoral work, arguing instead that social science students usually require a longer period for the full development of their thinking on their subject.

5.4.3 Q3 – What was the background of the student and the research that was done? How long did the writing of the thesis take? What kind of job did the student take up after completion of the doctorate?

Of the eight students represented in the corpus from the Department of Agricultural Economics:

- three were already employed as lecturers in the department,
- one was an economic consultant who worked away from the university
- three were Research Assistants at the department, on government funded projects
- one was on a MAFF (Ministry of Agriculture, Fisheries and Food) scholarship.

On completion, which took from four to eight years (substantially longer than in Agricultural Botany), one became a civil servant, six worked as university lecturers, and the economic consultant continued in her job. In general, it could be said that the doctorate acted in a gatekeeping role, providing fuller acceptance into an academic community, in contrast to the entry into Research and Development positions in industry that doctorates in Agricultural Botany provide.

5.4.4 Q4 – What is the essential nature of a thesis in your field of study? To what extent is a thesis different from a book, or a journal article?

The thesis was viewed by two supervisors to be an extensive display of knowledge of the subject, to prove one's academic credentials, rather than as a 'lens through which work is seen' (cf Agricultural Botany). The emphasis here seems to be on what the writer knows about the subject, rather than what is known about the subject, as in the case of the Agricultural Botany thesis. Additionally, the working out of a complex econometric model, or the recording and evaluation of the details of certain agricultural research and development programmes involves articulating the stages of the argumentation on paper, and it can be said that the writing is, to a much greater extent than is the case with Agricultural Botany theses, the *performance* of the research. Support for this is suggested in the relative lengths of theses in the two departments, as shown in Chapter 4, where Agricultural Economics theses were shown to be on average twice as long.

5.4.5 Q5 – What kind of audience do you think that a thesis writer should ideally have in mind when writing?

In terms of the differences between a thesis, a book and an article, it was stressed that a thesis, unlike the other two, is a piece of unfinished work. As noted above, the students in Agricultural Economics took between four and eight years to complete, and the thesis, by the end of the process, had become a 'millstone around the neck' that the student was 'desperate' (a word used by three supervisors) to get rid of.

One of the Agricultural Economics supervisors suggested the differences summarized in Table 5.1. The distinctions made are, on the whole, the same that the Agricultural Botany supervisors made, except for the perception that the purpose of the thesis is primarily to *convince the examiners*, a point of view shared by two other Agricultural Economics supervisors. A clear distinction can therefore be made between the Agricultural Botany supervisors who felt that the thesis writer should be, and is, addressing an audience of fellow researchers in the same general field, and the Agricultural Economics supervisors who felt that the PhD was written with the examiners primarily in mind. Further to this, it was observed that the Agricultural

Botany theses are often read by other members of the research project teams, thus enjoying a wider actual audience than the Agricultural Economics theses which will most likely only be read by the examiners, and, as one supervisor suggested, potential employers at other universities.

	Thesis	Book	Article
Purpose	convince examiners	share ideas with wider community	share ideas with a specific community
Requirements	meet the standards of the institution	meet expectations of publishers	meet standards of reviewers & editors

Table 5.1 Differences between a thesis, a book and an article (according to one Agricultural and Food Economics supervisor)

It is interesting to compare these with the observations made by the Agricultural Botany supervisors. As reported in 5.3.5 above, the audience for the Agricultural Botany thesis was recommended to be fellow researchers, addressed in a tone of parity, and not the examiners. While it is important to remember that the number of interviews conducted is small, and that therefore it is not justifiable to make generalisations, it is possible to hypothesize that the Agricultural Botany thesis is written with not only the examiners and supervisor in mind, but that the intended audience may include fellow researchers, especially those in the same research group. The Agricultural Economics writers, on the other hand, are working in relative isolation (O'Brien 1995 suggests that this is the typical condition of doctoral student work in the social sciences).

5.4.6 Q6 – Is there a conventional way of organizing/structuring a PhD thesis in your field of study?

The Agricultural Economics supervisors said that agricultural economists draw on a wide range of approaches and techniques, from philosophy, sociology, psychology, geography, anthropology, management theory, and so on, in addition to economics, to solve different problems. In answer to the question ‘Is Agricultural Economics a discipline?’, one supervisor replied that it was not; rather, it is ‘a problem-focused subject’. Although this view was not echoed by other supervisors, it was clear that the members of the department did not have a clear image of themselves as belonging to a discipline.

One supervisor explained that Agricultural Economics doctoral research is often componential, featuring a number of individual studies, and therefore theses tends to be diverse in form. This was placed in contrast to technical studies which, he stated, can be formulaic.

5.5 Macro-organisation of theses

As the Agricultural and Food Economics supervisors did not perceive there to be a conventional structure for a thesis, I looked instead at the theses in the corpus to see what patterns were evident, if any. In all cases there was an application of a framework, and the first part of the thesis formed the development of the framework. In one case (TAE-008), more than one framework was developed. The three econometric theses contain long discussions of theory and of the statistical methods used in that area of research, and this discussion is followed by the expression of the particular model of estimation to be used, then a description of the data set, the application of the model to the data, and finally the conclusion. The bulk of these theses is the theory and the development of the model. At the other extreme, we have TAE-006 which has a lengthy and substantial literature review (at 16,000 words, almost one third of the thesis), followed by a short outline of the analytical framework, and the studies carried out, and then the following five chapters present the analyses, organized thematically, which are derived from extensive field work.

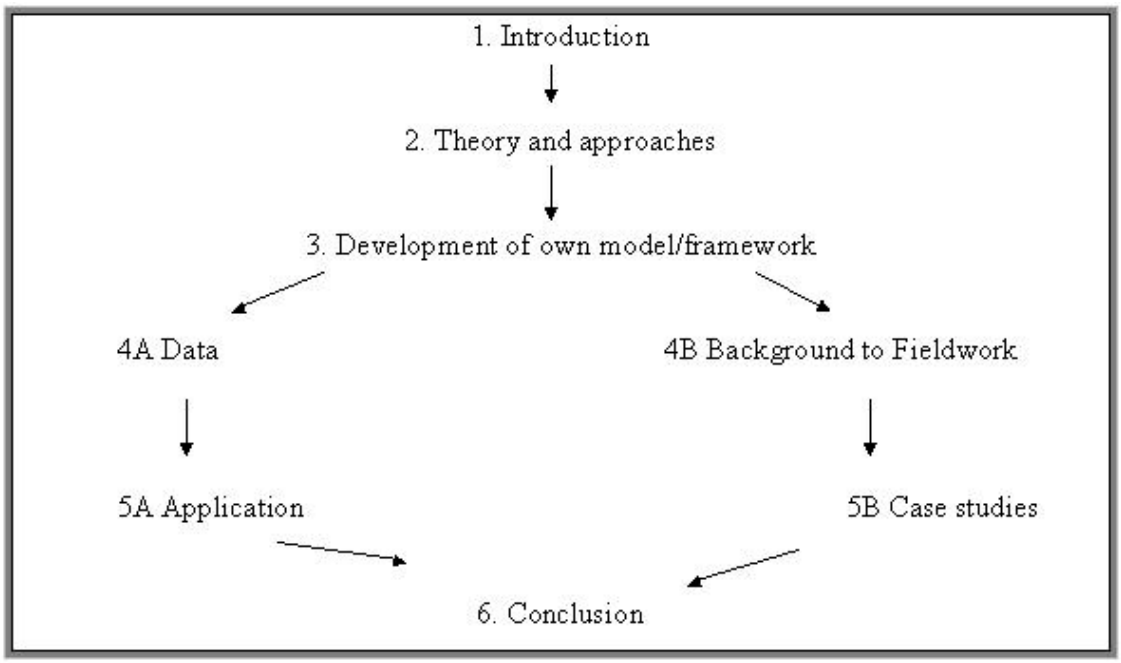

Figure 5.2 Macro-structures for a thesis in Agricultural Economics

Two other theses reflect the field work that the authors had conducted: TAE-001, and TAE-008. Both the theses are formed around the development of systems for use in the agricultural and food sector, one the description and classification of marketing systems, and the other the creation of a decision making support system for farmers involved with livestock disease control.

The remaining two theses are dissimilar in many respects, but the respect in which they are alike is that they both deal with the recounting and interpretation of past events. The longest chapter in TAE-004 is an historical review of agricultural and macro-economic policy in four sub-Saharan countries, while TAE-002 assesses the benefits to the UK of two research programmes, and this involves extensive discussion of decisions taken in the past, their implications, and speculation on what might have occurred if different decisions had been taken.

While it is clear that theses in Agricultural and Food Economics follow a wide range of patterns of organization, and that there is therefore no simple conventional system of organization, we can at least describe two general strategies for organization of a thesis in Agricultural and Food Economics. The diagram in Figure 5.2 represents two possible patterns of organization. In the Introduction, the relevance of the study to real

world issues is stated (the justification for the research), the aims of the study are stated, and general issues discussed. Normally the research questions are stated at the end of the chapter, and an overview of the thesis provided. The writer then discusses the theory and from the theory develops a framework for analysis. According to one of the supervisors, the review of theory and literature may draw on concepts from a wide range of disciplines, such as philosophy, sociology, psychology, marketing, or anthropology, as Agricultural and Food Economics work is often interdisciplinary. After this, there are two alternatives: the writer then establishes a data set, and applies the model to the data, or the writer then moves on to a set of case studies or some other set of componential studies. In the former case, the first two stages tend to form the bulk of the complete text; in the latter, the fifth stage, with the consequent discussion, are the major parts.

These macrostructures can be compared with the findings of Bunton (1998), Paltridge (forthcoming) and Ridley (2000). Bunton describes a three-stream model based on his analysis of 21 theses (both science and humanities) at M.Phil and Ph.D level from the University of Hong Kong (see Table 5.2 below). The first two streams are the simple and complex IMRD models that we have observed in the Agricultural Botany theses, and this indicates that the two models are used in a variety of science disciplines. The third stream is what Bunton terms a ‘topic-based’ structure, a description that Dudley-Evans (1999) also employs. The body of the thesis is a series of chapters that are focused around a particular topic. This topic-based structure does not seem to apply to the Agricultural and Food Economics theses in my sample.

<table>
<tr><td colspan="3">Introduction</td></tr>
<tr><td colspan="3">(Literature Review)</td></tr>
<tr><td colspan="3">(Pilot Study / Development of Instrument or Model / Theoretical Framework)</td></tr>
<tr><td>Topic:
Intro-(Lit)-Meth-Res-Disc</td><td rowspan="4">METHOD

RESULTS

DISCUSSION</td><td>Method</td></tr>
<tr><td>Topic:
Intro-(Lit)-Meth-Res-Disc</td><td>Topic:
Analysis-Discussion</td></tr>
<tr><td>Topic:
Intro-(Lit)-Meth-Res-Disc</td><td>Topic:
Analysis-Discussion</td></tr>
<tr><td>Topic:
Intro-(Lit)-Meth-Res-Disc</td><td>Topic:
Analysis-Discussion</td></tr>
<tr><td colspan="3">Conclusions</td></tr>
</table>

Table 5.2 Bunton's (1998:112) three stream model for the PhD thesis

Paltridge then examined 30 theses from different disciplines at an unnamed Australian university, and found only one example of the collection of published articles. This type of thesis appears to be becoming more common elsewhere, however, and is an interesting area of development in itself, as there are at least two sub-types possible. One thesis in the Reading Academic Text corpus (a biochemistry study submitted to the Department of Food Science and Technology) is a collection of four published papers, and it is remarkable in that there is little evidence of any effort to rewrite the chapters to make them cohere as a single text[6]. Personal communications with Swedish and German academics revealed that in some European universities the rule is that the published articles should not be rewritten at all, but that an introductory and concluding

[6] The University of Reading *Rules for Submission of Theses for Higher Degrees* states that '… a series of scientific papers appropriately supplemented to form an integrated whole counts as an acceptable format in which a thesis may be presented.' In this case, the rules do not appear to have been strictly followed.

chapter must be appended to the collection of papers. Clearly this type of thesis is quite different from the theses that are examined in the present study, as the rhetorical context for the papers is markedly different. The second sub-type is that where the articles are collected but are rewritten to form a literary whole. It would be interesting to investigate the revision strategies of writers who have to refashion their published articles for a different rhetorical situation.

Ridley (2000), working independently of Paltridge, built on the work of Thompson, Dudley-Evans and Dong, and developed a five pattern description, based on the examination of 50 theses submitted to a UK university, which is shown in Table 5.3 below. Her concern was with the literature review in each thesis and that is why there is detailed comment on the location of references to past literature in her account. The first two patterns are equivalent to the two patterns that I detected in Agricultural Botany theses, and Patterns 4 and 5 are similar to the macrostructures for Agricultural and Food Economics theses.

Pattern 1	• Introduction • Literature review • Methods • Results • Discussion • Conclusion
Pattern 2	• Introduction • Literature review • Different experiments are described in separate chapters each with their own ministructure: introduction, methods, results, discussion. Relevant aspects of the literature review may be revisited in the introduction to each experiment. • Overall discussion and conclusion
Pattern 3	• Introduction and overall contextualisation of the research • The main body of the thesis is based on an analysis and interpretation of texts with constant comparison and contrast with the way others have previously analysed these texts. • Overall discussion and conclusion
Pattern 4	• Introduction and overall contextualisation of the research • In the main body of the thesis, algorithms, theories, particular approaches or computer models are put forward and their development and implementation are described and evaluated. The whole thesis involves reference to and development of previous work in related areas. • Overall discussion and conclusion
Pattern 5	• Introduction • Literature review • The chapter divisions in the main body are determined by the precise nature of the empirical research of the study. This research usually involves field work using a combination and variety of different methods. Parts of the literature review are likely to be revisited in the context of different strands of the research. • Overall discussion and conclusion

Table 5.3 Five patterns for the structure of a PhD thesis, from Ridley (2000:65)

Her pattern 3 is the one pattern that was not evidenced in the theses I examined, and she comments that this pattern is typical of the Biblical Studies theses in her sample, and speculates that it might also be used in Literature studies, and other interpretive studies.

Bazerman (1995) makes a useful distinction between the *interpretive* fields, which 'attempt to understand the creations left by other human beings' (ibid: 390); *reconstructive* disciplines, such as history, which attempt to reconstruct what happened; *observational* disciplines that investigate 'what is happening now' (ibid:430); and experimental sciences that control events 'to generate specific types of data' (ibid:460). Pattern 3 can be seen to be a typical pattern of organization for the interpretive disciplines. Ridley did not, unfortunately, have any examples of reconstructive disciplines, and it may be dangerous to speculate but a guess would be that the history thesis would provide a different pattern, possibly along the lines of the topic-based structure, with extensive reference to the literature throughout, to indicate sources. The experimental sciences have Patterns 1 and 2 as their conventional macrostructures, and the Agricultural Botany theses fall into this category. While it is possible to say that the Agricultural and Food Economics theses follow Patterns 4 and 5, it is less easy to say that they belong exclusively to the observational category as some of the work conducted in the area is historical (TAE-002 is an example).

Ridley found that there was reasonable consistency in choice of pattern and discipline, so that the Medicine theses all followed Pattern 2, and the Architectural Studies theses used Pattern 5, but there were exceptions: 3 out of 13 Engineering theses, for example, did not follow Patterns 1 or 2. She concludes:

> Despite this overall tendency, my overall impression from the survey is that every PhD thesis is a unique and individual piece of writing (Ridley 2000:69).

5.6 General features of PhD theses

While the discussion of macrostructure in the theses points towards variation in the organization of theses, particularly those written in Agricultural and Food Economics, there are certain regularities across the majority of the theses in the RABET corpus, which could be identified as conventional features of theses in general. The Tables of

Contents for the theses in the corpus help to show the structure of the theses and they are reproduced in Appendices F and G.

5.6.1 Introductory chapters

Introductory chapters, for example, tend to include the following features:

- General context of research
- General aim of the research
- Justification of the research (in terms of how it relates to real-world issues, or to theoretical issues
- Objectives of the research
- Overview of the thesis

In the Agricultural and Food Economics theses, but not in the Agricultural Botany theses, the Introduction usually contains a statement of the hypotheses. The Agricultural Botany theses do not contain formal statement of hypotheses at any point in the text, although there may be reference to hypotheses that drive certain parts of studies. This difference between science and social science domain theses has also been noted by Bunton (1998) in his study of Hong Kong PhD theses.

5.6.2 Chapter previews and reviews

Chapters usually begin with an introduction which contains a preview of the chapter, and end with a summary of the main points of the chapter, with an option of a statement of what the next chapter deals with. Sections within chapters predominantly work from the general to the specific, as is to be expected.

5.6.3 Concluding chapters

The concluding chapter contains some if not all of the following:

- A restatement of the research questions and hypotheses
- A summary of the main findings, possibly chapter by chapter
- A discussion of the implications of the research
- A discussion of the limitations of the research
- Suggestions for further research

Interestingly, the conventional headings of ‘summary’, ‘policy implications’ and ‘directions for future work’ are more common in the Agricultural and Food Economics

theses, although this set of theses has been seen to be less conventionalised in overall structure.

The five components match well with Bunton's (1998) proposed set of moves in a thesis conclusion:

- Introductory restatement
- Consolidation of the present study
- Practical applications or implications
- Recommendations for future research (Bunton 1998:196-7)

Bunton included discussion of limitations as steps within the last two moves, along with a number of other steps that are optionally employed.

5.7 Differences within the departments

Discussion with the supervisors led to the following differentiation of research types within each department. In the case of Agricultural Botany, a distinction was made between field work and laboratory work. Field work is bound to the changing of the seasons and there is thus less flexibility in time constraints: crops must be planted in the sowing season, and schedules are planned accordingly. Variables cannot be easily controlled in the field, and the work tends to result in large amounts of data. There is a need for statistical analyses, and large-scale computation. Laboratory-based work, on the other hand, is independent of the seasons, uses fewer statistics, and has fewer variables to consider.

In Agricultural Economics, there appears to be a greater variety of research types, with the influences of econometrics, of microeconomics, marketing and management bearing on the work that is done within the department. There is a diversity of approaches, some more heavily quantitative and some mainly qualitative, and it is difficult to make simple distinctions. One supervisor, however, suggested that a distinction can be made between research based on 'strong theory' and that based on 'weak theory'. Strong theory informs statistically-driven research which restricts the number of variables. This research generates stronger predictions, and it involves the processing of huge quantities of secondary data (datasets). Weak theory, on the other hand, is typical of much social sciences research, where there are too many variables

involved to be able to make strong predictions. Such work tends to produce highly complex models without imposing on the data, which is typically *primary* data. There is an interplay between micro and macro perspectives. Of the two routes shown in Figure 5.2 above, the left side would be used for a thesis based on strong theory, and the right side pattern would be used for weak theory research.

5.8 Discussion

The limitations of this piece of research are clear: for one, the numbers involved are small – only eight supervisors were interviewed, and they may not be representative of all staff in their respective departments; secondly, the answers they gave may not necessarily represent their practice; thirdly, only the views of supervisors are represented, and it might be the case that the views of students themselves would tell a different story. The findings of these interviews can, therefore, only be taken to be suggestive, and certainly not conclusive.

From the data, there appear to be great differences between the two departments. These are summarised in Table 5.4 below. Where the Agricultural Botany doctoral programme functions as a training for science researchers in agricultural/food research and development centres, Agricultural and Food Economics is far more strongly oriented towards the preparation of individuals for a life in the academy. On an extremely broad level of community, this suggests that thesis writers in each department participate in markedly different communities, with different practices. The Agricultural Botany thesis may be read by other members of a research team, and has value as a record of the individual's contribution to an ongoing project, as well as being a submission for examination; the Agricultural Economics thesis, on the other hand, will have a far more restricted audience and this is expressed, it would seem, by the supervisors' perception that the thesis is written primarily for evaluation by the examiners.

Agricultural Botany	Agricultural and Food Economics
Conventionalized macro-structure: a simple IMRD model, or a complex cyclical IMRD model	Less conventionalized macro-structure: Broadly characterised as Introduction – Theory – Framework – Data – Application – Conclusion, with an alternative of Introduction – Theory – Framework – Background – Case studies – Conclusion Often componential *(as compared to technical subjects which can be formulaic)*
A report of research *(the lens through which research work is seen)*	A performance of research; the writing of the thesis is the achievement of the research
PhD research is a training in designing, conducting, managing, and communicating research, and it prepares the student for work in industry	PhD research is a training of the critical faculties, and in undertaking a major research project. Prepares students mainly for academia.
Core communicative purpose: reporting	Core communicative purpose: discursive
Lab-based vs Field-based	Strong vs weak theory

Table 5.4 Summary of main differences between the two departments

If genres develop through the repetition of communicative events, we would expect that members of the community would have exposure to exemplary texts, or receive repeated training in the conventions of the genre. In the case of the PhD, it is difficult to see that people have much exposure to texts. Because Agricultural Botany students may read the theses of other people in their department, at least the other members of their research team, there is presumably a higher likelihood that there will be conventionalisation than in the case of the Agricultural Economics theses. In both cases, though, much of the reading that students do is likely to be of articles (addressed to a different audience), and most students will only ever write one thesis. Comparing this to article writing, where writers have easy access to examples of the genre, and may well write several articles in their career, it would appear that the thesis is a peculiar genre, subject to considerable diversity between disciplines, and also *within* at least some disciplines.

The differentiation of research types in either department into ‘field vs laboratory’, and ‘strong vs weak theory’ provides a useful starting point for offering explanations of the choices thesis writers make in terms of macro-structure and language function. Field

work, and research in the social sciences based on weak theory, involve observation of relationships between a host of variables, and, because of the lack of predictive power inherent in such research, will feature greater degrees of hedging and discussion of alternative explanations. In addition, it is probable that stylistic choices will differ. For example, the field worker, like the ethnographer in Hansen's (1988) study (referred to in 2.4.4 above), is aware that the research conducted cannot be replicated (the conditions are unique), and that therefore the suppression of the first person pronoun is not necessary. Laboratory work, on the other hand, with its control of variables, aims to produce results that can be replicated (the experiments could theoretically be performed by any researcher) and therefore uses a writer-evacuated style. Similarly, Agricultural Economics research based on strong theory is capable of making more definite predictions, and therefore writers are likely to be more confident in their judgements, and to use a more assertive form of argumentation.

It is expected, therefore, that when we come to quantify and analyze the uses of modals in the theses (Chapter 7) that we will find more uses of modal verbs to make strong assertions in the Agricultural and Food Economics theses that are based on strong theory. The field-based studies in Agricultural Botany, on the other hand, will most likely be characterized by greater uses of modal verbs for the purposes of hedging, and in the discussion of alternative explanations.

The characterization of the central communicative purpose of Agricultural Botany theses as 'report of research work done' and of Agricultural and Food Economics theses as 'argument' may be dangerously reductive, but certainly not as reductive as the categorization of all theses as one genre. Theses constitute a genre in name, or in naming practices, but it is clear that theses perform a range of highly complex communicative functions. In the following chapters, we will examine the theses in the RABET corpus more closely and further develop our understanding of the complex diversity, as well as the common features of the members of this text category.

Chapter 6 Citation practices in the theses

6.1 Introduction

In the previous chapter, a set of interviews with supervisors of the theses in the RABET corpus was reported on, and discussed. These interviews established that the communicative purpose of a doctoral thesis in the two departments can be broadly characterized as *reporting* in the case of the Agricultural Botany theses, and *discursive* in the case of the Agricultural and Food Economics theses. The Agricultural Botany thesis is, to a degree, a report of work that has been done prior to the writing of the text, while the Agricultural and Food Economics thesis is, to an extent, the performance of the research work as it is the enactment of the argument. In both cases, however, it is necessary for the writers to place their own work, whether it be the mapping of DNA or the development of a model for calculation of the economic benefits of research, within the wider research literature. In this chapter, we will look at the citation practices of the writers in the RABET corpus, to examine the variation in practice between writers both between and also within disciplines, and also to see what types of citation are used in different sections of a thesis. This chapter examines the level and degree of manifest intertextuality (cf Chapter 3) in theses, and the roles that reference to other texts play in the construction of the thesis-as-text.

6.2 Previous research into citation practices

The ability to make appropriate references to the literature is an essential aspect of successful academic writing. The citation of other texts can be important at one of two levels: reference to the work or ideas of others can provide support for one's own position, and it can also show the novelty of one's work (Hyland 2000). References help to delineate what the present state of knowledge is, so that the writers can then locate their claims within the larger disciplinary framework, and establish a narrative context (Myers 1990).

Swales (1981, 1986) has pioneered the study of citation analysis from an applied linguistic perspective. In Swales (1986), he reviews the work on citations that has been carried out in other disciplines, such as in citation analysis and in the sociology of science (albeit for quite different purposes). One aim in citation analysis has been the

development of a means to count the number of positive citations of a paper, in order to rank its quality. Clearly it is not enough merely to count the number of references that a paper draws in other publications, because some of those citations could be negative, or they might be perfunctory (Moravcsik and Murugesan 1975). Swales reviews the various attempts to build a model of citation types and then proposes his own. His primary distinction is between citation forms that are **non-integral** and those that are **integral**: the former are citations that are *outside* the sentence, usually placed within brackets, and that play no explicit grammatical role in the sentence, while the latter are those that play an explicit grammatical role *within* a sentence. The citation at the beginning of this paragraph, for example, is an integral citation, and the reference to Moravcsik and Murugesan is a non-integral citation. He also used the terms 'short' and 'extensive', to describe citations that are at a single sentence level, or those that encompass more than one sentence (Swales 1986: 50). Extensive citations elaborate on the concepts or the details of the cited text.

Analysis of academic text corpora has the potential to inform our knowledge about the different forms and functions of citations in academic writing. Pickard (1995) used a small corpus of applied linguistics articles to investigate the citation practices of 'expert' writers. On the premise that novice writers tend to overuse particular items in their references, such as 'say', she investigated citation practices in the corpus to find out what expert writers do. Using concordancing software she was able to produce statistical information to identify preferences among her writers for *integral* or for *non-integral* citation forms, and to identify the different grammatical forms of integral citations (subject, agent, genitive noun phrase, etc). Pickard's research was a useful preliminary study; the limitations, however, were that the corpus was small (only eleven articles), and there was little discussion of the reasons why writers decide to choose one form rather than any other, because the categories are based on syntactic distinctions rather than functional. More importantly, however, it is not clear whether her discoveries about the practices of a small number of applied linguistics writers can be generalized to 'expert' writers across all the disciplines. It seems likely that writers in different disciplines follow different rhetorical conventions and have different preferences.

Bloch and Chi (1995) compared citation use cross culturally, between American and Chinese academics, and across disciplines, looking at writing in social sciences and in the sciences. They were interested in the dates of citations (eg, how recent were the texts referred to?), and also in whether the citations acted as background, support, and whether the use was critical. They found that Chinese citation practices are markedly different from those of the American academics: the Chinese tended to privilege older texts, they used proportionately fewer citations, and, though the science writers were similar in their use of citation strategies in both language groups, the Chinese social science writers used citations to support their arguments proportionately less than their American counterparts. These findings support our contention that writers from other cultures are likely to have some difficulties adapting to the writing conventions of British academia (cf, Chapter 2, Section 5 above) .

Hyland's (1999) investigation of citation practices in academic writing is the closest to the study reported below. He quantified and analysed the uses of integral and non-integral citation in a substantially-sized corpus of journal articles representative of eight disciplines. Among the features of citation practice that he looked at were:

- citation density in the texts
- surface form (integral vs non-integral)
- the number, form and process category of reporting verbs
- quotation from other texts

The purpose of Hyland's study was to explore the differences in citation practices between writers in different disciplines, and his analysis is based on a view of academic writing as social interaction – that is, writers' motivations to make citations and to choose particular citation types are seen to be rhetorically driven, with the aim of achieving ratification by the disciplinary community of the new claims contained in the text, and Hyland's interest is to find evidence of how writers respond to different rhetorical contexts by making different rhetorical choices. The results of his research are not discussed here, as they will be introduced in section 6.7 below.

Integral citations in which the cited author or text controls a verb (for example, 'Young (1997) has argued that …'), or, equally, non-integral citations in which the cited text

implicitly controls a verb (for example, 'It has been argued (Young 1997) …') create extra choices for the writer as to what kind of reporting verb to use, and also what tense and voice to employ. In the next two sections, I review relevant research into reporting verbs, and into the correlation of verb tense and voice in reporting verbs with function..

6.2.1 Reporting verbs

Thompson and Ye (1991) investigated evaluation in reporting verbs. Though evaluation is a fascinating feature to explore, I decided not to tag verbs for their evaluative load as it would have been necessary in numerous cases to check my interpretation with an expert informant, and the quantity of data made this impractical. The approach taken however does make limited use of the analytical framework that Thompson and Ye developed, as follows. Firstly, they distinguish between the *writer* and the *author*: the writer is the writer of the actual text and the author is the one who is cited by the writer. This distinction will be maintained in the following pages.

Further, Thompson and Ye categorize reporting verbs by process, as *research*, *textual* and *mental*. These three verbal processes relate well to the three subjects for the writer to bring into the text: data, actions, ideas (see the framework, Chapter 3 above). Research verbs are those that refer to mental or physical processes that are part of research work (such as *find*, *calculate* or *isolate*), textual verbs are those that refer to processes in which verbal expression is obligatory (such as *state*, *challenge* and *report*), and mental verbs are those that refer primarily to mental processes (such as *believe*, *think* and *consider*). Thomas and Hawes (1994), in a separate study, use similar categories, referring to activities rather than to verbs: experimental activities, discourse activities, and cognition activities. For our purposes, the classification of the reporting verbs used by the thesis writers in the two departments studied here according to these distinctions promises to reveal the preferred activities of researchers in different disciplines; we would expect that a discipline that grounds its research programme in experimental research in an empiricist paradigm will use research verbs predominantly, while a more discursive discipline that deals mainly with interpretations, for example, will use discourse verbs.

Thompson and Ye based their work on their reading of the Introduction sections of approximately one hundred articles drawn from a range of disciplines, while Thomas

and Hawes examined the use of reporting verbs in a single research area, that of psychosomatic medicine. Consequently, while the classification that Thomas and Hawes developed permits a sophisticated analysis of the use of reporting verbs in experimental research (for example, they created sub-categories which include 'pre-experiment' and 'post-experiment'), it does not seem apposite to the classification of reporting verbs in more openly argumentative writing. Thompson and Ye's classification, on the other hand, is based on a more diversified corpus, and it also makes the useful distinction between author acts and writer acts.

It is worth noting, at this point, that Thomas and Hawes equate experimental activities with real-world action. An interesting point is that Agricultural Botany thesis writers tend to refer backwards and forwards within their texts by referring to experiments rather than to chapters, sections and pages as Agricultural Economists tend to. The inference is that Agricultural Botany theses are largely reports of activities in the 'real world' ('the lens through which the research work is seen', Chapter 5), while Agricultural Economics writers construct their own worlds within text and discourse. It would be likely, then, that the former would use a large number of research reporting verbs (using Thompson and Ye's terminology) while the latter would use discourse and mental reporting verbs.

6.2.2 Verbs in academic writing: Tense and voice

Lackstrom *et al* (1972) were the first to claim that tense choice in scientific writing has its own special characteristics, being determined by rhetorical concerns, rather than by time lines (as in narrative), a claim that is supported by later studies (for example, Malcolm 1987; Gunawardena 1989; Salager-Meyer 1992). Lackstrom *et al* posited that the past tense is used to indicate lack of generality, while the present perfect gives a 'good generalization about past events', and the present simple makes a general claim.

Our interest with tense and voice is primarily with their deployment in the use of reporting verbs. Oster (1981) investigated the use of tenses in reporting past literature, in scientific writing, and her findings suggested the following:

—The present simple is used when referring to quantitative results that are supportive of the writer's work, or to refer to past literature when the intention is not to discuss it.

—The past simple is used when referring to quantitative results that are not supportive of the writer's work or to claim non-generality about the past literature.
—Present perfect[7] is used to indicate continued discussion, and also to claim generality about the past literature.

Caution is due in interpreting these claims, however, as the analysis was based on the reading of only two articles. A more extensive investigation of tense meaning in academic writing is Malcolm (1987), who, in a study of 20 experimental reports from the *Journal of Pediatrics*, found that:
—generalizations tend to be in the present simple
—references to specific experiments are in the past simple
—references to areas of enquiry are in the present perfect

It is important to note that these findings were based on statistical frequency counts which indicate *tendencies* not rules; the first statement was true in 74% of the instances of generalizations, while the second and third were 61% and 72% true. A partial explanation for this variability is provided by Malcolm's observation that writers are sometimes subject to *obligatory constraints* (for example, where the use of a temporal adverbial such as 'In 1995' obliges the writer to use past simple) while in other contexts the writer may make *strategic choices* regarding the tense to be employed. This suggests therefore that tense choice is sometimes non-temporal, as claimed by Lackstrom *et al*, but it is also sometimes temporal.

Swales (1990: 154) suggests that tense selection may indicate something of the author's stance towards the cited work, and this is where strategic choice comes in. The use of present perfect might imply closeness to the study while past tense would distance the writer from the cited work, for example. Shaw (1992) demonstrates that strategic choices are also affected, however, by higher level discourse concerns. Non-integrated reporting sentences, for example, with a present perfect verb in the passive are often used to introduce sub-topics, stating generalizations that will consequently be elaborated on. Alternatively, the choice of voice may be dictated by theme-rheme

[7] Following Shaw (1992), the three forms are referred to here as tenses, as we are not concerned with the distinction between tense and aspect in this discussion.

choices within a sentence, which decide which features to focus on or which provide coherence.

To summarise, the studies of tense and voice usage for reporting verbs in academic writing indicate that tense choice is linked to the writer's stance towards research work in the field, and also the degree of generality. They also suggest that some differences exist between disciplines, supporting the hypothesis that there are marked differences in language use across disciplines. However, it should also be noted that these studies were based on the analysis of *scientific* texts, and we cannot be sure that the observations will hold true for social science texts too.

6.3 The Tagging System

In order to analyze the uses of citation in the theses in the corpus, it was necessary to tag them, so that relevant information could later be extracted, using a concordancing program, and the features then quantified and examined.

Following Swales (1990:141) citations were divided into *integral* or *non-integral*.

- Integral citations appear within the sentence; if the citation is in the form of a name followed by year number, typically the name will be incorporated into the sentence as an integral part of the syntax of the sentence, and will not be separated by brackets.
- In a non-integral citation, the citation is separated from the sentence by brackets and it plays no *explicit* grammatical role in the sentence.

Citations can also take the form of a number (rather than name and year) reference but none of the writers in the corpus use this style, and thus it was not considered in this study.

The distinction between integral and non-integral is primarily formal, but there is also a functional aspect. A writer can choose either:

- to place focus on the researcher, the cited text, or the piece of research by including the citation in the sentence (*integral*), or

- to emphasize the proposition (the finding, the data, the concept, for example), and de-emphasize the researcher/text/piece of research by placing the citation outside the sentence (*non-integral*).

Weissberg & Buker (1990) describe these two types as 'author prominent' and 'finding prominent', and reference to these terms will be made in the analysis.

6.3.1 Procedure

Sample chapters from both sections of the RABET corpus were tagged as either researcher-integral (RI) or researcher-non-integral (RN). The tagged texts were then analysed, and a further set of distinctions created, which are described in the following sections. These distinctions are based on formal features, primarily, but, once again, the categorization also reveals something of the intention of the writer.

The categories that were devised were tested by three colleagues at the Centre for Applied Language Studies, who were given two chapters (one from an Agricultural Botany thesis, and one from an Agricultural and Food Economics thesis) to tag. The tagged texts were then compared and a final version of the tagging categories was devised, in consultation with the three colleagues. The tagging system is described in the following sections, with glosses provided to explain what the form and function of the citation type is.

6.3.2 Non-integral citations

The non-integral citations were classified and tagged according to the following criteria:

<RN Source>

The citation tells the reader where the information (verbal or numerical) or idea comes from. The function of the citation is that of *attribution.* The information is contained in a proposition, rather than in a single noun phrase. For example:

> *Ex:* `Both diseases are of economic importance, but black Sigatoka develops much more rapidly, causes more severe defoliation, and is more difficult to control than yellow Sigatoka (Stover and Dickson, 1976).` (TAB-005)

<RN Ident>

The citation identifies an actor in the sentence, where the actor is either explicitly or implicitly included.

> *Ex:* `It has been suggested (Wardlaw, 1972) that M. fijiensis might be a mutant of M. musicola ...` (TAB-005)

In this example, the sentence could also be expressed as 'It has been suggested by Wardlaw (1972) that …' but a choice has been made to de-emphasise the name of the researcher.

<RN Origin>

An example of this type is:

> *Ex:* `The LOD score (Ott, 1985) is defined as the log10 of …` [TAB-002]

Where Source citations attribute a proposition to a source, Origin citations indicate the originator of a concept, technique or product. The citation refers to a noun phrase within the sentence.

<RN Refer>

The citation refers the reader to a text to find further details. The details are not given in the writer's text. This form of citation usually has the word 'see' included, but not necessarily.

> *Ex:* `This equation can be rearranged to express Total Factor Productivity as a function of research spending (see Thirtle, 1988).` (TAE-002)

A 'Refer' citation usually functions as a shorthand device; the reader is directed to another text in which exact details can be found.

<RN Example>

The citation provides a number of examples of studies that are referred to in the sentence. 'e.g.' or 'for example' typically preface the name(s) but not necessarily.

> *Ex:* `The existing literature on the returns to research is considerable, and several summaries of the indicators found in the literature are available (e.g. Thirtle and Bottomley (1988) or Echeverria (1990))` (TAE-002)

6.3.3 Integral citations

Two types of integral citation can be distinguished. In the first, the citation controls a lexical verb in the clause, either as the subject *(X argues ...)* or as the controlling agent in a passive construction *(It is argued by X ...).*

6.3.3.1 Verb-controlling

Adapting Thompson and Ye (1991)'s distinction between research, textual and mental reporting verbs, three processes were identified in the lexical verbs controlled by the citation:

<RI Research verb>

These are verbs which refer to research work, and have the sense of 'real-world' activities. Verbs connected with experimental procedures, with findings, with measurement, with categorisation, with observation (in the sense of watching) are included in this group.

Ex: `Miller and Tanksley (1990a) found no such correlation when studying tomato genomic clones.` (TAB-002)

<RI Discourse verb>

These are verbs which describe processes in which verbal expression is involved, such as *suggesting*, *reporting*, *commenting*, or *arguing*. Thompson and Ye termed these 'textual' verbs but the term 'discourse' is preferred here.

Ex: `Nodari et al. (1992) suggest that this difference may be due to the fact that random clones mainly detect point mutations, whereas ...` (TAB-002)

<RI Other verb>

Thompson and Ye had a third category called mental processes which covered all the verbs of cognition (believe, consider, think). There are certain verbs in the corpus which do not fit neatly into either of the previous two categories, such as 'provide' or 'offer' (as in 'offer a view' or 'offer an alternative'), or 'visit' (as in *Young (1990) visited* ...). This third category was created, therefore, as a catch-all-else.

Ex: Bassett (1986) provides a comprehensive survey of the use of pyrethroids in UK agriculture. (TAE-002)

6.3.3.2 Naming

The other form of integral citation is one which is within the sentence, and which does not control a lexical verb form.

<RI Naming>

The citation works as a noun phrase, and is typically functioning either 1) as a modifier, as in *the work of Fuller (1997)* or *Fuller's (1997) work*, or 2) as a free-standing noun phrase followed by a linking verb, as in *Fuller (1997) is the best example of this approach.*

Ex: Surprisingly no attempt was made on publication of the work of Fukuda et al. (1989), to assay ACC oxidase from plant sources under these conditions. (TAB-007)

The citation in this case can refer to a person, as in *the work of Fuller (1997),* or to a text, as in *Fuller (1997) is the best example of this approach.*

6.3.4 Non-citation

<RN Non-cit>

<RI Non-cit>

Occurrences of a name in the text which did not appear as a citation (i.e. no year, or page, reference attached to the name) were also tagged. The exceptions to this were:

- instances where the name was used to identify a particular theorem, model, law, or other such commonly recognised construct.

- repeated use of the same name; in other words, a name was only tagged once as 'Non-cit'

These 'non-citations', occur, of course, after the researcher has already been cited.

Ex: These lower order moments potentially provide enough information to accurately specify an

appropriate lag structure (see Silver and Wallace) with minimal priors. (TAE-005)

Ex: Schmidt suggests the last expression can be regarded as the truncated remainder which although time dependant is asymptotically negligible and thus can be omitted in estimation. (TAE-005)

6.3.5 Quotation

<DQ>

To find out how often thesis writers in these two departments used direct quotation, a tag was inserted to indicate where a citation was followed by a direct quotation from the cited text.

Ex: In general 'the World Bank has tended to use the Internal Rate of Return as its principal discounted measure' (Gittinger 1992). (TAE-002)

6.3.6 Elaboration

<elab>

A citation may either be restricted to a single sentence, or it may be elaborated upon in the following lines of text (Swales 1986 terms these '**short**' and '**extensive**' citations). Where the latter is the case, the tag <elab> was added immediately after the first tag. For example:

Ex: Tapia et al., (1990) studied the anatomical features of leaf surfaces of Grande Naine (AAA), False horn plantain (AAB) and Pelipita (ABB), and the relationship with resistance to M. fijiensis. They identified an apparent relationship between stomatal density and resistance, however, this was confounded with genome, as studies were made across genomic compositions but not within AA and AB. (TAB-005)

In cases where more than one work was cited for a particular statement, only one tag was inserted. In other words, the tag indicates that a citation has been made; it does not indicate whether it is a single or a multiple reference citation.

6.3.7 Tense tags

In order to find out about preferred uses of tense and voice in the reporting verbs in citations, verbs that were controlled by an agent either in the sentence (the three Verb-controlling patterns: Research, Discourse, and Other) or outside the sentence (RN Ident) were tagged for tense and voice. Seven tags were employed as follows:

<present active>

<present passive>

<past active>

<past passive>

<perfective active>

<perfective passive>

<participle>

These tags should be self-explanatory and examples are therefore not given. The only comment that needs to be made, perhaps, is that the perfective tags cover both past and present perfect tense.

6.3.8 Tagging and quantification procedure

The text files for each corpus were opened in the NoteTab Pro® text editor. This text editor allows the user the option of creating a set of tags which are listed in a frame to the side of the screen and can be simply click-pasted in to the document. Each file was skim-read on the screen, and whenever a citation was located, it was tagged, with the RI/RN tag placed first, to the right of the citation, followed by a tense/voice tag where applicable, and then a <DQ> or <elab> where appropriate. An extra tag <vb> was added to the list and placed after the reporting verbs, so that the reporting verbs could later be identified, by searching through all the texts for the <vb> tag.

The following extract (taken from Chapter 7 of TAB-003) is an example of what a tagged file looks like:

> <P>Increased weed seed production is the main fear with reduced herbicide application rates and has been demonstrated <vb> by Fogelfors (1977) <RI Research verb><perfective/passive>. In a reduced herbicide system, the emphasis should be on the effect on seed production of sublethal rates and optimising timings which are conventionally recommended for plant kill (Andersson, 1992)<RN Source>. It would be of interest to fully understand the role played by maternal effects of inputs carried over from

previous generations, on the species composition and competitive ability. Maternal effects do exist as seen in Chapter 5 part 2 in <U>Viola arvensis</U> and other studies elsewhere, comprehensively reviewed <vb> by Roach and Wulff (1987)<RI Discourse verb><participle>.

As can be seen, the <vb> tags appear after the reporting verbs ('demonstrated' and 'reviewed') The verb-controlling tags, as with all RN and RI tags, follow the citation, and here we have examples of a 'Research' verb in 'demonstrate' and a 'Discourse' verb in 'review'. The problematic status of 'review' as a Discourse verb is acknowledged, as it can also be understood to have an evaluative aspect to it, and thus possibly also be considered a mental process), and further reference to this problem is made in 6.5.1 below. Tense and voice are shown in tags appearing after the 'Verb' tags, and the other tag shown in this example is a Source tag, used here because the citation is attributing the whole proposition to 'Andersson, 1992'. 'Andersson, 1992' provides the authority for the recommendation.

Once all the files had been tagged, ten of them were tagged by myself for a second time, and then checked against the first taggings. The accuracy rate was 96%, which was considered satisfactory. The tags were then quantified in the WordSmith Tools program, by performing concordances on each of the tags in each of the theses, sorting the results by file, and then entering the numbers into Excel spreadsheets. The figures then had to be converted into aggregate figures per 1,000 words, which was done by multiplying the figure by 1000 and dividing by the number of words in the chapter. In other words, if there were 21 <RN Source> tags in a chapter of 3,000 words, 21 is multiplied by 1000, to produce 21,000, and then divided by 3,000 to produce a final figure of 7 – meaning that the RN Source tag for that chapter occurred on average 7 times in every 1,000 words. That is,

21*1,000 / 3,000 = 7

6.4 Questions addressed in study of citations

The results of the tagging and quantification allow us to investigate variation in textual practices between disciplines, within disciplines, and also within different rhetorical sections of PhD theses. In the following sections, we look at the resulting data in the light of the following questions:

1. Are there differences in the quantities of non-integral, or integral, citations used by writers in the two departments?
2. Do writers within the same discipline tend to use the same types of citation?
3. Are different types of citation used in different rhetorical sections?
4. What reporting verbs are used most commonly in theses in Agricultural Botany and theses in Agricultural and Food Economics?
5. How common is the use of direct quotation from other texts in both parts of the corpus, and how frequently do writers elaborate on the citation beyond the sentence?
6. Are there differences between the two parts of the corpus in patterns of language around particular citation types?
7. Is there variation in the choice of voice and tense in different theses?
8. How does the data for the thesis corpus compare with the data from Hyland's corpus and the ResArt article corpus?

My expectation was that the Agricultural Botany theses, because they are scientific texts, would make greater use of non-integral citations, so that the focus would fall on the subjects of enquiry rather than the researchers themselves, and that they would feature higher use of research verbs, as the work of the researchers is mainly empirical experiment-based. As for the Agricultural and Food Economics theses, I did not have firm expectations as there is no literature on citation practices in economic or applied economics, but I expected that the Agricultural and Food Economics writers would elaborate on citations more because it is more of a discursive subject area (as indicated by the Agricultural and Food Economics supervisors who were interviewed in Chapter 5), where ideas must be discussed at some length. I also expected to find more quotation in these, because a discursive subject pays greater respect to the words in which ideas are articulated.

As for voice and tense, I expected the Agricultural Botany writers to use predominantly past tense, to **report** the studies of others, and their findings, and for there to be a preference for passive voice. The opposite, I anticipated, would be true of the Agricultural and Food Economics writers.

6.5 Data analysis

In the following sections, the results of the tagging are presented and analysed. The full sets of data can be found in Appendix D, in the form of tables, containing the raw figures for instances of each citation type, chapter by chapter, for each thesis. The figures presented in the sections below are aggregated figures (instances per 1,000 words) as this allows comparison between files.

After the data for the RABET corpus has been analyzed, the next section of the chapter presents discussion of a set of examples of chapters from theses in either part of the corpus, for illustration of the rhetorical functions that citations play in the texts, and this is followed by a comparison between the RABET corpus and two corpora containing articles.

6.5.1 Citation density between the two departments

Table 6.1 shows the ratio of non-integral to integral citations in the two sets of theses. As expected, in the Agricultural Botany theses, the preferred style of citation is non-integral, with the prominence placed on the information rather than the authors or researchers. The reverse is true of the Agricultural and Food Economics theses, where there is more attention paid to the names of the authors.

	RN %	RI %
Agricultural Botany	67	33
Agricultural Economics	38	62

Table 6.1 The ratio of non-integral (RN) to integral (RI) citations in the Agricultural Botany and Agricultural and Food Economics theses, shown as percentages

Figure 6.1 shows the relative frequency of use of the main citation types per 1,000 words of text. This shows very clearly the preference amongst Agricultural Botany writers for the RN Source and RN Ident types of citation. Contrary to expectation, the Discourse Verb type was more common than the Research type. However, this may be due to the difficulty in tagging verbs as Research or Discourse. The verb ‘find’, following Thompson and Ye, was tagged as a Research Verb, while the verb ‘report’ was tagged as Discourse Verb, because reporting involves verbal activity. It was clear, though, that for some passages of text, the two were used interchangeably, as in:

> Conlon *et al.* (1985) **found** that seed piece origin affected tuber size distribution but did not observe differences in the total yield or tuber specific gravity. McKeown (1990a,b) **reported** that basal (heel-end) seed pieces emerged later than shoots on apical (rose-end) pieces or whole tubers. Kleinhenz and Bennett (1992) **found** inconsistent responses to seed size in the cultivars Monona and Atlantic. Schotzko *et al.* (1984), Seyedbagheri and Neufeld (1987) and Nielsen <I>et al.</I> (1989) **reported** that cut seed performance depended upon cultivar and the size of seed from which the pieces were cut. (TAB-008, Chapter 2, my emphasis)

As shown in 6.5.4 below, *report* and *find* are the two most frequently used reporting verbs in the Agricultural Botany theses, so this problem of apparent interchangeability of the two verbs creates doubt over the validity of the distinction between research and discourse verbs. These categories need to be tested further, and one avenue for exploration is a test of what writers themselves perceive the differences between the reporting verbs to be. As an analyst, I suspect that the writer in the previous example was treating the two verbs as interchangeable, but it is not certain what the writer himself thought.

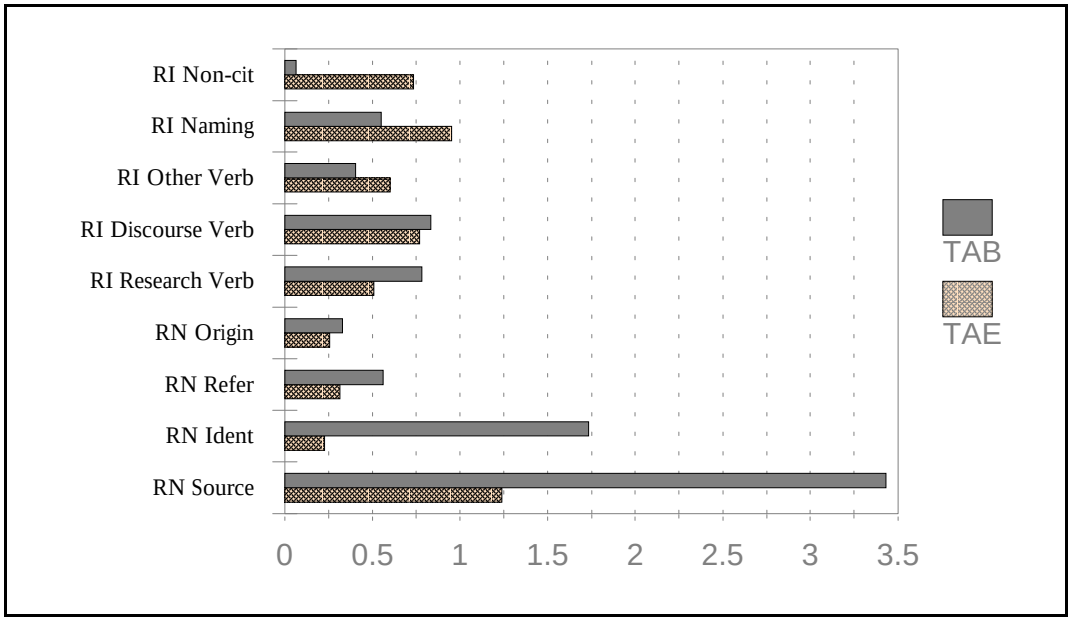

Figure 6.1 Proportion of citation types used in the two disciplines, using aggregated figures (instances per 1,000 words)

The Agricultural Economists make far greater use of integral Naming citations (reasons for which become apparent in 6.5.4 below) and also make more mentions of names without giving full citational information. At the same time, it should be noted that the

citation type most commonly used by them is RN Source, but that overall the Agricultural and Food Economics writers make fewer citations per 1,000 words than the Agricultural Botany writers.

6.5.2 Citation types in theses within same department

As can be seen in Figure 6.2 below, the density of citations in the individual Agricultural Botany theses varies from just under 5 per 1,000 words (TAB-005) to around 13 (TAB-002, and TAB-006). TAB-008 uses Verb-controlling citation types far more than any of the other writers, and far fewer non-integral citation types. Examination of this thesis reveals that the writer makes frequent reference to individual studies and compares their findings to his own experiments (X found this, and Y reported this. My findings were ...). TAB-007 by contrast, uses predominantly non-integral citation forms, and prefers to make information prominent through use of the Identification citation rather than the integral Verb controlling type. TAB-007 is a report of a laboratory-based investigation of innovative techniques for isolation of vacuoles, and therefore the emphasis is on the techniques, and the subject of study, that is, the vacuoles.

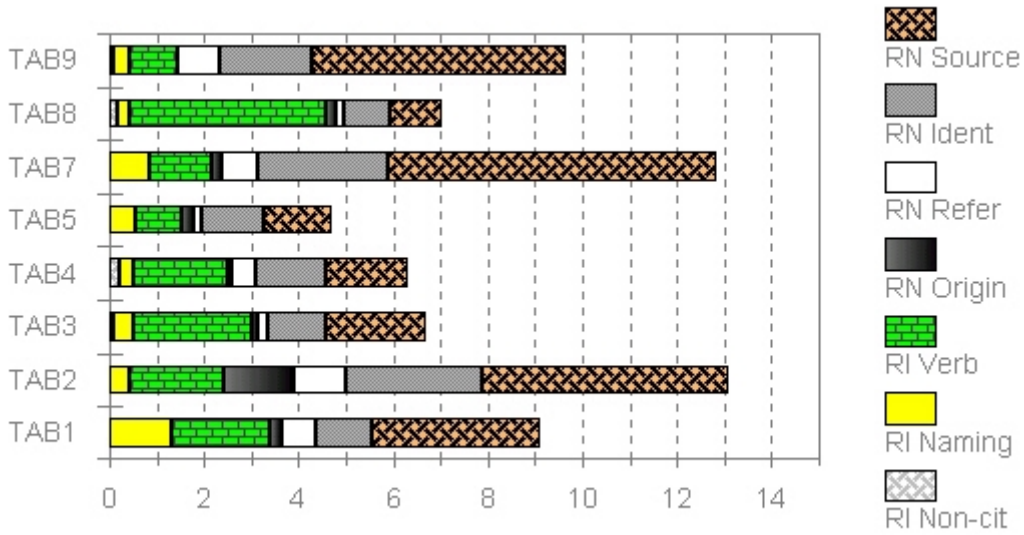

Figure 6.2 The average number of different citation types per 1,000 words of text found in the 8 Agricultural Botany theses. RI Verb contains all the RI Research, Discourse and Other verb instances

As for the Agricultural and Food Economics theses (see Figure 3), the density varies from just over 2 per 1,000 words to nearly 8 in TAE-007.

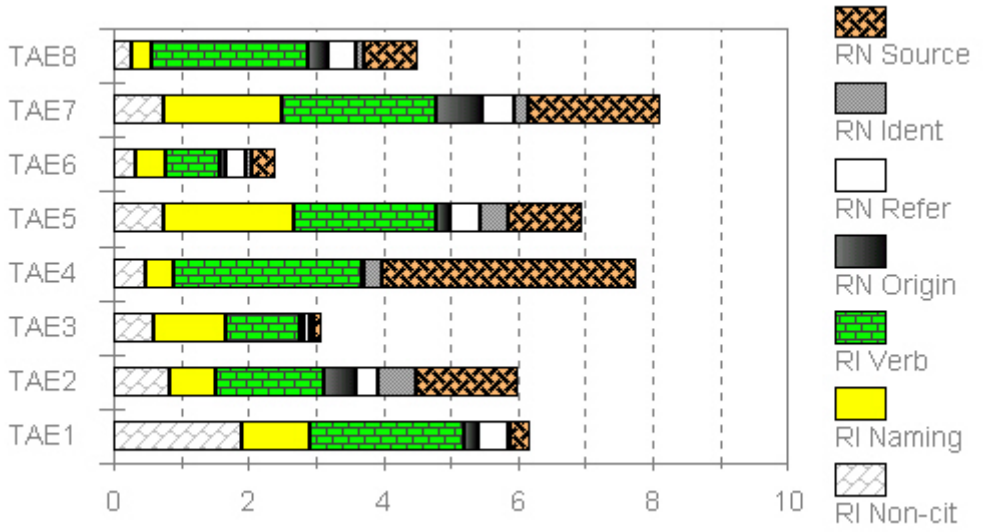

Figure 6.3 The average number of different citation types per 1,000 words of text found in the 8 Agricultural and Food Economics theses. RI Verb contains all the RI Research, Discourse and Other Verb instances

Here the variation between theses is perhaps even more marked than in the Agricultural Botany theses. TAE-004 is remarkable in having a very high ratio of RN Source citations, with a large number of Verb-controlling citations too and not many others, and TAE-005 and TAE-007 have a comparatively high number of RI Naming citations. TAE-001, TAE-003 and TAE-006 contain relatively low uses of the RN Source citation. Reasons for these differences will be suggested in the following sections.

Different writers within one discipline, then, take different approaches to research, and their rhetorical choices are, to a degree, determined by the nature of the research that they conduct.

6.5.3 Citation types by rhetorical section

In the Agricultural Botany theses, it was possible to divide the texts into four types of rhetorical section, following the conventions that are common in most scientific reports: Introduction - Methods - Results - Discussion. As can be seen in Table 6.2, there is considerable variation in the different sections of the theses, with relatively low use of citations in the Methods and Results sections of the thesis, and a markedly different set of citation types in the case of the Methods sections. To understand these variations, it is helpful to think of the hourglass model proposed by Hill *et al* (1982): the Introduction and Discussion sections of an article take a broad view, relating what is known in the field at large, while the Methods and Results sections are narrow,

focussing on the research itself. While the Introduction and Discussion sections contain many references to other studies to establish the current state of knowledge and where the current report fits in, the Methods section contains mainly references to the methods and techniques of others.

Section	*Density (per 1,000 words)*	*Most common types of citation*
Introduction	15.6	Source, Identification, Verb controlling
Methods	2.3	Refer, Origin, Naming
Results	2.4	Source (52%)
Discussion	10.1	Source, Identification, Verb controlling

Table 6.2 Citation types in different rhetorical sections of AB theses

This data shows that there is, then, variation in the density and type of citations used in different rhetorical sections of a thesis, and similar variation has been found across rhetorical sections in Physics, Chemistry and Biology masters' dissertations (Hanania and Akhtar 1985).

Because of the lack of a clear rhetorical framework that could be applied to all the Agricultural and Food Economics theses, a comparable analysis of the Agricultural and Food Economics theses could not be conducted. However, a comparison of the dispersion of citations across the whole length of a thesis in both sets of theses can be made.

Following the idea of the hourglass model, one would expect that the Agricultural Botany theses would contain a higher proportion of citations at the beginning and end of the thesis – the subject of the thesis is placed firmly within the literature at the beginning of the thesis and then the results and findings that emerge from the study are set within the literature at the end of the thesis to see which problems have been solved and which remain to be investigated. Figure 6.4 below indicates that this expectation is correct. In Figure 6.4, the comparative densities of citation per chapter are shown with the left most figures representing the first chapter and the right most being the final

chapter. The intervening chapters have been aggregated (in one theses, there were only two intervening chapters, while with others there were between four and eight). The introductory and final chapters clearly contain a higher density of citations.

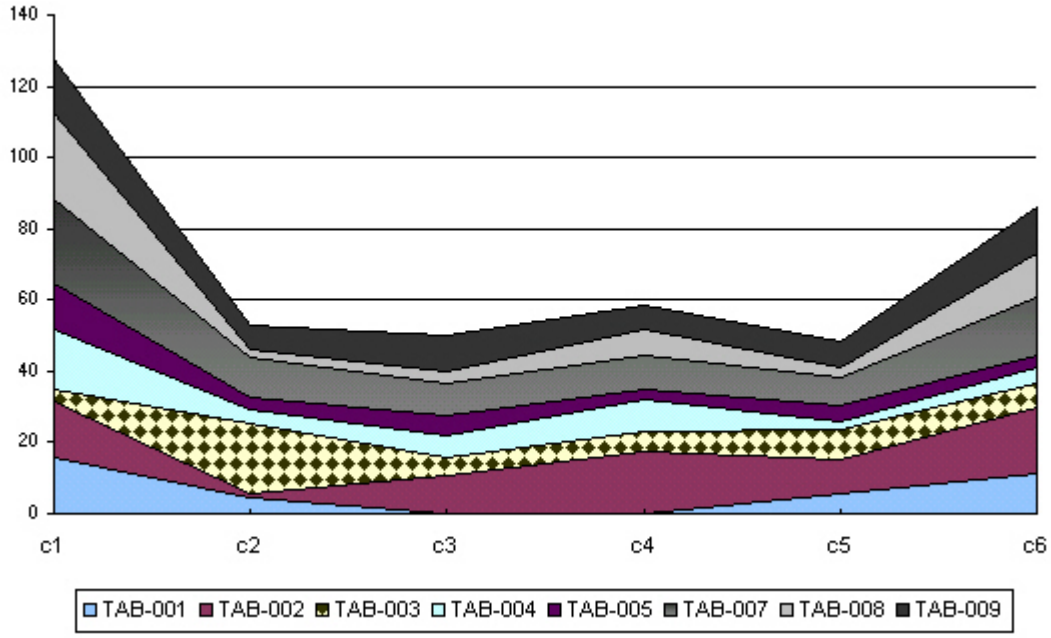

Figure 6.4 The density of citation by chapters in the Agricultural Botany theses. The first chapter on the left is Chapter 1, and the last one to the right is the Conclusions chapter

It should be pointed out, however, that the density of citations in the concluding chapter of the Agricultural Botany theses should not be taken to be typical of science theses in general, since Bunton (1998) found, on average, low use of citations in the concluding chapters of the science, medicine and engineering theses in his study. In three out of the six theses, there was no citation at all. As precise details are not available, it is difficult to know which disciplines were represented in his corpus.

Figure 6.5 shows the picture for the Agricultural and Food Economics theses, and the difference is remarkable. The first two chapters are dense with citation but the final chapters contain less and less, until the final chapter which contains on average 2 citations per 1,000 words.

Evidently, the Agricultural and Food Economics theses are constructed quite differently from the Agricultural Botany theses in terms of the way that they refer to other texts. How is this to be explained? In 2.4.7, reference was made to Bloor and Bloor's (1993)

study of hedging in economics articles, and their observation that in Economics it is the development of new models that is central, rather than the production of new knowledge. A possible explanation, then, is that the Agricultural and Food Economics theses are primarily concerned with model-building. Models (or frameworks) are developed out of investigation and criticism of the shortcomings of former models (a form of 'identifying the gap'). The model might be mathematical, or it might be conceptual or analytical. Once it has been developed, it is tested on data, and the data too has to be identified in its origin (therefore there must be further citation). On completion of the testing, however, there is little need to place the findings within the ambit of the larger literature on the subject. The model was developed to solve a problem, to answer some questions, and it can be assessed in terms of its success in helping to answer the questions.

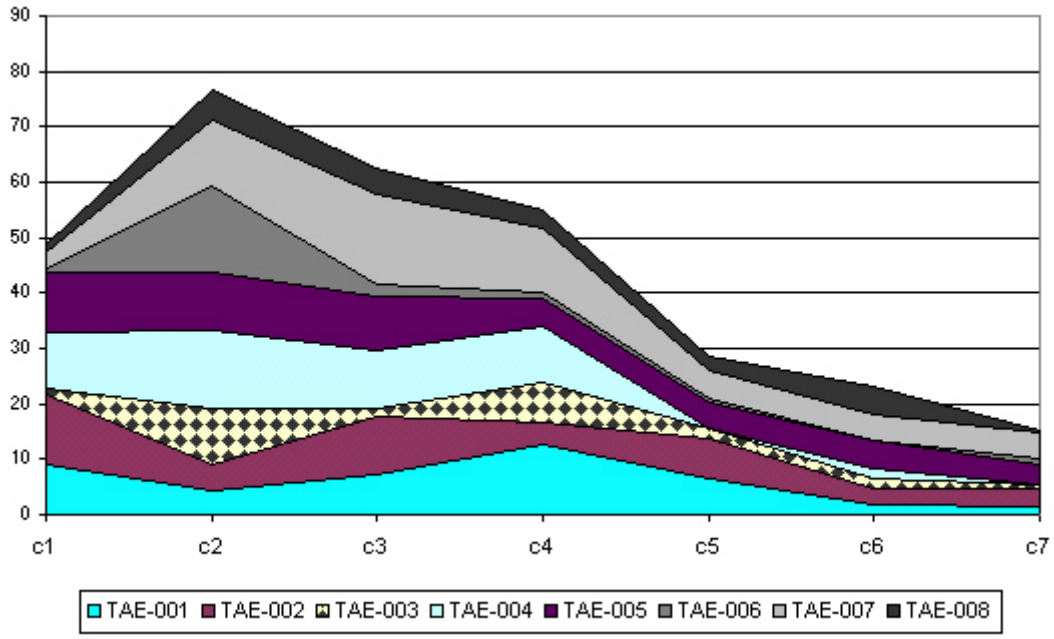

Figure 6.5 The density of citation by chapters in the Agricultural and Food Economics theses. The first chapter on the left is Chapter 1, and the last one to the right is the Conclusions chapter

It is also worth remembering that Holmes (1997) (as reported in 2.4.1 above) suggested, on the evidence of his move analyses of History, Political Science and Sociology research articles, that social science research articles, unlike those in natural sciences, tend towards complexity and elaboration at the beginning of the articles rather than at the end. The lower degree of manifest intertextuality in the Conclusions chapters of the Agricultural and Food Economics theses might reflect a more wide-

ranging difference in how knowledge is constructed in the two domains of the social and the natural sciences.

6.5.4 Reporting verbs

The Agricultural and Food Economics writers used a much wider variety of reporting verbs: 171 different verbs compared to 99 different verbs used in the Agricultural Botany theses. The most commonly used verbs in either section of the corpus are shown in Table 6.3 below.

Perhaps not surprisingly, the second most common verb in both sections of the corpus is *find*. In the Agricultural Botany theses, however, *find* is often used interchangeably with *report* (this is a good example of the problem of defining a verb as research or as discourse – *report* is typically thought of as a verbal process, and therefore a Discourse Verb, while *find* is tagged as a Research Verb even though for the writer the two may be virtually synonymous). If *find* and *report* are combined, this suggests that a quarter of the Agricultural Botany reporting verbs are reports of findings. Two verbs implying previously established knowledge (*demonstrate* and *show*) account for 14% between them.

Agricultural Botany			Agricultural Economics		
report	*85*	**12.6**	use/employ	*69*	**7.2**
find	*83*	**12.3**	find	*65*	**6.7**
show	*68*	**10.0**	suggest	*49*	**5.1**
suggest	*35*	**5.2**	describe	*43*	**4.5**
describe	*33*	**4.9**	note	*38*	**3.9**
use	*28*	**4.1**	estimate	*31*	**3.2**
demonstrate	*27*	**4.0**	provide	*31*	**3.2**
review	*18*	**2.7**	propose	*28*	**2.9**
propose	*17*	**2.5**	report	*27*	**2.8**
discuss	*15*	**2.2**	refer	*21*	**2.2**

Table 6.3 The ten most commonly used reporting verbs used in the theses. The first figure indicates the number of instances of the verb and the figure in brackets expresses this as a percentage of all the reporting verbs used in the group of texts.

In the Agricultural and Food Economics column, there are two closely related verbs, *use* and *employ*, which account for 7.2% (I have combined them in this table, as 'employ' is taken to be virtually synonymous to 'use' and it is only used by one writer

out of the eight), as opposed to 4.1% for *use* in the Agricultural Botany theses (there are no instances of *employ*). *Use* in the Agricultural and Food Economics theses is often used in relation to equations, functions and other measurement techniques, and this is indicative of the greater tendency of the Agricultural and Food Economics theses to deal with the discussion of frameworks and methodologies.

6.5.5 Quotation and elaboration

Table 6.4 shows the average use of direct quotation per thesis in either part of the corpus. These figures are slightly distorted as they do not take into account the unequal lengths of the theses; if we remember that the Agricultural and Food Economics theses are double the length of the Agricultural Botany theses, we could halve the ratio to get a more accurate indication of the relative uses of quotation and elaboration (1:8 and 2:3 respectively). The frequency of use of direct quotations from other texts is markedly higher in the Agricultural and Food Economics theses, and would be higher if not for an exceptionally high number of direct quotations in one Agricultural Botany thesis, which has 11 out of the total 19 quotations in that set of theses. This suggests that the words of other writers are considered important expressions of ideas in the Agricultural and Food Economics theses, which as we have already observed are far more discursive than the Agricultural Botany theses. Once again, we see the emphasis placed upon verbal expression and the texts of others, in a world in which models, ideas and issues are prime and the texts in which they are embodied.

	DQ	Elab
TAB	**1.88**	**12.75**
TAE	**28.63**	**36.13**

Table 6.4 The average numbers of direct quotations and elaborations in either part of the RABET corpus

6.5.6 Patterns of language

Close inspection of the different kinds of Naming citation in the theses revealed interesting differences in the discourses of the two disciplines. Firstly, in terms of simple frequency, it can be seen from Table 6.5 that this citation type is much more commonly used (by more than four times) in the Agricultural Economics texts.

RI Naming	Total occurrences
AB	116
AE	484

Table 6.5 The number of occurrences of Naming citations in each of the two groups of texts

In order to find out why this might be the case, concordance lines of the Naming citation type were examined. It was observed that certain patterns were regularly used, such as the three shown in Table 6.6.

	Agricultural Botany	***Agricultural Economics***
...in X (1991)	3 (12)	58
...of X (1991)	37 (154)	70
...by X (1991)	25 (104)	29

Table 6.6 The number of occurrences of a pattern in the thesis corpus of preposition + naming citation. In the middle column the figure in brackets shows an adjusted figure which would make the amount equivalent to the figure in the right column (n*484/116).

The pattern “in X (1991)” is clearly much more commonly used in the Agricultural Economics theses. The use of the preposition “in” indicates that the citation is a reference to a book, and this is supported by the examples given in Table 6.7.

Agricultural Botany		*Agricultural Economics*	
findings	of X	listed	in X
work		found	
method		extrapolated	
technique		quoted	
work	by X	views	of X
experiments		suggestions	
studies		model	
		function	
		exposition	

Table 6.7 Frequent patterns involving in ‘in’, ‘by’ and ‘of’ in the theses.

In the Agricultural Botany theses, on the other hand, “of” and “by” are more commonly used and these tend to refer to the research actions, findings, methods, and techniques of other researchers. Where Agricultural Economics thesis writers use “of”, it is noticeable that this also includes discourse nouns, such as views and suggestions, which are not found in the Agricultural Botany texts. The Agricultural Economics writers, therefore, appear to be concerned with the texts and concepts of others, while the Agricultural Botany writers make reference to the research activities and techniques of other scientists.

6.5.7 Voice and tense

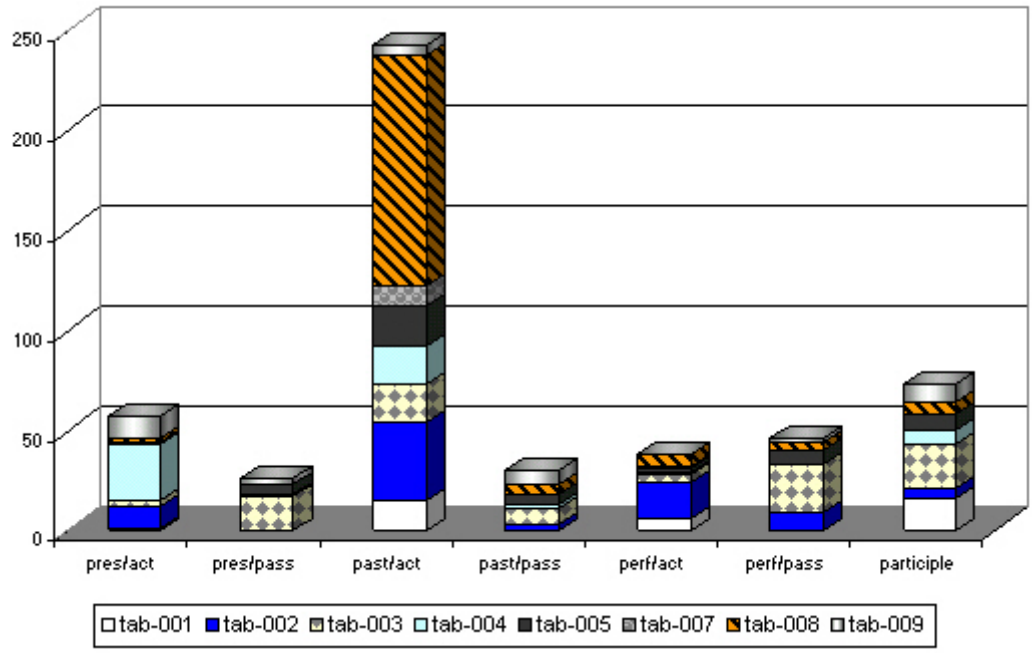

Figure 6.6 The numbers of instances of tense + voice combination in the Agricultural Botany theses

Figures 6.6 and 6.7 show the relative uses of tense and voice in the two parts of the corpus. It can be clearly seen in Figure 6 that, in the Agricultural Botany theses, by far the most frequent choice for the reporting verb is past/active (the third column from the left) which accounts for nearly half the reporting verbs. Although the participle form comes second, and the present active is third, half of the uses of present/active feature in one thesis: TAB-004. This thesis is remarkable among the Agricultural Botany theses for its discursive tone, with extended discussion of different perspectives on a complex set of relationships.

Counter to my expectations, the preferred voice was active rather than passive, and we only find passive being more common in conjunction with perfective aspect. A possible explanation for this is that the perfective is used to refer to general areas of enquiry (see 6.2.2 above) and the writer is likely to emphasise the area of enquiry, rather than the cited researchers, by placing it in theme position in the sentence. As would be expected, the predominant use of the verb controlling citations is to report what other researchers found or did.

In the Agricultural and Food Economics theses, on the other hand, the present/active is the most common choice, accounting for over half of the reporting verbs, and, though

the past/active is second most common, the majority of the instances occur in one thesis (TAE-004). Passive voice is a dispreferred choice in all tenses (40 for present, 23 for past, and 12 for perfective, out of a total of 960). In TAE-008, present/active accounts for over half of the instances, but past/active accounts for a third, which is a high figure comparatively.

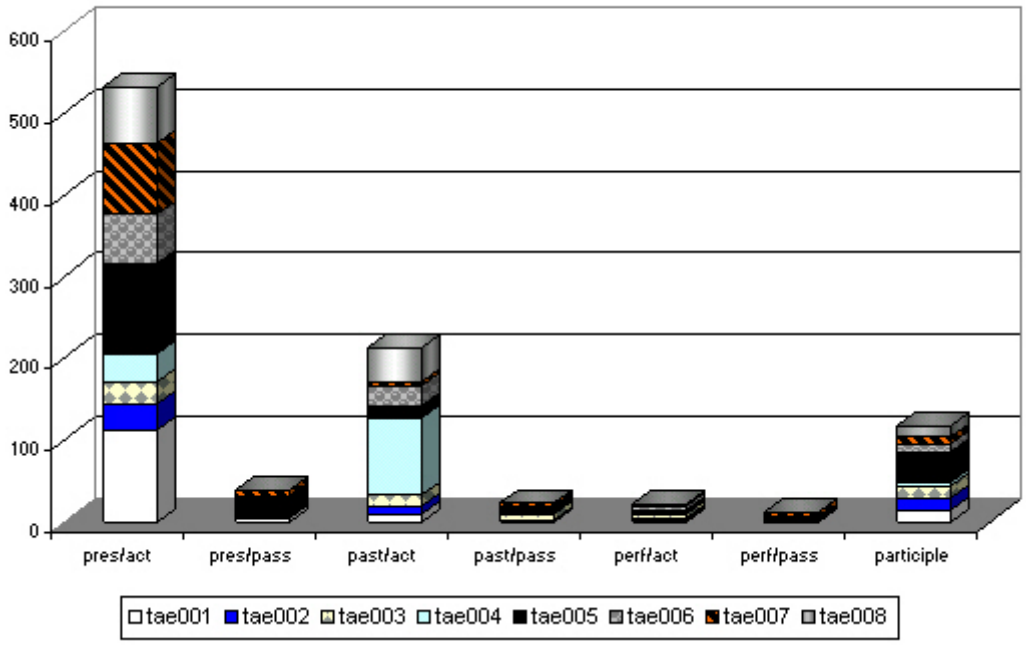

Figure 6.7 The numbers of instances of tense + voice combination in the Agricultural and Food Economics theses

The predominant use of the present simple, with active voice, in the Agricultural and Food Economics theses is worth some attention. As shown in the review above (6.2.2), the literature on tense choice in academic writing has focused mainly on scientific prose, where the tendency is to report experimental studies, and their findings. The Agricultural and Food Economics writers however are working in a different paradigm, where the citational present is the preferred choice. The use of the present tense can be seen to animate the discussion, to create the impression that the ideas are alive. Another aspect of the choice may be that the research is performed through the writing of it, and the emphasis is on a world of texts, rather than a world of experiments, and within texts, the textual world is always present, as it were, always now.

Agricultural Botany	Agricultural and Food Economics
Active : Passive :: **3.3 : 1**	Active : Passive :: **10 : 1**

Table 6.8 The ratio of active to passive voice constructions in reporting verbs in the two groups of texts in the RABET corpus

As for voice, the surprising finding was that in all cases the preference is for active voice but we find that the passive is more common in Agricultural Botany theses, as is shown in Table 6.8.

Nearly half of the uses of the passive voice in the Agricultural Botany texts occur in one thesis: TAB-003. This choice seems to be quite idiosyncratic: that is, the writer of this thesis has a personal preference for use of the passive, even though the information that is being reported is similar to that in TAB-009, where the preferred form is the active voice. In the case of the Agricultural and Food Economics theses, the most common use of the passive is with the present simple (40 cases out of 75 passive constructions).

6.6 Textual analysis examples

In this section I take three chapters from the Agricultural Botany, and three from the Agricultural and Food Economics theses, and illustrate the uses of the different citation types within the chapters. For the Agricultural Botany theses, I have chosen an Introduction chapter, a Literature Review chapter, and a Conclusions chapter. For the Agricultural and Food Economics theses, which do not have clearly conventionalised rhetorical organization, as we have seen, I have chosen a review of different approaches, an historical contextualisation, and a theory chapter. These particular chapters have been chosen because they elucidate some of the main functions that citations perform in both the Agricultural Botany and the Agricultural and Food Economics sets of theses. The purpose here is to demonstrate in more concrete form the different rhetorical purposes that the citation types can play. The complete tagged chapters can be found in Appendix B.

6.6.1 Agricultural Botany Introduction: TAB-005

An Introduction to an Agricultural Botany thesis typically summarises what is already known about the subject of the research, and then identifies a gap in the current state of knowledge, and establishes the niche for the current set of studies, following a Swalesian 'Create a Research Space' model. TAB-005 starts off with an account of the history and distribution of two banana leaf diseases, which are identified as having serious economic consequences for banana-growing countries. This is followed by sections on the symptoms, economic importance, epidemiology and control of the diseases. Throughout the sections, there is dense citation, with approximately half of the sentences carrying a citation. The most common type is RN Source, with the proposition (what is known) attributed to a source:

> Streaks usually appear 14-21 days after inoculation, first along the left margin of the leaf (Stover, 1987)

There is also frequent use of the RN Ident citation type, which allows the writer to focus attention on the information in the sentence rather than to the person that made the observation:

> A similar situation where pathogenic variability in M. *musicola* appears to be related to altitude has also been observed in Colombia and Costa Rica (Merchan, 1990).

Where the name of the researcher is introduced into the sentence, the preference is for passive voice, so that the particular topic is thematised, and the researchers, or texts, that provide the relevant information are in rheme position:

> Chemical control of Sigatoka up to the early 1970s has been described by Meredith (1970), Stover (1971a; 1972a), Wardlaw (1972), Long (1973) and most recently by Stover (1990), where a comprehensive historical account is given.

It is only in the final thirteen paragraphs of the chapter, when the writer summarises the research needs, presents the plans for the research, and then outlines the rest of the thesis, that citations are no longer needed. The three citations that do appear refer to other groups of researchers who have developed new methods, that have been highly important in the approach taken in the current work:

> Two groups published similar work almost simultaneously. One group described genetic mapping applications and called the new assay RAPD analysis, for random amplification of

> polymorphic DNA (Williams, *et al*., 1990), and the other group focused on genome fingerprinting and called their assay AP-PCR for arbitrarily-primed PCR (Welsh & McClelland, 1990).

Once again, the relegation of the names to the citation allows attention to rest on the method rather than the people involved.

6.6.2 A complete chapter: TAB-002

In many of the Agricultural Botany theses, the Introduction section to a chapter contains a review of the literature that relates to the particular set of experiments described in that chapter. In Chapter 4 of TAB-002, the use of RFLPs as a methodology to estimate genetic distance between inbred lines of sunflower is tested. The writer summarises what uses of this methodology have revealed about different plants, and also what the methodology is, and why it is used. Names for methods, theories or for phenomena are attributed to the originator (thus tagged 'RN Origin'):

> Heterosis (Shull, 1914) or hybrid vigour (Shull 1908, 1911; East 1908) is said to have occurred when the offspring of a cross exhibits a quantitative trait of higher value than either of the parental lines.

Compared to TAB-005, there is much more foregrounding of the names of researchers, as in:

> Numerous studies have now been carried out in maize to test this theory, but the results have been contradictory. Lee et al. (1989) and Smith *et al.* (1990) both reported a close association between hybrid performance and GD based on RFLP data. Conversely, Godshalk *et al*. (1990) and Melchinger *et al.* (1990) concluded that RFLPs are of limited value in predicting heterosis.

The reason for this difference is that the writer is comparing the findings of different studies. In the first sentence, 'numerous studies' is in theme position, and thus the names of the studies are placed in theme position in the following sentence. The RI sentences in this case are also examples to support a general observation and this is a frequent pattern in TAB-002, as in the following:

> RFLPs are ideal markers for germplasm identification, as they meet all these criteria for inbred crops. For example, Evola et al. (1986) showed that maize RFLP markers were stably inherited over a number of generations and Beckmann and Soller (1983)

> demonstrated the discriminatory power of RFLPs by calculating that the probability of distinguishing 20 inbreds from each other was 0.99, when considering only 2 alleles at 20 loci occurring with a frequency of 0.5.

The findings of previous studies are used here as evidence to support the general statement in the first sentence. We may note also the use of past/active in these reports of findings, which fits with the observations of Malcolm (1987) (see 6.2.2 above) that generalizations are stated in the present tense, and reports of specific experiments are in the past simple. Malcolm remarked of the present perfect that it tends to be used in general reports of work in a field of enquiry and we find an example of this in TAB-002:

> Molecular phylogenies have now been constructed for a number of genera including Brassica (Song *et al*. 1988), Lycopersicon (Miller and Tanksley,1990b), Solanum (Debener *et al.* 1990), Helianthus (Gentzbittel *et al.* 1992) and Beta (Jung *et al.* 1993).

This is a common pattern of reference in theses which report a number of studies on different plants (such as TAB-008 which compares work on different varieties of potato). The citation type used is tagged as RN Ident (because the citations identify the agent for the verb 'construct') and the use of RN Ident is common in these listings of the different studies. The advantage of this form of citation is that it is economical – a single main clause conveys a large amount of information, and this is typical of the condensed style of writing that is used in the most of the Agricultural Botany theses.

Finally, it should be pointed out that a third of the citations in this section of Chapter 4 are RN Source citations that attribute the proposition to other researchers.

In the following section, the materials and methods for the analyses that were made are described, in three paragraphs. There is a single RN Origin citation and one RN Refer citation, typical of Methods sections, to indicate the creator of a software program (SAS) and to explain where further information on a particular method can be found (thus obviating the need for a detailed explanation in the chapter).

In the Results section, there are six paragraphs, and only two citations. One is an RN Refer, again referring the reader to another text to find further details of a technique, and the second is an RN Source citation which acts as a warrant for a causal

explanation ('these "heterozygotes" may be due to a mixture of alternate fixed alleles as seen in rice inbreds (Wang and Tanksley, 1989)').

The Discussion section is dense with citations as the writer places the findings of his analyses within the research literature. The majority of the citations are RN Source and RN Ident, although variation is provided by the use of RI Research Verb constructions, and the occasional use of a Naming citation, as in:

> For example, Liu *et al* (1990), showed …
>
> A recent study by Tersac *et al* (1994) also showed …

The reason for choosing an integral citation form may be due to a desire to vary the patterns used in sentence structure, but an alternative explanation is the writer is constrained in his choice by theme-rheme relations, as the following example demonstrates:

> The mean PI for a sunflower RFLP marker was found to be 0.49 and this value can be used to compare the level of DNA polymorphism between different crop species. For example, the mean PI for a soybean RFLP probe on adapted germplasm was found to be 0.3 (Keim *et al* 1992).

It is necessary to place 'the mean PI for a soybean RFL probe on adapted germplasm' in theme position in the sentence (rather than 'Keim et al found that the mean PI …') because this measurement is being compared to the mean PI measure that the writer made. Again, we notice that the writer prefers to focus attention on the information rather than on the author or researcher.

6.6.3 Conclusions: TAB-007

The final chapter of TAB-007 is actually titled 'Discussion' but it can also be described as a concluding chapter. As we have noted in 6.5.3 above, the final chapter of an Agricultural Botany thesis, like the final section of an IMRD paper, has a high density of citations, comparable to the number of citations in the Introduction section. These are predominantly non-integral citations (29 out of 37 citations), and the RN Source type is once again the most common.

In this chapter the writer reviews the current state of knowledge of the subject of the thesis, the enzyme ACC oxidase. The main purpose is to state the problems that have not been resolved yet, both those that were recognised before, and those that have been identified by the writer's research work. It is interesting to note that there are few references to what the present set of studies have achieved other than this single example:

> In support of this proposal it has already been shown (Chapter 2) that the hydroxamate containing salicylhydroxamic acid is a potent inhibitor of ACC oxidase.

Other thesis writers, especially those in Agricultural and Food Economics, begin their concluding chapter with a summary of the previous chapters, and they stress the findings of their own work. In this thesis, as in TAB-001, the other notably short thesis, the writer presumably considers that the main findings of the research do not require recapitulation because the text is not excessively long. The expectation is that the readers will recognize and remember what the main findings have been. Instead the emphasis is on what is known about the enzyme and what remains to be investigated. In the following example, the first two sentences present the current state of knowledge about the inactivation of ACC oxidase, with typical RN Source citations as attributions for the origin of the information:

> A number of enzymes, are susceptible to inhibition by hydroxamate containing compounds, which may have several modes of action (Rich et al., 1978, Chamulitrat et al., 1992). In particular, hydroxamates have a high affinity for metal ions, especially Fe3+ (Rich et al., 1978) …
>
> Our understanding of ACC oxidase is still expanding, and the absence of an undiscovered cofactor could account for the inactivation observed … obviously considerably more work must be completed to elucidate the mechanism responsible for the inactivation of ACC oxidase observed.

The tone of the latter sentences is interesting. In choosing to talk of 'our understanding', the writer is suggesting a shared position, in a research community, with the reader. This is then strengthened by the use of the 'obviously', which implies that the writer is forestalling the reader's reaction, and is able to, through a shared understanding. This confident show of membership in the research community demonstrates that the writer is addressing a readership of fellow researchers, which

supports the observations by the supervisors (see Chapter 5 above) that theses in Agricultural Botany should be written for the research community, rather than for the examiners, and should be written in a tone of parity.

6.6.4 Reviewing different approaches: TAE-001, Chapter 4

This Agricultural and Food Economics thesis is an assessment of agricultural and food marketing systems. Chapter 4 briefly reviews some of the practical attempts to investigate marketing, starting with general approaches and then concentrating on particular work by French marketing system researchers.

The dependence on the words of the researchers reviewed for definitions and observations is evident in the fact that there are 13 direct quotations in the chapter, and eight of these are extended quotations. Many of the frameworks and ideas referred to appear to be textbooks rather than journal articles and there is frequent use of the RI Naming citation type in which the nominalised citation is a book or a person:

> The most comprehensive textbook in this regard, although again not suggesting methodologies for approach, is Rhodes (1987).

> Foxall (1980) is the other main writer to have addressed marketing systems in 'Agricultural Systems'.

Unlike the Agricultural Botany theses, in the Agricultural and Food Economics theses, researchers and their words are foregrounded. Researchers also lend their names to particular models:

> …Porter's (1980) model…

Because the chapter reviews the approaches taken by different authors, an author is introduced and then the following sentences elaborate on the review (there are 9 cases of elaboration in the chapter), and the following appearances of the name are not given citational detail:

> Koulytchizky (1983) gives the reasons for the use of the filière concept (especially in France and especially in French agri-food marketing chains) … Koulytchizky suggests this as a further reason why …As Koulytchizky puts it, the agri-food filière touches everyone …

This elaboration on an initial citation is typical in the Agricultural and Food Economics theses, which are principally concerned with approaches and models.

The verb-controlling citations, as can be seen in the last example, are in the present simple with active voice, so that the author's name is usually thematised.

6.6.5 The facts of the matter: TAE-002, Chapter 3

By way of contrast, the chapter that we examine in this section contains more non-integral citations than integral citations, and the most common type, like the Agricultural Botany theses, is the RN Source type. The reason for this is that this chapter assesses the costs and benefits of a research programme conducted in the past. The first section of the chapter provides the background to the study, explaining the context and the facts of the programme. As this is an historical account, the most common use of citation is to attribute a piece of information to its source:

> Permethrin is applied at a rate of 0.1 kg of active ingredient/ha while Monocrotophos (an organophosphate, OP) was applied at 0.5-0.8kg of active ingredient/ha (Hirano, 1989).

As with TAE-004, the need to recount past events and provide background information leads to a higher than normal (relative to all Agricultural and Food Economics theses) use of the RN Source type of citation.

The past is mainly reconstructed from other accounts, in other words, from other texts, and these events have been discussed, assessed and debated by various writers, as well as the different methods used to calculate costs and benefits. It is not surprising, then, that the second most common citation type in this chapter is the RI Discourse Verb, as in:

> Wood Mackenzie (1987) describes the price falls as taking them to 'commodity' status.

Even in this discussion of past events, the chosen tense for the reporting verb is present simple.

In the Methods section of the chapter, the number of citations drops markedly (note the similarity to the Agricultural Botany theses). The RI Naming citation is used to identify particular approaches with their proponents, and, if they are rejected, evidence to

support the decision may be supplied with a reference to the work of another researcher, who is given prominence with a Verb-controlling citation type:

> The best available approach to non-horizontal supply curves is that of Lindner and Jarret (1978). However, this was not used here … Wise (1984) has shown that the use of Lindner-Jarret supply curves does not lead to national benefit estimates greatly different from those found from horizontal supply curves.

6.6.6 Econometric theory: TAE-005, Chapter 3

Our final example of an Agricultural and Food Economics thesis chapter comes from an econometric study. Chapter 3 is the Theory chapter, an exceptionally long chapter at over 23,000 words. There are 221 citations in all, of which 164 are integral. Theories and definitions are identified with individuals:

> Early definitions of factor bias are attributable to Hicks (1932) and Harrod (1948). These definitions are originally with respect to an assumed two input production process, Labour (L) and Capital (C).

One might compare this with a putative rephrasing in scientific style: 'Early definitions of factor bias (Hicks 1932; Harrod 1948) are originally with respect to an assumed two input production process, Labour (L) and Capital (C).' This reformulation would contain the same information but would fail to emphasise the names of the researchers. In the following sentences the definitions of Hicks and Harrod are elaborated on, and that is why the formulation cannot be changed. Furthermore, the name of Hicks reappears later in the chapter in 'Hicksian regularity conditions'; once again, the name of the researcher is given to the model, the function or the definition articulated by that person.

A common feature in the theory sections of Agricultural and Food Economics theses is the tendency to refer the reader to other texts for further detail on a particular topic. This can be done through use of the 'Refer' type of citation:

> These lower order moments potentially provide enough information to accurately specify an appropriate lag structure (see Silver and Wallace) with minimal priors.

Or with a recommendation to turn to a specific text (using the Naming citation):

> In the agricultural economics literature, a simple introduction to the theory and an application (to land price determination) of the most common testing procedures can be found in Hallam *et al.* (1992) with Lloyd and Rayner (1993) providing a valuable supplement in highlighting the empirical difficulties and shortfalls of the common procedures.

The tone of recommendation adopted is remarkable. If the audience for the thesis is the examiners, it would seem strange to recommend particular texts to people who are presumably relatively expert in the literature, especially with the insertion of evaluative terms such as 'simple' and 'valuable'. The best explanation for this would seem to be that the tone adopted here is not that of doctoral student addressing examiners, but that of a lecturer addressing a class of students. Though not a common occurrence, this is not the only instance of a lecturing style being used in the Agricultural and Food Economics theses (TAE-008's 'There are a number of excellent texts on DSS …' for example), and it suggests that, while we may see a text as interaction between a writer and an audience, the audience being addressed at all stages of such a long and complex text as a PhD thesis may not always be the same. Myers (1989:2) wrote of research articles, 'It is often hard, for a published text, to say who is interacting and what interaction is involved', but this seems to hold true of unpublished PhD theses too. In the case of the Agricultural and Food Economics writers, it is possible to detect past experience manifesting itself in the adoption of this tone: as noted in Chapter 5, several of the doctoral students lectured within the department.

This challenges the opinion of the supervisors as reported in Chapter 5, that the audience for the thesis is the examiners. It is possible that in Agricultural and Food Economics it is customary to recommend texts to one's audience regardless of whether or not they are expert, but the more likely gloss is that the writer is adopting an expository voice that is common to the discourse of the discipline without taking into consideration the immediate audience for the thesis. However, at other crucial points, such as in the Conclusions chapters, a more circumspect tone is adopted, not through citation but through qualification of claims, as will be discussed in the next chapter. The lecturing tone can perhaps be explained as a display, a performance of one's authority as if in front of an audience of students, which confirms one's credentials as a member of the community. This, however, is speculation and cannot be proved without further research.

6.7 Comparison with other corpora

Question 8 for this study was: 'How does the data for the thesis corpus compare with the data from Hyland's corpus and the ResArt article corpus?'. In this section the findings of Hyland (1999) are compared with the findings reported above and with analyses of the data from the ResArt corpus.

Hyland (1999) looked at citations in a corpus of 80 research articles, composed of 10 journal articles from different disciplines (see Table 6.9 below for details).

Discipline	Av. Per paper	Per 1,000 words
Mechanical Engineering	27.5	7.3
Physics	24.8	7.4
Electronic Engineering	42.8	8.4
Marketing	94.9	10.1
Philosophy	85.2	10.8
Applied Linguistics	75.3	10.8
Sociology	104.0	12.5
Biology	82.7	15.5
ResArt corpus		
Agricultural Botany	24.2	7.8
Agricultural Economics	30.3	6.7
RABET corpus		
	Av. Per thesis	
Agricultural Botany	248.8	9.04
Agricultural Economics	333.5	5.25

Table 6.9 Number of citations in Hyland (1999), RABET and ResArt corpora

Table 6.9 shows the figures for instances of citation in three corpora, with the middle column showing the average number of citations per text, and the right column showing the average number of citations per 1,000 words of running text. The first set of figures relates to Hyland's corpus, while the next set refers to the ResArt corpus, the special corpus that I assembled from articles that had been referred to frequently by the thesis writers in their theses (see 4.3.6 above, and Appendix D, for further details). The lower density of citations in Hyland's corpus amongst the science and technology

articles (7.3-8.4) contrasted with higher incidence among the social science articles (10.1-12.5) while Biology stood out as exceptional with 15.5. Hyland postulated a difference in practice here between “hard” and “soft” disciplines, but speculated that Biology stood out from the other sciences because it is a relatively new discipline.

Comparing only Hyland’s data with the data for the RABET corpus, we can see that the density of citations in the PhD theses is much lower. If we presume that the Agricultural Botany theses should be roughly comparable to the Biology articles, the density is approximately three fifths lower, and while there is no easy comparison between the Agricultural Economics and any of the disciplines in Hyland’s study, the figure of 5.25 is substantially lower than any of the figures for the research articles. However, if we now include the data from the ResArt corpus for comparison, we see that there is not such a large difference between articles written in the equivalent areas for Agricultural Botany and Agricultural and Food Economics as would be suggested by the simple comparison between Hyland’s corpus and the RABET corpus. This contrast between the Hyland and ResArt corpora is puzzling, and at least three explanations are possible:

- the assumption that the Agricultural Botany theses should be comparable to the Biology articles is unfounded
- the articles in the ResArt corpus are not the same types of article as those in Hyland’s corpus
- economics, and agricultural economics, are more similar to physical sciences than to ‘softer’ social sciences such as marketing

Each of these explanations is plausible. Hyland’s biology articles are all molecular biology, and may be most closely related in subject matter and research type to TAB-002 and TAB-007, which are the two of the Agricultural Botany theses with the highest density of citations (see 6.5.2). The second explanation is supported by the nature of the corpora: the Hyland corpus contains articles dealing with original research reports, drawn from respected international journals, whereas the papers in the ResArt corpus are more varied, incorporating conference papers, comment papers, explanations of mathematical models, as well as research reports. One of the research reports in the ResArt corpus is remarkable in that it is a technical paper, which contains no references

to literature at all. In this sense, the Hyland and ResArt corpora are not comparable; it is interesting to note, on the other hand, that the ResArt and RABET corpora are quite similar in profile. As for the third explanation, this can be supported by reference to widely-held beliefs in the economics field that economics is the social science closest to the physical sciences (cf McCloskey 1985 who argues that economics should not try to emulate the physical sciences), or, using Biglan's terminology, is furthest to the 'hard' end of the spectrum.

These figures are difficult to interpret because the ResArt and Hyland corpora are not comparable (both contain articles but the purposes of the articles are clearly different), because they deal with different areas of enquiry, and, thirdly, because the corpora are not large enough. What is apparent, though, are tendencies:

- the Agricultural and Food Economics theses have a low density of citations, even by comparison to the ResArt corpus
- the theses have a low density of citations compared to the Hyland corpus of research articles
- the theses are more similar to the ResArt corpus articles in overall pattern

The variation is most probably due to the variation of content and communicative purpose. The finding that the Agricultural and Food Economics theses have a low density of citation can be explained by the observation made in 6.5.3 above that those theses are predominantly concerned with the development, and application of models, and that once the model has been developed there is less need for reference to the literature. In an extended text such as the thesis, there should therefore be less need for citation in the latter stages, especially where the focus is on application and, in the case of 'weak theory'-driven research, in the gathering of primary data. An example of this is in Chapter 5 of the present study, where there are very few citations because the focus is almost exclusively on the interview data.

Table 6.10 shows the relative percentages of the two types of citation, integral and non-integral, in Hyland (1999), the ResArt corpus and the RABET corpus. The Hyland figures show that there is considerable variation in citation practice between the different disciplines, with Philosophy being the only discipline that prefers the integral form over the non-integral, and with greater emphasis being placed on the arguments of

different individuals. Once again, though, we may note a disparity between Hyland's figures and the ResArt figures, which accord well with the figures for the RABET corpus. In both the ResArt and RABET corpora, the Agricultural and Food Economics writers show a clear preference for the integral form. Comparing this with the Hyland data, it would appear that Agricultural and Food Economics is closest to Philosophy! As we noted above, the corpora are composed of different types of text, and the ratios shown for Hyland's eight disciplines may well be influenced by the kinds of research articles included. All the same, the similarity of profile between the Agricultural Botany and Agricultural and Food Economics writers in the ResArt and RABET corpora is remarkable, and this implies that the preference for one form over another should not be seen to be genre-specific, but to be discipline-specific.

Discipline	**Non-integral**	**Integral**
Biology	90	10
Electronic Engineering	84	16
Physics	83	17
Mechanical Engineering	71	29
Marketing	70	30
Applied Linguistics	66	34
Sociology	65	35
Philosophy	35	65
ResArt corpus		
Agricultural Botany	73	27
Agricultural Economics	39	61
PhD theses		
Agricultural Botany	66.5	33.5
Agricultural Economics	38	62

Table 6.10 Ratios of non-integral to integral citations by discipline in the 3 corpora

In long texts, such as the Agricultural Economics theses, or in book length treatments of research, there is a higher likelihood that references to leading researchers in the field will be elaborated and give greater prominence to the author(s).

Table 6.11 below shows the percentage of citations in the two corpora that incorporate direct quotation form the source text. It is clear from these figures that quotation is a relatively common feature in the social science and humanities texts but that it is scarcely used in the science texts. Where quotation is used in the science texts, (viz. the 0.8% figure in the Agricultural Botany column), the citation is a definition, while many of the Agricultural Economics quotations are evaluative comments.

Articles (Hyland 1999)				RABET Corpus	
Biology	Electronic Engineering	Sociology	Applied Linguistics	Agricultural Botany	Agricultural Economics
0	**0**	**13**	**10**	**0.8**	**8**

Table 6.11 Sample percentages of citations in the Hyland and RABET corpora that include direct quotation

As with the data comparing uses of integral and non-integral citation form, this suggests that there are certain epistemological bases to research in the social sciences that underlie the practices of writers in those areas, and likewise for the sciences. The social science writers are more discursive and make reference to the words of others (and, thus, to their arguments and texts) while science writers tend to summarize the findings and propositions of others, or to refer readers to other texts for further detail, so that their own text can be as concise as possible. These fundamental differences in approach and perspective are manifested in research writing, whether for publication as an article or as a thesis.

6.8 Summary

In this chapter, the citation practices of thesis writers in two departments have been examined. To find out about the variation and the regularities in citation practice both between and within PhD theses in these two departments, the theses were coded with a specially developed set of tags, and then the data derived from the quantification of these tags was analysed. The questions that the analysis sought to answer were as follows:

1. Are there differences in the quantities of non-integral, or integral, citations used by writers in the two departments?

2. Do writers within the same discipline tend to use the same types of citation?
3. Are different types of citation used in different rhetorical sections?
4. What reporting verbs are used most commonly in theses in Agricultural Botany and theses in Agricultural and Food Economics?
5. How common is the use of direct quotation from other texts in both parts of the corpus, and how frequently do writers elaborate on the citation beyond the sentence?
6. Are there differences between the two parts of the corpus in patterns of language around particular citation types?
7. Is there variation in the choice of voice and tense in different theses?
8. How does the data for the thesis corpus compare with the data from Hyland's corpus and the ResArt article corpus?

The analysis of the texts suggests the following answers:

1. Writers in Agricultural Botany use predominantly more non-integral citations, while Agricultural and Food Economics writers use predominantly more integral citations.
2. While the overall tendency is for writers in one department to use a particular type of citation, certain writers may prefer one type to others, as this type is appropriate to their rhetorical purpose. An example of this is the heavy use of the Source type of citation by TAE-004 when writing a long historical account of exchange rate policies in different African countries.
3. The types of citation used do differ according to rhetorical section. For example, the density of citation is much higher in Introduction and Discussion sections, where the common citation types are Source, and Ident, while in Methods sections, the most common types are Refer, Naming and Origin. Another finding was that the overall dispersion of citations across whole texts was markedly different between the two departments: the Agricultural Botany theses tended to follow the hourglass model, with high levels of citation density in the final chapter, while the Agricultural and Food Economics theses were characterised by low levels of citation density in the final chapter (low relative to the density in early chapters).
4. The most commonly used reporting verbs for both sets of theses was not markedly different with frequent use of *use*, *report*, *find*, *suggest*, and *describe*, but it was noted that Agricultural Botany had a high proportion of use of the

Research verbs *show*, *find*, and *demonstrate*, while Agricultural and Food Economics used a far greater variety of reporting verbs than the Agricultural Botany writers did. The distinction between Research and Discourse verbs proved difficult to maintain in the course of tagging.

5. The use of direct quotation and of elaboration is far more common in the Agricultural and Food Economics theses. This reflects the greater value placed in Agricultural and Food Economics on the words and texts of others, and also the need to elaborate on the concepts and arguments of other researchers.
6. The patterning of language around the Naming type of citation was examined and the preference for Agricultural Botany writers for the 'by X' or 'of X', and of Agricultural and Food Economics writers for 'in X' was noted. The former indicates a concern with the research activities, findings and methods of others, while the latter suggests an inclination to identify models, equations and approaches with their originator, and to foreground texts.
7. In both sections of the corpus, active voice was preferred for the reporting verbs, and Agricultural Botany predominantly used past simple active, whereas Agricultural and Food Economics writers used predominantly present simple active. This confirmed the characterisation of Agricultural Botany theses as the lens through which past research work is seen, while the use of present simple in the Agricultural and Food Economics theses typified the concern with animation of the text, and also with a world view constructed through texts. Passive voice was only found to be the preferred option in the case of perfective aspect reporting verbs in Agricultural Botany, and this was attributed to the need to focus attention, through thematisation, on the subject of the generalised findings.
8. The comparisons with two other corpora, that if Hyland (1999) and the ResArt corpus were problematic, as the corpora are not truly comparable. The profiles of the Hyland corpus and the RABET corpus appeared to be quite different, with much lower levels of citation in the thesis than in the research articles. All the same, the level of citation density in the Agricultural and Food Economics theses was found to be low, and this was related to the fact that these writers make little reference to the literature in the closing stages of the thesis, and that the thesis writer elaborates on ideas and on references to the literature more in a thesis than in an article. Comparing the thesis profile with that of the ResArt

corpus, however, a high degree of similarity was revealed. The close similarity between the ResArt and the RABET corpora may be due to the diversity of article types in the ResArt corpus, which suggests that different chapters in a thesis can be compared to different types of article (ie, the theory chapters of a thesis are comparable to theoretical articles, the review chapter to a review article and so on).

In the next chapter, the uses of the core modal verbs in the theses are examined, to investigate how and where writers choose to modify their propositions, how they construct a persona within the text, and how they represent knowledge about the world.

Chapter 7 Study of the modal verbs in the theses

7.1 Introduction

In this chapter, we look at the roles that the core modal auxiliary verbs (*may, might, can, could, will, would, should, must*) play in the theses, and we also compare these to other corpora. The modal auxiliaries have been chosen because they are used to perform many important functions in academic text, not least that of qualifying statements (in their epistemic senses), but also in their roles in stating what is known, of indicating what outcomes can be expected, and for discussing alternative possibilities. As discussed in Chapter 3, they can be used in interactional aspects of academic writing, as in the use of review and preview statements, and in the use of expressions such 'it should be noted'; they are also used in hedged statements where the writer can be seen to be employing politeness strategies (cf, Myers 1989) or anticipating possible criticism (Crompton 1997). By examining the range of functions that writers use modal auxiliaries to perform, it is hoped that a better understanding of the rhetorical functions common to thesis writing can be achieved, and also a clearer description of what functions the modals perform in different rhetorical sections of theses.

In line with the general research questions driving this research (see Chapter 3, section 3), the following questions are posed:

1. Is there variation in the frequency of use of each of the core modal auxiliaries between:
    - the two sets of theses (Agricultural Botany and Agricultural and Food Economics)?
    - the 8 theses in each section?
    - different rhetorical sections within a thesis?
    - theses and journal articles?
2. What functions do modal auxiliaries help to perform in PhD theses?
3. Are some functions typically performed in certain rhetorical sections?
4. Are some uses of the modal verbs discipline-specific?

7.2 Contextualization

7.2.1 Frequencies of the modal verbs in written academic prose

The modal auxiliaries are frequently used in the thesis corpus: in the word list for the thesis portion of the RAT corpus, five of the modal auxiliaries appear among the first hundred most frequent words: *may* (position 38), *can* (46), *would* (51), *could* (73), and *will* (81). This high level of frequency of the modal auxiliaries in academic prose is noted also in the Longman Grammar of Spoken and Written English (Biber, Johansson et al. 1999). Figure 7.1 below shows the comparative incidences of 8 modal verbs in five sets of data. The first two are for the Agricultural Botany (TAB = Thesis Agricultural Botany) and Agricultural and Food Economics (TAE = Thesis Agricultural Economics) sections of the RABET corpus. Included for comparison are rough figures for 'academic prose' taken from the LGSWE (labelled 'Biber'), and also figures for the corpus of journal articles related to the theses (labelled 'ResArt'). The fifth set of data is calculated from a corpus of the transcripts of 30 academic lectures (part of the Reading/Warwick BASE corpus), which shows the comparative frequency of modal auxiliary use in *spoken* academic discourse: following the smoothed line in the graph, one can see a far greater use of *can*, *might*, *would* and *will*, with a markedly lower use of *may* in the lecture transcripts.

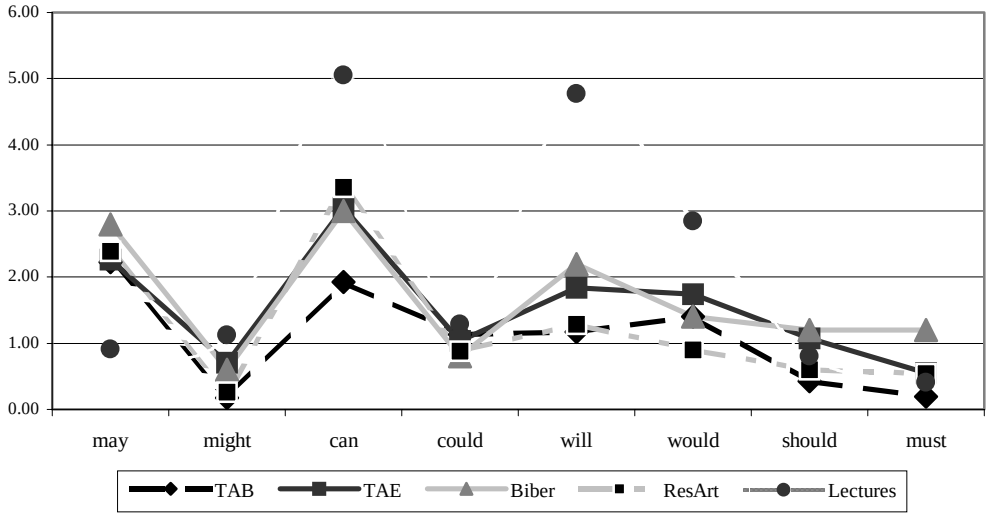

Figure 7.1 Comparison of occurrence of 8 modal verbs per 1,000 words in Agricultural Botany theses (TAB), Agricultural Economics theses (TAE), the LGSWE corpus, Research Article corpus (ResArt) and a lecture transcript corpus (Lectures)

Caution is due in drawing conclusions from the graph as it must be remembered that the TAB and TAE lines represent data drawn from only eight different writers each, while the other two written language corpora lines present data for much wider samples. What is most striking about this graph, though, is the similarity in profile of the four written corpora, particular in the incidence of *may*, *might*, and *could*. The Agricultural Botany theses have a distinctly lower use of *can* compared to the other three, just as there is lower use of the modal auxiliaries all together; in the Agricultural Economics theses; on the other hand, there is a slightly higher use of *would* than there is in the AB theses. In the spoken corpus data, there are far higher frequencies of *can*, *will* and *would*, with markedly less use of *may*.

7.2.2 Meanings of the modal auxiliaries in academic prose

These broad aggregate figures do not, of course, distinguish different senses of the modal auxiliaries; they merely report the number of times that a particular string of letters occurred in the data. The modal auxiliaries possess a wide range of meaning, as is well documented and discussed. Arguments have been put forward that the modals have a single core sense (eg, Perkins 1983), or that they have two core senses (eg Quirk *et al*. (1985) and Biber *et al*'s (ibid) distinction between *intrinsic* and *extrinsic* modality; Coates's (1983) *epistemic* and *root* modalities) or more (Palmer 1990 distinguishes epistemic, deontic and dynamic modalities).

7.2.2.1 Huddleston's categorization

Huddleston (1971) was one of the earliest studies of the uses of the modal auxiliaries in academic/scientific writing. He analyzed the uses of the modal auxiliaries in a small corpus of scientific textbooks and journal articles, and distinguished a range of senses, some of which are particular to scientific and technical writing, as shown in Table 7.1 overleaf.

modal verb	modal meaning
May	epistemic; qualified generalization; exhaustive disjunction; legitimacy; ability; general possibility; concession
Might	past of may (real sense); epistemic; legitimacy; ability; qualified generalization
Can	epistemic (in negative); qualified generalization; exhaustive disjunction; legitimacy; ability; general possibility
Could	past of 'can' (real sense); epistemic; exhaustive disjunction; legitimacy; ability
Will	futurity; induction; deduction
Would	past of 'will' (real sense); tentativeness or prediction ('unreal' senses)
Should	obligation; logical expectation; tentativeness
Must	obligation; logical necessity

Table 7.1 Senses of modal auxiliaries in academic writing (Huddleston 1971)

These meanings are explained as follows:

May, might, can* and *could

1 The epistemic or **uncertainty** modal meanings involve a judgement, on the part of the author, or of another person referred to in the text, or of people in general, as to the truth of a proposition.

2. In conversation, deontic modality expresses the granting of permission to perform an action. In academic prose, however, deontic modality refers to concepts of **legitimacy** – what is permissible given logical, legal, or other constraints. Usually this involves justification of the action, or the claim. It can also refer to the restrictions that are imposed by social laws.

3 **Ability** – refers to the inherent capability of the subject to perform an action. The paraphrase for this is '...are/is able to ...'. For example, 'We can put "signal" on the wave by increasing the power level of the oscillator for a time'.

4 **General possibility** - the easiest paraphrase of this is 'It is possible for ...' where it is not a question of ability, but of what is known to be possible in the

circumstances (the situation). For example, ‘This question can be answered in many ways.’ It is not a question of whether or not people have the ability to answer the question in different ways, but that there is a possibility for many different answers to be given.

5 **Qualified generalisation** suggests that a phenomenon occurs *sometimes*, or that *some* subjects do something. This is a statement of fact, and the modal verb qualifies the extent of the generalisation.

6 **Concession** is restricted to ‘may’ (and ‘might’ as the past sense of ‘may’). The writer admits a fact (or the possibility of something being a fact) but diminishes the significance of the fact in the following statement.

will

1 **futurity** - the ‘will’ signifies futurity relative to a present, either real or implied. The ‘will’ may be intratextual (referring to what will happen later in the text) or extratextual (referring to ‘real-world’ events).

2 **induction** ‘will’ is used for general, timeless truths, for example, ‘If you heat water to 100°C, it will boil’. The condition is either explicit or implicit (eg, ‘Oil will float on water’ = ‘If oil is mixed with water, the oil will float on the water’)

3 **deduction** will is used when one statement is a consequence of another. The implication is ‘It follows that...’

would

1 **tentativeness** in this sense ‘would’ usually collocates with ‘seem’ or ‘appear’.

2 **prediction** ‘would’ deals with unreal worlds, and is often used for forming hypotheses, or for speculating on alternative scenarios.

should/must

1 **obligation** covers both what the researcher or worker has to do, according to some imperative of the discipline, or social or natural law, and also what a person/people is are/ recommended to do. This is a deontic meaning.

2 **logical necessity** is epistemic. An example is ‘It should now be abundantly clear ...’ indicating the writer’s state of certainty, based on the evidence.

Butler (1990) used the same system of classification in his corpus-based study of the modal verbs. He quantified the uses of the modal verbs in these several senses in a corpus of 12 texts, consisting of 2 journal articles and 2 textbook extracts, from the following subject areas: Physics, Botany and Animal Physiology. Butler compared the frequency distributions of the modal auxiliaries in his corpus with those in Huddleston's corpus, and found that *may* and *might* were more frequent in his corpus, while *can*, *could*, *should* and *would* were more frequent in Huddleston's. Butler speculated that these differences can be attributed to the higher proportion of biology-related material in his corpus, which is in line with our expectation that uses of the modals vary across the disciplines (see 5.7 above where it was hypothesized that the uses of the modal verbs in theses based on strong theory would be different from those in theses based on weak theory). Butler also quantified the overall use of modal verbs in each of the IMRD sections, and found differing levels of density. This, he theorised, could be related to function: in the introduction, the writer informs the reader of what can be expected in the rest of the text, and therefore *will*/*shall* are frequently used (an observation that, incidentally, is not borne out in the thesis corpus data, where present simple is the preferred choice); in the results sections, the use of *could* refers to what was or was not possible to do in the experiments; in the discussion sections, claims are made about what it is legitimate to conclude, what *may* or *must* be the case, and so on. Butler's main concern in the paper was the quantification of the different senses of the modals, however, and the analysis of the functions that the modals help to perform is cursory.

7.2.2.2 Ewer's categorization

An alternative account of modal meanings in scientific writing is provided by Ewer (1979). Ewer divides the modal meanings into **primary meanings** and **sub-meanings**, as shown in Table 7.2 overleaf.

Primary meaning	Sub-meaning	Tentative / minimizing	Confident / emphatic
Capability	Realizable		can
	Hypothetical		could
Possibility	Realizable	could	can
	Realizable but not as frequent/certain as 'can'		may
	Hypothetical	might	may
Probability	Realizable	could well	can well
	Hypothetical	might well	may well
Impossibility	Realizable		cannot
	Hypothetical		could not
Choices / alternatives	Realizable	could	can
	Hypothetical	might	may
Concession	Realizable	could	can
	Hypothetical	might	may
Condition	Hypothetical but dependent on realizable future events	should (=if) if … should	
Prediction / expectation	Realizable	should	will
	Hypothetical	should	would
Inference / deduction			must
Recommendatory/ normative		should	must
Requirement / necessity			must, need
Arrangement / intention	Realizable		will
	Hypothetical		would
Willingness	Realizable		will
	Hypothetical		would
Permission / enablement			may
Directive / attention pointing			will
Counter-prediction			why … should

Table 7.2 Ewer's (1979) categorisation of modal meaning and function

The two sub-meanings are 'realizable' and 'hypothetical', which is an important distinction, as it reveals the difference between statements in which the writer talks

about what is achievable, and statements in which the writer forms hypotheses. The right-hand columns are used to show which modal verb (or *modal + adverb* combination) is used to express the confidence of the writer in the statement.

A further powerful distinction that Ewer makes is between what he perceives as two interacting rhetorics: the informational ('objective') and the attitudinal ('subjective'). These labels are similar to those employed by others: the ideational and interpersonal (Halliday 1994), or the transactional and interactional (Brown and Yule 1983).

7.2.2.3 Evaluation

Huddleston's and Ewer's categories have been explained in some detail as they are two of the most thorough accounts of the semantics of the modal auxiliaries in academic writing. There are problems with these classification systems, however, for our present purposes. Firstly, they are not as easy to use as the examples given in both papers suggest. There are many cases where it is difficult to know which sense to ascribe a particular instance to. For example, in the sentence 'Holtzman cautions in favour of a "careful choice of norms" by which the system can be analysed' (tae1c4), it is difficult to know whether the *can* should be tagged as an instance of ability, of legitimacy, or of general possibility; the action does not seem to depend on internal capability or on external law, nor really on possibility, as the analysis will be carried out following a particular set of norms. Secondly, it is not clear what use these categories can be either to a teacher (the categories are generally rather opaque and complex for the language learner), or to a researcher who is investigating rhetorical choices. I attempted to tag the theses in the corpus with a modified form of the Huddleston description but gave up after completing two theses, when I realized that the end result of the tagging would yield statistics that had little application. What would be the use, for example, of knowing that *may* was used 23% of the time in an epistemic sense, 12% as qualified generalization, 32% exhaustive disjunction, and so on, in a particular thesis? When used in an epistemic sense, would that reveal whose uncertainty was expressed through the modality – the writer's, the cited author's, or the discipline in general? In 7.2.2.1, we saw that Huddleston defined epistemic modal meaning as involving '… a judgement, on the part of the author, or of another person referred to in the text, or of people in general, as to the truth of a proposition'. The question of who is responsible

for the uncertainty is important – is the writer mitigating the proposition or is the writer indicating that no-one in the field is certain of the probability. If, in order to answer the question of whose responsibility was being expressed, I added in a further level to the tagging to identify the responsibility for the uncertainty, this would produce an increasingly complex model, but for what purpose?

The point to be made is that, while the classifications of modal meanings proposed by Huddleston and by Ewer help to distinguish the particularities of each sense, as they are used in academic writing, they do not produce a model that indicates **when** or **why** a writer might choose to use a modal auxiliary – a criticism that can also be made of the unitary or plural sense distinctions mentioned above. For pedagogical purposes, it is not enough to know the form but to know also why the form is used. An alternative approach is to develop a *functional* description of the uses that modal auxiliaries are put to, which describes both what the rhetorical purpose (albeit at a generalized level) of the writer is in choosing to use a modal auxiliary, and also in what rhetorical section of a text a writer might make such a choice. Such a project can reveal something of the choices that are available to academic writers to express their meanings, and something of the way that academic texts are constructed.

In the following sections, further quantitative data are presented and discussed, and these indicate the variation in practice between writers of different theses, and how the modals are used in different rhetorical sections. Following this, a reading of how modals are used in two of the theses, one from either half of the corpus, is given, to suggest the range of functions that the modal auxiliaries can perform in theses. Then a framework for a functional description of the uses of the modal auxiliaries in these theses is developed. Finally, an overview of the use of the modals in all sixteen theses is presented, with further discussion of the roles of the modals in particular sections of theses, and also of the uses of modals in the research article corpus, as a point of comparison, which suggests that some uses are discipline-specific. The relevance of this framework to the wider aim of this thesis, and also the possible application of this framework to pedagogical situations are discussed.

7.3. Data and procedure

For the purposes of this study, concordance searches were conducted on the corpus, using *WordSmith Tools*, on the core modal verbs (*may, might, can, could, will, would, shall, should, must*). *Cannot* and reduced forms of the negative were also searched for. *Shall* occurred extremely rarely, and is omitted from the figures reported here, and *cannot* is added to *can*. Occurrences of similar word strings, such as the month 'May' were removed from the concordance output, and so were occurrences in quoted text (parts of the text that were quoted from other sources). Frequency of occurrence was then quantified for each modal verb in every chapter of each thesis, entered into a spreadsheet, and these figures were adjusted to show the average occurrence of the modal per 1,000 words of text, so that comparisons could be made between theses/chapters of differing lengths.

In addition to the quantification of modal verb use, the sets of concordance lines have been studied intensively, and the texts have been read. As the texts are also held in HTML format, all occurrences of the modal verbs were given colour tags, so that the texts could be read in a browser with each of the modal verbs highlighted in a different colour. This allowed a swift reading of the texts which made clear the dispersion and clustering of the modal verbs, and it is on the basis of this reading that Section 7.5 below is set.

7.4 Frequency data

7.4.1 Between Agricultural Botany and Agricultural and Food Economics

	may	might	can	could	will	would	should	must	all
TAB (all)	2.16	0.15	1.77	1.01	1.03	1.24	0.40	0.19	7.93
TAE (all)	2.26	0.71	3.05	1.04	1.84	1.74	1.07	0.55	12.26

Table 7.3 Modal auxiliaries per 1000 words in theses in two departments

The figures in Table 7.3 show the average occurrence of each of the modal verbs per 1,000 words in theses in the two departments. TAB represents (Thesis) Agricultural Botany, and TAE stands for (Thesis) Agricultural and Food Economics.

These simple aggregates indicate a greater tendency for thesis writers in Agricultural and Food Economics to use modal verbs (over 12 for every 1,000 words of text). The most marked differences are in the use of *can*, *will*, *must* and *should*, which the AE writers use far more frequently, and *may* and *could* which are used more, in relative terms, by the AB writers. These figures indicate variation between the subject areas, and a preliminary hypothesis might be that the AE writers prefer *can*, and *will* because they make more predictions, and explications of mathematical procedures and recommendations (*should*), while the AB writers make more hedged claims (*may* and *could*).

7.4.2 Within sections

As the figures are simple aggregates they do not reveal the variation between theses in one department, nor the dispersion of the modal verbs over a whole thesis. Figure 7.2 shows the relative frequencies of use of each of the 8 modal auxiliaries in the 8 Agricultural Botany theses. As can be seen, there is considerable variation in density of use, with the heaviest use appearing in Theses numbers 002, 003 and 004 (approximately 10–11 per 1,000 words) while TAB-008 is very low with approximately 3 per 1,000 words. While we noticed in Figure 7.1 above a general similarity in profile of the incidence of each modal verb across the 4 corpora, we can see in Figure 2 above marked variation between different theses within this sub-corpus. TAB-002, for example, makes far greater use of *can* than other theses, while TAB-003 makes proportionately higher use of *could* and *would*. This can be explained by reference to the nature of the research being done in each case: TAB-002 explores innovative methodology in gene mapping and is concerned with what can and what cannot be done through different methods; TAB-003 is an investigation of plant-weed interactions, and an assessment of extensive farming policies, which requires lengthy discussion of possible outcomes, amongst a range of variables.

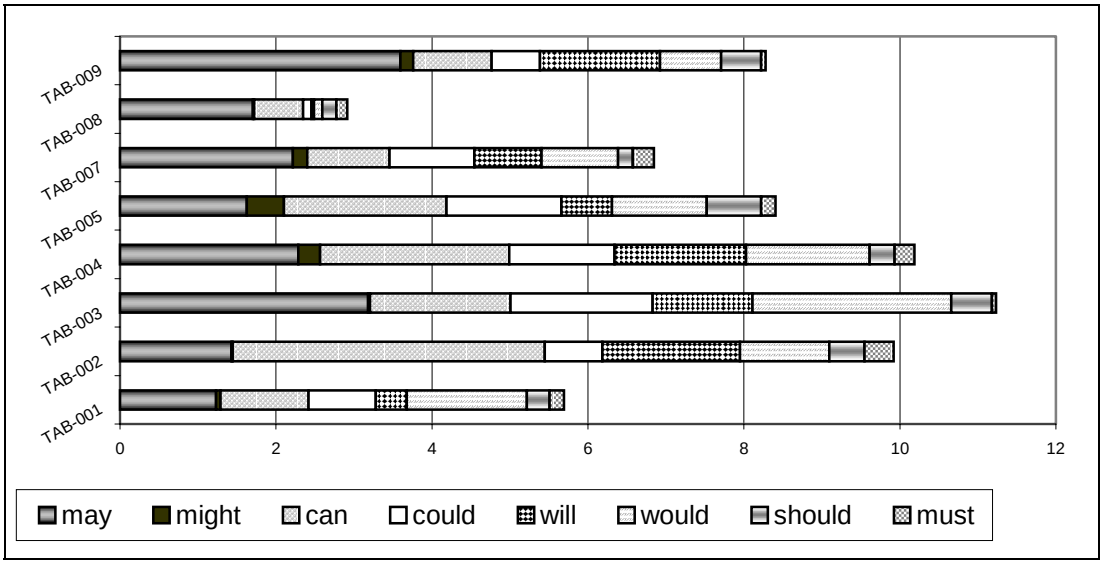

Figure 7.2 Relative occurrence of 8 modal verbs per 1,000 words in individual Agricultural Botany theses

Figure 7.3 below shows the relative occurrence of the same modal verbs amongst the Agricultural Economics theses. Variation is again clearly evidenced with much greater use of *should* in TAE-001, for example, or of *might* and *would* in TAE-002. TAE-001 develops a systems model for real-world application and a large portion of the thesis consists of, amongst other things, recommendations or guidelines for how to operationalize this model. TAE-002 assesses the long-term benefits of research and development work to the agricultural sector, focussing on two case studies, and evaluating what could have happened in different circumstances, and what the outcomes of taking different methodological approaches would be, and therefore the modals used are *could* and *would*.

It can be seen, then, that there is variation between the theses. This can be attributed, to a large degree, to the differing rhetorical purposes of the writers, and also to their research orientations. Research that is strong on theory and which has strictly controlled variables is capable of making stronger predictions, with more use of *will* and *can*, for example, than research work, like that in TAB-003, which is conducted in the field, and which has to consider complex interactions between variables, and make greater use of hypothetical and qualified statements.

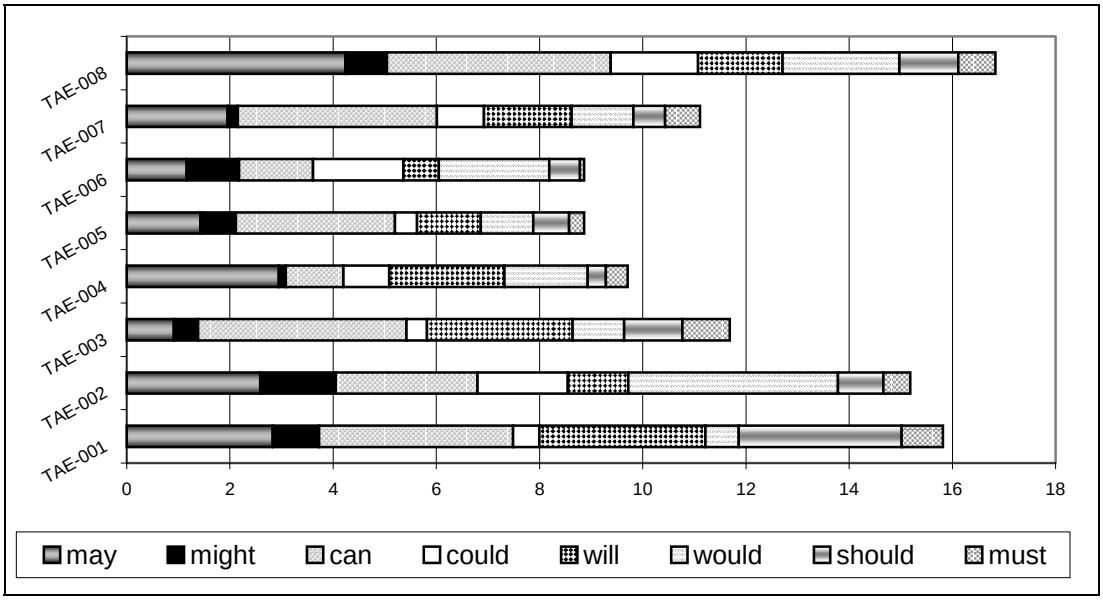

Figure 7.3 Relative occurrence of 8 modal verbs per 1,000 words in individual Agricultural Economics theses

It must be acknowledged, too, that another factor at play in the variation between theses is that of different personal backgrounds. Table 7.4 below shows the occurrences per 1,000 words of the modal verbs in two theses from both departments. These theses were selected as they show clear variation between theses within the same department. TAB-008 has a remarkably low use of modal auxiliaries at 2.91 per 1,000 words, with *may* and *can* accounting for the majority of these instances. In the interview data (see Chapter 5), I learned that several of the modals were added in the final draft at the request of the supervisor who recommended that there needed to be some qualification of the claims made in the Conclusion. The writer, who had been working in industry for a long time, was used to writing for a different audience, one which required directness in recommendations, and who was therefore requested to make adjustments to the strength of his conclusions for an academic audience, and also to speculate on possible future directions for research in the area, which is a move that he never had to make for his industrial readership.

TAB-003, by contrast, has approximately four times the density of modal verbs per 1,000 words, with *could*, *will* and *would* appearing quite frequently. In the case of the two AE theses, there is roughly the same frequency of use of modal verbs per 1,000 words, but we see that TAE-002 uses *would* much more, while TAE-004 uses *may* and *will* more than other modal auxiliary verbs, but has a much lower use of modal auxiliary verbs on average overall.

	may	might	can	could	will	would	should	must	all
TAB-003	3.18	0.03	1.80	1.82	1.28	2.55	0.53	0.05	11.23
TAB-008	1.70	0.02	0.63	0.11	0.03	0.10	0.18	0.14	2.91
TAE-002	2.59	1.46	2.74	1.75	1.17	4.06	0.88	0.52	15.18
TAE-004	2.95	0.14	1.12	0.89	2.24	1.61	0.36	0.42	9.71

Table 7.4 Occurrences of modal verbs per 1,000 words in 4 theses

In other words, the frequency of occurrence of particular modal auxiliaries can differ considerably from one thesis to another. Furthermore, if we look at the chapters within one thesis, as in Table 7.5, we find frequencies change from chapter to chapter.

chapter 1	chapter 2	chapter 3	chapter 4	chapter 5	chapter 6	chapter 7
3.23	4.03	6.22	3.49	2	3.58	4.47

Table 7.5 Use of 'can' per 1,000 words in thesis TAE-007

Such changes are not surprising when we remember that in each chapter the writer has different rhetorical aims. In this thesis (TAE-007), Chapter 3, in which *can* appears over six times every 1,000 words, is entitled "Theory" and it is the part of the thesis in which the writer develops the mathematical framework for the analysis of data in a later chapter. From concordance lines, it can be seen that 'can' is repeatedly used in this chapter in expressions such as *can be modelled/calculated/computed/measured* and *can be understood/interpreted/judged.* Similarly, we can understand the differences in use of *would,* for example, between TAE-002 and TAE-004 (Table 4), if we know that TAE-002 is a retrospective evaluation of the returns to agriculture from two research programmes, and thus consists of detailed discussion of different approaches to evaluating past events, and speculation on what alternative decisions might have caused (see 7.5.2 below for an example).

7.4.3 Modal use in different rhetorical sections: IMRD

Another aspect of variation that can be investigated is that of modal use within different sections of a thesis. Figure 7.4 shows the dispersion of modals across the conventional rhetorical divisions of a scientific experimental report, as found in TAB-001. This

thesis, it should be remembered, is not typical of theses in Agricultural Botany as it follows a simple Introduction - Methods - Results - Discussion (IMRD) pattern of organisation, with one chapter devoted to each of these in turn, while other theses follow a cyclical pattern (chapter 1: General introduction - following chapters each contain IMRD in themselves - final chapter: Conclusion). Because TAB-001 is already divided into an IMRD structure by chapter, though, it provides a clear illustration of the dispersion of modal verb use across rhetorical sections.

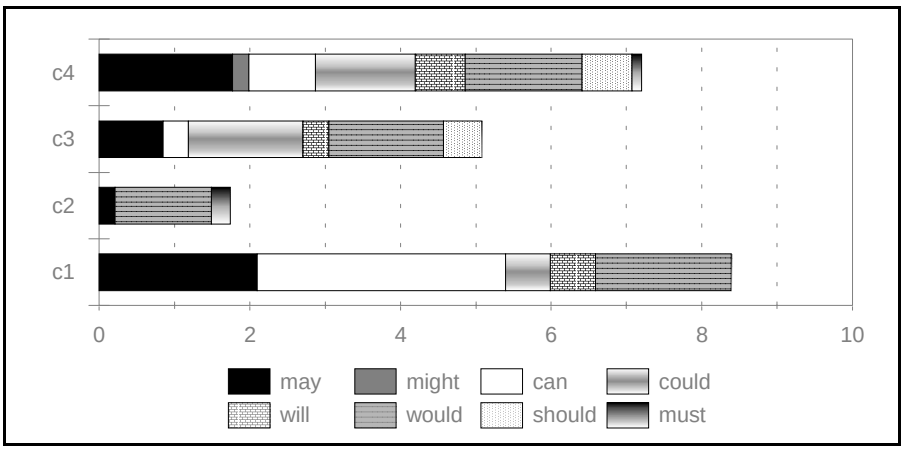

Figure 7.4 Modal verbs per 1,000 words in each chapter of TAB-001

The highest use of modal auxiliary verbs occurs in the Introduction (c1), with *may*, *can*, *would* being the most commonly used. The Introduction includes a literature review element, and establishes what is known about the subject already, what is not, and also, in this case, relates what researchers can or cannot do using different techniques. The Materials and Methods section (c2) is characterised by low use of modal verbs, with the majority of sentences being simple, unqualified statements of what was done, and which method was used. The use of modal verbs increases over the following two chapters, as results are reported (c3) and the inferences to be drawn from the findings are discussed (c3, and c4). *Can* is no longer so predominant, giving way to a more speculative *could.*

To determine the use of modals in different rhetorical sections in greater detail, I then extracted all the sections of the Agricultural Botany theses that were labelled Methods (or equivalent), Results or Discussion and made small sub-corpora of them. Figure 7.5 shows the frequency of use of the 8 modal auxiliaries in these three sub-corpora. Densities of use, as might be expected, are far higher for the modals in the Discussion

sections than they are in the other two, except in the case of *might* which is very rarely used at all. The slightly higher use in the Results sections may be due to the tendency for some writers to offer interpretations of specific results within that section, and this can involve hedging and hypothesizing. Compared to the dispersion for TAB-001, shown in Figure 7.4 above, the overall pattern across all Agricultural Botany theses shows more pronounced use of *may*, suggesting that there is a higher degree of hedging (*may be due to, may be caused by*) in the Discussion sections of other theses than in TAB-001.

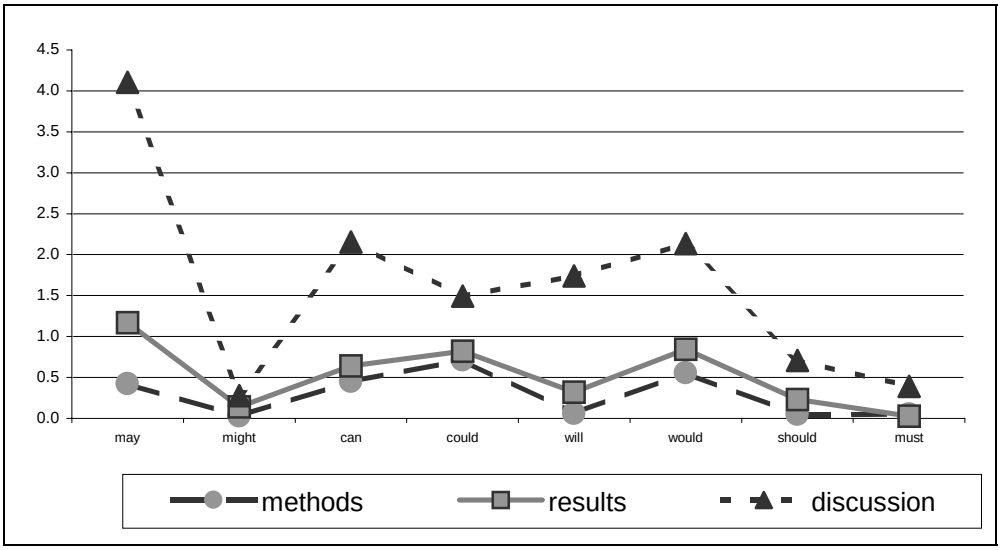

Figure 7.5 Occurrence of the eight modal verbs per 1,000 words in the Methods, Results and Discussion sections of the Agricultural Botany theses

The greater use of *may* in the Discussion sections of some theses can be seen in Table 7.6 below.

Discussion	TAB-001	TAB-002	TAB-003	TAB-004	TAB-005	TAB-007	TAB-008	TAB-009
may	1.82	3.57	5.77	4.81	3.17	3.84	3.90	5.97
might	0.23	0.08	0.00	0.49	1.00	0.00	0.00	0.43
can	0.91	4.56	2.88	2.59	3.17	1.05	1.30	0.77
could	1.36	1.52	2.77	1.97	2.44	0.70	0.33	0.85
will	0.68	3.11	1.73	1.85	1.54	2.80	0.00	2.22
would	1.59	2.20	4.96	3.33	2.17	1.40	0.22	1.19
should	0.68	0.76	0.81	0.37	1.90	0.35	0.11	0.68
must	0.45	0.53	0.12	0.37	0.45	0.35	0.43	0.43
total	7.72	16.33	19.03	15.79	15.85	10.49	6.28	12.54

Table 7.6 Occurrences per 1,000 words of the 8 core modal verbs in the Discussion sections of the Agricultural Botany theses

The three theses with the highest occurrences of *may* in the Discussion sections are TAB-003, TAB-004, and TAB-009, which are all theses that investigate crop-weed interactions in field studies. As has been suggested in Chapter 5, a distinction can be made between Agricultural Botany theses that are lab-based and those that are field studies, in that the number of variables in the field are far greater than the variables in the laboratory and there is consequently a need for greater qualification of interpretations in the latter group of theses. The figures for use of the modal *may* shown in Table 6 indicate that the distinction is valid. *May* is often used in the qualification of statements of causality, as in 'may be due to'.

Finally, a comparison of the overall use of the modal verbs in the Methods, Results and Discussion sections of the Agricultural Botany theses to the findings of Butler (op cit) shows a broadly similar pattern, as shown in Table 7.7. The one disparity is in the figures for the Methods sections, which is far lower for the Butler corpus (0.65 per 1,000 words). However, to test the figures given by Butler, I quantified the uses of the modals in my ResArt corpus and obtained a different figure: 1.96. This figure is much closer to that of the Agricultural Botany theses, and suggests that Butler's results were extraordinary (there are 20 Agricultural Botany articles in the ResArt corpus, compared to 6 articles in the Butler corpus).

	Methods	Results	Discussion
TAB	**2.32**	**4.20**	**13.00**
Butler	**0.65**	**3.14**	**12.51**

Table 7.7 Modal use per 1,000 words in the Agricultural Botany and in the Butler (1990) corpus

7.4.4 Comparison of uses of modal verbs in the theses and in journal articles

Table 7.8 below shows the overall frequency of use of each modal in the thesis corpus and the research article corpus. As will be remembered, the research article corpus contains articles that were frequently cited by the writers of the theses, and therefore are in the same subject areas.

	may	might	can	could	will	would	should	must	*sum*
AB Theses	2.16	0.15	1.77	1.01	1.03	1.24	0.40	0.19	*7.93*
AB Articles	2.29	0.47	2.39	1.36	0.99	1.24	0.39	0.44	*9.57*
AE Theses	2.26	0.71	3.05	1.04	1.84	1.74	1.07	0.55	*12.26*
AE Articles	2.69	0.13	4.49	0.61	1.66	0.72	0.79	0.68	*11.77*

Table 7.8 Comparison of the use of the modal verbs in the Agricultural Botany texts in the RABET and ResArt corpora

In the two sets of theses, there are opposite tendencies, relative to their article corpus counterparts: the Agricultural Botany theses contain less use of the modals than the articles, while in the Agricultural and Food Economics theses, there is a higher density of modal use. In both cases, the theses writers make greater use of *will* which may be explained by the length of the texts and the need to explain what will be presented later in the chapter or the thesis. It is difficult to know how to interpret the lower density of the Agricultural Botany theses, (almost 1.5 per 1,000 words lower) other than to suggest that the thesis writer has more space available than the article writer and may therefore report the experiments in greater detail, therefore making more unmodalised statements. The exceptionally low use of modal auxiliaries in TAB-008 (under 3 per 1,000 words) skewed the data, and may have created a distorted picture.

If we pick out specific theses, however, and compare them to the articles that they often referred to, some similarity in dispersion can be seen. Let us take TAE-005, a primarily econometric study, and the corresponding articles (taken from the *Journal of Econometrics*, and the *American Journal of Agricultural Economics*) for both TAE-005 and for TAE-003, another econometric study. The figures for each chapter of TAE-005, and the figures for six articles (3 articles referred to in TAE-003, and 3 in TAE-008, as indicated by the first 4 characters of the name) can be seen in Table 7.9 below:

	may	might	can	could	will	would	should	must	Sum
tae5c1	0.00	0.00	1.06	0.00	0.53	0.00	0.53	1.06	3.17
tae5c2	3.22	0.67	4.03	0.81	1.61	1.07	0.27	0.13	11.81
tae5c3	1.60	0.48	6.59	0.22	1.52	0.61	1.00	0.26	12.27
tae5c4	1.25	0.42	1.49	0.54	0.42	1.25	0.90	0.18	6.45
tae5c5	1.52	1.34	3.49	0.22	0.85	0.67	0.76	0.13	8.99
tae5c6	1.59	1.09	3.33	0.60	1.14	1.29	0.60	0.10	9.73
tae5c7	0.80	0.80	1.61	0.60	2.61	2.21	0.80	0.20	9.63
Average	1.43	0.69	3.08	0.43	1.24	1.01	0.69	0.30	8.86
tae3mukh	0.6	0.0	2.3	0.6	1.7	0.3	0.0	0.0	5.4
tae3shum	1.1	0.3	9.2	0.3	1.1	1.9	0.5	1.9	16.2
tae3swam	5.1	0.0	6.3	0.0	2.5	0.1	0.9	0.1	15.1
tae5antl	3.4	0.1	7.9	0.3	0.4	0.7	0.8	1.1	14.8
tae5bern	1.0	0.2	5.7	0.0	0.8	2.2	0.5	0.0	10.4
tae5lutk	0.0	0.5	7.8	1.5	2.4	0.0	0.5	0.0	12.6
Average	1.9	0.2	6.5	0.4	1.5	0.9	0.5	0.5	12.4

Table 7.9 Uses of the modal verbs per 1,000 words in TAE-005 (thesis) and 6 articles (tae3mukh to tae5lutk)

The average density of modal use per 1,000 words is much higher in the articles (12.4) than in the thesis as a whole (8.86). The most frequently used modal in the articles is *can*, and if we look at the thesis to find chapters in which the use of *can* is comparable to that in the articles we notice that chapters 2 and 3 are the closest in general dispersion, and also in overall average. These two chapters are discussions of theory, and chapter 3 presents theory and the mathematical functions that can be employed to estimate technical change (the subject of the thesis). The articles referred to are mainly concerned with the elaboration of new approaches to estimation and that is why the profiles are so similar: chapter 3 of the thesis and these articles are alike in broad purpose.

While this may seem an obvious point to make, the implication is worth considering. Whereas we can see that in the case of the science theses, the conventional IMRD organization is a pattern that runs through journal articles and through the individual chapters of a thesis, in the case of the Agricultural and Food Economics theses we can see similarities between a particular chapter and a particular type of journal article. In other words, when a writer writes a chapter on theory, theory articles are useful models, and when writing a chapter on theoretical concerns in developing an applied system of management, then comparable articles may be the closest models (other than chapters

from similar theses), provided that the difference in rhetorical aims of each writer is recognised (in other words, the difference in rhetorical situation of a thesis and an article). This point will be taken up again and discussed in Chapter 8.

7.5 Towards a functional account: readings of two theses

The description of frequencies has indicated that there is variation in use of the modal auxiliaries between writers, between fields of research, and also between rhetorical sections within a thesis. As argued above, these observations, while useful indicators of the diversity of conventions and rhetorical choices available to academic writers, do not by themselves contribute much to the development of a pedagogically useful model. We need to know what functions academic writers use the modal auxiliaries to perform, whether the use of modal auxiliaries is conventionalized in lexical patterning that can be discipline-specific, and what uses of modal auxiliaries are likely to feature in different sections of a particular category of text, such as the PhD thesis. The next section of this chapter therefore turns to a reading of two of the texts in the corpus, in order to identify some of the key functions played by expressions containing modal auxiliary verbs within distinct sections of the texts.

In the following, two theses in the corpus are briefly summarised and the chief uses of modal verbs in different sections outlined. To keep the account readable, the analysis given here must be brief, and the account of the content of the two theses perhaps grossly simplified. It must be stressed also that the analysis given below can be said to inform our understanding of how modal auxiliaries are used in these two theses, and, to some extent, the theses contained in the two subsections of the RAT corpus, but that it cannot be claimed that they constitute models for how PhD theses *in general* are composed. As Dudley-Evans (1999) has indicated, our knowledge of the distinctive uses of language, and the range of organizational structures found in the texts that are constructed as 'PhD theses' is still very limited, and this work should be seen as an exploratory foray into relatively unknown territory.

7.5.1 An Agricultural Botany thesis: TAB-005

The typical structure of theses in the department of Agricultural Botany, as described in Chapter 5, is that they either follow a simple IMRD format, with one chapter for each section (one thesis follows this model), or they have a repeated IMRD structure, in

which the first chapter provides a general introduction, the second optionally describes the general methods used, and then there is a set of chapters describing different experiments with an IMRD cycle (possibly 2 or 3) for each chapter (the other seven theses follow this model). As we have already observed, methods sections typically feature low incidence of the modal verbs and this can be seen in chapter 2 of the selected thesis, TAB-005, as shown in Table 7.10. Chapter 2 describes the overall methodology of the research, and only the modal *could* is used, as past tense of *can* to indicate what was or was not possible during the experiments. In chapters 3 to 9, however, there is a repeated pattern of IMRD and so the figures represent the density of modal verb use for all four sections within a chapter, and are not distinguished by rhetorical purpose.

	may	might	can	could	will	would	should	must	SUM	Words
Chapter 1	2.26	0.38	1.89	0.50	0.88	0.25	0.63	-	6.79	7,949
Chapter 2	-	-	-	2.02	-	-	-	-	2.02	989
Chapter 3	4.58	-	4.58	1.02	-	1.53	-	0.51	12.21	1,966
Chapter 4	1.67	-	2.14	2.62	1.19	2.38	1.67	-	11.68	4,197
Chapter 5	1.17	0.23	1.87	0.47	1.64	0.23	0.94	0.70	7.26	4,269
Chapter 6	1.35	0.54	1.08	1.35	0.27	1.62	0.54	-	6.74	3,710
Chapter 7	0.76	0.51	3.30	-	0.25	0.51	0.76	-	6.09	3,941
Chapter 8	0.83	0.42	1.25	1.53	0.28	1.53	0.42	-	6.25	7,199
Chapter 9	1.99	2.21	2.66	3.76	1.33	2.88	1.33	0.44	16.60	4,518
SUM	1.62	0.48	2.09	1.47	0.65	1.21	0.70	0.18	8.40	38,738

Table 7.10 Modal verb use per 1,000 words by chapter in TAB-005. The left column shows the number of the chapter, and the right column shows the word count for each chapter.

The subject of this thesis is the development of molecular methods for the identification of the fungi that cause two particular types of banana leaf disease. In Chapter 1, the nature of these diseases, and the means of identifying them are described, with a justification for the research given in terms of the economic importance that attaches to the control of the diseases from banana-producing areas in tropical countries. On one page of the introduction the research needs are expressed in what is a clear example of a situation-problem-solution-evaluation structure:

- the disease has spread in the last 25 years
- need for more spraying ("spraying **will** have to be started on plantains if yield and quality are to be maintained, otherwise production **will** decline"; many small-scale farmers **cannot** afford the costs

- an alternative is to develop knowledge of genetic variability
- this **would** provide invaluable information for development of resistant cultivars, and for the prediction of changes in the diseases.

The use of modals in this section is essential: *will* establishes the certainty with which the problem can be stated, whereas *would* indicates the as yet unrealizable potential of the proposed solution.

The higher-than-average use of *may* in this chapter can be explained in two ways. Firstly, it is used to indicate the current state of knowledge about a phenomenon, as is common in introductory sections:

> Examination of herbarium specimens has since indicated that it **may** have been present before this time

This is an epistemic use of *may*, indicating that the examination has shown that "it is possible that it was present before this time". The key point to note here that this is an instance of *objective* modality in which the judgement on the truth of the proposition is not made by the writer but is made by others. Whereas subjective epistemic modality *qualifies* an 'assertion of the factuality of the proposition', objective epistemic modality *states* an 'unqualified assertion of the possibility of a proposition' (Lyons 1977:750).

Secondly, *may* is used to indicate the conditionality of an outcome (cf, "qualified generalization" in Table 7.1 above):

> there **may** also be mass spotting along the mid vein

In other words, 'it is possible, but not definite, that this will occur'. This, again, is an instance of objective modality, as the qualification of the statement is due to previous reports. It is worth emphasizing that objective modality is a central feature of the use of modal auxiliaries in these texts, establishing as they do what is already known, for certain, and also not so certainly, about the phenomenon under study.

The first chapter, then, among other things, establishes what is known about the subject, what the properties of the subject are, and also indicates what is not known, thereby creating the niche that that the present research is to occupy (cf Swales's (1990) CARS model for introductions).

Chapter 2, as noted above, is a description of the general methods used, and has a low use of the modal auxiliaries. Chapter 3 assesses a particular approach, in order to reject it:

> although isozyme analysis **can** provide …, it has also been shown to have some limitation

The use of *can* here is concessive. The incidence of both *may* and *can* in this chapter is particularly high, partly because the shortness of the chapter exaggerates the density, but also because the discussion focuses on what the approach enables one to do, and also what circumstances are not accounted for by the approach:

> the variation present **may** be too low

In the following five chapters, a series of experiments using DNA-level investigations are reported. The IMRD format is used for each chapter. Each chapter begins with a 3-5 paragraph introduction: in Chapter 4, this is a contextualization of the research that follows, in terms of what the 'real-world' applications of the test developed (a serological assay) could be. The heavy use of *could* and *would* in this chapter is to be found in this section, where hypothetical potential is discussed. In the following chapters, however, the introductions focus on technical issues rather than real-world applications, and the most common use of an auxiliary is *can*, to express the enabling features of the technique.

While not uniform, the discussion sections of the 5 chapters tend to include statements of what was expected but did not happen (*would, should have*), and also on what remains to be done, either by the writer herself or by other researchers (*should, would have to be done*). The identification of what remains to be done by the researcher is important for constructing connections between chapters; invariably, these are the questions that are investigated in the following chapter and the whole text is given a coherence by the progressive cycle of experimentation, evaluation, and positing of new questions to be explored by the next round of experiments.

Conclusions are often stated with a degree of tentativeness in the final paragraphs of a discussion, such as the following example is from Chapter 7:

> These data **would** tend to suggest that the pathogen was introduced into Latin America from the Pacific islands …

It is important to note that the tentativeness is not carried by the modal auxiliary alone, but through the combination of the auxiliary, catenative and main verbs. In many cases throughout the corpus, hedging is expressed by tentative verbs such as "tend to" or "suggest", without a modal, and it is therefore, as Hyland (1999) has indicated, wrong to restrict notions of hedging to the use of modal verbs. In the concluding chapter of TAB-005, for example, the writer mitigates a potentially face-threatening statement without using a single modal verb:

> Although these results do not explain the precise means of this introduction, they **tend to** refute the hypothesis put forward by Buddenhagen (1987) that …

Chapter 9 has the highest density of modal auxiliary use in the thesis. In many cases, these are past tense forms of *may*, *can* and *will*, that appear in the summaries of the preceding chapters. There are also statements of what needs to be done (*should*) and also of the enabling features of the main techniques used, expressed through timeless *can*.

7.5.2 An Agricultural Economics thesis: TAE-007

We will now look at the uses of the 8 modal auxiliaries in an Agricultural Economics thesis.

	may	might	can	could	will	would	should	must	Sum	Words
Chapter1	1.21	-	3.23	1.61	2.02	1.61	-	0.81	10.48	2,480
Chapter2	2.81	0.61	4.03	-	2.81	2.93	0.49	0.86	14.54	8,186
Chapter3	2.54	0.35	6.22	0.81	2.19	0.69	0.35	0.58	13.72	8,676
Chapter4	1.94	-	3.49	0.65	2.07	0.26	0.52	0.65	9.57	7,735
Chapter5	0.86	0.43	2.00	0.29	0.43	0.57	0.57	0.57	5.71	7,008
Chapter6	1.34	-	3.58	0.67	0.89	1.12	0.89	0.67	9.16	4,475
Chapter7	2.98	-	4.47	2.38	1.49	1.19	1.49	0.60	14.60	3,357
Sum	1.95	0.20	3.86	0.91	1.70	1.20	0.62	0.67	11.11	41,917

Table 7.11 Modal verb use per 1,000 words by chapter in TAE-007

TAE-007 has an introduction in which the research is contextualized within real-world issues, the key field-central theoretical questions are raised, the research problems are stated, and the shape of the thesis outlined. The following chapters present the broad, and then the more specific, theory out of which the research framework is developed.

Following this, the data set to be used is explained, the data are tested and the outcome is evaluated.

This thesis presents an attempt to measure the costs of undesirable outputs in agricultural processes, using the case of leaching of nitrates into groundwater. The author develops a framework, using econometric techniques for estimation of the parameters of the output distance function, to apply to his data sets. This thesis is comparable to two other theses in the corpus, TAE-003 and TAE-005, in its use of econometric methods, and in the macrostructure of the thesis.

In Chapter 1, the Introduction, the problem is stated in real-world terms, and the value of the thesis, both in its relation to a real world problem and in its contribution to economic theory, is claimed. The importance of econometric measures and what such calculations make possible is asserted with frequent use of *can* in an enabling sense. The short chapter concludes with a statement of the hypotheses that is expressed in terms of predictive *will*, and an overview of the thesis. These metadiscoursal overview statements are made in the present simple, a feature that is typical in most of the AE theses, both in the overview of the thesis and also in the preview and review sections of each chapter. It is only in the preview comments within the body of the text in a chapter that *will* is used (eg, "as **will** be detailed later in this chapter" tae7c2).

Chapter 2 presents a review of the known and possible effects of nitrate contamination. The role of the modal auxiliaries in expressing degrees of probability, as to processes, is crucial. For example, in the following, *may* could be paraphrased as "it is likely that X will …"

> a point is eventually reached after which further quantities of nitrogen **may** decrease yield

In the following, the use of *will* indicates that this is a law of nature:

> In areas with higher rainfall total levels of nitrate leached **will** be higher, but nitrate concentrations **will** be lower due to dilution

The various options that can be adopted are then outlined, and evaluated. In the case of strategies, what can be done is replaced by speculation on the possible effects of taking an action, leading to frequent use of *would*:

> if such a measure was applied unilaterally, UK farmers **would** be at a competitive disadvantage (tae7c2)

Additionally, there is a noticeable use of *must* and *should* in the summarising of EC directives about water quality, where the obligation is due to social laws.

Chapter 3 presents the broad theoretical background to the study, with an historical review of frontier function methods of estimation. As Table 7.11 shows, nearly 50% of the modals used in this chapter are *can* and the use is mainly in terms of what can be achieved through mathematical procedures. The verbs that *can* collocates with in this chapter, as with all the Theory chapters for the AE theses (there are no comparable chapters in the AB theses), are *measure*, *estimate*, *classify*, *compute*, *calculate*, usually in passive constructions. *Can* appears both in statements that imply logical legitimacy (the law of logic permits the action, or the proposition), and in presentations of available options, as in the following:

> Generally, stochastic frontiers are estimated using ML although the MOLS approach **can** also be employed and **may** be preferable because of its ease of computation.

The difference in use of *can* and *may* in this example is interesting: *can* is used to indicate general possibility, while *may* is used in an epistemic sense, to qualify the proposition, for concessive effect.

It is useful at this stage to consider the roles of *may* and *can* in general in both sets of theses. Approximately half of the occurrences of both *may* and *can* in the two sets of theses are followed by the string *be*, but in the case of *may*, the majority of these instances are of *may* + *be* (copula) + *noun/adjective*, as shown in Table 7.12, where only a little over 10% of *may* + *be* instances are found to be part of a passive construction. 61% of *can* + *be* combinations, on the other hand, are followed by a verb participle. This is consistent with the findings of Butler (1990: 155), who makes the further point that *may*, in academic writing, tends to be used for epistemic and what he terms existential uses (following Huddleston 1971), whereas *can* is used for legitimacy, ability and general possibility uses. *May* often is paraphrasable as "it is possible/likely that …" whereas *can* is usually paraphrasable as "it is possible to …". *May* collocates with stative verbs such as *be*, *have*, and *appear*, which are intransitive (and therefore do not admit passive voice). *Can*, in the Agricultural Botany theses, collocates fairly

strongly with *be used, be seen* and *lead to*, while *can* in the Agricultural Economics theses collocates strongly with *be used, be found, be seen, be considered, be derived*, *be applied* and *be written*.

	TAB	TAE
may	**11**	**14.5**
can	**61**	**61**

Table 7.12 Percentage of *may/can* + *be* occurrences that are passive constructions

Chapter 4 continues the discussion of theory, focusing in on the different approaches to the estimation of distance functions, in order to justify the particular choice of method that the writer will make. This requires evaluation of the methods, and some of this must be negative, as methods are to be rejected. The expression of negative evaluation is mainly hedged, in phrases such as "it could be suggested that …", and "the study could be further criticised …" but again we find that not all hedging is conveyed through modal auxiliaries:

> Although true, according to the results of the analysis, the comparison is, perhaps, invalid.

The question of whether or not criticisms need to be hedged seems, however, to depend on the individual writer's confidence and assertiveness, as we find in TAE-005, for example, a forthrightness that is in marked contrast to the attitude of TAE-007:

> …all of the above approaches can be criticised as assuming full static equilibrium… This is obviously an unrealistic, restrictive and potentially distorting assumption when measuring technical change. (tae5c7)

In the discussion of the various approaches, the writer of TAE-007 often gives a shorthand account of the workings of a particular equation, referring the reader to the source text with "(further details) can be found". This is a noticeable feature of the Agricultural Economics theses: the implicit enjoinder to the audience to do something. Readers are addressed with phrases such as "should be noted/remembered" (37 instances in TAE theses), "can be found" (33) and "can be seen" (35), which is one of several manifestations of a more discursive tone of voice in the way that these writers address their audience, compared to that used in the TAB theses. One of the initial questions at the beginning of this chapter was 'Are some uses of the modal verbs

discipline-specific?'; these uses of the modal are not discipline-specific, as they may well also occur in other subject areas which take mathematical approaches to the solution of problems, such as Meteorology, but they are uses that do not occur in Agricultural Botany.

In Chapter 5 the compilation of the data sets to be analysed is described. Much of the data comes from government and industry records, and there is reference to the legal requirements placed on different bodies (eg, "Farms included in the FBS **must** be full-time commercial farm businesses"), which is another use of the legitimacy type modal, only in this case the laws referred to are social laws. This chapter has the lowest density of modal auxiliaries of all, and is comparable to a Methods section in a scientific experimental research report, in that it is primarily factual. In contrast, however, the data available is often not ideal for the purposes of the analysis, and explanation of the means adopted to compensate for gaps and inconsistencies in the data must be given.

Though Chapter 6 is labelled Results, the first half of the chapter contains further discussion of estimation approaches, and explains the reasons for choosing a particular approach ("Both tests decisively reject the assumption of uncorrelation and indicate that the within model **should** be the model employed"). Statistical tests allow strong statements:

> show that such a statement **cannot** be made with any confidence
>
> Rejection of symmetry … **must** throw extreme doubt on all the results reported here.

Interpretation of some features of the results still require a degree of tentativeness:

> This indeterminacy **may** be caused by two factors

The final chapter, Chapter 7, contains the highest density of modals in the whole thesis. While this appears predictable, in that concluding chapters typically contain interpretation and recommendations, it should be noted that in two of the Agricultural Economics theses (TAE-003 and TAE-005, discussed above), the theory chapters have a higher degree of modal use than the conclusions chapter, because the theory chapters contain lengthy discussion of mathematical techniques and reasoning, with frequent use of *can* and *will*.

The chapter summarises the main findings of the research, indicates the limitations, discusses the policy implications and makes recommendations for further research, all of which appear to be conventional elements of a final chapter to a thesis in this applied subject. In this particular case, the writer of the thesis is obliged to state that the real-world implications of the research are severely restricted by flaws in the model, and a large portion of the chapter is devoted to discussion of the limitations of data available and also to the suggestion of an ideal scenario:

> it **would** be preferable to estimate the frontier function directly rather than the average function

> investigation of the performance of flexible functional forms in addition to the translog **should** ideally have been carried out here

There is a higher than average use of *could* in this chapter. Half of the occurrences are past tense forms of *can* in the summary of the thesis, while the other half are markers of tentativeness in the recommendations for improving the estimation methodology:

> this **could** be adapted for the distance function case

7.6 A functional framework

In the remainder of this chapter I will describe a tentative categorization framework developed from qualitative analysis of sections of the theses, and analysis of concordance lines. This will be followed by a brief discussion of the different uses that writers in the two disciplinary areas make of the core modal auxiliary verbs, and what these differences reveal about the ways that writers in these two groupings construct their texts.

The following list of categories is not exhaustive. As stated in Chapter 1, the approach to the study of theses taken in this thesis is concerned with whole texts, and the detection of regularities and variation patterns across large quantities of data. It is, therefore, unrealistic to claim that the analysis of uses of citations or of modals can possibly be exhaustive. The following is a preliminary framework for the analysis of function areas in which the core modal verbs are used in the texts in the corpus. This framework is useful, it is hoped, in viewing the uses of modal verbs in different sections of PhD theses, and thus deepening our understanding of the organization of

such texts, and of how writers in different disciplinary areas construct their texts. It also has potential applications in pedagogy, both for materials developers and also for teachers.

7.6.1 Describing properties of subjects of enquiry

The modal verbs can be used to show what is typical of the phenomena under scrutiny, as in the following example:

(Ex. 1) As the disease develops, and if infection is heavy, the dead leaf tissue becomes greyish and defined lesion borders **may** or **may** not be evident. *(TAB5c1)*

(Ex. 2) However, in the case of some pathogens, plant cell exudates **may** stimulate the growth of competitive micro-organisms. *(TAB4c1)*

This is a key element of the Introduction sections within the Agricultural Botany theses in the RABET corpus, which, in all cases, are reports of experiments conducted either in a laboratory in the field. Before the experiments are reported, the current state of knowledge within the discipline regarding the subject of study is reviewed. The chief modal verbs used in this function are ‘may’ and ‘will’, whereas ‘can’ is used mainly to qualify actions that are possible for the researcher, analyst or other actor.

The use of the modal verb here is not epistemic as it does not express, in itself, a judgement on the truth of the proposition; the proposition made is that defined lesion borders will sometimes be evident, but not always. The truth of the statement is not in doubt. Huddleston’s (1971) ‘qualified generalisation’ and ‘exhaustive disjunction’ senses would apply to this use.

7.6.2 Considering alternatives

The modal verbs can be used when the writer considers different strategies. This might be at a theoretical level as in:

(Ex. 3) An alternative strategy **would** be to use hypervariable probes as core markers, such as SSRs/microsatellites (Litt and Luty, 1989; Weber and May, 1989). *(TAB2c7)*

It is interesting to note here that the use of the hypothetical *would* here sets up the expectation that this alternative method will be rejected. In the following sentence, the expectation is fulfilled:

(Ex. 4) SSRs, such as $(GACA)_4$, have been shown to be more polymorphic than both RFLP (Phillips *et al*, 1994) and RAPD markers in tomato (Rus-Kortekaas *et al*, 1994); however the number of detectable loci is small and their genomic distribution is unknown. *(TAB2c7)*

The positive features of the alternative are conceded in the first half of the sentence, but the *however* signals the change to the conclusive drawback of the method. A modal auxiliary verb can also be used in a directly concessive sense, as in the following, where the *can* modifies the statement of potential benefit, which is then countered by the opposing negative.

(Ex. 5) Although isozyme analysis **can** provide a great deal of useful information and has been important in the genetic analyses of many organisms, including fungal plant pathogens, in some circumstances it has also been shown to have limitations. *(TAB4c5)*

Another example of modal auxiliary verbs being used in the presentation of alternatives is that of their use in the discussion of different interpretations of past events. This is particularly frequent in TAE-002, which, as stated above, is an evaluation of past research programmes. One feature of this thesis is the speculation on how things might have turned out if the circumstances had been different:

(Ex. 6) In this case the UK manufacturers **would** not have enjoyed their major share of this sector and licence income **would** have been reduced by many millions of pounds. *(TAE2c4)*

In this instance, we again find that the modal auxiliary verb is being used in a statement of the alternative scenario, the scenario that is to be rejected - this reinforces the argument that UK manufacturers enjoyed considerable benefits as a result of what actually happened.

The writer may postulate, or refer to, alternative explanations for certain phenomena, which act as hypotheses to be tested in the research:

(Ex. 7) A number of associations were identified which will be subjected to further study. One example is the observation that increased slug damage occurred in crops with more weeds. This association **may** be because slugs and weeds are linked to another unmeasured or unassociated factor. *(TAB9c1)*

In this example, the writer is challenging the claim made in a previous study (crops suffer greater slug damage if there are more weeds) and suggests an alternative reason. The thesis writer may also contrast the real with the ideal, in discussing which method to use:

(Ex. 8) In theory, the ideal screen for detecting polymorphic probes **would** consist of a wide range of breeding lines digested with as many enzymes as possible; however, in practice, this **would** be prohibitively expensive […] *(TAB2c3)*

7.6.3 Hedging

'Hedging' is taken here to mean the mitigation of a proposition by the writer. Hedging can be used as a defensive strategy; the writer qualifies a statement in order to reduce the strength of a claim, lest the claim be vulnerable to a strong challenge. In a thesis, where the writer is being assessed by examiners, this is an important concern. However, hedging can also be a frank admission that there are insufficient grounds to make a strong claim, and therefore that an alternative to the proposition may also be valid.

The modal verbs are often associated with hedging and many examples exist in the corpus. Generally speaking, instances occur in Discussion sections, where the writers draw conclusions from the results, and impute causal relations. In the Agricultural Botany theses, in particular, a frequent pattern is the modal *may* + *be due to* / *be attributed to* / *cause* / *lead to*:

(Ex. 9) The reduction in specific weights on highly fertilized plots **may** have been due to poor grain packing because of smaller less evenly filled grains on these plots. *(TAB3c4)*

It should be noted, however, that other lexical choices are available to writers, and that hedging is frequently expressed, as Hyland (1999) has observed, by verbs such as 'tend to', 'appear' and 'seem', nouns such as 'possibility,' and adjectives such as 'possible' and 'plausible'. Within the corpus, it is common to find tentative verbs collocating with modal verbs in hedged statements:

(Ex. 10) Withers (1987) concluded from his work and that of others that specific weight undergoes little change with added nitrogen and this **would** also **appear** to be the case from the present study. *(TAB3c4)*

Hedging is a clear example of the epistemic sense of the modal verbs, where the uncertainty is identified with the writer. The qualification of the statement is an index of the writer's level of certainty as to the truth of the proposition. Palmer (1974) described the epistemic and deontic functions of the modals as being *discourse-oriented*, while the dynamic functions of the modals are *subject-oriented.* While this distinction between epistemic and deontic, on the one hand, and dynamic on the other, is questionable (a deontic use of a modal need not be performative, for example, as is shown in 7.6.4 below), the concept of 'discourse-oriented', as an indication of the writer/speaker's attitude towards the subject of the discourse (as opposed to a focus on the subject itself) is a useful one. The hedged statement is a manifestation of the writer within the discourse, in that it signals a judgement, and it forms a part of the interactional level of the discourse. Another use of modal verbs in discourse-oriented statements is in metadiscourse, which will be discussed next.

7.6.4 Metadiscourse

Swales (1990) suggests that a distinguishing feature of PhD theses might be in the use of metadiscourse. His reason for this is that theses are usually long texts, and writers need to guide their readers around the parts of the whole text, and need to ensure that readers take note of the key points, by signalling their importance. Modal auxiliary verbs can help to serve these functions, as in the simple future use of *will* in the preview:

(Ex. 11) These **will** be described in detail in Chapter 4, but by way of introduction a few notes follow describing how they came to be carried out. *(TAE6c1)*

Interestingly, the previews in the beginning section of most chapters are written in the simple present, while preview statements within later sections of a chapter tend to use *will*. Modals are also used in enjoinders to the reader, such as:

(Ex. 12) It **should** also be noted that the harrowing was undertaken before spraying and in the work of Blair & Green, 1993, this sequence was less effective than the converse. *(TAB9c3)*

Although the example given here is taken from an Agricultural Botany thesis, this device is much more commonly found in the Agricultural Economics theses, with the occasional variation shown of a simple direct 'Note that …'. A characteristic of the Agricultural Economics theses is a tendency on the part of the writer to adopt a lecturer's style of exposition (as noted in Chapter 6) which can be seen more clearly in the following examples:

(Ex. 13) It **should** be noted that with ADF(p) representations, there may be potentially p unit roots which should be tested for by appropriate differencing. *(TAE5c4)*

(Ex. 14) We **might** note that the VAR and ECM forms of the GC test are exactly equivalent and are used interchangeably to convenience, both here and in Chapter 6. *(TAE5c5)*

Overall there is a far higher use of metadiscourse in the Agricultural Economics theses than there is in the Agricultural Botany theses, and this can be explained by:

- **length of the theses**: the Agricultural Economics theses are twice the length, and thus, following Swales' reasoning, require more metadiscourse
- **the essential nature of the thesis**: a thesis in Agricultural Botany tends to be essentially a report on experimental work, while a thesis in Agricultural Economics is essentially the development and elaboration of an argument (a model, or a theory), and consequently the predominant style of writing is *discursive* rather than *reporting*.

There are also highly conventionalized forms used in the expression of mathematical reasoning in the development of a model:

(Ex. 15) The dependency variable **can** be represented by the following equation: [EQUATION] *(TAE4c4)*

(Ex. 16) The production function, in general form, **can** be written as [EQUATION] *(TAE3c2)*

This style, as noted above, is most common in the three Agricultural Economics theses that use an econometric approach, and, while not directly addressed to the reader, invokes a virtual audience through its lecturing expository style, with its frequent references to 'we':

(Ex. 17) We **can** thus clearly reject TFP from the RND equation. *(TAE5c6)*

Another highly conventionalized set of expressions, guiding the reader's attention, is those used for referring to tables and figures:

(Ex. 18) **As can be seen** from the diagrams in chapter five, key nodes can usually be identified for any marketing system, and can help in locating the problems and main areas of the chain. *(TAE1c3)*

This use of the modal can be described as 'directive'. While commonly used in most of the theses, it is not used in the two examples of laboratory based research reports (the two shortest theses), where the emphasis is on brevity and the projection of a depersonalized voice. This is an interesting example of differences not between departments, but between research types or paradigms (controlled variables experimental research vs complex variables experimental research).

7.6.5 What is possible or necessary, given the circumstances

Conditions can determine what results, or inferences, are possible/likely, and can also constrain what is possible:

(Ex. 19) As initial weed densities between plots were not uniform, only trends **can** be inferred. *(TAB3c4)*

(Ex. 20) If the relationships described above account for most of the variation between total plant and tuber dry weight, then the harvest index **can** be seen to vary with time *(TAB8c4)*

(Ex. 21) The second part of the chapter explores an opinion much espoused by Chambers (1983) that in certain situations the collection and analysis of information **must** be adapted so that [...] *(TAE1c6)*

Ex. 19 indicates a constraint placed upon the analyst by the nature of the results, and this is clearly a useful sentence pattern to be used in the discussion of findings.

Certain effects are revealed to be evident at specific times or periods of time:

(Ex. 22) During this phase, the biosynthetic enzymes SST and FFT **can** be found in the tubers. *(TAB1c1)*

The observation can also be retrospective:

(Ex. 23) If more isolates had been available, it **would** be (*sic*) possible to compare the variability at different locations *(TAB5c9)*

7.6.6 Indicating legitimacy

This is drawn from Huddleston's (1971) semantic category of 'legitimacy'. The modality here is deontic, but, as Huddleston observes, it is not a question of the writer giving permission to the audience (highly unlikely in a PhD thesis!) but of actions or claims being warranted by reference to natural or social laws, or the laws of reason. An example of this is:

(Ex. 24) What **may** reasonably be concluded from this review? *(TAE6c2)*

A reformulation of this question is: what is it permissible to conclude from this review, if we follow the laws of what constitutes reasonable inference?

In sections where the writer is explaining how a model or method can be used in practice, the legitimacy may be of a more pragmatic nature, tending more towards a statement of what is acceptable in the circumstances:

(Ex. 25) At this stage, the identification of intermediate levels **can** be relatively simplistic. *(TAE1c6)*

The *can* in the following example does not suggest possibility or ability but rather a claim that the approach taken is a legitimate one, based on reasoned argument:

(Ex. 26) The general theoretical position taken is that a marketing chain **can** be viewed as a series of interacting institutional forms. *(TAE1c5)*

As with most claims, the responsibility for the claim may lie with the writer (in which case the writer is claiming that the proposition is a permissible one, as appears to be the

case in Ex. 25), or the question of permissibility may be external as in the following case, where the law invoked is a social law, that applies to extratextual matters:

(Ex. 27) Specifically, this means that up to 98,000 tonnes of potatoes **may** be imported into the Community at 0% duty between 1st January and 31st March. *(TAE1c6)*

7.6.7 Expressing enabling functions

Models and methods enable the researcher (or others) to achieve desired ends, and the modal auxiliary verbs are often used to express what is, or is not, possible.

(Ex. 28) The isomerization of intramolecular disulphide bridges **can** be catalysed *in vitro* by the enzyme protein disulphide isomerase (PDI) (Noiva and Lennarz, 1992). *(TAB7c4)*

(Ex. 29) The number of probes that can be mapped on a single population is dependent on the level of intraspecific polymorphism and this **can** be assessed by fingerprinting a range of breeding material (see Chapter 4). *(TAB2c5)*

In other words, in 28, 'The enzyme PDI makes it possible for us to catalyse the isomerization', and in 29, 'fingerprinting makes it possible for us to assess the level of intraspecific polymorphism'. This use of the modal *can* is common in the Agricultural Botany theses in sections dealing with techniques for solving problems, and in the Agricultural and Food Economics theses in sections which deal with models and their application. The past tense form *could* is used sometimes in Agricultural Botany Methods sections, usually in the negative, to describe what was not possible and then introduce the alternative technique.

7.6.8 Referring to required actions or conditions

Writers may use the modal auxiliary verbs in statements of what are required conditions:

(Ex. 30) In the case of the dioxygenases of gibberellin biosynthesis, dithiothreitol (DTT) or its isomer dithioerythritol (DTE) **must** be present during purification (Graebe and Lange, 1989). *(TAB7c3)*

Or to specify necessary actions:

(Ex. 31) As was mentioned in chapter five, the activities of the chain **must** be isolated geographically in order to be able to identify boundaries to the system and its analysis. *(TAE1c7)*

These two examples indicate what is obligatory, as is signalled by the modal *must*. (31) is a typical statement in the theses that deal with the construction of complex non-mathematical models that describe relationships between parts of a system (TAE-001 and TAE-008), and it indicates necessary action on the part of the researcher or analyst. The tone is often changed to that of recommendation:

(Ex. 32) This diagram is particularly important, as it will form the basis of many of the diagrams to follow. It **should** be revised and checked in order for it to finally form an accurate description of the system. *(TAE1c7)*

(Ex. 33) As Egypt is able to produce earlier, their exporters **should** be encouraged to supply the UK market as early as possible ... *(TAE1c7)*

In several instances the tone is one of caution:

(Ex. 34) The indirect effects of nitrogen supply **should** not be underestimated (Mahn 1988). *(TAB3c7)*

7.6.9 Stating expectations and making predictions

An important task for academic writers, especially those in applied subjects such Agricultural Botany and Agricultural Economics, is that of making predictions. This can involve the introduction of expected outcomes in one's own experiments, prior to the textual enactment of the experiments:

(Ex. 35) To check that the ligated linkers **would** also digest, an aliquot of the above ligation reaction was subjected to EcoRI cleavage for 1 hour at 37°C […] *(TAB2c3)*

(Ex. 36) It was anticipated that this work **will** lead to a greater understanding of the factors that are important for developing strains of Jerusalem artichoke that maintain a high degree of polymerisation of inulin over the winter months. *(TAB1c1)*

(Ex. 37) This **should** have encouraged the production of antibodies which were present in only the mycelial homogenate of M. fijiensis. *(TAB5c4)*

It is interesting to note that the use of a modal auxiliary, with perfective aspect, in (37) leads the reader to anticipate the failure of the expectation. Another feature to be noted is the place of the modal verbs within productive sentence patterns, such as 'It was anticipated that X will/would', in which the verb 'anticipated' can also be replaced by 'expected'. An extra modal can also be inserted before the verb as in:

(Ex. 38) Although fuel prices increased dramatically and hence affected all outputs, we **might** expect that crop production **would** be most adversely affected, being machinery intensive. *(TAE5c5)*

where the use of *might* (as with *should* in (37)) predicts, intratextually, a counter statement introduced by 'However'.

Predictions should be made with a warrant provided, and a typical form of predictive statement is the conditional statement:

(Ex. 39) They utilise the nitrogen when the soil would otherwise be bare and if correctly used **will** release it in the following season (Schroder, Ten Holte & Janssen, 1997).

A prediction could also be a firm expectation of what will happen based on general wisdom, the laws of nature, of society, or of disciplinary consensus.

(Ex. 40) It is said that any field with traces of brown rot **will** ensure the whole field, or even farm, is rejected. *(TAE1c7)*

7.6.10 Suggesting the potential

In Discussion sections, and more commonly in Conclusion chapters, writers reflect on the value of models, findings, programmes, and so on. They suggest the potential benefits, with some qualifications:

(Ex. 41) Such a DSS **can** be 'relatively quickly' built, at least to a basic level of usefulness, using a minimum of resources. *(TAE8c8)*

(Ex. 42) A programme of alternative weed control strategies **could** prove to be more suitable in many situations, as would the use of reduced herbicide application rates. *(TAB3c7)*

and also the limitations:

(Ex. 43) It is very unlikely that this approach **could** ever be used to detect the fungi directly in host tissue due to the presence of host enzymes that would confound such a test. *(TAB5c9)*

Can and *could* are used for this function, with *can* carrying the meaning 'realizable' and *could* the meaning 'hypothetical' (cf Ewer's categories in 7.2.2.2 above).

Writers suggest what is needed, and for what benefit:

(Ex. 44) There is a need for a regular survey of livestock producers and livestock populations in Great Britain which **could** be used to better estimate the prevalence and incidence of various diseases and conditions and the distribution of disease effects in terms of both their physical and financial impacts over time. *(TAE8c10)*

They also suggest further possibilities for research:

(Ex. 45) These fragment patterns **could** also be analysed by similar methods to those used above to give information on the phylogenetic relationships between the species. *(TAB5c9)*

7.6.11 Stating principles

Principles are generally expressed in declarative terms, with *will* and *would* (as the past form of *will*) used:

(Ex. 46) Farmers **will** increase quantities of fertiliser inputs up to the point where the marginal cost of that nitrogen is equal to the marginal revenue derived from the crop output.*(TAE7c2)*

(Ex. 47) The principle of this method was that vacuole lysis **would** result in the release of both fructose and polymeric fructan into the incubation medium, while carrier mediated fructose export **would** result in the selective release of fructose and the retention of polymeric fructan. *(TAB1c1)*

7.6.12 Summary

To recap, the functional categories are as shown in Table 7.13:

Function	Typical exponents
1. Describing properties of subjects of enquiry	*may, will*
2. Considering alternatives	*would*, concessive *can* and *may*
3. Hedging	*may, might, could, would*
4. Metadiscourse	*will, can, should* [should be noted, can be written]
5. What is possible or necessary, given the circumstances	*can, would*
6. Indicating legitimacy	*may, might, can, could*
7. Expressing enabling functions	*can, could*
8. Referring to required actions or conditions	*should, must*
9. Stating expectations and making predict-ions	*will, would, should,* concessive *may*
10. Suggesting the potential	*can, could*
11. Stating principles	*will, would*

Table 7.13 The functions of the modals in the theses and their typical exponents

The right hand column shows the typical exponent of each function. It should be reiterated that the list of categories is not exhaustive. The analysis made here is to be seen as heuristic, rather than as definitive. It could be argued that there is overlap between some categories. 'Suggesting the potential', for example, may well involve hedging. This possibility is acknowledged – statements can often perform several functions at once.

It has not been possible to determine precisely which rhetorical sections each of these functions is most likely to appear in, across a number of texts, as such a task would require a separate book length treatment, and also because a single rhetorical model for the Agricultural and Food Economics theses cannot be described. In the following section, we turn to a discussion of the general patterns of modal function in rhetorical sections across the whole corpus of 16 theses.

7.7 Discussion

In the description of modal verb use in two theses above (Section 7.5), most of the functions have been evidenced. The first was found in the introduction section of TAB-005, for example, where the current state of knowledge was established, in terms of what *may* happen, and what *will* happen (degrees of probability). Hedging can be found in the conclusions made in the discussion sections, in recommendations for further research and also in criticisms of other research work, but this is frequently a combination of a minimizing modal auxiliary such as *would* with a tentative verb, such as *appear*, *tend* or *seem*, and hedging can often be expressed without the use of a modal auxiliary. As in the first clause of the previous sentence, a number of expressions are regularly used to guide the reader ("can be found", "may be noted"), and these tend to be used in the "body" of the chapter, rather than at the beginning or end (the preview statements at the beginning of a chapter use present simple tense). The use of *can* in sentences that express enabling functions is frequent, in discussion of methods and techniques, particularly.

7.8 Summary

At the beginning of this chapter, the following questions were posed:

1. Is there variation in the frequency of use of each of the core modal auxiliaries between:
    - the two sets of theses (Agricultural Botany and Agricultural and Food Economics)?
    - the 8 theses in each section?
    - different rhetorical sections within a thesis?
    - theses and journal articles?
2. What functions do modal auxiliaries help to perform in PhD theses?
3. Are some functions typically performed in certain rhetorical sections?
4. Are some uses of the modal verbs discipline-specific?

It was found that there is considerable variation in the uses both between the two sections of the RABET corpus, and also between theses within the same section. There were clear differences in uses of modal auxiliaries between rhetorical sections as shown

in 7.4.3. In the Methods section, for example, there is little use of the modals, and those that are used tend to be ‘could’ and ‘would’. The overall uses of modal verbs in theses and in research articles were not found to be markedly differently, but it was suggested that certain styles of writing are particular to a research field, such as the uses of modals in mathematical reasoning in econometric studies. The uses of expressions such as ‘can be written thus’ and ‘X will mean/ refer to Y’ are not discipline specific, however, but are specific to what might be described as a register of logical exposition used in mathematical reasoning, a register which may be instantiated in a number of different disciplines.

A difference in tone was noted, particularly in the adoption of a lecturing tone (in expressions such as ‘we might note’ or ‘it should be noted’) in some of the Agricultural and Food Economics theses. It was suggested that this might relate to the experience of some of the writers as lecturers within the department, and this also supported the observation that the audience as constructed in the texts is not a unitary entity but it is likely to change in different parts of the thesis. The Agricultural Botany theses do feature uses of impersonalised expressions such as ‘it should be noted’ but not the direct ‘we might note’. This variation is indicative of a difference in rhetorical constructs between writers in two areas; it cannot be said, however, that this is discipline-specific, as the variation may be due to broader differences between science and social science writing, rather than the disciplines in question.

A functional description of uses of the modal auxiliaries in the theses has been presented, and this answers our initial question about the functions that modal auxiliaries can perform in theses. These are not exclusive to thesis writing and we would to expect to find modal auxiliaries performing the same functions in research articles. Within the description, we observed that some functions are typically used in certain rhetorical sections (such as the uses of 1 ‘Describing properties of subjects of enquiry’ in the introduction sections).

An additional point is that the research reported in laboratory studies is of a very different nature from that conducted in field experiments. In the laboratory, variables are controlled, and it is possible to make strong inferences from the findings. The experiments can be replicated in another laboratory with the same results, provided that

the procedures have all been described in sufficient detail. In the field, however, it is usually not possible to control variables, and it is not possible to claim that the results from the experiments can be replicated in different locations. It is necessary, therefore, to exercise greater caution over the strength of claims, and there will also need to be greater discussion of the possible influence of different variables. The nature of the research conducted, therefore, requires different uses of modal verbs, and this can be extended also to the Agricultural and Food Economics theses, with the separation into work based on strong theory and that on weak theory.

Most, if not all, the functions that the modal auxiliaries have been seen to help to perform in these texts are common also in research article writing, and, as such, they do not define the class of text. On the other hand, examination of the uses that the modals are put to in different parts of a complete text has helped to elucidate some of the organizational features and particularities of PhD theses, at least as they are written in two areas of academic enquiry, Agricultural Botany and Agricultural and Food Economics.

Chapter 8 General Discussion and Conclusions

8.1 Introduction

As stated in Chapter 1, the motivation for this thesis is primarily pedagogical. The declared aim is 'to achieve a richer understanding of the nature of the texts that are produced as PhD theses, in order to support the improvement of the substance and quality of tuition for non-native speaker doctoral students who are preparing to write a thesis'. Consequently a large portion of this final chapter is devoted to the discussion of the pedagogical implications of the study. In the first half of the chapter, the research questions are restated, and the main findings of the study are summarised, in terms of how they relate to the research questions. The implications of the findings, both for pedagogy and for further research, are then discussed.

Following this, the study as a whole is evaluated and limitations of the study are indicated. The thesis is then brought a conclusion with suggestions for further research.

8.2 Summary and discussion of findings

8.2.1 Research questions

In Chapter 2 the relevant literature was reviewed and working definitions of 'discipline' and 'discourse community' were proposed, as well as statement of the purposes of genre analysis. The main thrust of the review was to establish that the notions of discipline, genre, discourse community and communicative purpose should be conceived of as complex and plural. Disciplines contain conflict and diversity, genre does not imply a single fixed model, discourse communities exist at different levels of coherence and attraction, in both physical and in virtual forms, and communicative purpose in genre exemplars is complex not unitary. Understanding that form is bound to content and to communicative purpose (cf Coe 1987, 1994) allows us to see that genres can change and vary according to the content to be expressed and the rhetorical aims of the writer, within the bounds of the expectations and rules of the community.

In Chapter 3, the framework of context and relations within which the PhD theses, as texts and as communication, could be conceived was developed. Intertextuality and modality were identified as major features of the rhetorical situation in the writing of a thesis, and it was decided to focus on instances of manifest (or explicit) intertextuality (Fairclough 1992) and on the roles that the modal auxiliaries play in the positioning of the author within the text and in the construction of a textual world, and of an audience.

The research questions, stated at the end of Chapter 3, were:

1. Is there variation in rhetorical organization, at the macro-level, between theses within each department, and between theses in the two different departments?
2. If there is variation in rhetorical organization at the macro-level, what factors contribute to the variation?
3. Is there variation between theses in the two disciplinary areas, and between theses in the same disciplinary area, in the types of citations used? If there is, what do these differences reveal about the purposes of the writers, the ways that they structure their texts, and the ways that the writers position themselves in relation to other texts?
4. Within and between the two disciplinary areas, is there variation, in quantity and rhetorical purpose, in the uses of specific language items that are acknowledged to be of central importance in academic writing: modal verbs?
5. How do PhD theses compare to research articles in the uses of modal verbs and in the uses of citations?

In the following sections, the results of the interviews, and the findings as regards the macro-organisation of the theses, and the uses of citation and the core modal verbs, in the RABET corpus are summarised and discussed, in relation to the research questions.

8.2.2 The interviews (Questions 1 and 2)

Interviews with eight supervisors, four from each department, led to a preliminary distinction between the core rhetorical purposes of theses in the two departments: Agricultural Botany theses tend to be *reports* of what was done in a set of experiments, while Agricultural and Food Economics theses tend to be *discursive*, developing models or frameworks from a consideration of previous models, and then testing them on data. Theses in Agricultural and Food Economics were described as typically

componential by one of the supervisors, and thus liable to have a less conventionalized structure.

In the case of Agricultural Botany, a further distinction was made between studies that are conducted under laboratory conditions, and those conducted in the field. For Agricultural and Food Economics, a contrast was made between work based on strong theory and that based on weak theory. In the former case, the emphasis is on the theory and the data used is usually secondary data, while the latter deals with primary data, and focuses more on the discussion and analysis of the data, and is more tentative in its assertions.

Doctoral study in Agricultural Botany was seen to be a training for work in industry or in government research stations, while Agricultural and Food Economics is far more strongly oriented towards the preparation of individuals for a life in the academy. This implies that thesis writers in each department participate in markedly different communities, with different practices. Furthermore, the Agricultural Botany thesis may be read by other members of a research team, functioning as a record of the individual's contribution to an ongoing project, while the Agricultural Economics thesis will probably have a far more restricted audience, and is written primarily for examination purposes.

It was also suggested, in the discussion of the interview data, that the thesis was a peculiar genre, in that writers have comparatively little exposure to examples of the genre, compared to the genre of book reviews, for example, which are relatively easily accessed and which writers tend to have encountered numerous examples of before they are asked to write a review themselves. As writers are only likely to write one PhD thesis in their lifetime, it is made even more peculiar, in that so much effort is put into writing a single instance of an uncommon genre. This, in turn, could lead to there being a lower degree of conventionalisation in the realisations of the genre than there would be in research articles that have been published in international journals, for example, where conventions for the communication of new knowledge claims have been developed over years of interaction, and have thus streamlined the process (Bazerman 1988).

8.2.3 Macrostructure (Questions 1 and 2 continued)

In Chapter 5, two models each for Agricultural Botany and Agricultural and Food Economics theses were proposed. These are reproduced in Figure 8.1 below.

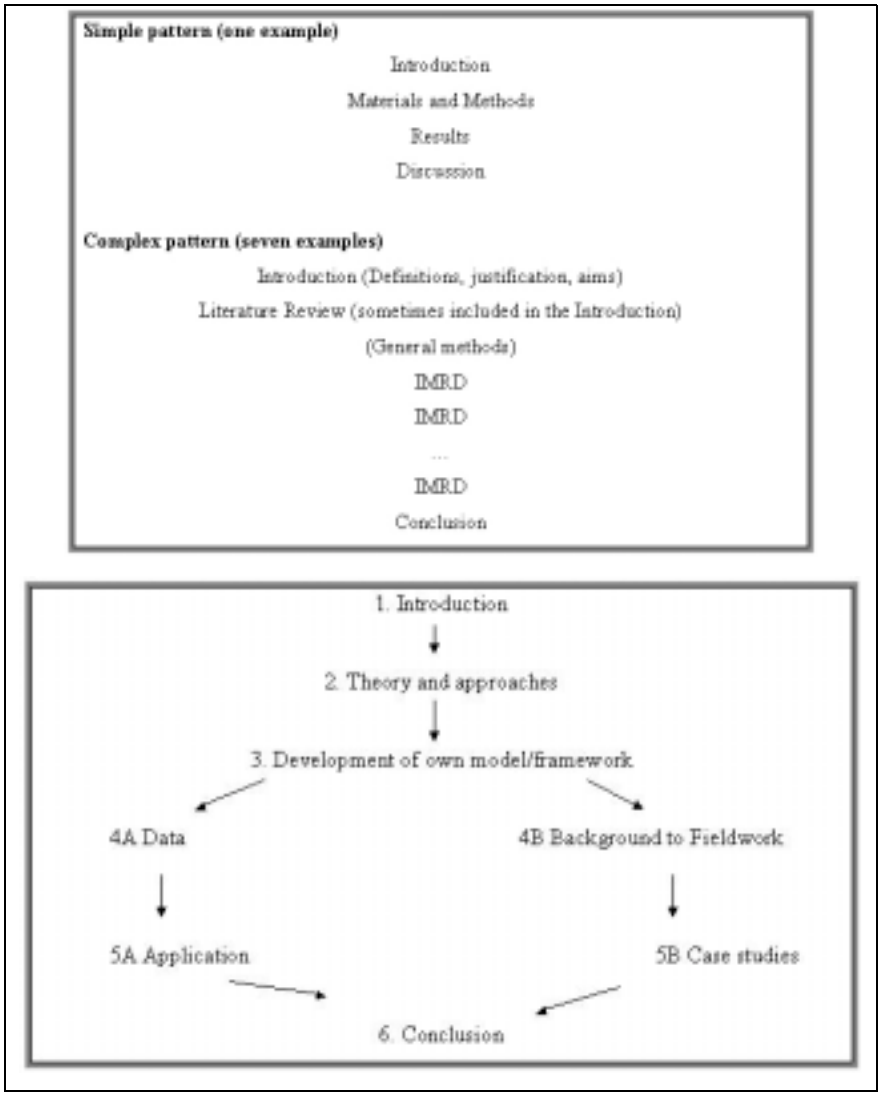

Figure 8.1 The patterns for macro-structure of theses in Agricultural Botany (upper) and Agricultural and Food Economics (lower)

These four patterns were compared with the five patterns described by Ridley (2000) and found to correspond to four of the five. Ridley's fifth pattern was one used in interpretive studies, such as in Literature or Bible Studies. The macrostructures of theses in Agricultural Botany, representative of science, are more conventionalised than those in the Agricultural and Food Economics theses, and so it was possible in the empirical studies of citation and modal use to divide the Agricultural Botany theses into sections according to the IMRD model, following either the simple or the complex pattern, but it was not possible to divide the Agricultural and Food Economics theses into such comparable sections. Instead it was possible only to look at common types of

chapter within certain groups of texts, such as the theory sections of comparable research type theses (the three econometric studies).

The citation study, as observed below, showed that the Agricultural and Food Economics theses contained fewer references to the literature as the thesis drew to a conclusion, which was counter to the hourglass model of the Agricultural Botany theses. The modal auxiliary study extended our understanding of the differences between the rhetorical sections of an IMRD (simple or complex) model, where the use of modal auxiliaries was high in the introduction and discussion sections, with heavy use of modal auxiliaries in the introduction to explain what is known already (the given properties), and in the discussion section for discussion of possible causes, and the hedging of claims. In the case of Agricultural and Food Economics theses, because macrostructure is less conventionalized, it was difficult to connect particular modal uses to rhetorical sections, but the uses of *should* in TAE-001 and TAE-008, for example, to advise agricultural researchers on how best to implement a model emphasized the distinct nature of these more applied theses. Conversely, the frequency use of 'we + *can/will* + see/define/write' in certain chapters of the econometric theses helped to distinguish the particularities of style and tone in these theses, and to identify these chapters as theory chapters.

In answer to Questions 1 and 2, then, there is variation both between theses in one department and the other, and there is also variation between theses within one department. Alternative macrostructures for either department have been described, and the structure of Agricultural and Food Economics theses was found to be less conventionalized than the structure of Agricultural Botany theses.

Reasons for the variation can be found in the different orientations of research in the two departments. The Agricultural Botany thesis tends to be a report of work done in the laboratory or in the field[8], while the Agricultural and Food Economics theses are

[8] This is based on the evidence of the theses included in the corpus. One of the Agricultural Botany supervisors pointed out that two of his colleagues were involved in computer modelling research (see 5.3.6 above). This might lead to a different form of thesis, perhaps closer to the first of the two Agricultural and Food Economics thesis structures

primarily discursive. Research in Agricultural Botany, typical of a science, aims to add to a bank of knowledge (O'Brien 1995), while Agricultural and Food Economics research, as a social science, tends to be concerned with the construction of models that explain social processes and relations, and the development of an original *thesis* (argument). This latter point is exemplified in the low use of citations in the final chapter of the Agricultural and Food Economics theses, while in the Agricultural Botany theses it is important for the writer to show the new knowledge fits into the accumulated knowledge base that is already established, and also to identify what yet remains to be added.

8.2.4 Citations (Question 3)

For the purposes of this research, a refinement of the basic distinction between integral and non-integral citations was made. The basic distinction allows writers to choose to focus on the **proposition** or on the cited **authors**. The framework that I have developed also:

- identifies whether the integral citation controls a verb (and is thus agentive) or whether it appears in the sentence as a nominalised form, either as the head of a noun phrase, or as a modifier
- identifies whether a non-integral citation acts as an attribution for the whole proposition, whether it identifies an actor in the sentence (therefore substituting for the verb-controlling integral citation), whether it identifies the originator of a term/model/technique, or whether it acts as a form of shorthand by referring the reader to another text for more detail.

Additional features of the framework are:

- that it identifies the reporting verbs used with the citations, and also the tense and voice choices made when using a reporting verb
- that it indicates whether a citation is *brief* or *extensive* (terms taken from Swales, 1986)
- that it indicates how many citations involve direct quotation from another text

The analysis of the citation data supported, and extended the observation made in Chapter 5, that there is variation in rhetorical organization, at the macro-level, between theses within each department, and between theses in the two different departments. The most remarkable difference noted between the two set of theses was that the

Agricultural Botany theses followed the hourglass model with a high density of citations in the final chapter of the thesis, while this was not true of the Agricultural and Food Economics theses, where the density of citation use per chapter actually declined towards the end of the thesis. An instance of the variation between theses in one department was that citation density in the final chapter of (lab-based) biochemistry theses (TAB-001 and TAB-007) was low relative to the field study theses (TAB-003, TAB-004, TAB-008, TAB-009). It would be dangerous to extrapolate from these findings, however, that density of citations in the concluding chapter is typical in science theses generally; Bunton (1998) found, on average, very low use of citations in the concluding chapters of the science, medicine and engineering theses in his study.

Direct quotation is rare in the Agricultural Botany theses (on average, 1.9 quotations per thesis), and citations tend to be brief while the converse is true of the Agricultural and Food Economics theses (a finding that is supported by Hyland 1999). The Agricultural Botany researchers appear to be more interested in brief summary statements that report the findings of other researchers, than in either the words in which the findings were expressed, or in elaborating on the report. In Agricultural and Food Economics, the writers place value on the word, both their own words and the words of others, and they need to make extensive citations in order to elaborate on the ideas and arguments of others.

The tense and voice data showed that in both sections of the corpus, active voice was preferred for the reporting verbs. Agricultural Botany writers predominantly used past simple active, whereas Agricultural and Food Economics writers used predominantly present simple active. This confirmed the characterisation of Agricultural Botany theses as ‘the lens through which research work is seen’ (as one of the supervisors interviewed in Chapter 5 put it), while the use of present simple in the Agricultural and Food Economics theses is typical of discursive text in which ideas are animated through textual realisation.

8.2.5 Modals (Question 4)

Quantification of the use of eight core modal auxiliary verbs in the corpus showed considerable variation in the use of these verbs between the two sections of the corpus. Thesis writers in Agricultural and Food Economics were found to use modal verbs

(over 12 for every 1,000 words of text) much more than their counterparts in Agricultural Botany (approximately 8 per 1,000 words). The most marked differences were in the use of *can*, *will, must* and *should*, which the Agricultural and Food Economics writers used far more frequently, and *may* and *could* which are used more, in relative terms, by the Agricultural Botany writers.

Variation was also evident within each department and between different rhetorical sections of theses. In the Methods section, for example, there was little use of the modals, and those that were used tended to be *could* and *would*. Comparing the overall uses of modal verbs in theses and in research articles, no marked difference was detected, but it was suggested that certain styles of writing are particular to a research field, such as the uses of modals in mathematical reasoning in econometric studies. The uses of expressions such as 'can be written thus' and 'X will mean/ refer to Y' are not discipline specific, however, but are specific to what might be described as a register of logical exposition used in mathematical reasoning, a register which may be instantiated in a number of different disciplines.

The use of modal verbs in expressions that are directed at the reader were of interest as they made it clear that the writers had multiple audiences in mind, or rather that they might address parts of a thesis to one audience, and parts to another audience. The frequent invocation by Agricultural and Food Economics thesis writers to the reader to 'See X, Y, and Z for further discussions on data analysis' in particular was noted. This supported the notion of discourse community as a plural entity, as suggested in the discussion in Chapter 2, at least as far as the community is constructed through the text.

A framework for the description of the uses of the modal verbs in the theses was developed, around a set of eleven broad functions. The exemplification of each these categories with sentences drawn from the corpus showed that some of the functions are typically found in certain rhetorical sections of a thesis, such as the function 'Describing properties of subjects of enquiry' which tends to be found in the introduction section, but a clear indication of where each function appears was not possible, as the functions are not all necessarily tied to a certain section, nor was it possible to define the possible rhetorical sections for Agricultural and Food Economics theses.

The functional framework, shown in 7.6.12, is not intended to be a definitive categorisation of the core functions of the modal auxiliaries in thesis writing, as the corpus is not diverse enough, nor has the analysis been sufficiently exacting, but the framework is presented primarily for its heuristic value. It may well have pedagogical value too, for materials developers who wish to identify the range of functions that the modal auxiliaries can play in academic text, and also for teachers who want to present a set of functional categories to their students for exploitation in a range of analytical and production activities. Suggestions for activities are presented in section 8.4 below.

The different nature of research based on strong theory and that based on weak theory, a distinction made by one of the supervisors as reported in Chapter 5, was evident in the strength with which conclusions could be drawn from the results of experimentation. In the econometric studies and in the biochemistry theses, the writers were able to state the inference of their results with far greater confidence, and the use of modal verbs was thus correspondingly different from the field-based studies and the interpretive, or primary data based theses.

8.2.6 Theses and articles (Question 5)

The comparison of theses and articles was complicated by the fact that the two article corpora were not properly comparable. The Hyland corpus contains 80 research articles taken from international journals (one article from each journal, and a total of eight different disciplines). The ResArt corpus that I compiled was an assortment of different kinds of papers, from international journals and from conference proceedings. These papers were all related to the theses in the corpus, in that they had been cited more than once in a given thesis. No control had been placed on the papers to restrict the collection to only those papers that reported original research, for example, and that is why the corpus cannot be compared, like for like, with the Hyland corpus.

The comparison of citation practices in the theses with the two article corpora showed that:

- the Agricultural and Food Economics theses have a low density of citations, even by comparison to the ResArt corpus

- the theses have a low density of citations compared to the Hyland corpus of research articles
- the theses are more similar to the ResArt corpus articles in overall pattern

The variation observed is most probably due to the variation of content and communicative purpose. The finding that the Agricultural and Food Economics theses have a low density of citation can be explained by the observation that those theses are predominantly concerned with the development, and application of models, and that once the model has been developed there is less need for reference to the literature. In an extended text such as the thesis, there should therefore be less need for citation in the latter stages, especially where the focus is on application and, in the case of 'weak theory'-driven research, in the gathering of primary data. Furthermore, there is a higher likelihood that, in an Agricultural and Food Economics thesis, references to leading researchers in the field will be elaborated upon at length.

The similarity of profile between the Agricultural Botany and Agricultural and Food Economics writers in the ResArt and RABET corpora suggests that the preference for one form over another should not be seen to be genre-specific, but to be discipline-specific.

The Agricultural Botany theses contained less use of the modals than did the articles, while in the Agricultural and Food Economics theses, there was a higher density of modal use. In both cases, the theses writers made greater use of *will* due to provide metadiscoursal preview statements. If particular chapters are picked out and the statistics for that chapter then compared with the articles in the ResArt corpus that relate to that thesis, it is possible to find similar profiles in terms of modal use between them, so that the theory section of a thesis, for example, closely resembles a theory article. This implies that the various of the chapters in an Agricultural and Food Economics thesis may correspond closely to a particular article type. In the case of Agricultural Botany theses, the conventional form of the IMRD structure means that most chapters are similar in structure, at least, to research articles, but in the Agricultural and Food Economics theses it may be that the closest analogue for a chapter might be an article that deals with similar content.

8.3 Pedagogical implications

To see how citation and the use of modal auxiliary verbs are treated, a number of EAP writing textbooks were examined. The textbooks examined were:

- *Academic Writing for Graduate Students* (Swales and Feak 1995)
- *Writing Up Research: Experimental research report writing for students of English* (Weissberg and Buker 1990)
- *Exploring Academic English: A workbook for student essay writing* (Thurstun and Candlin 1997)
- *Academic Writing Course* (Jordan 1992)
- *Study Skills for Academic Writing* (Trzeciak and Mackay 1994)

These books are written primarily for general EAP teaching purposes, and therefore do not address the more genre-specific problems that ESL/EFL doctoral students might face in the writing of their theses. These are books, however, that are used on pre-sessional and in-sessional language courses in the UK, and are thus, to a certain extent, representative of the materials used in EAP courses.

Of the five, Thurstun and Candlin's *Exploring Academic English* is probably the least well-known. I have chosen to include it here, because it follows an innovative approach to academic writing, that relates to the ideas proposed in section 8.4 below. In the workbook, the language used to perform a set of six rhetorical functions in academic writing is presented through asset of concordance-based activities. For each function, three or four lexical items are focussed on. The students first analyse lexical and grammatical patterning in examples of the items in a set of concordance lines, and then complete a number of tasks related to the function of these items, leading finally to practice writing activities. A review of the book can be read in Thompson (2001).

8.3.1 Advice on citations

Jordan (1992) offers little explicit advice on how to cite other writers, and concentrates mainly on quotations to provide models for learners to work from. Furthermore, the emphasis is on the form of the citation rather than on the purpose. He exemplifies three kinds of citation: the non-integral Source (as attribution for a quotation), and the integral Naming and Verb-controlling types, again with reference to the utterance of others' propositions.

Trzeciak and Mackay (1994) comment on three instances where citation is used:

- Reporting using paraphrase
- Reference to source
- Direct quotation

This distinction between summary, attribution and quotation is fundamental, but there is little in the way of clear guidance to the apprentice writer and no discussion of disciplinary differences (although discussion of such differences could be made, using the wide range of text samples included in the book).

Swales and Feak (1994) and Weissberg and Buker (1990) take a genre-based approach: the patternings and language used in moves common to the main rhetorical sections of experimental research articles are analysed and practised. Swales and Feak give a relatively full range of advice and examples, and discuss the contrast between non-integral/integral and footnote styles. They discuss the choices of tense for reporting statements, in terms of past tense for single studies, present perfect for reference to areas of enquiry, and the present simple for reference to present states of knowledge. As we have seen, the present simple is the most common choice of tense for reporting verbs in Agricultural and Food Economics theses where it does not act to indicate present state of knowledge so much as to animate the discussion of ideas. The account given in Swales and Feak seems to privilege the research report type of paper by presenting this account of tense usage.

Weissberg and Buker provide a useful set of exercises on the difference between information-focussed citations (non-integral) and author-focussed citations (integral), and they also present three basic methods of ordering citations: chronological, by approach and by ‘distant to close’ (page 46). While the book generally has limitations because it is aimed at the writing of experimental research articles, and not for article writing outside the empirical experiment paradigm, nor for thesis writing, the insights offered into the types and the ordering of citations are useful.

In none of these books is there any discussion of variation in use of types of citation in different rhetorical sections of an article or thesis, such as the tendency, noted in

Chapter 6, for ‘Refer’, ‘Origin’ and ‘Naming’ types to occur in Methods sections. In Swales and Feak and in Weissberg and Buker, citation is treated in the part of the book that deals with the ‘Literature Review’ as a rhetorical section.

8.3.2 Modal verbs

Generally the textbooks focus on the use of the modal verbs for expressing degrees of certainty and for hedging. In Swales and Feak, for example, modal verbs are mentioned in connection with qualification and strength of claim (page 86), qualifying one’s conclusions (pages 97-8), and in a discussion of the use of unreal conditionals in writing critiques, and indicating an alternative that is preferable to the actual (pages 134-5). Jordan contains a set of exercises for practising the modals in connection with degrees of probability, but also has examples of the modals used for implying legitimacy in the conclusions drawn, and also for expressing caution. There is no treatment of the modal verbs as used in referring to required actions or conditions, for example, or in stating expectations and making predictions.

The distinction between subjective and objective modality is an important one, but it is not always made clear. Novice writers have difficulty in indicating status of knowledge and need to be able to distinguish between what is generally held to be possible (objective modality) and what the writer feels is likely (subjective modality). A problem that I found in the account of the modal *may* given in Thurstun and Candlin was a failure to make this distinction between tentativeness on the part of the writer, and the indication of established possible outcome. The modal *may* is presented in the book as indicating tentativeness, and ‘tentative’ is paraphrased as meaning that ideas are ‘not absolutely certain’. However, several of the examples given in the book are propositions that are certain: for example, ‘This alteration may be either excitatory or inhibitory’, which states clearly that there are two possible outcomes, and which it is important to identify as a typical use of *may* in establishing what *is* known about an entity.

An example of a subject-specific textbook is Mountford (1977), which focuses on English for Agriculture. It is built around a functional-notional syllabus, includes exercises on the modals which focus on the functions of *recommending*, *giving instructions* and *expressing necessary actions, or conditions*. This offers a wider

account of the range of modal auxiliary uses than is found in the general academic writing coursebooks, but the account is not related to genre.

8.4 Pedagogical recommendations

8.4.1 Macrostructure

As shown above, textbooks give some coverage to the uses of modals and of citations but the advice is not sufficient to help a thesis writer to understand all the options available, both for the structuring and for the writing of a text, nor to understand how the choices may vary from one discipline to another. The emphasis in academic writing textbooks has tended to be on the research article as a model for student writing (as in Swales and Feak, and in Weissberg and Buker), and the primary model is that of the IMRD structure. The motivation for this may well be that the IMRD model is the most common for the majority of students, and the books are aiming at the widest possible market. An alternative gloss might be that the IMRD structure is the one that is most easily described, and for which rules can most easily be ascribed. As we have seen, however, there is a range of different patterns for the organization of a PhD thesis, and the choice of a particular pattern then implies in each case a different range of options to writers in different disciplines, which makes the description of 'best practice' far more complex than it is if only the IMRD model is treated.

If we are to provide a better preparation for students in different disciplines to understand what rhetorical choices are available to them, a good starting point is to look at the various models of organization that are available to them. Ridley's five structural patterns for PhD theses could be presented to students, with the topic-based structure as an optional sixth. Students could then be asked to look at theses in their subject area, in the library, and identify which of the patterns are used. The emphasis here would be on identifying what the conventional forms in their discipline are. Alternatively students could be presented with printouts of the Contents pages for a number of subject-related theses. The analysis could then be refined, after a consensus had been reached on which pattern was most applicable to a particular student's needs, by picking out sections of the macrostructure, such as the development of the framework, and determining the organization of this part of this thesis, on a few examples. The organization of a conclusions chapter for a thesis may be different from

the equivalent section of a research article, as Bunton (1998) has indicated, and students could examine a range of examples from relevant theses, and determine what the rhetorical sections of a conclusion chapter are, and how much variation there is between theses. It would be important for them also to explore the relations between content, form and purpose in the texts, and so they could also be asked to speculate on why variation exists (why one thesis contains discussion of 'policy implications' while another does not).

8.4.2 Working with corpora

If the teacher has access to a corpus of sample theses, an extension to the analysis could involve the use of concordance lines derived from particular sections of the thesis. The language of the chapters in which the theoretical framework is developed could be examined through the compilation of a wordlist for these chapters. The most frequent words could then be studied by looking at the patterning of language around sets of concordance lines for each word. Gledhill (1995), for example, took the most frequent words from different articles and examined the lexicogrammatical patterning around these words (words such as *the* and *of*) to show how writers move from verbal phrase to nominalization in their articles. Students could work on similar but simpler analyses. Alternatively, Tribble (2000) used the technique of comparing the wordlist for a particular exemplar corpus with that of a reference corpus, using the Keyword facility in WordSmith Tools (Scott 1996) to identify the words that are distinctive of the exemplar corpus (*salient* rather than *frequent*), and then investigating the collocates of each key word to identify the key phraseology of that genre. These techniques could easily be adapted for use by students.

For the study of citations, students can be presented with concordance line output from an appropriate corpus that shows the citations with the year reference in the centre of the page, as shown in Figure 8.2. These concordance lines can be expanded (in WordSmith Tools) so that more lines of context can be shown for each example, if required. Alternatively they could work with a few short texts that are held in electronic format and extract the concordance lines themselves by searching for occurrences of 19* or 20* (where the asterisk acts as a wildcard and therefore would locate all year references from 1900 to 2001) in the texts. They could then be asked to put the instances into the set of categories that I have used in Chapter 6, and then to speculate

on reasons for the choice of a particular citation type (further details can be found in Thompson and Tribble 2001).


riculture is a long one. In his brief history of rotations, Wibberley (1996) reports their use in Roman times with the alternate growi
d emergence were compared by Buhler, Mester & Kohler, (1996). Presence of crop residues prolonged the emergence pe
ams, Chancellor & Drennan, 1983b). Feldman <I>et al</I>. (1997) speculate that diversity was greater under non-inversion
microflora are shrouded in uncertainty. <P>Liebman and Dyck (1993b) in their extensive review highlight three areas deservin
esign of rotation experiments are reviewed in the paper by Cady (1991) and a number of these will now be discussed in relation t
f a cereal crop and fallow. Karlen, Varvel, Bullock & Cruse, (1994) trace their use back to the Han dynasty in China 3000 ye
ertaken for the major weeds of major crops. Francis and Clegg (1990) highlight gaps particularly in relation to the contribution o
n relation to P and K levels measured by Chalmers <I>et al</I>., (1997) on a range of soil types. They note that crops are unlikel
e only partially successful in organic systems reported by Leake (1996). This was due to difficulties in destroying emerged weed
ressions, following the method outlined in Mead & Curnow (1983). Data were not included where crop populations were &lt
cultivars were also noted for <I>V. persica</I> in work by Cosser (1996). The density of <I>V. persica</I> emerging varied betwe
spot cycle have been considered regular and periodic until Yule (1927) demonstrated that it might in reality be an averaged sum
undertaken. One of the earlier reviews by McWhorter and Shaw (1982) point out the need for research in almost all areas of the
viously thought to be a long term process, Feldman <I>et al</I>. (1997), report differentiation of the seedbank under different till
lls this population prior to drilling the crop. Reviewed by Cussans (1995) stale seedbeds were considered still relevant in today's
discussed in relation to this experiment. <P>Quoting Patterson (1964), Cady observes that rotation experiments fall into one of
e still able to emerge. Stale seedbeds were also used by Leake (1996) who notes that their success, particularly in an organic s

Figure 8.2 Concordance lines form TAB-009 showing citations in the centre

In the present study, large quantities of data have been used and it is obvious that one could not present students with so much data. Selected concordance lines could be used, however, in the way that Thurstun and Candlin (1997) do in their concordance line-based EAP writing textbook. They take the auxiliary *may* as one of three examples of ways of expressing possibility, and 32 selected KWIC concordance lines are presented, which illustrate the collocation patterns, and sense uses that the authors then guide the learner to 'notice'. Their account does not relate uses of *may* to particular rhetorical sections, however. An exercise that could relate the use of the modal auxiliary to specific rhetorical sections of text genres would be the matching of selected concordance lines into rhetorical section categories, such as Introduction, Methods, Results or Discussion. Another, more local, level of rhetorical section upon which to work is the Situation–Problem–Solution–Evaluation sections, as described in 7.5.1. above. The relevant section of TAB-005, Chapter 1 is reproduced here:

> The expansion of black Sigatoka has been a major feature of banana production in the last 25 years. The spread of black Sigatoka through Central America has dramatically increased the costs of production due to the increased spraying required. Spraying will have to be started on plantains if yield and quality are to be maintained, otherwise production **will** decline. Marginal levels of production **will** no longer be economic and the increased costs **will** only be met by maximising production on first-class soils. Many of the basic questions that need to be answered concerning the biology of these species **can** be considered under the general heading of genetic variability. In general terms, a greater understanding of the

genetic variability is important in order to predict the possibility of future changes in the population in response, perhaps to newly introduced control measures, whether these are resistant varieties or novel fungicides. It **would** also have important implications for quarantine restrictions. An approach which investigates the genetic variability of the pathogens, by whichever means is most appropriate, **would** therefore provide valuable information to help to answer some of the major questions which are still outstanding concerning the epidemiology and spread of the diseases.

Students could be asked to separate the passage into the 4 sections, and then they would be asked to explain the choices of the modal verbs, and what functions the modal verbs perform. As a follow-up exercise at a later time they could be presented with a copy of this passage from which the modal verbs had been removed, and asked to decide where the insertion of modal verbs would be appropriate. Alternatively, the materials designer can take examples from an appropriate range of academic texts and ask students to assign concordance lines drawn from this data to the different headings (Situation–Problem–Solution–Evaluation), noticing which modal verb is used in which section, and then to create analogous texts.

To explore the uses of *can* and *may*, and to see what kinds of verbs the auxiliaries most commonly collocate with, students could be shown selected concordance lines, as in Thurstun and Candlin, but for contrastive purposes, and asked to find patterns of collocation; this could then be followed up with a gap-fill set of concordance lines, with the modal auxiliaries blanked out. Examples are given in Figures 8.3 and 8.4 below.

nal assessment, in the absence of nitrogen and herbicide treatment, can be explained by more protracted emergence on these plots.
of x has been obtained for a specific plot, the alpha index of diversity can be calculated using the equation; -EQUATION-. The higher the i
ely variable Crop Equivalents (Courtney and Johnstone, 1988). It can be concluded from these data that the increased availability of n
s initial weed densities between plots were not uniform, only trends can be inferred. The general survey at the onset of the trial gave a d
flour, this being quoted as 220 (Wibberley, 1989). HFN of wheat can be increased by the application of nitrogen during stem extensi
lready discussed, the HFN is prone to variability in its precision and can be altered by crop husbandry, grain handling and lack of adhere
0; 1961) outlines the method involved. There are several factors that can give rise to unrepresentative and irregular HFN's (Vaidyanathan,
t phytotoxic effect of the treatment employed. The use of herbicides can have a deleterious effect on crop performance which may have b
duction in biomass observed on 5 July for the 160 KgN/ha treatment can probably be attributed to weed senescence, the weeds having c
d also appear to be the case from the present study. The differences may be an artefact of an exceptionally dry year. These results and t
highly competitive. The negative effect of herbicide on crop yield may be accounted for by a slight phytotoxic effect of the treatment
of Polygonum aviculare compared to the other species present, may be an indication of a greater degree of plasticity in this species
e predominance of Viola arvensis on plots receiving 160 KgN/ha may further account for the reduction of less nitrophilous species, re
ising of early maturing weed species such as the Veronica spp , may have contributed to early senescence on the highly fertilised plo
ain fractions of the samples that were weighed. The unfertilised plots may have higher specific weights because the generally smaller grai
o herbicide may have been due to the abnormally dry summer which may have reduced weed competition. In addition the weed flora of th
var and field in a subsequent year. The lack of response to herbicide may have been due to the abnormally dry summer which may have
arvensis which due to its disproportionate contribution to the flora, may have masked underlying patterns. Without the inclusion of this
1991). The reduction in specific weights on highly fertilised plots may have been due to poor grain packing because of smaller less e
d with increased nitrogen rate. Reductions in nitrogen rate may improve the malting quality of some barley cultivars through lo
ed nitrogen level demonstrates how continued high inputs of nitrogen may promote an undesirable weed flora dominated by competitive ni
iting (1990). They proposed that late competing weeds in particular, may reduce both thousand grain weights and specific weights if left

Figure 8.3 Concordance lines for *can* and *may* from TAB-003, chapter 4

The task for the students would be to look for similarities in the patterning of language around the two modals. The sorting of the concordance lines by the first word to the right helps to focus attention on the collocational (lexical relation) and colligational (grammatical relation) features, and one would hope that students find that there is a degree of interchangeability between *may* and *can*, but that *can* tends to be used with verbs that indicate what the researcher is able to do (the enabling sense), and *may* is often used to express caution in proposing causal relations, and is also the only option available in this context for the pattern 'modal + have + participle'. At a later stage, the students could be presented with selected lines in which the modal verb is blanked out, and their task would be to decide which of the two modal auxiliaries to place in the blank, or whether it would be possible to use either, and in what circumstances. These are challenging activities, designed for advanced students.

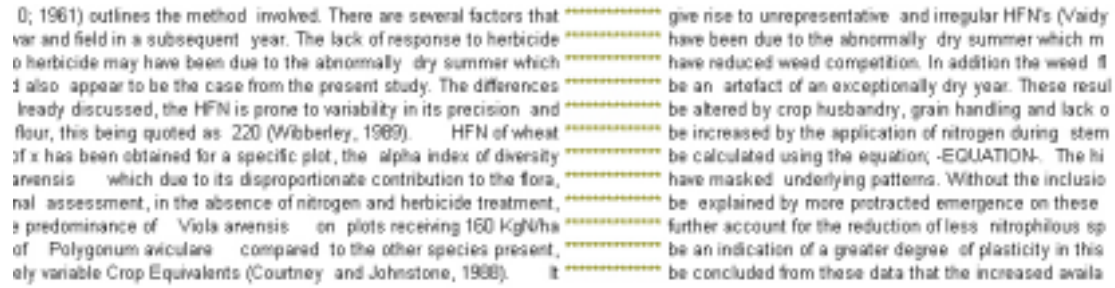

Figure 8.4 Selected concordance lines for *can* and *may* in TAB-003, Chapter 4, with the search terms blanked out

8.4.3 Text in context

In this study, the supervisors of the theses in the corpus were interviewed before the analyses of the texts were conducted (as reported and discussed in Chapter 5). These interviews indicated some differences in the contexts of situation and culture in which the texts were produced, and these insights enabled us to explain some of the variation that was found in the use of citations and of modal verbs. An understanding of the contextual factors, as Malinowski (1960) argues, is necessary if we are to understand the text. If students are to be able to make informed rhetorical decisions they need to understand not only the language but also the culture.

In Chapter 2, the literature on contrastive ESP rhetoric was briefly reviewed. Trzeciak (1996) treated the particular problems that NNS students face under the headings of: macro-discoursal patterns; coherence and style; degrees of commitment and

detachment; use and attribution of source material. The contrastive ESP rhetoric literature has made it clear that writers from other cultural and linguistic backgrounds may have problems with providing sufficient amounts of metadiscoursal guidance to the reader, as these levels of explicitness are not acceptable in their own culture, where reader-responsible prose is the norm. They may also understand the nature of a rhetorical section, such as an introduction or a conclusion, in a different way. In the previous two sections we have looked at ways in which these can be dealt with in the classroom, through textual analysis.

The problems are not only linguistic, but also cultural, and there is a need to explore the contexts in which writing takes place. It is useful to refer back to the framework that was presented in 3.2.4 above. The rhetorical situation of writer-text-reader relationships takes place within the contexts of situation and of culture. Culture can be approached on a range of levels: disciplinary cultures, local institutional cultures, and ethnic cultures. As studies such as those reported in 2.4.4 show, students need to learn to think in the accepted ways of the discipline, and understand the local culture of the place of study (Swales' 'Place Discourse Community'). While this holds true for native speaker students too, the difficulties for international students are often compounded by their language problems, and their ethnic background.

For helping students to perceive and come to a better understanding of the local and broader cultural factors that need to be considered when planning their rhetorical strategies, the concept of 'student-as-researcher' that Ann Johns has advocated is useful. Johns (1997) describes a number of activities, such as interviewing members of faculty, or learning how to conduct participant observation, that involve the student in investigating the norms, the activities and the expectations of the department that they belong to. While the activities that she describes do not refer to L2 students specifically, they could be adapted for use in an in-sessional class. A group of doctoral students could, for example, work together on devising a questionnaire that has the aim of helping them to understand the values and expectations of their supervisors (the teacher would need to seek the supervisors' agreement in advance), and then the individual students would carry out the interviews and report back to the group at their next meeting and collate the results. To investigate the sociocultural context of their place of study, they could be asked to keep an observation journal detailing aspects of

the relations between peers and staff in their department, and the ways that individuals express agreement or disagreement with each others.

In Chapter 2, it was stated that the aims of genre analysis are:

- to identify the regularities of form, of rhetorical organization and of linguistic features within the genre
- to relate these regularities of form, of rhetorical organization and of linguistic features to communicative purpose
- to establish which features are obligatory and which are optional, within given discourse communities
- to understand why and how the genre has developed into its present form, and what functions the genre plays within the community

In these pedagogical recommendations, an underlying assumption is that not only teachers and researchers but students also can benefit from conducting genre analysis, albeit with the guidance of the teacher. The exploration of differences in enactments of a genre from one discipline to another, and also within different parts of the same discipline can help to focus the attention of students on understanding the textual and research practices of a situation, and of how linguistic and rhetorical choices are both general but also specific to local contexts. Because of this, it is essential also that forms should not be presented as models, but, as argued by Coe (1987, 1994), that the interrelatedness of form and content should be understood.

Genre analysis is a powerful heuristic for exploring texts and searching for both the general and the specific: the general in terms of regularities that exist across texts and discourses, and the specific in terms of how communities, at the local, the institutional, the disciplinary, and other, levels, adapt the genre for their own purposes. As Yunick (1997:331) observes:

> An informed pedagogy is ideally aware of both the specific, disciplinary uses of language, and how these are particular instances of more general processes

The analysis of genre is a productive means towards that informed pedagogy.

8.5 Resources

The EAP teacher and the materials writer need to have adequate resources available to them to support and inform their teaching. These resources may take the form of representative texts, or they may take the form of an accessible research literature that extends our knowledge of the textual practices, the processes, and the variation in the genres that students are preparing to write. As shown in Chapter 1, the research literature on the PhD thesis is still sparse, and there is a need for much more research in this area. Suggestions for possible new directions in research into the thesis genre are made in Section 8.7 below. The following paragraphs consider the resources required to support teachers, materials developers and researchers.

What constitutes a good exemplar text for students preparing to write a thesis? What texts should teachers and materials writers use as sources of language data to base their teaching/materials on? To what extent are language practices generalizable across the different genres of academic writing?

Such questions are fundamental in our consideration of what texts are required as resources for teachers and materials writers. It should be clear from the foregoing that the uses of language and the rhetorical patterns used in the performance of a genre do vary markedly both between and within disciplines, but this does not necessarily imply that the best exemplar texts have to be those that are in the same genre, domain and research paradigm. In some cases, it may be that certain features can be made more accessible through use of familiar genres, because the target genre is so complex. For example, it might be easier to exemplify the difference between fast and slow methods sections (cf Bloor 1999, referred to in 2.4.1 above) by using examples from research articles dealing with subjects that students are familiar with, rather than by using methods sections from PhD theses that may be on rather obscure, specialised topics. For some purposes, however, such as the analysis of macro-structure, and also for analysis for research purposes, it is important to be able to work with authentic examples of the genre in question.

8.5.1 Access to representative texts

A problem, as stated in Chapter 1, with researching PhD theses, is that they are often difficult to access. Theses (within the UK, at least) are not published and restricted access conditions may have been placed on them by the university library. Even when they are accessible, in hard copy form, the teacher or materials developer has the problem of dealing with the size of the text, both physically and also in the amount of text, which may well be written about a highly specialised topic in a research area which the teacher has little knowledge about. Furthermore, there is no indication with a thesis as to the quality of the ideas, or whether it was a bare pass or a highly-regarded piece of scholarship.

A solution to this problem is to build more corpora of PhD theses, such as the RABET corpus and the other theses in the RAT corpus. As Upton and Connor (2001:329) observe, 'specialized corpora allow for a more thorough understanding of how language is used in particular contexts or in a particular genre', providing that they contain whole texts and that researchers can see how language is used in different parts of texts. The development of specialized corpora may only be a partial solution, however, as copyright constraints make it difficult to allow full open access to such corpora, and the question of indication of quality is not resolved.

Burnard (personal communication) suggests that theses could be made more accessible for linguistic analysis by an institutional requirement that all theses should be submitted to university examination offices in both paper copy and also in electronic format. Virginia Polytechnic Institute and State University in the United States has already established a Networked Digital Library of Theses and Dissertations, and it now requires all students to submit their dissertations and theses in electronic format, not on paper. As of the 17th September 2001, there were 3505 dissertations and theses in the electronic library[9], which constitutes a massive resource.

These documents are available in PDF (Portable Document File) format. With appropriate copyright clearance, it would be possible to convert these files

[9] Details can be found at: http://scholar.lib.vt.edu/theses/ [Viewed 17/9/01]

automatically into HTML files (Adobe offers this service free of charge online[10]), and the HTML files can easily be analysed using a concordancer programme, such as WordSmith Tools. This would allow the individual teacher or researcher the opportunity to access theses and conduct concordance searches on the texts with relative ease. The drawback is that the texts cannot be analysed by linguistic and structural features unless the researcher is able to spend time on the preparation and mark up of the texts.

8.5.2 XML corpora

An alternative approach, which requires more technical expertise and the allocation of appropriate resources, is to develop a corpus of theses with markup. In the present study, the texts were tagged with an adaptation of HTML but the likelihood is that HTML is soon to be superseded by XML (eXtensible Markup Language). XML is already becoming the markup language of choice for corpora (the second version of the British National Corpus, for example, is in XML). Where XML differs from HTML is that it does not restrict the number of tags that can be used. For the present study, I was obliged to create extra tags for certain features of the texts, and these tags are not standard HTML and cannot be parsed by standard browsers. In XML, however, the set of tags (which describe elements, possible attributes and values) for a document is defined within a DTD (Document Type Definition), which is referred to by the browser. One can either use a pre-specified DTD (all of the files in the BNC follow a single DTD, the Text Encoding Initiative DTD) or one can create a customised DTD for a particular document. In the case of corpus development, it is clearly advantageous to follow standard DTDs so that corpora are marked up in the same ways. The University of Iowa is currently working with Michigan University to develop XML standards for the CIC association of Big Ten Schools.[11]; these standards are designed primarily with the aim of disseminating research findings and information, in a clear, consistent format, and not for purposes of linguistic analysis.

[10] http://access.adobe.com:80/simple_form.html

[11] http://vedavid.org/xml/docs/mla99_jgardner.html [Viewed 17/9/01]

8.5.3 Categorization of texts: genre

The XML standards for markup of theses and dissertations, as expressed in the University of Iowa's Thesis and Dissertation Markup (TDM) language[12], go some way towards meeting the needs of the language researcher but not far enough. This thesis has investigated variation within a text category, that of the PhD thesis genre. The investigation has been conducted on the assumption that texts within a genre can, and do, vary in form, depending on the content and communicative purposes of the writer(s), and that genre analysis is primarily heuristic, rather than taxonomic or categorical in orientation, and this assumption has been supported by the analysis. The intention has not been to determine core defining features of the genre, but to explore variation.

How useful is 'genre' as a text category for corpus studies, though? Lee (2001) describes an innovative classification of texts within the British National Corpus by genre, which he claims is of greater usefulness to researchers than the existing classification by domain. He proposes that 'genre' is a category, following culturally-recognisable criteria' (Lee 2001:11), for complete texts that have been collected for inclusion in a corpus. His definition of genre is broad, including fictional prose and professional letters as examples of genres. Academic prose (termed a 'super-genre') is divided into academic domain areas rather than by whether the text was an essay, a thesis or a lab report, for example. The problem here is the familiar problem of variation in understanding of what constitutes a genre. In the case of the present study, the problem was resolved by stating that we were only concerned with genre as a means for classing the set of texts that are known as PhD theses, and excluding broader issues of naming. When creating text categories in a corpus designed for use by a research community, however, decisions about classification are complex and important as they will, to differing degrees, place constraints upon the users of the corpus. As Aston (2001) points out, a further problem with Lee's approach is that the British National Corpus was not designed with Lee's classification in mind and there are not sufficient representative texts in several of Lee's genre categories to make a corpus-based study of the genre possible. General corpora like the BNC, Aston

[12] http://vedavid.org/xml/tdm_dtd.html [Viewed 17/9/01]

contends, can only provide a few sample texts from which researchers can develop hypotheses, and the hypotheses will then have to be tested out on specialised corpora that have been designed to represent the variance within the genre.

A possible solution to the problem, then, is to develop genre-specific corpora (such as the RAT corpus). The problem still remains of how to enable users of the corpus to make searches of the corpus that explore the variation possible within the genre. Returning to the more specific question of dealing with a corpus of texts that represent the PhD thesis genre, we can pose the following questions:

- For language teachers and researchers who want to use the Reading Academic Text corpus to be able to find out about the linguistic and discoursal features of PhD theses in order to understand better what texts their students are likely to write in future, what information about the texts would be required and how should the texts be classified?
- What information needs to be included in the Head section of a document in the corpus (the section where background data is included), and what kinds of tagging are needed in the text itself?

The RAT corpus contains both articles and PhD theses, so one requirement is that the genre (*article* or *thesis*) is indicated. As it has been shown, there is variation within the genre, so there must then be further information which allows the researcher to refine their investigation. For example, if the researcher is trying to find out about the use of signposting expressions in Biochemistry theses, it would be necessary for them to be able to identify firstly which theses were Biochemistry theses, and then also to be able to know what the structure of each thesis was and to be able to look at uses of the signposting expressions in different sections of the thesis. At present, the Head section of theses in the RAT corpus includes the name of the Department (Agricultural Botany, Agricultural and Food Economics), the year of graduation, and other factual information. This does not actually indicate whether the thesis is in the area of Biochemistry; it could just as easily be a Crop Ecology study. A possibility, therefore is that new additions to the corpus should also include a description of the research field that the thesis is perceived (by the thesis writer) to belong to. This suggestion does, however, have its problems, as some research work might be of a highly

interdisciplinary nature and thus difficult to place in a simple category. Therefore it might be better to ask the writer to supply as many names of disciplines that they feel the work to be highly representative of, rather in the manner that HTML documents can contain a list of key words in their Head section, and these would be placed in a category called 'DiscAreas', or similar. Users would then be able to identify any texts in the corpus that had 'biochemistry' in the <DiscAreas> description, and then select texts from that initial search.

It would be useful, for analysis purposes, to formulate a set of tags that indicate what the chief rhetorical intent of a chapter or a section of the chapter is. For example, it would be useful to place tags at the opening and closing of sections that deal with the description of methods. This would allow teachers, students and researchers to focus their linguistic searches on the methods sections of all theses in their corpus, if they were interested in analysing the language of methods sections.

This study has taken a top-down approach to the description of macrostructure, by trying to identify the purpose of a chapter, and sections within a chapter, by the names given to them in the headings. A possible alternative would be to conduct variation analyses in the style of Biber's (1988) register analyses. Firstly, the corpus would require POS (Part of Speech) parsing, using an automatic parser. Each section of each chapter of the thesis could then be analysed for particular lexicogrammatical patterning and then correlation analyses conducted to determine which sections were similar to each other. The next stage of the analysis would then involve close readings of the sections that appear to be related in order to establish whether they are also related in broad function. If the correlational analyses could be used to identify sections of broadly similar rhetorical function, this method would facilitate the automated preparation of a tagged corpus which included tagging for rhetorical section.

8.6 Evaluation

This thesis has explored the nature of texts that are, to some degree, representative of the genre of texts known as PhD theses. The aim of the thesis was deliberately broad, as the purpose was to explore, and describe, rather than to test hypotheses. The analyses have contributed to a more detailed understanding of the variation that is possible within the genre, and it has indicated that the simple heading <Genre 'PhD Thesis'> is

not adequate. Genre has been seen to be a productive concept for investigative purposes but, because of the fuzziness of the concept (which is perhaps what makes it so productive) a problematic category label. In order to differentiate between the different types of thesis that it is possible to write, we need to have finer categories. This thesis is an initial step towards the definition of such categories.

The primary motivation of the study was pedagogical and it is hoped that the frameworks for the description of citations and of the functions of modal auxiliary verbs have value, in pedagogical terms, for the teaching of English for Academic Purposes. The study approaches genre analysis from a relatively new perspective, combining the use of interviews with expert informants with the analysis of data drawn from a corpus, attempting to combine the contextual with the quantitative, in order to identify regularities and relate them to purposes. The corpus also is original in that it contains whole texts of PhD theses, and the use of whole text corpora for genre analysis is still relatively new. In an article that was published this month, Upton and Connor (2001) advocate a new approach to corpus-based studies of genre, one that ties the study of lexicogrammatical patterns to discourse moves, in whole texts.

The methodology used in this study combined the quantitative with the qualitative, the broad horizontal perspective with the narrow, vertical focus, following the recommendations of Yunick (1997), and in an established ESP/EAP research tradition. The interviews with supervisors provided contextual information which was valuable in the interpretation of the quantitative data. Ideally, interviews with the writers themselves would have been added, but this was not possible for reasons of accessibility. Although it has not been made explicit in the account given of the empirical studies in Chapter 6 and 7, the analysis of the data, particularly of the modal auxiliaries, was extremely complex, involving a repeated cyclical movement between the statistical output of word and tag counts, and the close reading of the texts and concordance lines. The decision to keep all the files in two formats was an important one, as it turned out; the text files were used for tagging and computer-driven analysis, while the HTML files could be read (as clear, clean text) on a browser. Moreover, as described in Chapter 7, it was also possible to highlight certain features of the text by simple ‘Find and replace’ operations; the modal verbs could all be colour coded in all files by changing the string ‘may’ to ‘may’ with a

single command (#7FFF00 is a colour code). Such manoeuvres made it easier to read the text as text, while also highlighting the dispersion of each of the modal auxiliaries across a chapter, allowing the researcher to approach the text both from a textlinguistic point of view, while also allowing access to broad quantitative analytic tools.

The citation study was more satisfactory than the modal study, in that it was possible to define a restricted set of tag categories which were reasonably easy to apply. The modal auxiliaries, on the other hand, are a notoriously difficult class of verb, and it proved impossible to devise a classification scheme that could be productively applied to all instances of the modals in the corpus. In hindsight, either the number of modals should have been reduced to, say, 'may' and 'can', with their past tense realizations in 'might' and 'could', or a certain use of modality should have been focused on, such as the use of hedging devices. In the case of hedging, the linguistic exponents of hedging could be identified and then tagged, with a scheme devised to indicate what the position of the hedge was within the rhetorical section, and whether the hedge related, for example, to field-central claims or to substantive claims (cf, Bloor and Bloor's 1993 study, referred to in 2.4.3 above, which found that economists qualified field-central but not substantive claims). Such an approach would have been much tidier and more focused than the approach that I eventually took, although it might have produced a less extensive set of insights into the PhD thesis as a whole.

There are clearly a number of limitations to the study. A variable that has not been discussed in this thesis is the possible influence of idiosyncrasy. In such a small sample, there is a high likelihood that certain features are strongly idiosyncratic. For example, in Chapter 6, the ten most commonly used reporting verbs used in each of the two sets of theses were listed in order of frequency. In this case, it was pointed out that the uses of one verb (*employ*) were all found in one thesis, and this usage can be considered idiosyncratic as it only appears in one thesis. If we generalise from this small sample to a larger sample, it may well be that the features that have been observed to be common to several theses in the sample are actually not common in the larger sample, and that we would therefore have to treat these as idiosyncratic uses, not as general features. It is possible that the low density of modal use in the Agricultural Botany theses, for example, is due to these particular writers' personal stylistic preferences. At present, we have no means of controlling for this possibility and so it

must be admitted that the findings of this study remain tentative, pending further investigation.

The theses in the corpus were all written by native English speaker students. A possible criticism of the design of the corpus is that, although this was not the intention, it effectively promotes the idea that the native speaker thesis is the norm, and is the model to be emulated. Furthermore, it is possible that the types of research undertaken by native English speaker students are different from the research that international students carry out. In the interviews with supervisors, I asked two of the supervisors whether this might be the case, and was shown examples of theses by recently graduated international students that appeared to use similar macrostructures to theses in the RABET corpus, but there is still a need to investigate further the possibility that the theses in the RABET corpus are not representative of the kinds of theses that international students in the two departments are likely to write, because they may be concerned with different approaches to research and with different problems.

From one point of view, it could be argued that the sample of theses included in this study is too small: there are a mere sixteen theses from two departments within a single university. In the academic year 1999/2000, 11,550 doctorates were awarded in the UK[13], which means that the sample of 16 theses for a ten-year period is not at all representative. Furthermore, the departments chosen are not typical of major disciplines, and they are both applied subjects. While this argument carries some weight, it should be remembered that the choice of one discipline from the biological/ life sciences, and one from the social sciences allowed a preliminary comparison of theses in different broad disciplinary areas, which should generate new hypotheses about differences in disciplinary discourses, and that the research involved both textual analysis and the exploration of the context through interviews with supervisors, which would have been difficult to conduct on a wider scale.

An opposite point of view might be that too much has been attempted in this thesis, and that there is insufficient in-depth analysis of the data. Closer textual analysis of the uses of the modals, for example, might have led to a more rigorous description of the

[13] http://www.hesa.ac.uk/holisdocs/pubinfo/student/quals90.htm [Viewed 19/9/01]

different functions that they perform. Alternatively, detailed analysis of the uses of citations in, say, the discussion sections might have provided a much richer account of the options available to a writer. These shortcomings in the methodology adopted are acknowledged, with the rejoinder that the thesis was intended as an exploratory study, attempting to view the theses as complete texts, and that it is hoped that further research will extend, test and refine the insights and understandings that have been achieved through this study.

8.7 Recommendations for further research

This study has been concerned with the thesis-as-product. It would be valuable, however, to supplement the findings of this kind of research with investigation of thesis-as-process. As indicated in Chapter 1, there have been a number of case studies of second language thesis writers being socialized into a disciplinary community (for example, James 1984; Swales 1990; Bloor and Bloor 1991; Dudley-Evans 1991), which have provided useful insights into the expectations of individual supervisors, and the language related problems that individual students faced. While these studies inform our understanding of some of the processes of adaptation to a new culture, there is still a need for more longitudinal studies, similar to Berkenkotter et al's (1991) study of a native speaker student's 'initiation', of how international students learn what the expectations of the disciplinary community are, in a foreign culture, and how they negotiate their position and identity in a language that is not their first language. These could be compared with co-temporaneous case studies of home students' experiences, to compare and to determine to what extent, if at all, the international students are disadvantaged, and in what ways.

As has been argued in 8.5 above, there is a need for more corpora of PhD theses, if we are to advance our understanding of the range of textual practices and of the expectations of supervisors and examiners in different disciplines. To enable exchange of corpora and comparability it is important to develop a standardised approach to the collection, preparation and tagging of these corpora. The University of Iowa's Thesis and Dissertation Markup language is a starting point, but it is not designed with the requirements of language and discourse analysts in mind. Further research is required to identify what markup of the theses is required.

One possibility is to tag for the beginnings and endings of rhetorical sections, and possibly also of rhetorical moves. The tagging of such features could be of great value to analysts and also to materials developers and teachers, but the challenge is to develop a set of rhetorical section categories of sufficient reliability, bearing in mind the problems that Crookes (1986) found in operationalizing Swales's (1981) original 4-move model for description of the moves in research article introductions. The categories would need to be thoroughly validated, through extensive testing and piloting.

Bazerman (1994) has argued that genres do not operate in isolation but they belong to systems of genres. The PhD thesis is a part of a system of texts which may include an initial proposal, a formal proposal, a further proposal at the end of the first year, draft chapters, letters to the university examinations office, and so on. A corpus could be developed which incorporated many, if not all, of the texts representative of the whole complex system of genres involved in the PhD process. This would allow for studies which, for example:

- compare the various drafts of the chapters with the final version to establish what changes were required, when, for what reasons, and at whose behest
- compare the end-of-first-year proposal with the final thesis to compare the positioning of the writer within these two text-types, and the purposes of citation – for display purposes, or for support of the argument?
- investigate the influence of the supervisor on the writing of the candidate, through analysis of revision following supervisor feedback

Obviously, as such a corpus would contain confidential information, it would require highly restricted access.

At the University of Reading, the RAT corpus already contains theses from other departments (Psychology, Food Science, Applied Linguistics), and an extension of the present study should be made, using the same frameworks, to determine whether there is further variation in the other disciplinary areas represented. To test for generalizability, the findings of this study also need to be tested on other theses in Agricultural Botany and Agricultural and Food Economics, both at the University of Reading, and at other universities. Lastly, but certainly not least, there is a need for

studies of the theses that UK-based international students write so that comparisons can be made between these texts and those produced by home students.

The PhD thesis as a genre has been neglected for a long time, as Dudley-Evans (1999) has argued, while excessive attention has been given to the research article. That a change of focus in EAP research direction is occurring is evident in the fact that John Swales is currently working on a major revision of his work on genre analysis. His 1990 book was, and still is, the seminal text on the analysis of academic text genres, but there was only a page of discussion of the PhD thesis genre, while the bulk of the book dealt with the genre of the research article. In the new book, however, a whole chapter is devoted to an extensive discussion of the thesis.

The research literature on the thesis genre is still sparse, though, and it is hoped not only that this thesis has satisfactorily performed its role as a text in the genre, but that it has also contributed to the development of a fuller description of the rich variation that exists in this genre, and that it will engender further research along similar lines.

List of References

Aijmer, K. (1999). "Epistemic possibility in an English-Swedish contrastive perspective". In H. O. Hasselgard, S. *Out of corpora: Studies in honour of Stig Johansson*. Amsterdam, Rodopi**:** 302-323.

Al-Ali, M. and R. Holme (1999). "The flight from a perfect world: Rethinking the notion of genre in language teaching". In P. Thompson. *Issues in EAP writing research and instruction*. Reading, UK, CALS, The University of Reading.

Allison, D. (1996). "Pragmatist discourse and English for Academic Purposes." *English for Specific Purposes*(2): 85-103.

Ashmore, M. (1989). *The Reflexive Thesis: wrighting sociology of scientific knowledge*. Chicago, University of Chicago Press.

Askerhave, I. and J. Swales (2001). "Genre identification and communicative purpose: A problem and a possible solution." *Applied Linguistics* **22**(2): 195-212.

Aston, G. (2001). "Text categories and corpus users: A response to David Lee (commentary)." *Language Learning and Technology* **5**(3): 73-76.

Aston, G. and L. Burnard (1998) *The BNC Handbook* Edinburgh, University of Edinburgh Press

Bakhtin, M. (1986). *Speech Genres and Other Late Essays*. Austin, University of Texas Press.

Barnett, R., Ed. (1994). *Academic community: discourse or discord?* London, Jessica Kingsley.

Bazerman, C. (1984). "Modern evolution of the experimental report in physics: spectroscopic articles in Physical Review, 1893-1980." *Social Studies in Science* **14**: 163-196.

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

Bazerman, C. (1994). "Systems of genres and the enactment of social intentions".In A. Freedman and P. Medway. *Genre and the New Rhetoric*. London, Taylor and Francis.

Bazerman, C. (1995). *The Informed Writer: Using sources in the disciplines (Fifth Edition)*. Boston, Houghton Mifflin.

Becher, A. (1987). "Disciplinary discourse." *Studies in Higher Education* **12**(3): 261-274.

Becher, A. (1989). *Academic tribes and territories*. Milton Keynes, Open University Press.

Belcher, D. (1994). "The apprenticeship approach to advanced academic literacy: Graduate students and their mentors." *English for Specific Purposes* **13**(1): 23-34.

Benesch, S. (1993). "ESL, ideology, and the politics of pragmatism." *TESOL Quarterly* **27**(4): 705-716.

Berkenkotter, C. and T. Huckin (1993). "Rethinking genre from a sociocognitive perspective." *Written Communication* **10**(4): 475-509.

Berkenkotter, C. and T. Huckin (1995). *Genre knowledge in disciplinary communication*. Hillsdale, IL, Lawrence Erlbaum Associates.

Berkenkotter, C., T. Huckin, et al. (1991). "Social context and socially constructed texts: The initiation of a graduate student into a writing research community". In C. Bazerman and J. Paradis. *Textual Dynamics of the Professions*. Wisconsin, University of Wisconsin Press.

Bex, A. (1996). *Varieties of English: Texts in society, societies in text*. London, Routledge.

Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. London, Longman.

Bhatia, V. K. (1999). "Integrating products, processes, purposes and participants in professional writing". In C. Candlin and K. Hyland. *Writing: Texts, Processes and Practices*. Harlow, Addison Wesley Longman.

Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge, Cambridge University Press.

Biber, D. and E. Finegan (1994). "Intra-textual variation in medical research articles". In N. Oostdijk and P. de Haan. *Corpus-based research into language*. Amsterdam, Rodopi**:** 202-221.

Biber, D., S. Johansson, et al. (1999). *The Longman Grammar of Spoken and Written English*. Harlow, Pearson Education.

Biglan, A. (1973). "The characteristics of subject matter in different scientific areas." *Journal of Applied Psychology* **57**(3): 195-203.

Bloch, J. and L. Chi (1995). "A comparison of the use of citations in Chinese and English academic discourse". In D. Belcher and G. Braine. *Academic Writing in a Second Language*. Norwood, NJ, Ablex**:** 231-274.

Bloor, M. (1999). "Variation in the Methods sections of research articles across disciplines: the case of fast and slow text". In P. Thompson. *Issues in EAP Writing Research and Instruction*. Reading, Centre for Applied Language Studies, University of Reading.

Bloor, M. and T. Bloor (1991). "Cultural expectations and socio-pragmatic failure in academic writing". In P. Adams, B. Heaton and P. Howarth. *Socio-cultural Issues in English for Academic Purposes*. London, MacMillan.

Bloor, M. and T. Bloor (1993). "How economists modify propositions". In W. Henderson, A. Dudley- Evans and R. Backhouse. *Economics and Language*. London, Routledge.

Borg, E. (2000). "Citation practices in academic writing". In P. Thompson. *Patterns and perspectives: Insights for EAP writing practice*. Reading, UK, CALS, University of Reading.

Brett, P. (1994). "A genre analysis of the Results section of Sociology articles." *English for Specific Purposes* **13**(1): 47-59.

Brown, G. and G. Yule (1983). *Discourse Analysis*. Cambridge, Cambridge University Press.

Bunton, D. (1998). *Linguistic and textual problems in Ph.D and M.Phil theses: an analysis of genre move and metatext*. Unpublished PhD thesis, University of Hong Kong.

Bunton, D. (1999). "The use of higher-level metatext in PhD theses." *English for Specific Purposes* **18**: S41-S56.

Burgess, S. (1997). *Discourse Variation Across Cultures: A Genre-Analytic Study of Writing on Linguistics*. Unpublished PhD thesis, Reading University.

Butler, C. (1990). "Qualifications in science: modal meanings in scientific texts". In W. Nash. *The writing scholar: studies in academic discourse*. Newbury Park, CA, Sage**:** 137-170.

Campbell, C. (1990). "Writing with others' words: using background reading text in academic compositions". In B. Kroll. *Second Language Writing: Research insights for the classroom*. Cambridge, Cambridge University Press.

Carne, C. (1996). "Corpora, genre analysis and dissertation writing: An evaluation of the potential of corpus-based techniques in the study of academic writing".In S. Botley, J. Glass, T. McEnery and A. Wilson. *Proceedings of Teaching and Language Corpora 1996*. Lancaster, UCREL.

Charles, M. (2000). "The role of an introductory *it* pattern in constructing an appropriate academic persona". In P. Thompson. *Patterns and Perspectives: Insights into EAP writing practice*. Reading, CALS, The University of Reading.

Clark, G. (1994). "Rescuing the discourse of community." *College Composition and Communication* **45**(1): 61-74.

Clyne, M. (1987). "Cultural differences in the organisation of academic texts." *Journal of Pragmatics* **11**: 211-247.

Cmejrkova, S. (1996). "Academic writing in Czech and English". In E. Ventola and A. Mauranen. *Academic Writing: Intercultural and Textual Issues*. Amsterdam, John Benjamins.

Coates, J. (1983). *The semantics of the modal auxiliaries*. London, Croom Helm.

Cope, B. and M. Kalantzis, Eds. (1993). *The Powers of Literacy: A Genre Approach to Teaching Writing*. London, Falmer Press.

Coe, R. (1987) "An apology for form; or, who took the form out of the process?" *College English* 49/1:13-28

Coe, R. (1994) "Teaching genre as process" In A. Freddman and P. Medway *Learning and Teaching Genre* Portsmouth, NH, Boynton/Cook

Couture, B. (1986). "Effective ideation in written text: A functional approach to clarity and exigence". In B. Couture. *Functional Approaches to Writing: Research Perspectives*. Norwood, NJ, Ablex**:** 69-91.

Creme, P. and M. Lea (1997). *Writing at University: A guide for students*. Milton Keynes, Open University.

Crismore, A. and R. Farnsworth (1990). "Metadiscourse in popular and professional science discourse". In W. Nash. *The Writing Scholar: Studies in Academic Discourse*. Newbury, CA, Sage.

Crismore, A., R. Markkanen, et al. (1993). "Metadiscourse in persuasive writing: a study of texts written by American and Finnish university students." *Written Communication* **10**(1): 37-71.

Crompton, P. (1997). "Hedging in academic writing: Some theoretical problems." *English for Specific Purposes*.

Crookes, G. (1986). "Towards a validated analysis of scientific text structure." *Applied Linguistics* **7**(1): 57-70.

Currie, P. (1993). "Entering a disciplinary community: Conceptual activities required to write for one introductory university course." *Journal of Second Language Writing* **2**(2): 101-117.

Devitt, A. (1993). "Generalising about genre: new conceptions of an old concept." *College Composition and Communication* **44**(4): 573-86.

Dong, Y.-R. (1998). "Non-native graduate students' thesis/dissertation writing in Science: Self-reports by students and their supervisors from two U.S. institutions." *English for Specific Purposes* **17**(4): 369-390.

Dudley-Evans, A. (1991). "Socialisation into the academic community: linguistic and stylistic expectations of a PhD thesis as revealed by supervisor comments". In P. Adams, B. Heaton and P. Howarth. *Socio-cultural Issues in English for Academic Purposes*. London, MacMIllan.

Dudley-Evans, A. (1993). "The debate over Milton Friedman's theoretical framework: an applied linguist's view". In W. Henderson, A. Dudley- Evans and R. Backhouse. *Economics and Language*. London, Routledge.

Dudley-Evans, A. (1994). "Variation in communication patterns between discourse communities: the case of highway engineering and plant biology". In G. Blue. *Language Learning and Success: Studying Through English*. London, MacMillan.

Dudley-Evans, A. (1999). "The dissertation - a case of neglect?". In P. Thompson. *Issues in EAP writing research and instruction*. Reading, UK, CALS, The University of Reading**:** 28-36.

Dudley-Evans, A. and W. Henderson (1990). "The organisation of article introductions: evidence of change in Economics writing". In W. Henderson and A. Dudley- Evans. *The Language of Economics*, ELT Documents 134 Modern English Publications**:** 67-78.

Dudley-Evans, A. and M. St John (1998). *Developments in ESP: a multi-disciplinary approach*. Cambridge, Cambridge University Press.

Eggins, S. (1994). *An Introduction to Systemic Functional Linguistics*. London, Pinter.

Evans, D. (1995). *How to Write a Better Thesis of Report*. Melbourne, Melbourne University Press.

Ewer, J. (1979). *The modals in formal scientific discourse: Function, meaning and use*. Santiago, Chile, University of Chile, Department of English Research Report Mimeograph.

Fairclough, N. (1992). *Discourse and Social Change*. Cambridge, Polity Press.

Field, Y. and L. Yip (1992). "A comparison of internal cohesive conjunction in the English essay writing of Cantonese speakers and native speakers of English." *RELC Journal* **23**(1): 15-28.

Freadman, A. (1994). "Anyone for tennis?". In P. Medway. *Genre and the New Rhetoric*. London, Taylor and Francis.

Freedman, A. and P. Medway (1994). "New views of genre and their implications for education".In A. Freedman and P. Medway. *Learning and teaching genre*. Portsmouth, NH, Boynton/Cook.

Geertz, C. (1983). "Blurred genres". In *Local knowledge*. New York, Basic Books.

Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge, Polity Press.

Giles, A. K. (2000). *From 'Cow College' to Life Sciences*. Reading, The University of Reading.

Gledhill, C. (1995). *Scientific Innovation and the Phraseology of Rhetoric. Posture, Reformulation and Collocation in Cancer Research Articles*. Unpublished PhD thesis University of Aston.

Gledhill, C. (2000). "The discourse function of collocation in research article introductions." *English for Specific Purposes* **19**: 115-135.

Granger, S. and S. Tyson (1996). "'Connector usage in the English essay writing of native and non-native EFL speakers of English." *World Englishes* **15**(1): 17-28.

Gunawardena, C. (1989). "The present perfect in the rhetorical divisions of biology and biochemistry journal articles." *English for Specific Purposes* **8**: 265-273.

Halliday, M. (1994). *An Introduction to Functional Grammar (2nd Edition)*. London, Arnold.

Hanania, E. and K. Akhtar (1985). "Verb form and rhetorical function in science writing: a study of MS theses in Biology, Chemistry and Physics." *ESP Journal* **4**(1): 45-58.

Hansen, K. (1988). "Rhetoric and epistemology in the social sciences: a contrast of two representative texts". In D. Joliffe. *Advances in writing research, Volume 2: Writing in academic disciplines*. Norwood, Ablex**:** 167-210.

Henderson, W. and A. Dudley- Evans (1990). "Introduction: The analysis of economics discourse". In W. Henderson and A. Dudley- Evans. *The Language of Economics*. ELT Documents 134, Modern English Publications.

Hewings, M. (1993). "The end! How to conclude a dissertation". In G. Blue. *Language learning and success: Studying through English*. London, MacMillan.

Hinds, J. (1987). "Reader versus writer responsibility: A new typology". In U. Connor and R. Kaplan. *Writing Across Languages: Analysis of L2 Text*. Reading, Mass, Addison Wesley**:** 141-152.

Hoey, M. (1984). "Persuasive rhetoric in Linguistics: A stylistic study of some features of the language of Noam Chomsky." *Forum Linguisticum* **8**(1).

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

Hopkins, A. and A. Dudley-Jones (1988). "A genre-based investigation of the discussion sections in articles and dissertations." *English for Specific Purposes* **7**: 113-121.

Huddleston, R. (1971). *The sentence in written English: a syntactic study based on an analysis of scientific texts*. Cambridge, Cambridge University Press.

Hudson, R. (1980). *Sociolinguistics*. Cambridge, Cambridge University Press.

Hyland, K. (1994). "Hedging in academic writing and EAP textbooks." *English for Specific Purposes* **13**(3): 239-256.

Hyland, K. (1996). "Talking to the academy: forms of hedging in science research articles." *Written Communication* **13**(2): 251-281.

Hyland, K. (1996). "Writing without conviction? Hedging in science research articles." *Applied Linguistics* **17**(4): 433-454.

Hyland, K. (1999). "Academic attribution: citation and the construction of disciplinary knowledge." *Applied Linguistics* **20**(3): 341-367.

Hyland, K. (1999). *Hedging in scientific research articles*. Amsterdam, John Benjamins.

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

Hyon, S. (1996). "Genre in three traditions: implications for ESL." *TESOL Quarterly* **30**(4): 693-722.

Ivanic, R. (1998). *Discourse and Identity*. Amsterdam, John Benjamins.

James, K. (1984). "The writing of theses by speakers of English as a foreign language: the results of a case study".In R. Williams, J. Swales and J. Kirkman. *Common Ground: Shared Interests in ESP and Communication Studies*. Oxford, Pergamon.

Johns, A. (1997). *Text, role and context: Developing academic literacies*. Cambridge, Cambridge University Press.

Johnson, K. (2001). *Is it a wood or are they trees? (Pit Corder Memorial Lecture)*. BAAL 34th Annual Meeting, Reading, UK.

Joliffe, D. and E. Brier (1988). "Studying writers' knowledge in academic disciplines". In D. Joliffe. *Advances in Writing Research, Volume 2: Writing in Academic Disciplines*. Norwood, NJ, Ablex**:** 35-88.

Jordan, R. (1992). *Academic writing course*. London, Nelson.

Kachru, Y. (1988). "Writers in Hindi and English". In A. Purves. *Writing Across Languages and Cultures: Issues in Contrastive Rhetoric*. Newbury Park, Sage**:** 109-137.

Kaplan, R. (1966). "Cultural thought patterns in inter-cultural education." *Language Learning* **16**: 1-20.

Kennedy, G. (1998) *An Introduction to Corpus Linguistics* London, Longman

King, A. and J. Brownell (1966). *The Curriculum and the Disciplines of Knowledge*. New York, John Wiley.

Knorr-Cetina, K. (1981). *The Manufacture of Knowledge*. Oxford, Pergamon.

Kress, G. (1993). "Genre as social process". In B. Cope and M. Kalantzis. *The Powers of Literacy: A Genre Approach to Teaching Writing*. London, Falmer Press.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago, Chicago University Press.

Lackstrom, J., L. Selinker, et al. (1972). "Grammar and technical English." *English Teaching Forum* **X**(5): 3-14.

Latour, B. and S. Woolgar (1979). *Laboratory Life: the Social Construction of Scientific Facts*. Beverly Hills, California, Sage Publications.

Lea, M. and B. Street (1998). "Student writing and faculty feedback in higher education: An academic literacies approach." *Studies in Higher Education* **23**(2): 157-172.

Lee, D. (2001). "Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle." *Language Learning and Technology* **5**(3): 37-72.

Littlemore, J. (2001). "An empirical study of the relationship between cognitive style and the use of communication strategy." *Applied Linguistics* **22**(2): 241-265.

Lyons, J. (1977). *Semantics*. London, Longman.

MacDonald, S. (1992). "A method for analyzing sentence-level differences in disciplinary knowledge making." *Written Communication* **9**(4): 533-569.

Malcolm, L. (1987). "What rules govern tense usage in scientific articles?" *English for Specific Purposes* **6**: 31-44.

Malinowski, B. (1935). *The Language of Magic and Gardening*. London, Allen Unwin.

Malinowski, B. (1946/1923). "The problem of meaning in primitive languages". In C. Ogden and I. A. Richards. *The Meaning of Meaning (Eighth Edition)*. London, Kegan Paul.

Martin, J. (1985). "Process and text: two aspects of human semiosis". In J. Benson and W. Greaves. *Systemic Perspectives on Discourse*. Norwood, NJ, Ablex. **1:** 248-274.

Martin, J. (1991). "Nominalisation in science and humanities: distilling knowledge and scaffolding text". In E. Ventola. *Functional and Systemic Linguistics: Approaches and Uses*. Berlin, Mouton de Gruyter**:** 307-337.

Mauranen, A. (1993). "Contrastive ESP rhetoric: metatext in Finnish-English Economics texts." *English for Specific Purposes* **12**(1): 3-22.

McCloskey, D. (1985). *The Rhetoric of Economics*. Brighton, Wheatsheaf Books.

McEnery, T. and A. Wilson (1996) *Corpus Linguistics* Edinburgh, University of Edinburgh Press

Miller, C. (1984/1994). "Genre as social action". In P. Medway and A. Freedman. *Genre and the New Rhetoric*. London, Taylor & Francis.

Miller, C. (1994). "Rhetorical community: the cultural basis of genre". In P. Medway and A. Freedman. *Genre and the New Rhetoric*. London, Taylor Francis.

Milton, J. and E. Tsang (1993). "A corpus-based study of logical connectors in EFL students' writing: directions for future research". In R. Pemberton and E. Tsang. *Studies in lexis*. Hong Kong, Hong Kong University of Science and Technology.

Miyahara, A. (1986). "A need for culture-particular criteria for observation and evaluation of persuasiveness in public communication: The case of Japan." *Studies in English Language and Literature* **XXXVII**(1-2): 85-102.

Moravcsik, M. and P. Murugesan (1975). "Some results on the function and quality of citations." *Social Studies in Science* **5**(1).

Myers, G. (1989). "The pragmatics of politeness in scientific articles." *Applied Linguistics* **10**(1): 1-35.

Myers, G. (1990). *Writing Biology: Texts in the Construction of Social Knowledge*. Wisconsin, University of Wisconsin Press.

Nash, W. (1990). "The stuff these people write". In W. Nash. *The writing scholar: studies in academic discourse*. Newbury Park, CA, Sage.

O'Brien, K. (1995). "The reform of doctoral dissertations in humanities and social sciences." *Higher Education Review* **28**(1): 3-19.

Oster, S. (1981). "The use of tenses in "reporting past literature" in EST". In L. Selinker, E. Tarone and V. Hanzeli. *English for Academic and Technical Purposes*. Rowley, MA, Newbury House**:** 76-90.

Palmer, F. (1990). *Modality and the English modals*. London, Longman.

Paltridge, B. (1997). *Genre, Frames and Writing in Academic Settings*. Amsterdam, John Benjamins.

Paltridge, B. (forthcoming). “Thesis and dissertation writing: An examination of published advice and actual practice.” *ESP Journal*.

Pennycook, A. (1997). “Vulgar pragmatism, critical pragmatism and EAP.” *English for Specific Purposes* **16**(4): 253-269.

Perkins, M. (1983). *Modal expressions in English*. London, Pinter.

Phillips, E. and D. Pugh (1994). *How to Get a PhD: Second Edition*. Milton Keynes, Open University Press.

Pickard, V. (1995). “Citing previous writers: what can we say instead of 'say'?” *Hongkong Papers in Linguistics and Language Teaching* **18**: 89-102.

Quirk, R., S. Greenbaum, et al. (1985). *A Comprehensive Grammar of the English Language*. London, Longman.

Rafoth, B. (1990). "The concept of discourse community: descriptive and explanatory adequacy". In G. Kirsch and D. Roen. *A Sense of Audience in Written Communication*. Newbury Park, Sage.

Ramanathan, V. and R. Kaplan (2000). “Genres, authors, discourse communities: Theory and application for (L1 and) L2 writing instructors.” *Journal of Second Language Writing* **9**(2): 171-191.

Régent, O. (1992). “Pratiques de communication en médicine: Contextes anglais et français.” *Langages* **105**(Mars): 66-75.

Ridley, D. (2000). "The different guises of a PhD thesis and the role of a literature review". In P. Thompson. *Patterns and Perspectives: Insights into EAP writing practice*. Reading, CALS, The University of Reading.

Robinson, P. (1991). *ESP Today: A practitioner's guide*. New York, Prentice Hall.

Rosch, E. (1975). “Cognitive representations of semantic categories.” *Journal of Experimental Psychology: General* **104**: 192-233.

Salager-Meyer, F. (1992). “A text-type and move analysis study of verb tense and modality distribution in medical English abstracts.” *English for Specific Purposes* **1**(2): 93-114.

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

Shaw, P. (1992). “Reasons for the correlation of voice, tense, and sentence function in reporting verbs.” *Applied Linguistics* **13**(3): 302-319.

Simpson, P. (1990). "Modality in literary-critical discourse". In W. Nash. *The Writing Scholar*. Newbury Park, CA, Sage**:** 63-94.

Spack, R. (1988). “Initiating ESL students into the academic discourse community: How far should we go?” *TESOL Quarterly* **22**(1): 29-51.

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

Swales, J. (1986). “Citation analysis and discourse analysis.” *Applied Linguistics* **7**(1): 39-56.

Swales, J. (1993). “Genre and engagement.” *Revue Belge de Philologie et d'Histoire* **71**(3): 687-698.

Swales, J. (1998). *Other Floors, Other Voices*. Mahwah, NJ, Lawrence Erlbaum Associates.

Swales, J. (2001). "EAP-related linguistic research: An intellectual history". In J. Flowerdew and M. Peacock. *Research perspectives on English for Academic Purposes*. Cambridge, Cambridge University Press**:** 42-54.

Swales, J. and C. Feak (1995). *Academic Writing for Graduate Students*. Ann Arbor, Michigan University Press.

Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge, Cambridge University Press.

Tarone, E., S. Dwyer, et al. (1981/1988). "On the use of the passive in two astrophysics journal papers".In J. Swales. *Episodes in ESP*. New York, Prentice Hall**:** 188-209.

Thetela, P. (1997). “Evaluated entities and parameters of value in academic research articles.” *English for Specific Purposes* **16**(2): 101-118.

Thomas, S. and T. Hawes (1994). “Reporting verbs in medical journal articles.” *English for Specific Purposes* **13**(2): 129-148.

Thompson, G. and Y. Ye (1991). “Evaluation in the reporting verbs used in academic papers.” *Applied Linguistics* **12**(4): 365-382.

Thompson, P. (1999). Exploring the contexts of writing: Interviews with PhD supervisors. *Issues in EAP writing research and instruction*. Reading, UK, CALS, The University of Reading.

Thompson, P. (2001). “Review of Thurstun and Candlin 'Exploring Academic English'.” *Language Learning and Technology* **5**(3): 28-31.

Thompson, P. and C. Tribble (2001). “Looking at citations: Using corpora in English for Academic Purposes.” *Language Learning and Technology* **5**(3): 91-105.

Thurstun, J. and C. Candlin (1997). *Exploring Academic English: A workbook for student essay writing*. Sydney, NCELTR.

Tribble, C. (1997). "Improvising corpora for ELT: quick-and-dirty ways of developing corpora for language teaching". In M. J and B. Lewandowska-Tomaszczyk. *PALC '97 Proceedings*. Lodz, Lodz University Press.

Tribble, C. (2000). "Genres, keywords, teaching: towards a pedagogic account of the language of Project Proposals". In L. Burnard and T. McEnery. *Rethinking Language Pedagogy from a Corpus Perspective*. Hamburg, Peter Lang.

Trzeciak, J. (1996). *Cultural Factors in English Academic Writing - The Problems of Non-Native Speaker Students*. Unpublished MA Dissertation, The University of Reading.

Trzeciak, J. and S. Mackay (1994). *Study Skills for Academic Writing*. Hemel Hempstead, Phoenix ELT.

Upton, T. and U. Connor (2001). “Using computerized corpus analysis to investigate the textlingustic discourse moves of a genre.” *English for Specific Purposes* **20**(4): 313-330.

Valero-Garcés, C. (1996). “Contrastive ESP rhetoric: metatext in Spanish-English Economics texts.” *English for Specific Purposes* **15**(4): 279-294.

Vande Kopple, W. (1985). “Some exploratory discourse on metadiscourse.” *College Composition and Communication* **36**: 82-93.

Ventola, E. and A. Mauranen (1991). "Non-native writing and native revising of scientific articles". In E. Ventola. *Functional and Systemic Linguistics: Approaches and Uses*. Berlin, Mouton de Gruyter.

Weissberg, R. and S. Buker (1990). *Writing Up Research: Experimental research report writing for students of English*. Englewoods Cliff, NJ, Prentice Hall Regents.

Williamson, M. (1988). "A model for investigating the functions of written language in different disciplines". In D. Joliffe. *Advances in Writing Research, Volume 2: Writing in Academic Disciplines*. Norwood, NJ, Ablex.

Yunick, S. (1997). “Genres, registers and sociolinguistics.” *World Englishes* **16**: 321-36.