The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

A corpus based investigation into the relationship between propositional content and metadiscourse in student essay writing

by

Mark Morgan

2011

A Dissertation presented in part consideration for the degree of:
MA Applied Linguistics and ELT
University of Nottingham, Ningbo-China campus

Summary

During the 2009/2010 academic year, a team from the Continuing Support Centre (CSP), a division of CELE, were involved in a project designed to gather information from UNNC divisional staff into their perceptions of academic shortcomings amongst their students. A common issue raised across the disciplines was that students commonly fail to adequately marshal arguments when writing essays. It was this realisation, together with the importance for students to be able to write argumentative essays, which prompted the course of this research project. However, far from considering this a problem related to students I approach the subject by investigating how metadiscourse and propositional content interact when arguments are marshalled with the aim of informing teaching and materials development.

Chapter one broaches this question by considering the type of language students are exposed to in the EAP classroom. It compares and contrasts language features of professionally written research articles with those commonly found in introductory textbooks. Its aim is to bring into focus the defining characteristics of the opposing genres. The chapter also points out that extracts from introductory textbooks commonly form the content, and therefore language of EAP textbooks, and concludes by arguing that students are not exposed to the type of language necessary for writing an effective argumentative essay.

Chapter 2 introduces metadiscourse and hypothesises that students fail to marshal arguments because of a lack of awareness of how to interact with the reader by anticipating and responding to potential reader queries or criticisms. This hypothesis is upheld by the research findings. Whilst this is likely to be of no surprise to the reader, the research compares the metadiscourse features from two corpora: a learner corpus compiled for the purpose of this research project and a commercially available corpus of essays written by native speaker students. The findings are set against a backdrop of work by Karl Maton into the sociology of education and reveal that successful native speakers not only interact with their readers but that they simultaneously interact with the propositional content of their essays. This allows them to effectively marshal arguments: a three way interaction that is absent from the literature on metadiscourse. It concludes by arguing that CELE students are taught how signal the structure of their writing to the reader but not how to interact with the reader or the content of their essay and that their learning remains segmented.

Contents

Acknowledgements

I would like to thank my supervisor, Dr John McKenny, for inspiring me to take a corpus based approach and to my friend and former office mate, Gina Roach, for introducing me to the work of Karl Maton. God bless you both.

Total word count 13,191 (inclusive of the summary)

Chapter 1: Introduction

Background

The University of Nottingham, Ningbo, China (UNNC), is a Sino-British university offering a range of undergraduate and postgraduate degrees. All courses, with the exception of modern languages, are taught and assessed through the medium of English: degrees being conferred by the University of Nottingham, UK. Students who apply to study at UNNC without the stipulated level of English language competency for their course are required to undertake a one year English for Academic Purposes (EAP) programme (although the course has changed since the start of this project, data were collected from students who had completed general EAP studies). The modules are designed and taught by tutors from the Centre for English Language Education (CELE). The CELE student body is substantial, with the intake for 2010/2011 standing at 1385. Of these, 1237 enrolled as undergraduates and 148 as Master's students; the majority of whom are Chinese nationals.

EAP and ESAP (English for Specific Academic Purposes) are sub-branches of English language teaching that focus on the needs of students who are about to enter tertiary education and study through the medium of English. Given these parameters, EAP and ESAP do not focus on language for the sake of acquiring language, but on the '…cognitive, social and linguistic demands of specific academic disciplines' (Hyland & Hamp-Lyons, 2002:2). One of the key roles of EAP and ESAP is to help students adjust to socio-cultural and linguistic expectations (Currie, 2005) of their chosen discipline (Hyland, 2003). The profession, therefore, requires teachers, materials developers and course designers to be cognisant of the socio-linguistic demands that students have to face as well as the needs of their students in terms of expectations imposed upon them and past educational experiences. In the case of UNNC, this means focusing primarily on the needs of Chinese students.

During the 2009/2010 academic year, a team from the Continuing Support Centre, a division of CELE, carried out a survey entitled English Language Competencies and Academic Skills (ELAS). The project involved interviewing

faculty members and conducting class observations under the remit of identifying the needs of postgraduate students at UNNC. ELAS flagged up a number of issues, one of which was students' failure to adequately marshal arguments: that they rely on knowledge telling and rarely engage in knowledge construction (Scardamalia & Beretier, 1987). A claim mirrored by Hood (2004) who, reporting on Chinese students in Hong Kong, suggests it is likely due to a lack of exposure to, and understanding of, how knowledge is constructed within academic disciplines. Although ELAS was aimed at postgraduate students it is inferred here that undergraduates have similar problems given that they share the same socio-linguistic background; with the exception that lecture's expectations of undergraduates, in terms of argumentative essay writing, are likely to be less exacting (Swales, 1995).

Marshalling arguments is taken to mean the ability to construct or develop a reasoned argument to answer an essay question. The ability to construct a reasoned argument, however, relies on a number of factors including an awareness of audience and purpose (Hyland, 2009; Johns, 1993; Park, 1986) and a mastery of necessary linguistic resources. This raises the important issue of raising students' awareness to the fact that writing is the product of an author's will to communicate with a specific readership for a specific purpose (Halliday & Hasan, 1976; Johns, 1993), and to recognise the discourse features used to achieve these aims (Hyland, 2003; 2004a).

EAP students exposure to rhetorical styles of academic language

Introductory textbooks commonly form the core reading, at least for the first year, of undergraduate course in the humanities and social sciences. In the hard sciences, they are commonly used right up until the time students are ready to write their dissertations (Myers, 1992). They also inform EAP and ESAP course books (Hyland, 1998b) which means that the language used by textbook writers filters down through the system such that not only are L2 students exposed to textbook language at the inception stage of their studies, but that it constitutes their main access to academic language (c.f. the example reading from a published EAP course book contained in appendix 12).

When writing academic essays students are commonly expected to interact with the reader, evaluate positions (both their own and others) and anticipate the reader's perspective on issues (Charles, 2007; Hyland, 2004b). It appears, however, that introductory undergraduate textbooks are written in a way that makes the knowledge they contain accessible to the neophyte by presenting knowledge as facts, or as Brown puts it a "single voiced' reading' (Brown, 1993:67) which, by doing so, omit to represent how the current state of knowledge has been developed through the discourse of 'past voices' (*ibid*). They tend to use strategies '…to convince their readers of the certainty of what they are describing' (Paxton, 2007:113) which can give the impression that knowledge is codified (Hyland, 1998b; Myers, 1992) as opposed to dynamic and constantly changing (Myers, 1992; Scardamalia & Beretier, 1987). This is observable in the following extract taken from an introductory textbook on business communication.

> *According to Professor Francis W. Weeks, Executive Director Emeritus of the Association for Business Communication, the most prevalent problem in business communication is that writers think only of themselves and their problems, not the reader. Putting what you want to say in you-attitude is a crucial step both in thinking about the reader's needs and communicating your concern to the reader* (Locker, 1989:98).

This short extract demonstrates to students how not to reference a source! There is no date of publication and the researcher's credentials and social ranking take prominence over the propositional content of the passage. This gives a clear and distinct impression that Weeks is the sole author of this claim and does not stand in opposition to other arguments or has gained either acceptance or criticism from other researchers. The use of the definite article in the comment '*the most prevalent problem*', as well as the use of the present simple form of the verb '*is that writers think*' give a distinct impression that the claim is to be taken as an accepted fact and is not open to discussion or critique by the student-readers. The use of the active voice and the pronoun *you* helps it

speak directly to the reader. It presents, in other words, the statement as a fact rather than a claim grounded in literature (Hyland, 2009).

In contrast to textbooks, research articles attempt to persuade a readership of knowledge peers to accept new ideas (Hyland, 1998b; Myers, 1992; Park, 1986). They rely on hedging devices (Hyland, 2009) and readily acknowledge the author's contribution to a growing body of knowledge (Myers, 1992). So, whilst research articles address the disciplines, textbooks, which tend to use more directives than research articles (Swales, et al., 1998), 'speak principally to students' (Hyland, 2002a). Once again, this phenomenon can be observed in this short extract from a journal article:

> *Studies (Gee, 1996; Heath, 1983) show that middle class urban literacy practices acquired in the home (i.e., primary discourses) coincide with those taught in school and at university (secondary discourses), while other literacy practices do not; this means that schooling creates inequalities and can lead to low attainment of certain groups in the educational context* (Paxton, 2007:111).

Both extracts are discussing the concept of communicating ideas. Yet the discourse features between the two differ greatly. Whereas the textbook presents a claim as a fact not to be challenged, the research article acknowledges the work of other researchers so as to form a backdrop for the author's own work. Not only does the style of referencing used in the research article make it possible (and arguably invites) the reader to follow up on the sources used, but the status of Gee and Heath (eminent professors or Ph.D. students) is not given. This gives a greater sense of equality, not only amongst researchers (Hyland, 2002a), but also the reader and the writer. This humbling move allows the reader to take a questioning stance by not making him or her feel intimidated by the writer's status in the community, which is the antithesis of the underlying social message conveyed by the textbook author. Moreover, the use of the modal verb *can* opens up the possibility of the claim being flawed and thereby asks for acceptance of the claim from the academic community (Myers, 1992).

So, whereas research articles aim to achieve harmony amongst peers (Hyland, 1998b; Swales, et al., 1998) textbooks place the student in a position of novice reader (Hyland, 1998b). This means that the linguistic forms used when writing textbooks differ in kind to those when writing research articles. Textbooks, then, tend to orient students towards a concept of knowledge telling by not marshalling arguments surrounding claims. This disallows students easy access to the type of rhetorical devices that make it easy for them to refer to the literature in order to marshal arguments in their own writing (Hyland, 1998b; Paxton, 2007). As Hyland puts it:

> *Understanding the written genres in one's field is essential to full acculturation and success, but introductory textbooks are … not representative of academic discourse in general* (Hyland, 1998b:4).

Textbooks, therefore, not only represent a canonised version of currently accepted knowledge but also frame academic language within a knowledge telling monologue. Once this type of discourse becomes embedded within EAP and ESAP course material it becomes students' point-of-reference for assimilation into the language of the discourse community.

Textbooks are written for students who are new to a discipline. They aim to introduce the main theories and concepts of the discipline in order that students will be able to tackle complex research articles that assume this knowledge. Moreover, knowledge of one's audience shapes the rhetorical choices in writing (Charles, 2007; Hyland, 2009; Park, 1986) which is the reason textbook authors use the rhetorical style they do. This, however, can mislead students into believing this to be the style of academic writing that should be mimicked in their own essays (Paxton, 2007), and which is likely to become their default rhetoric (Canagarajah, 2002; Fox, 1994).

This highlights the need to expose students to the type of language they need in order to write an effective essay (Evans & Green, 2007) and reiterates the point that EAP and ESAP should raise students' consciousness of audience awareness. However, if, as Hyland suggests, textbooks commonly form the basis for EAP and ESAP course material then problems may arise because the purpose for which the textbooks were originally written, and for which they are being used in EAP/ESAP classrooms, remain markedly different in the sense that the notion of audience remains an '…inherently situational concept…' (Park, 1986:480). This means the raising of audience awareness requires explicit instruction so that students know exactly what it is that *their* readers expect of *them* (Charles, 2007). This is what Searle refers to as 'extra linguistic institution' which he defines as '…a system of constitutive rules in addition to the constitutive rules of language …' (Searle, 1976:14).

In way of a conclusion, it has to be said that whilst students are expected to marshal arguments when writing essays, the models they are exposed to, in terms of language and academic discourse, are often far removed from the rhetorical style they should be using in their own writing. In fact, if features of academic language are represented along a cline, then professional research articles represent one end of the spectrum in terms of the language used, the style of discourse employed as well as the assumed knowledge of their audience, whilst textbooks represent the opposing extreme. In other words, whilst research articles talk to subject experts, textbooks address subject novices. Neither, therefore, are idealised models for an EAP and ESAP writing class in that essay writing, as a genre, has requirements that would fall somewhere in between these two extremes. Materials developers, course designers and teachers, therefore, need to be cognisant of the potential for misalignment between teaching practices and student needs. This in turn underscores the need for research that focuses on exposing the underlying features of high scoring essays that can, in turn, form the basis of good teaching practice.

So far, the terms EAP and ESAP have been used interchangeably, giving the impression that both are common elements of university pre-sessional language courses. The real world reality is, however, that most EAP courses are general in

nature because of the constraints on financial, material and physical resources, as well as teaching staff that are required to run discipline specific ESAP programmes. For this reason, this dissertation will focus on general EAP requirements.

Chapter 2: Metadiscourse

Argumentative essays are one of the more common genres that students in tertiary education have to produce (Hyland, 1998b; 2009; Johns, 1993). They are central to many disciplines (Kuteeva, 2011) and amongst the most perplexing for non-native speakers (Johns, 1993) as they involve the writer interacting with the reader (Hyland, 2004b). This means that argumentative essay writing is primarily a social practice (Kuteeva, 2011) that requires the writer to have a grasp of the reader's expectations of how ideas are communicated, as well as a mastery of the linguistic features that are used to covey meaning. This is where metadiscourse comes into the mix as it is, essentially, language that allows the writer to achieve these aims.

Metadiscourse is defined by Crismore *et al.* (1993:39) as: 'refer[ing] to writers' discourse about their discourse — their directions for how readers should read, react to, and evaluate what they have written about the subject matter'. This definition, however, seems somewhat imposing in that it places the writer in a position of dictating the reader's reaction to the text. Hyland takes a slightly different view in that he sees metadiscourse as part of a process of negotiation of meaning between the writer and reader. He defines metadiscourse as:

> *the cover term for the self-reflective expressions used to negotiate interactional meaning in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community (Hyland, 2005:37)*

He argues that metadiscourse features are 'crucial rhetorical devices' (Hyland, 1998b:5) that allow the writer to engage with the reader according to socially accepted norms. This allows the author to appear 'credible and convincing' (Hyland, 1998b:5). They also empower the writer with stylistic choices, within the framework of their discipline, so that writers can present their own voice (Hyland, 1998b; 2005); what Johns (1997:58) refers to as 'sociolinguistic practices' which are a 'defining feature of successful writing' (Hyland, 2005:xxiii).

Hyland (2005) aligns himself with Vande Kopple's (1985) textual and interpersonal categorisation of metadiscourse by broadly categorising metadiscourse features into interactive and interactional, where interactive devices serve to 'guide the reader through the text, while interactional resources involve the reader collaboratively in the development of the text' (Thompson, 2001:58). So whilst interactive metadiscourse is used to structure a text, the interactional dimension denotes the writer's voice through commentary on propositional content. An overview of these features is presented below.

Metadiscourse functions

Interactive	**Function: to guide the reader through the text**	**Resources**
Transitions	Express relations between main clauses	In addition; but; thus; and
Frame markers	Discourse acts, sequences or stages	Finally; to conclude; my purpose is
Endophoric markers	Reference to information in other parts of the text	Noted above; see figure; in section 2
Evidentials	Reference to information in other texts	According to X; Z states
Code glosses	Elaborations on propositional meaning	Namely; e.g.; such as; in other words
Interactional	**Function: to involve the reader in the text**	**Resources**
Hedges	Withhold commitment and open dialogue	Might; perhaps; possible; about
Boosters	Emphasise certainty or close dialogue	In fact; definitely; it is clear that
Attitude markers	Express writer's attitude to proposition	Unfortunately; I agree; surprisingly
Self mentions	Explicit reference to author(s)	I; we; my; me; our
Engagement markers	Explicitly build relationship with reader	Consider; note; you can see that

Table 1: showing the different metadiscoursal features. Reproduced from Hyland (2005:49).

Hyland cautions, however, that rhetorical features only serve as metadiscourse if they refer to information within the text, and not if they are external to the writing (Hyland, 2005). Frame markers (first, then, at the same time) for example, function metadiscoursally if they synthesis the argument and not if they refer to time sequences. Any study of this nature, therefore, requires a manual analysis of text to compliment a corpus approach.

Reader-writer engagement

When students write argumentative essays they are expected to take a stance and persuade the reader to accept a position (Hyland, 2002a; Wu, 2007) by evaluating the course content in an appropriate manner (Wu, 2007), for which metadiscourse is essential (Hyland, 2005). This means having to anticipate, and respond to, reader responses to propositional content (Thompson, 2001). Yet as Park (1986) points out, to achieve this a writer first needs to understand his or her audience. It also requires him or her to assess claims made against a backdrop of subject knowledge (Wilson & Sperber, 2004) which the initiate into a subject, by definition, lacks. This means that the effective use of metadiscourse devices to achieve a rhetorical aim hinges on understanding the social norms, the relationship with the reader and the purpose for writing (Hyland, 1998a) which hints at a more complex interaction than Hyland's categories suggest. It also depends on a shared knowledge of disciplinary practices between reader and writer, as well as understanding and familiarity with the genre (Hyland, 2005) which can be highly problematic for L2 writers who may lack the cultural insight as well as the necessary linguistic sophistication (Aijmer, 2002; Park, 1986).

The skill of the writer in engaging the audience has been shown to have a significant effect on the grades students are awarded for their essays (Mei, 2007). Hyland (2005), for example, discovered that high scoring GCSE essays written by Chinese speaking students in Hong Kong tended to exhibit metadiscourse features closely associated with L1 students' writing. He goes on to argue that metadiscourse is the language of a community of practice in that it shapes the discourse so as to conform to the knowledge building norms of that community (Hyland, 2005). He thus concludes that 'a lack of familiarity with the metadiscourse conventions central to many expository genres in English may be detrimental to learners' academic performance' (Hyland, 2005:136) and that interactional features of metadiscourse are a 'defining feature of successful academic writing' because they allow the writer to 'claim solidarity with readers' whilst displaying self reflection on claims made and 'acknowledging alternative views' (Hyland, 2005:219). Thompson (2001) concurs, stating that interactive metadiscourse needs to form the focus of training in academic writing whilst Wu

(2007) argues that students' mastery in evaluating evidence, as measured by their grades and marker's comments, is positively correlated with their ability engage with the audience.

Culturally-based rhetorical styles have also been shown to exist (Biggs, 1996; Fox, 1994; Liu, 2008). Asian students, for example, tend to overuse frame markers such as *firstly, secondly, finally* (Hyland, 2005), whilst Swedish (Aijmer, 2002) and Chinese (Hyland, 2005) speakers of English have shown to exhibit an over-reliance on the model verb *will* compared to the over use of *can* and *could* by German speakers, and *may* by French (Aijmer, 2002). In contrast, native English writers are more likely to favour alternative devices such as attitude markers, hedges and boosters. It appears, therefore, that L2 writers show a preference for interactive metadiscourse devices, whereas native speakers appear to be achieving a balance between interactive and interactional.

There could be a number of reasons for this. Hinds (2001), for example, puts forward the thesis that English is a writer responsible language as opposed to the reader responsible style of Chinese. Yet, at the same time, it has been claimed that not only is there little evidence to show a major divergence in rhetorical patterns between Chinese and English writing styles (Kirkpatrick, 1997; Mohan & Lo, 1985) but that '…language transfer seems more likely to help them [Chinese writers of English] than to interfere' (Mohan & Lo, 1985:515); so the evidence for L1 transfer, in terms of writing styles, remains opaque.

Whilst L2 writing is undoubtedly influenced by an author's L1 writing styles many of these 'cultural nuances' may have as much to with the way in which students are taught to write English than it does with L1 transfer which, as argued in chapter one, may well be compounded by the language they are exposed to. Kuteeva (2011), in fact, concludes from a study involving a multicultural group of students on a writing course that cultural differences have far less an impact on students use of interactional resources than their educational background does. She found, in fact, that students from a humanities background tend to engage the audience more than students from a science background. This suggests that

to simply attribute errors in student writing to L1 influence may, actually, mask the whole truth. A number of writers, however, contest that this does not have to be a barrier because students can be taught how to engage an audience with a degree of effect (Charles, 2007; Mei, 2007; Thompson, 2001).

Zarei and Mansoori (2007), however, claim that Persian academics tend to downplay reader involvement in favour of textuality suggesting that L1 could influence writing styles more so than is being acknowledged here. Notwithstanding, metadiscourse analyses can provide information about the way in which writers create a relationship with readers according to the genre and discipline so that students can be explicitly taught the concept of reader-writer engagement within their discipline. Ajmer (2002) picks up on this point arguing that an L2 writer's lack of engagement with the audience is likely due to learner's uncertainty about linguistic choices when developing an academic argument. She claims, moreover, that ESL (English as a Second Language) textbooks put too much emphasis on the use of modal verbs, and neglect alternative strategies. Whilst her reference to ESL is slightly confusing, the fact that she is discussing the teaching of academic writing justifies the addition of her comments in this study. An important aspect of EAP is, after all, to socialise students into an academic culture which means exposing them to the rhetorical features that lecturers expect to see in their essays.

Are there problems specific to Chinese students?

Evans and Green (2007) conducted an extensive survey of Cantonese speaking students in Hong Kong studying in tertiary education through the medium of English. The survey took a multi-dimensional approach using self-reported questionnaires, interviews with staff and students and focus-group discussions with programme leaders. Their data are presented in terms writing criteria, including expressing ideas clearly, referring to sources and summarising. They report that students largely experience difficulty with discipline specific writing. That their problems in reading derive from limited linguistic resources rather than problems understanding content or structure and that they experience difficulty when communicating ideas clearly and succinctly in writing. Whilst this

is an important starting point in identifying problems students' experience, the data do not identify specific problems that students have with academic writing or the root cause of those problems.

Much of the literature, then, focuses on the interaction between reader and writer and/or addresses socio-cultural influences on L2 writing. However, the relationship between metadiscourse and propositional elements in essay writing remains a neglected area (Hyland, 2005). This study, therefore, intends to investigate this relationship in student essay writing.

Metadiscourse and semantic gravity

At this point I feel it is necessary to introduce the work of Maton (2009; forthcoming) on knowledge construction in education and to link this to the thesis of the dissertation: the interface between metadiscourse and propositional content.

Maton uses the term *semantic gravity* as a measure of the extent to which knowledge is context dependent. Knowledge that is highly context dependent in the mind of the student and cannot easily be applied to contexts outside of that in which it has been learned is said to have a strong semantic gravity. Knowledge that can easily be applied to other contexts is said to have a weak semantic gravity. His thesis is based on the observation that knowledge learned in the classroom is highly context dependent. Not because of the nature of the knowledge but because of the way it is taught which isolates students from situations whereby they have the opportunity to use their knowledge in contexts other than those in which it is presented (Maton, 2009). Knowledge, in other words, is not automatically or easily transferable from one context to another. However, far from being a black and white issue, Maton (forthcoming) places knowledge, or more accurately the learning of knowledge, or what he terms 'forms of learning' (Maton, forthcoming: Ch.6:4) along a cline of semantic gravity ranging from weak to strong. This is shown in table 2 below.

Forms of learning

Semantic gravity	Coding of responses	Form taken by student responses	Example quote taken from student answer
Weaker ↑ ↓ Stronger	Abstraction	Presents a general principle or procedure that moves beyond the cases to address wider or future practice	Legal intellectual property issues are a major consideration when developing a product
	Generalisation	Presents a general observation or draws a generalising conclusion about issues and events *in* the case	Precious time would be wasted and deadlines not met when members did not have a full concept of the project
	Judgement	Goes beyond re-presenting or interpreting information to offer a value judgement or claim	While each metaphor provides a realistic learning environment …, I felt that the *Nardoo* metaphor assists with navigation, while the *StageStruck* metaphor was a barrier to effective navigation
	Interpretation	Seeks to explain a statement by interpreting information from the case or adding new information. May include use of other literature or personal experience	While not alluded to in the interviews, this may have caused problems for the team, as there would have been a new software to work with, and transferral of information from *Hypercard* to *MediaPlant*
	Summarising description	Descriptive response that summarises or synthesises information presented in the case, including re-wording and re-structuring of a number of events into one statement. Does not present information from beyond the case	This involved creating the overall structure and content of the project, with design briefs and statements being forwarded to the client, with the final design statement being signed off by the client, giving a stable starting position for the project
	Reproductive description	Reproduces information directly from the case (i.e. quotations)	The NSW Department of and Water Conservation (DLWC) approached the Interactive Multimedia Learning Laboratory (IMMLL) at the University of Woolongong to develop an educational multimedia package

Table 2: showing Maton's 'a language of description for semantic gravity'. Reproduced in its entirety from Maton (2009:49).

Metadiscourse forms the underlying rhetorical features that enable the writer to interact with the propositional content of the essay. It operates by allowing the author to strengthen the semantic gravity when concrete examples are required or to weaken it when judgments or generalisation are called for (Maton, forthcoming). Kuteeva (2011:47), in fact, argues that metadiscourse makes a text more 'reader-oriented' which is a mark of good academic writing. Yet as Hyland states:

> *it is rare for metadiscourse to be either explicitly taught or adequately covered in writing materials in a way which either shows the systematic effect of particular options or reveals the important interactive nature of discourse (Hyland, 2005: 178).*

Wu (2007:256), therefore, calls for 'the critical practices expected of students to be unmasked'. This is a key point because writer-reader interaction is a social process that is often neglected in EAP writing classes. It is also for this reason why I anticipate that CELE students will show a leaning towards interactive metadiscourse, which is easily taught, and less so towards interactional metadiscourse that relies on sociolinguistic understanding as well as an ability to apply the knowledge to contexts outside of classroom exercises.

Maton (2009; forthcoming) conceptualises this as cumulative versus segmented forms of learning. The problem is that he comes from a sociological perspective and works with epistemological as opposed to linguistic models. Nevertheless, the issues he addresses deal with the way in which packages of knowledge are taught and learned. He also addresses the issue of whether or not these packages of knowledge remain, in the minds of the students, entrenched in the context in which they were initially learned, what he terms 'segmented learning' (Maton, forthcoming: Ch.6:5), or whether learners can step outside of the context in which they were learned and apply them to other situations; what he terms 'cumulative learning' (ibid). Given that this dissertation is considering the interface between metadiscourse and propositional content, Maton's (2009; forthcoming) theoretical model is highly relevant when it comes to interpreting the findings.

In my endeavour I have taken a corpus-based analysis of student writing by developing a learner corpus of CELE essays. Although the essays were submitted without grades CELE students rarely achieve more than 60%, with the average grade being in the region of 45-50%. Therefore, whilst the majority of students manage to pass the EAP course successfully, and progress onto their academic programmes, more could arguably be done to assist them in understanding the expectations of academic essay writing in terms of engaging with the reader and the content of their essays.

The corpus exhibits a weakness in that it is not discipline specific. Instead, it is a corpus of essays from a general EAP course loosely based around the humanities and social sciences. Subsequently it has not been possible to control for linguistic differences between genres and subjects (Hyland, 2002a). In an attempt to compensate for this the control corpus consists of essays from disciplines that fall under the general remit of the arts, humanities and social sciences.

Hypothesis

CELE students use interactive metadiscourse effectively but fail to engage the audience at the interactional level.

Research questions

1. How is metadiscourse used by successful student writers?
2. How do writers orient themselves to their audience?
3. What are the key differences in the use of metadiscourse between the CELE and BAWE corpora?
4. Is there a link between metadiscourse and propositional content?

Chapter 3: Corpus based studies

Corpus linguistics is defined by Granger (2002:4) as '…a methodology which is founded on the use of electronic collections of naturally occurring texts, vis. Corpora'. Corpora, then, are computer databases of naturally occurring language (McEnery, Xiao, & Tono, 2006). They enable researchers to observe enormous amounts of data in a relatively short time with relative ease (Gilquin, Granger, & Paquot, 2007) which renders corpora powerful tools for discovering language features (Granger, 2002). The language they contain has been produced naturally for the purpose of real life communication, whether casual conversations between workmates or academic essays written by students. It has not, in other words, been produced under controlled conditions for purposes such as teaching or research. This allows researchers to draw comparisons across a range of co-texts which, in turn, allows for the analysis and description of linguistic features (Gilquin, Granger, & Paquot, 2007). This has led, amongst other things, to the realisation that academic phraseology is not generic but that it varies across genres and is affected by the communicative purpose it serves (Gilquin, Granger, & Paquot, 2007) which has had a major impact on language teaching (Hunston, 2002; Johns, 2002; Mukherjee, n.d.).

Concordance tools are computer programmes designed to sift out and read specific language features, at the behest of the researcher, from the language contained in the corpus: data are presented in the form of concordance lines. This enables the researcher to observe specific language functions as well as the framework of norms that are defined by the community in which the communication is taking place, and which shapes features of the discourse such as formality and vocabulary. In terms of student essay writing, these norms are defined by the academic community that the lecturer who set the task belongs to, and into which the student is being apprenticed. A corpus approach, therefore, constitutes a powerful way to observe specific language features that are part of the community's discourse practices.

A generic weakness with corpora studies is that the language observed, at least initially, can only be viewed within the very limited context of the concordance

lines, which is particularly problematic when studying features such as metadiscourse which are context dependent (Hyland, 2005). In defence, concordance programmes have a feature that allows the researcher to expand on a chosen concordance line to view the context in which the language operates. A number of writers (Charles, 2007; Granger, 2002; Luzon, 2009; Weber, 2001) also caution that corpus studies compliment, and do not replace, other research methods. Luzon (2009) and Weber (2001), for example, call for corpora approaches to teaching academic writing to be combined with genre analysis whilst Charles (2007) argues for the teaching of specific rhetorical functions by combining corpora studies with discourse analysis. A corpus based study, therefore, does not necessarily define the boundaries of a research project yet forms, in many respects, a potential starting point for further analysis.

Corpus design considerations

Corpus design is a crucial factor. Whilst purposes such as vocabulary studies require corpora to be balanced by drawing data from a variety of different mediums (lectures, essays, research reports) (Schmitt, 2000), others need to be specialised (Hyland, 2002b; 2009). Regardless of the design, language samples will always be of a restricted range (Cook, 1998) because the number of possible linguistic variations is potentially infinite. This aside, provided corpora represent, as far as possible, a cross section of the discourse of the community being studied in terms of genre, discipline and even task, then a fairly true picture of how language operates within that community of practice can emerge (McEnery, Xiao, & Tono, 2006).

Language contained in corpora is authentic in terms of language, purpose and task (used here to refer to essay writing in general, rather than a specific question type such as 'discuss'). This constitutes one of the strengths of a corpus approach. Yet as Gilquin *et. al.* remind us, EAP researchers and materials developers tend to draw their data from native speaker corpora. They report, in fact, that '…the overwhelming majority …' of corpus based studies published in the Journal of English for Academic Purposes between 2005 and 2006 were

'…native-corpus based' (Gilquin, Granger, & Paquot, 2007:323). This can be problematic because it does not necessarily identify language that learner's are producing (Sinclair, 1991).

Flowerdew (2001) suggests this may be due to the absence of published learner corpora vis-a-vis the accessibility of those compiled from native speaker data. Whatever the reason, reliance solely on native speaker corpora only allows writers and researchers to view language from the perspective of desired *input*, that is what the syllabus dictates that students should learn, and fails to consider the *production* stage during which the needs of specific groups of language learners can be identified (Ellis, 1994). This can lead to assumptions being made about learner needs (Nesselhauf, 2004) for whilst novice student writers may share similar problems regardless of whether they are native or non-native speakers, some difficulties remain specific to L2 writers which can have '…far-reaching methodological and pedagogical implications…' (Gilquin, Granger, & Paquot, 2007: 323). A native speaker corpus is fine, therefore, so long as the purpose of the research is to study how language is used by native speakers. However, if the same corpus is used to identify areas in which teaching of language to L2 speakers could be improved then the *research approach* lacks authenticity because student needs, as identified through an analysis of their writing, has not been assessed.

In terms of research for the purpose of materials development, *authenticity of needs* is a crucial consideration. This requires the specific needs of a group of L2 learners to be identified by matching features of their language against a benchmark set by native speaker *students* (Luzon, 2009; Weber, 2001) for which a corpus based study is ideally suited. A key consideration, therefore, is the purpose of the corpus. If it is to identify L2 writer needs then a learner corpus is required that needs not be as specific in terms of genre and task than if the purpose was to investigate discipline specific language features when an L1 or even professional writer corpus, that has a high degree of specificity, would be required. Yet whilst the corpus may not need to be *as* specific it would be wrong to assume that specificity, especially discipline specificity, is not important. Specificity is, after all, necessary to control for variables given that

disciplines provide the framework for socially constructed meaning reflected in the language choices (Hyland, 2008; 2009) which are of interest in this study. But then so is representativeness. That is to say with a project such as this, which studies student essay writing, data have to represent what the cohort being studied understand in terms of academic writing as well as the nature of the course they have completed. It is important, therefore, that a corpus represents the language of the community which is being studied (Mukherjee, n.d.). This dissertation contrasts the language features of essays written by native speakers with those of L2 writers to add authenticity to the research and help compensate for a lack of specificity.

Learner corpora

Learner corpora are built from work produced by L2 writers. They are a way of reviewing language at the output or production stage; the findings of which are fed directly into syllabus design and material development (Mark 1998. Cited in Granger, 2002). A comparison of data from both L2 and L1 essays, then, can shed light on the specific needs of L2 learners which may not be achievable using other methods such as genre analysis or contrastive rhetoric (Gilquin, Granger, & Paquot, 2007; Nesselhauf, 2004), what Gilquin, Granger & Paquot (2007:322) term contrastive interlanguage analysis. This being said, considering Hyland's (2002b; 2009) insistence that specificity is the key to successful data collection, it would be wrong to assume that the so-called errors, or features of Chinese speakers' written academic English are transferable to L2 writers from other socio-linguistic backgrounds (Biber, Conrad, & Reppen, 1998). Conclusions drawn from this study, therefore, can only lay claim to features of Mandarin-Chinese speaking student writing.

Sinclair (1996. Cited in Granger, 2002) points out that data collected for a learner corpus has to come from naturally occurring acts of communication and discounts the use of data gathered from experimental designs. The definition of natural speech acts is, however, a fuzzy concept in that coursework essays are *de facto* exams which arguably puts them into a category of an artificial communication act; a point that Sinclair acknowledges (Sinclair, 1991). Yet if

the purpose of the study is to investigate ways in which students can improve their essay writing skills then the *purpose* for the act of communication remains authentic in that essays are a communicative act in their own right (Ede & Lunsford, 1984; Hyland, 2004b; Johns, 1993; Park, 1986; Scollon & Scollon, 2001; Wollman-Bonilla, 2001). This, in turn, authenticates the language being used. Comparisons need to be made, therefore, between native and non-native speaker corpora data gathered under similar conditions.

Corpora size

The more words a corpus contains the more it will show norms of language use, making it a more reliable research tool. This can be better explained in terms of the social sciences that may conduct a study of IQ scores. If the participant pool is large enough then data from the IQ tests will form a bell shaped curve that is representative of the wider population, and from which the average IQ score can be deduced. If the sample is too small anomalies such as very high or very low scores could easily skew the results and give an untrue picture of the average for the wider population. In terms of corpus studies, therefore, a large corpus is required so that a true picture of language features can emerge which can then be claimed to represent the norm for that community in terms of the purpose for which the language serves (i.e. essay writing).

It is difficult to obtain clear figures on idealised corpora sizes because the goalposts are constantly shifting as the technology for collecting, sorting and retrieving data develops. Indeed, many modern corpora now run into multiple-millions of words. The CELE (learner) corpus has a running total of 7,555 words whilst the BAWE sub-corpus has a running total of 32,746 words. The learner corpus, therefore is comparatively small. Even too small, one may argue. Yet it must be remembered that it is highly specialised in that it has been compiled from a relatively small homogenous group of students sharing a common linguistic, cultural and educational background who have all undertook the same EAP course. It should be sufficient, therefore, to reveal anomalies specific to this cohort.

Chapter 4: Research Design

This research project uses a hypothesis as its starting point. The hypothesis, however, is intended to give the research direction as opposed to defining the parameters to be studied. This is because the preconceptions that normally define a hypothesis need to be highly conceived so as to narrow the focus of the research. This research project is treading relatively new ground in that it focuses on the relationship between metadiscourse and propositional content (Hyland, 2005). A heuristic is necessary, therefore, to ensure that data are viewed objectively and not distorted by pre-conceived ideas. It also reduces the possibility of committing a type one or type two error by falsely accepting or rejecting the hypothesis on the basis of the findings.

This study makes a comparison between two general, as opposed to discipline specific, corpora. This could prove to be problematic when it comes to making claims based on the findings because language features are affected by genre, discipline and even task (a 'discuss' essay, for example, will exhibit different features from an 'analyse' essay). University degrees, however, are becoming 'increasingly interdisciplinary' (Krishnamurthy & Kosem, 2007:363) which blurs the boundaries somewhat when it comes to teaching discipline specific language, especially on an EAP course.

The research design takes a top-down approach by reviewing a wide range of metadiscourse before narrowing the focus of study onto features that have been identified as being of interest during the first stage of data analysis. It does not, as many projects do, take a bottom-up approach by focusing on pre-identified features, such as boosters or hedging devises. This, hopefully, will work in its favour. Indeed, as Biber (1993) points out, general as opposed to highly specific corpora include slight variations across the texts making them more suited to a purpose such as this that seeks to inform materials development and teaching on a general EAP programme. The lack of subject specificity, therefore, may reveal patterns of language features that may otherwise have been missed if the corpora were too specific (McEnery, Xiao, & Tono, 2006) and may even help to

shed new light on features of academic written discourse that could or should be taught on a general EAP programme.

Developing the learner corpus

Learner corpora allow researchers access to language features that would otherwise prove difficult, if not impossible, to access (Aijmer, 2002). However, metadiscourse needs to be viewed in context of the essay (Hyland, 2005). A learner corpus (UNNCELE) was therefore compiled using whole texts, as opposed to chunks of text, to allow for the analysis of features at the essay level. These were drawn from a central UNNC email account to which UNNC students were asked to voluntarily submit their essays by Dr John McKenny, UNNC's resident corpus linguist, and from whom permission was gained to access the account and use the data. The email account contains a wide range of essays and reports, both undergraduate and postgraduate, from across the disciplines (not just CELE). For the purpose of this research project, however, only CELE essays, written by Chinese students who had completed the undergraduate EAP course, were used.

The corpus was refined to ensure that data constitute a representative sample of work produced by CELE students at the culmination of their course after completing two semesters of EAP instruction. This was achieved by removing essays that were either produced during semester one of the EAP programme, that were written for one of the first year content courses, were submitted by students who had completed the pre-Master's EAP course or were written by non-Chinese students. This was to ensure that variables such as course content, educational background, maturity level, essay type, level of study, student ethnography and word length were controlled for. Essays were then cleaned up by removing titles, title pages, reference lists and graphs or charts to ensure these features did not interfere with the concordance analyses. UNNCELE, therefore, consists solely of cleaned up essays submitted by Chinese students at the completion of their EAP course for summative assessment.

The control corpus

The British Academic Written English (BAWE, 2004-2007) corpus was used as the control. BAWE was an Economic and Social Research Council funded project undertaken conjointly by the University of Warwick, the University of Reading and Oxford Brooks for the purpose of developing a corpus for researching higher education student writing. The data contained in the corpus are from upper-second (60-69) or first class (70+) essays written by British students and was chosen on the basis that the essays have, by definition of their grades, successfully engaged their readers.

BAWE, in its entirety, contains essays from across a number disciplines including the Social Sciences, Arts and Humanities, Physical and Life Sciences as well as different genres such as essays, research reports, proposals and literature reviews. In an attempt to develop a 'general' corpus consisting of the type of subjects most CELE students were expected to go on to study, and which most closely matched the content of the EAP course at UNNC (as it was at the time of data collection), a sub-corpus was compiled from BAWE. The following subjects were therefore not included: science subjects, engineering, law, creative writing for publication, anthropology, histography, medicine, computing and philosophy.

It was not possible to control for the various genres in the BAWE sub-corpus, with the exception of reports which were omitted as these were considered too different from the essays in terms of structure and language. This means that there is a misalignment between UNNCELE and the BAWE sub-corpus. Whilst this should be taken into consideration when interpreting the data it should also be remembered that this is largely a heuristic research project designed to identify potential areas of weaknesses in CELE student writing in terms of marshalling arguments. For the purpose of this research, therefore, the difference between the two corpora should not pose too much of an issue, even though a closer alignment between the two corpora would probably be necessary if a more focused project was to follow on from the findings.

Corpora statistics

Both corpora remained untagged on the basis that it takes a heuristic approach. Data were analysed using *WordSmith Tools 5.0* (Scott, 2009).

	UNNCELE	BAWE sub-corpus
Number of essay	92	435
Running word total	107,913	198,024
Number of tokens	7,555	32764
Standardised type-token ratio	40.70	39.65

Table 3: showing the statistics for the two corpora.

As can be seen from table 3, the two corpora were not matched according to size. Although matching the two corpora for size was considered it was rejected to avoid anomalies which may have been present in the BAWE sub-corpus from overly affecting the findings. To address this balance data analysis progressed in stages beginning with a quantitative analysis and following up with a qualitative analysis of features indentified as being of interest. The stages are outlined below.

Approach to data analysis

Occurrences of metadiscourse features in different categories were calculated as a percentage of the total count for metadiscourse features of that category. For example, UNNCELE displayed 943 code glosses of which 'say' accounted for 0.42%. BAWE showed a total of 7,523 code glosses of which 'say' accounted for 1.44%. This meant that authors of the BAWE sub-corpus essays used 'say' approximately 3.5 times more than writers of the UNNCELE essays. This difference was considered significant from a quantitative perspective and therefore worthy of further investigation from a qualitative point of view.

The mean (*M*) was calculated for each metadiscoursal feature to show anomalies between the two corpora. *M* was calculated as a percentage total of the total number of discourse markers within a given category for each individual corpus as outlined above. Next, the range was calculated to indicate differences within

and between the two corpora. The Standard Deviation (SD) was also computed to observe the differences between the two sets of data (one from each corpus). Some of the SD figures, however, may have been skewed by extreme scores. This is especially so in categories such as transition markers which include common words such as 'and'. The final step, therefore, was to identify significant difference between the two corpora. This was done by manually checking for differences within categories, as well as between the two corpora, that appeared significant. Items were judged by their percentage of use as outlined above. Select anomalies were then subjected to a qualitative analysis.

Chapter 5: Data Analysis and Discussion

Quantitative data summary

The data show a disparity in the ranges and SD scores. This suggests that BAWE sub-corpus tends to favour a wider range of linguistic devices compared to UNNCELE which shows a more even distribution of score. UNNCELE data are grouped around a norm indicating that students may rely on a restricted range. This may, however, be as much due to the disparity between corpora sizes as it is indicative of CELE students relying on a narrow range of features they had learned.

Summary of quantitative data

	UNNCELE	**BAWE sub-corpus**
Running word total	107,913	198,024
Total number of metadiscourse features	85,493	116,967
Range	3,501	30,746
Standard deviation	188.0742	1389.452
Metadiscourse features as a %age of word total	79.23	59.1
Total number of interactive metadiscourse	80,739	63,176
Interactive metadiscourse as a %age of the total	94.44	54
Total number of interactional metadiscourse	4,483	52,705
Interactional metadiscourse as a %age of the total	5.3	45.1

Table 4: showing a summary of the quantitative data

UNNCELE shows a higher total percentage of metadiscourse features (79.23%) than BAWE sub-corpus (59.1%) which is indicative of an overuse by CELE students. The data also show that CELE students rely, predominantly, on interactive metadiscourse (94.44%) and less so on interactional devices (5.3%). This is compared to the BAWE sub-corpus where the writers use interactive and interactional devices in roughly equal proportions (54% and 45% respectively). Clearly, therefore, CELE students are not interacting with the reader anything like as much as the authors of the BAWE sub-corpus essays. The hypothesis, therefore, has been upheld. The research questions, however, can only be answered following a qualitative analysis of the language features.

Qualitative analysis of select metadiscourse features

Interactive devices: Code glosses

Code glosses are important rhetorical functions that assist the reader in 'grasp[ing] the writer's intended meaning' (Hyland, 2005:32). This is achieved through a variety of means such as rewording, defining or clarifying. It also requires writers to be able to anticipate reader's knowledge of the subject and to anticipate the response to claims made (Hyland, 2005).

Code glosses				
	UNNCELE		**BAWE sub-corpus**	
	Frequency	**%age of category markers**	**Frequency**	**%age of category markers**
Or X	235	25	3510	46.66
Such as	278	29.5	1131	15.03
For example	122	12.93	741	9.85
Indeed	7	0.74	339	4.51
In fact	8	0.85	261	3.47
For instance	94	9.97	239	3.18
Called	27	2.86	190	2.53
e.g.	2	0.2	179	2.38
i.e.	1	0.1	211	2.08
Say	4	0.42	108	1.44
Known as	15	1.6	97	1.29
Specifically	4	0.42	83	1.10
Defined as	57	6	78	1.04
	Range 277		Range 3509	
	SD: 73.86208		SD: 728.2643	

Table 5: showing the top 13 code glosses from BAWE, ranked according to their percentage for the category. Range and SD figures pertain to the full data set (see appendix 1).

Analysis of 'say'

Say makes up 1.44 percent of the code glosses identified in the BAWE sub-corpus. The strongest collocate of *say* is *to say that* which appears significantly more often than any other trigram in this metadiscourse category.

N	Cluster	Freq.
1	TO SAY THAT	46
2	SAY THAT THE	20
3	THAT IS TO	16
4	IS TO SAY	16
5	IS NOT TO	13

6	NOT TO SAY	13
7	WE CAN SAY	11
8	TO SAY THE	10
9	THIS IS NOT	9

Table 6: showing the top 9 trigram code glosses for 'say', ranked according to frequency, evident in BAWE.

The cluster data show that *to say that* occurs more than twice as often as the next most frequently used phrase and that the proportion to which each phrase occurs is fairly uniform throughout the range. The data are, however, simply conforming to Zipf's law which states that a word's frequency rating is inversely proportional to its rank. This means that the top ranking datum will appear roughly twice as often as the second, which will appear twice as often as the third and so on. Caution, therefore, needs to be exercised before reading too much meaning into the frequency scores between the phrases. What is of interest is the fact that whilst *to say that* is fairly low in the percentage rankings of code glosses (as shown in table 5) it appears top in the ranking of trigrams (as shown in table 6) which shows it to be significant. 'To say that' also takes on the role of being a 'single entry' (Schmitt, 2000: 97), or single lexical unit, but, more importantly, collocates with other metadiscourse items to alter its meaning. This makes it worthy of further investigation.

The examples below show the trigram tends not to be used as a code gloss in its own right, that is to say as interactive metadiscourse, but is commonly used with hedging devices as evident in examples 1, 2, 3, and 5, and attitude markers, as evident in examples 4 and 6. This shows they are functioning as *interactional* devices by engaging the reader with the propositional content (as opposed to simply describing it) either, as is the case with the examples below, by presenting a different take on a claim or by downplaying the strength of a reasoned argument that is familiar to the reader.

1. A more reasonable argument would be <u>to say that</u> hoplite warfare did contribute to the development of poleis but as part …

2. Therefore it is reasonable to say that it is the reaction of the government that is the deciding factor …

3. I believe it is not a great exaggeration to say that the L-shaped model of the SRAS/LRAS curve is too rigid for the present …

4. It might be an exaggeration to say that the Commission is trying to understate the importance of such agreement …

5. … it may be enough to say that a parent who is involved with their child also has a secure attachment …

6. … to scientifically test Freud's theory does not make it incorrect, because to say that his work falls short of scientific standards is to …

In contrast, there is only one occurrence of the trigram *to say that* in UNNCELE. This is exampled below.

- It is hard to say that whether the environment can endure it.

It is unclear from this single concordance line whether or not the trigram is used as an interactive or interactional device. Yet when viewed in context of the paragraph (shown below) the phrase, whilst engaging with the reader by presenting what appears to a valued judgement, is not followed through by a sufficient explanation. It is, instead, followed up by a fairly descriptive take on the literature.

> *Due to rapid development, people are always exploiting the energy while producing litter. It is hard to say that whether the environment can endure it. According to Seitz (2008), the amount of solid rubbish in America has approached approximately 1600 pounds per year, which causes significant increase of landfills. Consequently, recycling is quite necessary.*

Extract 1: showing an extended version of the UNNCELE concordance line

This is probably better seen in contrast to an extract from the BAWE sub-corpus. The one below is a contextualised extension to the first concordance line.

> *A more reasonable argument would be to say that hoplite warfare did contribute to the development of poleis but as part of a combination - the 'polis developed due to military and sociopolitical change'. This lasting influence that hoplite warfare had on poleis and the Greek empire in general was that it redefined the social boundaries within a polis. No longer was a polis separated between the elite and the citizens, hoplites gave rise to the emergence of the middle class, ending the phase of elite domination and ushering in an age of more egalitarian constitutions in which the free played a decisive role. Kurt Raaflaub argues …*

Extract 2: showing an extended version of one of the BAWE concordance lines

Here the writer makes a valued judgement and extends it with an explanation. This metadiscourse item, then, seems to provide a commentary on the text (Hyland, 2005) as it allows the writer to engage in a socially appropriate way with the reader. It conveys writer attitude, acknowledges writer position in relation to the claim made, and engages the reader with the wider community. It is, however, the explanation that is supported with evidence from the literature rather than the claim itself. This shows that the trigram *to say that* is an important persuasive rhetorical device in that it is used to respond to an anticipated reader response. This, in turn, makes it an important feature for students to be aware of (Charles, 2007; Thompson, 2001). It also suggests that certain metadiscourse devices may act as an interface between the propositional content, the writer and the reader and have the effect of weakening the semantic gravity relevant to the propositional content.

In comparison, the one example from UNNCELE of *to say that* anchors the propositional content in the interpretation category which could well be the result of the student utilising content from the input material without adequate understanding how to interact with the reader. Put another way, it would appear

that certain metadiscourse devices, whether they be interactive or interactional, play an essential role in shifting what Maton (2009) refers to as semantic gravity away from reproductive description towards abstraction. What appears to be happening, and what may be indicative of a dangerous trend, is that students have segmented the knowledge they have gleaned of metadiscourse and how it functions. I use the term 'dangerous trend' in reflection of Maton's (forthcoming) claim that knowledge often fails to become part of a broader aspect of understanding and remains, instead, a series of discrete items.

Data from the BAWE sub-corpus show how students anticipate and respond to reader reactions to the text (Thompson, 2001) so as to persuade for a position (Wu, 2007). This engagement with the audience in the BAWE sub-corpus essays, and, conversely, apparent lack of audience engagement in CELE essays can be further observed through an analysis of the frame marker 'focus'.

Frame markers: announce goals

Frame markers are interactional rhetorical devices used by a writer to map out an argument within a text. This may include phrases such as *to sum* or *my purpose is to* which help to summarise or consolidate an argument for the reader. Frame markers also include stage labelling such as *next I would like to come onto* or *the next point I would like to make* which stage the argument, again for the reader, as well as topic shifters such as *now* or *let us move onto*. They appear, therefore, to be independent of propositional content in that they act to signal to the reader the direction the writer's thoughts are going. This requires the writer to consider the act of written communication from the dual perspective of the both the reader and themselves.

Frame markers: announcing goals				
	UNNCELE		**BAWE sub-corpus**	
Frequency	**%age of category markers**	**Frequency**	**Frequency**	**%age of category markers**
Aim	0	0	50	26.32
Focus	12	66.7	38	20
(in) this section	2	11.1	33	17.37
Purpose	4	22.2	31	16.32
Intend to	0	0	8	4.21
	Range 10		Range 48	
	SD: 3.189268		SD: 16.5083	

Table 7: showing the top 5 frame markers: announcing goals from BAWE, ranked according to their percentage for the category. Range and SD figures pertain to the full data set (see appendix 3).

What makes this feature of interest is how it is used in the respective corpora as shown by the quantitative data. In total, *focus* appears 29 times in UNNCELE, of which 17 occurrences are propositional and 12 are interactive. Contrast this with the BAWE sub-corpus where it occurs 389 times: 38 times as an interactive device and 351 as *interactional* rhetoric. Writers of the BAWE sub-corpus essays, therefore, appear to favour the interactional use of the lexis whereas UNNCELE indicates it is being used fairly mechanically as a guide for the reader by CELE students.

Analysis of 'focus'

Focus makes up 20 percent of the frame markers sub-group - announcing goals - identified in the BAWE sub-corpus. The strongest collocate of *focus* is *focus on the* which, Zipf's law notwithstanding, appears significantly more often than other trigrams in the sub-group.

N	Cluster	Freq.	Length
1	FOCUS ON THE	44	3
2	THE FOCUS OF	26	3
3	TO FOCUS ON	24	3
4	WILL FOCUS ON	19	3
5	A FOCUS ON	17	3
6	FOCUS ON FORM	14	3
7	FOCUS ON MEANING	13	3
8	THE FOCUS ON	10	3

Table 8: showing the top 8 trigram for 'focus', ranked according to frequency, evident in BAWE.

Focus on the appears top in the ranking of trigrams and, once again, appears to adopt the status of a single entry lexical item. However, unlike *to say that*, which collocates with other metadiscourse items to alter meaning, the trigram *focus on the* has an anaphoric referencing function.

Judging by the examples below, *focus on the* functions to map out the structure of the essay for the reader, as evident in examples 1, 3 and 6, to present an interpretation on events, such is the case with example 4, or to present a judgement as in example 5. Whilst the concordance data show the trigram *focus on the* is most commonly used to guide the reader through the structure of the discourse, as is Hyland's (2005) observation, it does so in conjunction with a reference to either the writer (I will) or the text itself (the last section). Otherwise, it is used as an exophoric referencing device to refer to something outside of the text, either the wider community, as in example 5 and 7 below, or as example 2 shows, as an evaluative tool.

Data from the BAWE sub-corpus

1. In the last section I would focus on the strategies employed by unions for …
2. However, the theories that will be applied do not focus on the classical oratory techniques discussed by Aristotle but around …
3. I will particularly focus on the introduction and development of English; its role in a post colonial …
4. Elias as a German Jew writing prior to the Second World War places a strong focus on the place of aggression in society …
5. Were we to focus on the details of modern life we would become …
6. In the last section I would focus on the strategies employed by unions …
7. Historians have tended to focus on the emergence of the 'middle class' as something that occurred …

If the definite article is omitted, as in *to focus on* (see below), the phrase modifies the rhetorical function. Examples 8 and 9 show that it still functions in the same way as *focus on the*, that is to say to map out the structure of the discourse, yet, as examples 10 to 13 indicate, it is predominantly used to present a judgement.

Data from the BAWE sub-corpus

8. The pronouns I intend to focus on can be grouped as shown below: I aim to find out …

9. This essay intends to focus on 'child-directed speech,' a form of language input and will examine …

10... greater use of the cassette in the course would have given a target to focus on and Spolsky's fourth condition of …

11.… view my Persian learning in a more informed light. On the surface it appears to focus on meaning because it is organised by topic and aims to teach …

12.… as a beginner learning Persian, the hardest thing was that lessons appeared to focus on two methods: focus on meaning and focus on forms.

13. Early sociological responses to studying chronic illness tended to focus on Parsonian sick role theory and labelling perspectives which …

This shows the extent to which the authors of the BAWE sub-corpus essays are using metadiscourse skilfully to create a triage of interaction between the reader, the writer and the content: a key factor that is missing from the CELE essays.

In contrast, cluster data from UNNCELE show only four multi-word phrases, three of which are trigrams of *focus*. Whilst *focus on the* appears in the cluster groupings, *to focus on* does not. In other words, CELE students seem to use the definite article to limit the function of the phrase to mapping out the discourse

which is also the case for three of the four trigrams, including the two with the highest frequency ratings.

N	Cluster	Freq.	Length
1	WILL FOCUS ON	9	3
2	THIS ESSAY WILL	8	3
3	FOCUS ON THE	7	3
4	ESSAY WILL FOCUS	6	3

Table 9: showing the only 4 trigrams for 'focus', ranked according to frequency, evident in UNNCELE.

A manual check of the concordance lines, however, shows four instances of *to focus on* being used, two of which are given as examples below.

Data from UNNCELE

> 14. Based on the threat of global warming, people tend to focus on the beneficial characteristics of nuclear power, namely …
>
> 15. Therefore, it is important to focus on further questions and try to get the real and honest answers.

A review of the concordance lines indicates the trigram may be used to engage with the content which shifts the focus of the discourse away from what Maton (forthcoming) refers to as the reproductive description level. But when viewed in context it becomes apparent that it is being used at the summarising descriptive level (example 14 above) or to make a judgement about a research process; that is the authors' own research rather than propositional content. This can be better seen when viewed in context (see extract 3 below). To this end it is functioning to interact with the reader to justify decisions made.

> *There are two methods. One is the semi-structured interview and another one is holding a focus group. All interviewees are first year and the second year students in UNNC. They will be asked questions related to opinions of essay cheating and attitude of looking others' cheating. However, there may not have obvious and exterior difference between interviewees. Therefore, it is*

important to focus on further questions and try to get the real and honest answers.

Extract 3: showing an extended version of one of the CELE concordance lines for the trigram 'to focus on'.

Whilst this appears important in the context in which it is used there is no evidence to suggest CELE students interact on a three point focus: writer, reader, content. There is sufficient evidence, therefore, to substantiate the claim that CELE students commonly fail to engage with their readers beyond the structural level, which does as much to expose a weakness in teaching as it does student's lack of linguistic and socio-cultural knowledge. The data presented here, however, are limited by the constraints of the dissertation that does not allow for a broader analysis of metadiscourse features. This being said, the analyses show a clear disparity in the way in which the authors of the BAWE sub-corpus essays and the authors of the CELE essays interact with the audience and the content of their writings.

Research questions revisited

It was hypothesised in chapter two that L2 writers may tend to overuse interactive metadiscourse and underuse interactional features when compared to native speaker writing (Aijmer, 2002; Hyland, 2005). This assumption has been upheld by the data. The discussion below, therefore, addresses the three research questions:

1. *How is metadiscourse used by successful student writers?*
2. *How do writers orient themselves to their audience?*
3. *What are the key differences in the use of metadiscourse between the CELE and BAWE corpora?*
4. Is there a link between metadiscourse and propositional content?

The data show that the metadiscourse features evident in the BAWE sub-corpus are not limited to writer-reader interaction but represent a three way process

during which the authors interact with the reader *and* the propositional content of the essay. This can be represented diagrammatically:

The three way interaction as evident from the BAWE sub-corpus data

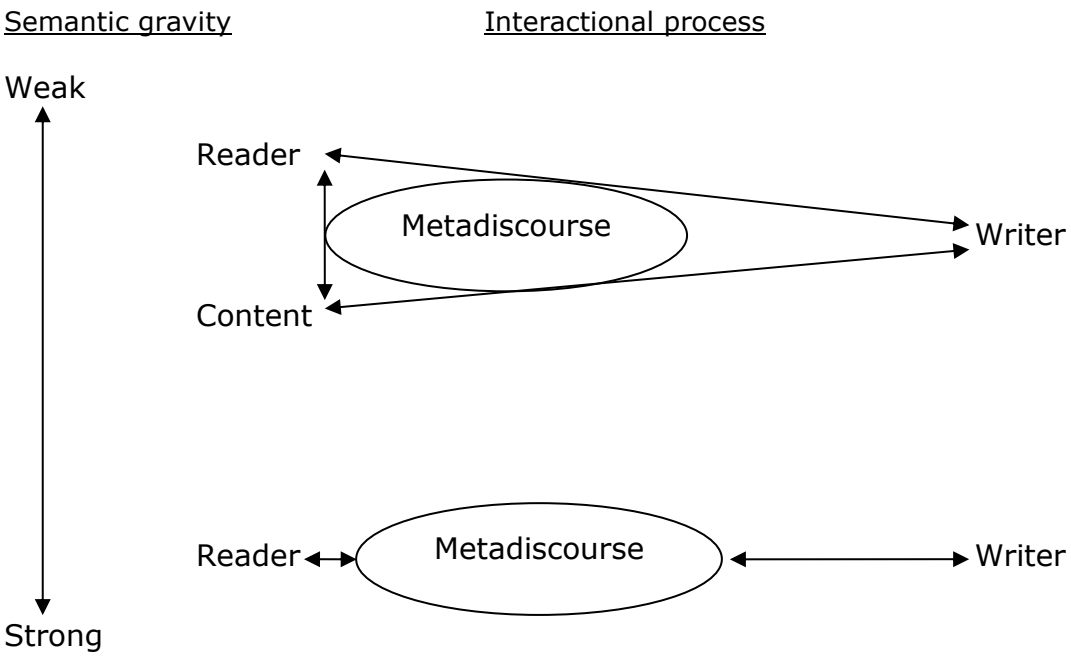


Fig.1: showing the triage of interaction between content, writer and reader that is achieved through metadiscourse in the BAWE sub-corpus essays. This has the effect of weakening the semantic gravity.

Writer's of the BAWE sub-corpus essays, who have all achieved upper-second or first class grades, use metadiscourse flexibly by modifying the features according to the desired effect. This is to say their learning appears to have been cumulative. A case in point is *to say that*, an interactive code gloss that appears to be acting as an interactive device by collocating with hedges, attitude markers etc. Similarly the subtle changes in the use of *focus* from *focus on the* to *to focus on* allows the author to make a judgement. This flexibility in the use of metadiscourse allows the writers to engage with both the audience and the propositional content and marshal well reasoned academic arguments.

In seemingly stark contrast, authors of the CELE essays appear to be using metadiscourse features in a fairly fixed, inflexible way. That is to say, their *knowledge of* metadiscourse (as opposed to metadiscourse itself) has a strong semantic gravity when compared to that of the BAWE sub-corpus authors' which appears much weaker and therefore easily transferable to contexts outside of which they have been learned. What they have learned of metadiscourse features, therefore, appears to remain segmented. This culminates in a correlation between the students' usage of metadiscourse and their ability to engage with the reader and the content beyond the reproductive descriptive level.

The reasons however remain unclear, especially as a correlation is no indication of cause and effect. The extent to which CELE students' fail to marshal arguments in the way that the writer's of the BAWE sub-corpus essays do may result from inadequate material or teaching practices. Whilst Maton (forthcoming) would allude to this claim in relation to mainstream education, it has to be noted that the CELE essays were written by L2 speakers. Such a claim, therefore, stands on rocky ground and could easily lead to biased and ill thought through conclusions. It may simply be that students lack the linguistic knowledge to interpret content of their academic readings beyond the reproductive level or be unaware of the criteria that define an argumentative essay. After all, in order to achieve the level of interaction that is evident in the BAWE sub-corpus essays writers needs to be immersed in the subject matter to the degree that they can make claims, judgements or generalisations whilst appreciating that the reader may have alternative views or questions. This, of course, was the case with the BAWE sub-corpus essays and arguably less so with CELE students who may be simply jumping through the hoops asked of them by their EAP course and which may well have skewed the findings. Future studies that compare high and low scoring essays from a narrowly defined student cohort may help to clarify this nagging concern.

Notwithstanding, there is still room at this stage of the investigation to argue for a need to align the teaching of metadiscourse features with Maton's (2009) concept of semantic gravity, not only for academic curiosity but also to better

inform materials developers, course designers as well as good teaching practice. The data used in this dissertation are, nevertheless, drawn from a general (i.e. non discipline specific) corpus, and therefore call for caution in that a discipline specific analysis could well shed a very different shade of light on the claims being made. The findings do, however, support those of Kuteeva (2011) who discovered that L2 writing styles has far less an impact on students use of metadiscourse features than their educational background does.

Conclusion

Hyland was quoted in chapter two as saying that students' performance in academic writing is largely due to 'a lack of familiarity with the metadiscourse conventions central to many expository genres' (Hyland, 2005:136). The findings of this dissertation take his concept to another level by arguing that students' problems lie, not necessarily with a lack of exposure to metadiscourse (although this is undoubtedly a significant factor), but that their learning of metadiscourse remains segmented and therefore not easily transferable to contexts outside of that in which it was learned. Moreover, the apparent absence of such a notion in the CELE essays is tantamount to identifying a weakness in the way in which students are taught to understand, and use, academic language to engage with the reader in a socially appropriate manner. Embedding the teaching of metadiscourse in cumulative learning practices could consequently empower students to develop both linguistically and intellectually.

As far as academic writing is concerned metadiscourse is a collective term given to rhetorical devices that serve one of two purposes: to map out the development of an argument so that the reader can follow the progression of the writer's thoughts and to express understanding of propositional content so as to persuade a reader to accept the writer's stance. This dual purpose is noted by a number of researchers, amongst whom Vande Kopple (1985) appears influential. He views metadiscourse as being either interpersonal (features such as hedges) or textual (features such as text markers), whilst discourse primarily fills an ideational function. Vande Kopple (1985), therefore, appears to disassociate metadiscourse from propositional content. Crismore *et al*. (1993) also uses Vande Kopple's distinction, as does Abdi (2002), who carries on this tradition by

focusing on the cultural influences that affect a writer's use of interpersonal metadiscourse. Hyland (2005) adopts a similar view that metadiscourse is either interactive (Vande Kopple's textual) or interactional (Vande Kopple's interpersonal).

This bifurcation of metadiscourse represents a break from Halliday's (1978, 1985. Cited in: Abdi, 2002) concept of linguistic behaviour that is dependent upon three factors: ideational, interpersonal and textual and reduces it to reader-writer interaction. I have tried to argue, however, that when it comes to the teaching of metadiscourse this provides an incomplete picture because it places metadiscourse in a framework of what Maton (2009) refers to as segmented learning. The result can be that students fail to recognise that interaction is actually a three way process between the writer the reader and the propositional content of the essay and remain unaware of the crucial role that metadiscourse plays in achieving this aim. I therefore conclude by stating that metadiscourse needs to be taught in a manner that facilitates cumulative learning so that students can understand the role metadiscourse plays in the interaction between the writer, reader and propositional content when marshalling arguments.

Scope for further research

Unfortunately the limitations of this dissertation do not allow a broader or deeper investigation into metadiscourse and its association with Maton's (2009) semantic gravity that is needed to add credence to these claims. There is, however, enough raw data available in the appendices should other researchers wish to take up the challenge. Indeed, as the raw data show, many other anomalies could be investigated. It was also mentioned in chapter one, during the discussion on the language of textbooks and research articles, that textbooks tend to use more directives than research articles (Swales, et al., 1998). This is further grist for the researcher's mill as a contrast could be established between the CELE essays and features of textbook language to test this claim and to evaluate the effects it has on student essay writing.

Appendices

Overview of the appendices

These appendices contain raw data from the quantitative research stage. Parenthesised comments are the original unedited notes made during the research process and have been included here as they may be of interest to the reader.

Some tables are accompanied by, again original and unedited, comments and or examples of concordance lines. Where this is the case it indicates that these features were deemed of interest but were not included in the main data analysis chapter because of word length constraints. Being part and parcel of the appendices may allow them to inspire further research.

Appendix 1: code glosses

Code glosses (If these date were plotted on a chart both corpora would show a correlation) SDs very close correlation)					
	UNNCELE			**BAWE sub-corpus**	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
As a matter of fact	0	0		3	0.04
Called	27	2.86		190 (+36 propositional)	2.53
Defined as	57	6		78	1.04
e.g.	2	0.2		179	2.38
For example	122	12.93		741	9.85
For instance	94	9.97		239	3.18
I mean	0	0		5	0.07
i.e.	1	0.1		211	2.08
In fact	8	0.85		261	3.47
In other words	7	0.74		67	0.89
Indeed	7	0.74		339	4.51
Known as	15	1.6		97	1.29
namely	29	3.1		61	0.81
Or X	235 (mostly propositional)	25		3510	46.66
Put another way	0	0		4	0.05
Say	4	0.42		108 (+100 propositional and 16 that is to say)	1.44 Used to clarify meaning (see examples below)
Specifically	4	0.42		83	1.10
Such as	278	29.5		1131	15.03
That is	26	2.76		74 (+277 propositional)	0.98
That is to say	2	0.2		16	0.21
That means	3	0.3		5	0.07
This means	3	0.3		70	0.93
Viz	0	0		1	0.01
Which means	19	2		50	0.66
	Total 943	**Total**		**Total 7523**	**Total**
	Range 277			**Range 3509**	
	SD: 73.86208			**SD:** 728.2643	

Appendix 2: engagement markers

Engagement markers (some significance, adds to argument that BAWE engage/consider the audience)					
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
(the) reader's	0	0		0	0
Add	2 (+1 propositional)	0.22		48 (+2 propositional)	0.4
Allow	3 (+2 propositional)	0.34		172 (+2 propositional)	1.44
Analyse	8	0.89		93 (+2 propositional)	0.78
Apply	3	0.34		87 (+2 propositional)	0.73
Arrange	0	0		7	0.06
Assess	0	0		51	0.43
Assume	2	0.22		83	0.7
By the way	0	0		6	0.05
Calculate	0	0		42	0.35
Choose	10	1.12		120	1.01
Classify	0	0		10	0.08
Compare	2	0.22		77	0.65
Connect	0	0		13	0.11
Consider	7	0.78		196	1.61
Consult	0	0		7	0.06
Contrast	14	1.57		33	0.28
Define	0	0		94	0.79
Demonstrate	7	0.78		57	0.48
Determine	8	0.89		123	1.03
Do not	18	2.01		351	2.94
Develop	18	2.01		151	1.27
Employ	1	0.11		38	0.32
Ensure	5	0.56		161	1.35
Estimate	4	0.45		31	0.26
Evaluate	43	4.81		61	0.51
Find	17	1.9		318 (+3 propositional)	2.67
Follow	2	0.22		124	1.04
Go	11	1.23		169 (+2 propositional)	1.42
Have to	18	2.01		189	1.58
Imagine	0	0		17	0.14
Incidentally	0	0		3	0.03
Increase	83	9.28		517	4.33
Input	1	0.11		10 (+129 propositional)	0.08
Insert	0	0		1	0.008
Integrate	0	0		13	0.11
Key	23	11.19		246 (+75 propositional)	2.06
Let us	1	0.11		7	0.06
Let x = y	-	-		-	-
Let's	0	0		0	0
Look at	1	0.11		113 (+119 propositional)	0.95

Mark	0	0		14 (+82 propositional)	0.12
Measure	12 (+1 propositional)	1.34		19 (+119 propositional)	0.16
Mount	0	0		(5 propositional)	0
Must	20	2.24		789	6.61
Need to	16	1.79		306	2.57
Note	9	1.01		100 (+33 propositional)	0.84
Notice	1 (+1 propositional)	0.11		12 (+24 propositional)	0.1
Observe	(1 propositional)	0		47	0.39
One's	0	0		0	0
Order	72 (71 in order to)	8.05		339 (+603 propositional)	2.84
Ought	3	0.34		36	0.30
Our (*inclusive*)	18	2.01		122 (+401 propositional or not inclusive of the reader)	1.02
Pay	13 (+9 propositional)	1.46		18 (+147 propositional)	0.15
Picture	1	0.11		38 (+74 propositional)	0.32
Prepare	3	0.34		3 (+13 propositional)	0.03
Recall	0	0		2 (+17 propositional)	0.02
Recover	(2 propositional)	0		1 (+21 propositional)	0.008
Refer	2	0.22		120	1.01
Regard	5	0.56		126	1.06
Remember	0	0		23 (+23 propositional)	0.19
Remove	(1 propositional)	0		11 (+25 propositional and 1 directive)	0.09
Review	0	0		26 (+116 propositional and 1 directive)	0.22
See	3	0.34		550	4.61
Select	0	0		42	0.35
Set	5 (+19 propositional)	0.56		35 (+430 propositional)	0.29
Should	115	12.86		1,105	9.26
Show	13	1.46		310	2.6
State	21 (+17 propositional)	49.32		668 (+259 propositional)	5.6
Suppose	0	0		7	0.06
Take (a look/as example)	4 (+65 propositional)	0.45		2	0.02
Think about	1	0.11		13	0.11
Think of	0	0		16	0.13
Turn	(12 propositional)	0		143 (+198 propositional)	1.2
Us (*inclusive*)	9	1.01		329 (+127 propositional)	2.76

Use	216	24.2		1,163	9.75
We (*inclusive*)	19	2.13		1,162	9.74
You	1	0.11		392	3.29
Yours	0	0		1	0.008
	Total 894	**Total**		**Total 11929**	**Total**
	Range 215			**Range 1662**	
	SD: 29.8471			**SD:** 254.0707	

The data show that the engagement marker 'key' is more evident in CELE essays than in BAWE.

Key	23	11.19		246 (+75 propositional)	2.06

It appears that the engagement marker 'key' is used to introduce a concept and map out the discourse for the reader and serve, in effect, to function as interactive devices. This is event in both the CELE and the BAWE corpora.

Appendix 3: frame markers: announce goals

Frame markers: announce goals ('Focus' is interesting')					
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
(in) this chapter	0	0		(they pertain to dissertations)	0
(in) this part	(1 propositional)	0		6	3.16
(in) this section	2	11.1		33 (+17 propositional)	17.37
Aim	(3 propositional)	0		50 (+107 propositional)	26.32
Desire to	(2 interactional)	0		(62 interactional)	0
Focus	12 (+17 propositional)	66.7		38 (+351 interactional)	20
Goal	(5 propositional)	0		(89 propositional)	0
Intend to	0	0		8 (+11 interactional)	4.21
Intention	(1 propositional)	0		2 (+38 interactional)	1.05
Objective	(1 propositional)	0		6 (+137 interactional)	3.16
Purpose	4 (+16 propositional)	22.2		31 (+142 interactional)	16.32
Seek to	0	0		4 (+39 interactional)	2.11
Want to	(5 interactional)	0		2 (+83 interactional)	1.05
Wish to	0	0		2 (+28 interactional)	1.05
Would like to	0	0		8 (+8 interactional)	4.21
	Total 18	**Total**	**90859**	**Total 190**	**Total**
	Range 10			**Range 48**	
	SD: 3.189268			**SD:** 16.5083	

Frame markers: announce goals are greatly underrepresented in CELE essays

Appendix 4: endophoric markers

Endophoric markers (If these date were plotted on a chart both corpora would show a correlation)					
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
(in) chapter x	0	0		72	2.93
(in) part x	6 (+171 propositional)	1.9		19 (+657 propositional)	0.77
(in) section x	3	0.94		188 (+269 propositional) but only corpus includes dissertations	7.66
(in) the x chapter	0	0		12 (2 of which pertain to dissertations)	0.49
(in) the x part	0	0		16 (+76 propositional)	0.65
(in) the x section	0	0		(they pertain to dissertations)	0
(in) this chapter	0	0		(they pertain to dissertations)	0
(in) this part	(1 propositional)	0		15 (+76 propositional)	0.61
(in) this section	2	0.63		(they pertain to dissertations)	0
Example x	184	57.9		1231	50.16
Fig. x	10	3.14		26	1.06
P. x (page x)	55	17.3		(they pertain to dissertations)	0
Page x	0	0		(they pertain to dissertations)	0
Table x (ignore-too dependent on task)	0	0		174 (+3 propositional)	7.09
X above	45 (+6 propositional)	25 = mentioned above		471	19.19
X before	9 (+30 propositional)	2.83		35 (+104 propositional)	1.43
X below	1 (+2 propositional)	0.3		140 (+60 propositional)	5.7
X earlier	3 (+2 propositional)	0.94		63 (+73 propositional)	2.57
X later	(+8 propositional)	0		18 (+214	0.73

				propositional)	
	Total 318	**Total**		**Total 2454**	**Total**
	Range 183			**Range 1219**	
	SD: 43.34851			**SD:** 289.7394	

The SD for BAWE indicates there are certain norms for endophoric markers that CELE students may be unaware of and which should be taught on an EAP course. It also indicates they extent to which authors of the BAWE essays engage the reader through the use endophoric markers, especially when directing the reader in to aspects of the text outside of the immediate focus of attention.

Examples of 'earlier' from BAWE (contrast this to how they are used in CELE essays) – allow the reader to follow a thread through the essay

1. … although is typically realised as the Indirect Object as will be later discussed. In the sentences (2), (3) and (4) above, the Subject lacks a …
2. This can be linked to the earlier discussion of the way in the medicalisation of society has resulted in a …
3. As stated earlier, the employees will be taking on more responsibility and a larger range …
4. … but its role in mitigating some of the risks identified earlier in the report make it a worthwhile venture. Since the reorganisation is …
5. As stated earlier, it was actually found to be about 3.2 V. Once new, larger diameter …
6. … is not necessary for its continuation- as I mentioned earlier the nation-state is not as sovereign as it used to be …
7. Thus, as stressed earlier, what determine comparative advantage in the H-O model are merely …
8. As hinted earlier it was the extent not the nature of Bevan's housing policy which is most …
9. … the somewhat silent and static atmosphere of which I spoke earlier. About the type of people, I can say the following.

Example of 'later' from BAWE

1. … I shall return to this point later.
2. This will be discussed further later in the essay.
3. … we can put this down to gender differences as we shall see later.
4. During the 1850's, as we will see later on, the South saw themselves in an increasingly minority position within …

Endophoric markers not listed by Hyland (2005)

1. Indeed, the example just given shows that to being to switch codes in this way is a valuable …
2. According to the previous discussion, it has been proved that …
3. … is one of the limitations according to its definition above.

Part *

1. … will be evaluated in the following part of the essay. In most homes, lighting accounts for 10 to 15 percent of the…
2. … the process of deforestation in the first part of essay …
3. … and climate change in the second part of it.
4. … let us now turn to the second part of the effects…

Appendix 5: evidentials

Evidentials (If these date were plotted on a chart both corpora would show a correlation)					
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
(date) (name)	-	-		-	-
(to) cite x	0	0		0	0
(to) quote x	0	0		4	0.62
[ref. no.]/[name]	-	-		-	-
According to x	122 (+27 propositional)	75.78		265 (+117 propositional)	40.77
cited	34 (10 x cited, 24 x cited in)	21.1		320 (47 x cited, 273 cited in)	49.23
quoted	5	3.1		61	9.38
	Total 161	**Total**		**Total 650**	**Total**
		72.68			
	Range 117			**Range 316**	
	SD: 52.15554			**SD:** 151.5437	

Appendix 6: frame markers

	Frame markers: sequencing (There were no significant differences – numbers too low)				
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
(in) chapter x	-	-	(they pertain to dissertations)	-	-
(in) part x	-	-		-	-
(in) section x	-	-	(they pertain to dissertations)	-	-
(in) the x chapter	-	-	(they pertain to dissertations)	-	-
(in) the x part	3	0.79		21 (+55 propositional)	1.61
(in) the x section	-	-	(they pertain to dissertations)	-	-
(in) this chapter	-	-	(they pertain to dissertations)	-	-
(in) this part	(1 propositional)	0		6	
(in) this section	2	0.53		34	2.61
Finally	80 (+23 propositional)	21.11		179 (+38 propositional)	13.73
First	43 (+64 propositional)	11.35		24 (+772 propositional)	1.84
First of all	11	2.9		20	1.53
Firstly	77	20.32		222	17.02
Last	10 (+28 propositional)	2.64		26 (+176 propositional)	1.99
Lastly	11	2.9		43	3.3
Listing (a, b, c etc)		0		0	0
Next	4 (+13 propositional)	1.06		126 (+54 propositional)	9.66
Numbering (1, 2, 3 etc)		0		0	0
Second	15 (+23 propositional)	3.96		131 (+202 propositional)	10.05
Secondly	33	8.71		137	10.51
Subsequently	11	2.9		49	3.76
Then (not included in calculations)	73 (+16 propositional)	19.26		191	14.65
Third (not included in calculations)	2 (+36 propositional)	0.53		0	0
Thirdly	2	0.53		28	2.15
To begin	2	0.53		31 (+2 propositional)	2.38
To start with	0	0		4	0.31
	Total 379	**Total**	**90859**	**Total 1304**	**Total**
	Range 78			**Range 218**	

	SD: 27.28307		**SD:** 71.82171

Frame markers: label stages ('now' is the only significant result)					
UNNCELE				**BAWE sub-corpus**	
CELE essays	**Frequency**	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
All in all	3	2.31		6	0.97
At this point	0	0		18	2.89
At this stage	0	0		9	1.45
By far	2	1.54		9	1.45
For the moment	0	0		0	0
In brief	7	5.4		2	0.32
In conclusion	52	40		130	20.93
In short	18 (+8 propositional)	13.8		24 (+8 propositional)	16.1
In sum	0	0		7	1.13
In summary	3	2.31		11	1.77
Now	2 (+51 propositional)	1.54		57 (+416 propositional)	9.18
On the whole	0	0		9	1.45
Overall	15 (+2 propositional)	11.54		189	30.43
So far	3	2.31		55	8.86
Thus far	0	0		15	2.42
To conclude	7	5.4		48	7.73
To repeat	0	0		6	0.97
To sum up	15	11.54		16	2.58
To summarize (ise)	3	2.31		10	1.61
	Total 130	**Total**	**90859**	**Total 621**	**Total**
	Range 50			**Range 187**	
	SD: 12.31649			**SD:** 48.80352	

Examples of 'now' – they are used in mid text in BAWE to walk the reader through the process. CELE use them predominantly in the introduction only.

1. However, I now turn to the reasons why the Gold Standard had worked well before 1914.

2. I will now explain how grammar would be impossible to learn through imitation alone.

3. … they do not suggest the fundamental idea of stages to be wrong, so we will now take a look at Vygotsky, another stage theorist.

4. A guideline of policy recommendations. Up to now I have addressed the vulnerabilities and factors that triggered recent …

5. We now turn to consider the present day scenario and the constraints faced by union

Or, to show development of reasoning

1. I now realise that the models applied during the seminar presentations are best …

2. I am now satisfied with the variables I have considered for the model: ABILITY, ALEVE

3. If reductionism is not used within psychology, the question will now have to be addressed of where the boundary of the self and the world should

Frame markers: shift topic This is interesting. I would have expected more topic shifters in the BAWE text					
	UNNCELE			**BAWE sub-corpus**	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
Back to	0	0		11 (+72 propositional)	1.91
Digress	0	0		0	0
In regard to	0	0		3	0.52
Move on	2	7.41		5	0.87
Now	2 (+51 propositional)	7.41		52 (+364 propositional)	9.01
Resume	0	0		2	0.35
Return to	(1 propositional)	0		8 (+51 propositional)	1.39
Revisit	0	0		0	0
Shift to	0	0		1 (+4 propositional)	0.17
So	4 (+69 propositional)	14.8		18 (+1398 propositional)	0.17
To look more closely	0	0		1	0.17
Turn to	1 (+1 propositional)	3.7		9 (+9 propositional)	1.56
Well	17 (+97 propositional)	63		392 (+461 propositional)	67.93 'as well as'
With regard to	1	3.7		75	12.99
	Total 27	**Total**	**90859**	**Total 577**	**Total**
	Range 16			**Range 391**	
	SD: 4.497252			**SD:** 103.345	

Transition markers (need to look at the function of the ones that are significant)					
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
Accordingly	9	0.01		51	0.1
Additionally	42	0.053		73	0.14
Again	1 (+4 propositional)	0.001		290 (+3 propositional)	0.57
Also	406	0.51		2403	4.72
Alternatively	0	0		29	0.06
Although	79	0.1		677	1.33
And	3502	4.43		30,747	60.36
As a consequence	28 (CELE Ss are taught the language of cause and effect)	0.04		35	0.07
As a result	77 (CELE Ss are taught the language of cause and effect)	0.1		257	0.50
At the same time	21	0.03		85	0.17
Because	181	0.23		1,269	2.49
Besides	15	0.02		57	0.11
But	165	0.21		2,389	4.69
By contrast	4	0.005		20	0.04
By the same token	0	0		3	0.01
Consequently	41 (CELE Ss are taught the language of cause and effect)	0.052		166	0.33
Conversely	0	0		28	0.05
Equally	1	0.001		96	0.19
Even though	6	0.008		80	0.16
Further	3 (+31 propositional)	0.004		579	1.14
Furthermore	110	0.14		308	0.60
Hence	21	0.03		349	0.69
However	229	0.29		2,094	4.11
In addition	150	0.19		240	0.47
In contrast	4	0.005		90	0.18
In the same way	0	0		36	0.07
Leads to	48 (CELE Ss are taught the language of cause and effect)	0.061		81	0.16
Likewise	1	0.001		44	0.09
Moreover	111	0.14		271	0.53
Nevertheless	50	0.063		164	0.32
Nonetheless	13	0.016		89	0.17
On the contrary	4	0.005		28	0.05
On the other hand	32	0.04		197	0.39
Rather	16	0.02		643	1.26
Result in	53 (CELE Ss are taught the language of cause and effect)	0.07		76	0.15

Similarly	20	0.03		166	0.33
Since	153	0.19		655	1.29
So	73	0.09		1,373	2.7
So as to	0	0		43	0.08
Still	136	0.17		589	1.16
The result is	1	0.001		13	0.03
Thereby	4	0.005		72	0.14
Therefore	86	0.11		1,294	2.54
Though	58	0.13		311	0.61
Thus	58	0.13		1,040	2.04
Whereas	30	0.03		196	0.38
While	89	0.11		751	1.5
Yet	6	0.008		396	0.78
	Total 79064	**Total**		**Total 50,943**	**Total**
	Range 3501			**Range 30744**	
	SD: 503.2009			**SD:** 4416.998	

Frame markers: transition markers are overly (?) represented in CELE essays

Transition markers were taught in CELE at the time when the essays were collected (the syllabus has since changed). As a result, CELE students appear to be using the markers in similar proportions to the BAWE students. What this quantitative analysis fails to show however, is the way in which they are used; a qualitative analysis revealing that BAWE students use the devices in response to anticipated, or imagined, audience reactions to the text, in contrast to the mechanical application favoured by CELE students.

Example of 'further' from BAWE – Ss anticipate the reader's response and counter potential critiques.

1. This does not mean, however, that the speech they produce is normal. Indeed, it is far from it …

2. … could be used in therapy to help patients with communicative disorders. However, with the development of psycholinguistics and pragmatics the last 20 years …

3. … as practice for the patient and to test the syntactical level of their ability. However, they must also remember to simplify their language …

4. … they show the need for linguistic knowledge of all the levels of language. However, if we consider the numerous other communication disorders that a speech …

5. … to the therapist's questions and gives appropriate information in his answers. However, it can certainly be shown that the problems that aphasic patients …

Example of 'further' from BAWE – further evidence of Ss interacting with the reader to anticipate questions/doubts they may have regarding the evidence give. The writers address this by extending the evidence they give to make their argument more persuasive.

1. It also leads to a further question whether brain has any effects on …
2. There is also a further question arising at the end. It is assumed that …
3. Moreover while looking at language use by apes it leads to a further consideration on language acquisition by human beings.
4. In order to understand language further it is necessary to investigate why humans are capable to acquire and use
5. A further example of using concordancing software to determine patterns is the use …

Appendix 7: attitude markers

Attitude markers (this is surprising. I would have anticipated significant differences here)					
	UNNCELE			BAWE sub-corpus	
	Frequency	%age of category markers	%age of total discourse markers	Frequency	%age of token
Admittedly	3	0.74		5	0.15
Agree	4	0.98		74	2.19
Agrees	7	1.72		28	0.83
Amazed	1	0.25		2	0.06
Amazing	3	0.74		2	0.06
Amazingly	0	0		1	0.03
Appropriate	13	3.2		174	5.15
Appropriately	1	0.25		10	0.3
Astonished	0	0		1	0.03
Astonishing	1	0.25		0	0
Astonishingly	0	0		2	0.06
Correctly	0	0		24	0.71
Curious	0	0		4	0.12
Curiously	0	0		1	0.03
Desirable	3	0.74		39	1.15
Desirably	0	0		0	0
Disappointed	0	0		0	0
Disappointing	0	0		4	0.12
Disappointingly	0	0		0	0
Disagree	2	0.49		19	0.56
Disagreed	0	0		10	0.3
Disagrees	0	0		7	0.21
Dramatic	18	4.42		86	2.54
Dramatically	27	6.63		46	1.36
Essential	33	8.12		159	4.7
Essentially	3	0.74		94	2.78
Even x	112	27.52		867	25.65
Expected	19	4.7		253	7.49
Expectedly	0	0		0	0
Fortunate	0	0		7	0.21
Fortunately	2	0.49		3	0.09
Hopeful	1	0.25		2	0.06
Hopefully	1	0.25		6	0.18
Important	113	27.8		881	26.07
Importantly	5	1.23		61	1.8
Inappropriate	10	2.46		18	0.54
Inappropriately	0	0		5	0.15
Interesting	0	0		153	4.53
Interestingly	0	0		42	1.24
Prefer	5	1.23		27	0.8
Preferable	0	0		5	0.15
Preferably	0	0		1	0.03
Remarkable	6	1.5		14	0.41
Remarkably	0	0		16	0.47
Shocked	1	0.25		2	0.06
Shocking	0	0		2	0.06
Shockingly	0	0		2	0.06
Striking	1	0.25		21	0.62

Strikingly	1	0.25		4	0.12
Surprised	0	O		5	0.15
Surprising	1	0.25		26	0.77
Surprisingly	0	0		20	0.59
Unbelievable	0	0		1	0.03
Unbelievably	0	0		0	0
Understandable	0	0		17	0.5
Understandably	0	0		2	0.06
Unexpected	0	0		25	0.74
Unexpectedly	0	0		7	0.21
Unfortunate	0	0		7	0.21
Unfortunately	4	0.98		44	1.3
Unusual	3	0.74		21	0.62
Unusually	0	0		8	0.24
Usual	3	0.74		23	15.65
	Total 407	**Total**		**Total 3380**	**Total**
	Range 112			**Range 880**	
	SD: 20.35715			**SD:** 157.0131	

CELE essays use a restricted range of attitude markers.

Appendix 8: boosters

Boosters (disparity adds weight to BAWE considering the audience)					
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
Actually	15	3.07		243	2.95
Always	20	4.09		284	3.45
Believe	9	1.84		203	2.47
Believed	27	5.52		150	1.82
Believes	1	0.2		55	0.67
Beyond doubt	0	0		0	0
Certain	34	6.95		421	5.12
Certainly	3	0.61		133	1.62
Clear	18 (+3 propositional)	3.68		370 (+9 propositional)	4.5
Clearly	10	2.04		279	3.39
Conclusively	0	0		2	0.02
Decidedly	0	0		2	0.02
Definite	0	0		36	0.44
Definitely	3	0.61		31	0.38
Demonstrate	7	1.43		57	0.69
Demonstrated	12	2.45		60	0.73
Demonstrates	3	0.61		49	0.6
Doubtless	0	0		0	0
Established	15 (+9 propositional)	3.07		216	2.62
Evident	5	1.02		123	1.49
Evidently	1	0.2		17	0.21
Find	17	3.48		321	3.9
Finds	2	0.41		37	0.45
Found	27	5.52		535	6.5
In fact	8	1.64		261	3.17
Incontestable	0	0		3	0.04
Incontestably	0	0		0	0
Incontrovertible	0	0		0	0
Incontrovertibly	0	0		0	0
Indeed	7	1.43		339	4.12
Indisputable	0	0		5	0.06
Indisputably	0	0		2	0.02
Know	4	0.82		16 (+9 propositional)	0.19
Known	25	5.11		189	2.3
Must	20	4.09		789	9.6
Never	5	1.02		235	2.86
No doubt	9	1.84		30	0.36
Obvious	17	3.48		102	1.24
Obviously	16	3.3		57	0.69
Of course	0	0		76	0.92
Prove	2	0.41		80	0.97
Proved	14	2.86		82	1
Proves	2	0.41		30	0.36
Realise (ize)	5	1.02		50	0.61
Realised (ized)	3	0.61		58	0.7
Realises (izes)	0	0		6	0.07
Really	11	2.25		125 (+6	1.52

				propositional)	
Show	13	2.66		310	3.77
Showed	16	3.3		139	1.69
Shown	7	1.43		366	4.45
Shows	23	4.7		407	4.95
Sure	1	0.2		35	0.43
Surely	0	0		42	0.51
Think	7 (+9 propositional)	1.43		196	2.38
Thinks	1	0.2		13	0.16
Thought	15	3.1		207 (+11 propositional)	2.52
True	13	2.66		235	2.86
Truly	9	1.84		61	0.74
Undeniable	1	0.2		13	0.16
Undeniably	0	0		6	0.07
Undisputedly	0	0		0	0
Undoubtedly	6	1.23		36	0.44
Without doubt	0	0		5	0.06
	Total 489	**Total**		**Total 8230**	**Total**
	Range 26			**Range 533**	
	SD: 8.465707			**SD:** 156.4635	

As with the other data reviewed, a quantitative analysis reveals little in the way of significant differences with the exception that CELE students tend to use a restricted range. A qualitative analysis, therefore, is necessary.

Appendix 9: self mentions

Self mention (this is significant – audience engagement)					
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
I	0	0		1893	36.31
Me	0	0		173	3.32
Mine	(4 propositional)	0		10	0.19
My	0	0		406	7.79
Our	19 (+5 propositional)	100		523	10.03
Us	(30 propositional)	0		527	10.12
We	0	0		1612	30.92
The author	0	0		38	0.73
The author's	0	0		11	0.21
The writer	0	0		17	0.33
The writer's	0	0		3	0.06
	Total 19	**Total**		**Total 5213**	**Total**
	Range 19			**Range 1890**	
	SD: 5.728716			**SD:** 667.7942	

Self mentions are rare in CELE essays. Possibly because Ss have been taught not to use them.

This is the only category of metadiscourse features that reveal any significant differences between the two corpora. However, the data themselves can be misleading because the low representation of engagement markers in CELE essays may be the result of students being taught not to use personal pronouns.

Appendix 10: engagement markers

Engagement markers (some significance, adds to argument that BAWE engage/consider the audience)					
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**Frequency**	**%age of token**
(the) reader's	0	0		0	0
Add	2 (+1 propositional)	0.22		48 (+2 propositional)	0.4
Allow	3 (+2 propositional)	0.34		172 (+2 propositional)	1.44
Analyse	8	0.89		93 (+2 propositional)	0.78
Apply	3	0.34		87 (+2 propositional)	0.73
Arrange	0	0		7	0.06
Assess	0	0		51	0.43
Assume	2	0.22		83	0.7
By the way	0	0		6	0.05
Calculate	0	0		42	0.35
Choose	10	1.12		120	1.01
Classify	0	0		10	0.08
Compare	2	0.22		77	0.65
Connect	0	0		13	0.11
Consider	7	0.78		196	1.61
Consult	0	0		7	0.06
Contrast	14	1.57		33	0.28
Define	0	0		94	0.79
Demonstrate	7	0.78		57	0.48
Determine	8	0.89		123	1.03
Do not	18	2.01		351	2.94
Develop	18	2.01		151	1.27
Employ	1	0.11		38	0.32
Ensure	5	0.56		161	1.35
Estimate	4	0.45		31	0.26
Evaluate	43	4.81		61	0.51
Find	17	1.9		318 (+3 propositional)	2.67
Follow	2	0.22		124	1.04
Go	11	1.23		169 (+2 propositional)	1.42
Have to	18	2.01		189	1.58
Imagine	0	0		17	0.14
Incidentally	0	0		3	0.03
Increase	83	9.28		517	4.33
Input	1	0.11		10 (+129 propositional)	0.08
Insert	0	0		1	0.008
Integrate	0	0		13	0.11
Key	23	11.19		246 (+75 propositional)	2.06
Let us	1	0.11		7	0.06
Let x = y	-	-		-	-
Let's	0	0		0	0
Look at	1	0.11		113 (+119 propositional)	0.95

Mark	0	0		14 (+82 propositional)	0.12
Measure	12 (+1 propositional)	1.34		19 (+119 propositional)	0.16
Mount	0	0		(5 propositional)	0
Must	20	2.24		789	6.61
Need to	16	1.79		306	2.57
Note	9	1.01		100 (+33 propositional)	0.84
Notice	1 (+1 propositional)	0.11		12 (+24 propositional)	0.1
Observe	(1 propositional)	0		47	0.39
One's	0	0		0	0
Order	72 (71 in order to)	8.05		339 (+603 propositional)	2.84
Ought	3	0.34		36	0.30
Our (*inclusive*)	18	2.01		122 (+401 propositional or not inclusive of the reader)	1.02
Pay	13 (+9 propositional)	1.46		18 (+147 propositional)	0.15
Picture	1	0.11		38 (+74 propositional)	0.32
Prepare	3	0.34		3 (+13 propositional)	0.03
Recall	0	0		2 (+17 propositional)	0.02
Recover	(2 propositional)	0		1 (+21 propositional)	0.008
Refer	2	0.22		120	1.01
Regard	5	0.56		126	1.06
Remember	0	0		23 (+23 propositional)	0.19
Remove	(1 propositional)	0		11 (+25 propositional and 1 directive)	0.09
Review	0	0		26 (+116 propositional and 1 directive)	0.22
See	3	0.34		550	4.61
Select	0	0		42	0.35
Set	5 (+19 propositional)	0.56		35 (+430 propositional)	0.29
Should	115	12.86		1,105	9.26
Show	13	1.46		310	2.6
State	21 (+17 propositional)	49.32		668 (+259 propositional)	5.6
Suppose	0	0		7	0.06
Take (a look/as example)	4 (+65 propositional)	0.45		2	0.02
Think about	1	0.11		13	0.11
Think of	0	0		16	0.13
Turn	(12 propositional)	0		143 (+198 propositional)	1.2
Us (*inclusive*)	9	1.01		329 (+127 propositional)	2.76

Use	216	24.2		1,163	9.75
We (*inclusive*)	19	2.13		1,162	9.74
You	1	0.11		392	3.29
Yours	0	0		1	0.008
	Total 894	**Total**		**Total 11929**	**Total**
	Range 215			**Range 1662**	
	SD: 29.8471			**SD:** 254.0707	

The data show that the engagement marker 'key' is more evident in CELE essays than in BAWE.

Key	23	11.19		246 (+75 propositional)	2.06

Engagement markers are used by CELE students in roughly equal proportions to those used by the authors of the BAWE essays. The question that remains, therefore, is how effectively are they used to persuade the reader to accept their argument?

It appears that the engagement marker 'key' is used to introduce a concept and map out the discourse for the reader and serve, in effect, to function as interactive devices. This is event in both UNNCELE and BAWE.

UNNCELE: Introducing or clarifying a concept

1. … water pollution as acid rain, which is caused by the significant amount of carbon emissions is another key factor linked with loss of biodiversity.

BAWE: Introducing or clarifying a concept

2. Ken Pryce's investigation into West Indian society in Bristol highlights some key issues within the study of ethnography. Pryce aims to create a study with no

3. … ideas of Liberals, between the normative basis and its implementation, is key to the argument of this essay.

UNNCELE: Mapping out the discourse approach

1. This essay will examine the key causes of air pollution and water pollution separately and consider how they affect human health in large cities.

2. Water, as one of the most essential resources, is considered the key ingredient for the existence of life forms.

BAWE: Mapping out the discourse approach

1. One of the key aspects of the holistic approach is the addressing of imbalances in the …
2. One key difference between feminist and most male based theory is that subjectivity
3. … this is a key strength of the theory.
4. Key to this argument, is an appreciation of the ever-changing and non-static …

The major difference between the two corpora is that the authors of the BAWE essays use the term 'key' as an interactional device by commenting on propositional claims. They anticipate potential audience reactions and counter and critiques as a persuasive technique. This use of the term is not evident in UNNCELE.

BAWE: Commenting on propositional claims

1. … as 'in the first person plural using we/us/ourselves/ours.' However, there is a key distinction between the ways in which these two types of pronouns are used.
2. … understanding but also in terms of correspondence to reality. Nevertheless, the key notation clarifies the problem with Brazil's choice for the position of his …
3. … at pre-intermediate to intermediate levels. Harmer clarifies that 'one of the key issues in adolescence ... is the search for individual identity …
4. ... is the search for individual identity, and this search provides the key challenge for this age group' (2001, p. 39).
5. … to draw a critical comparison between single and multi-strategy research, three key questions should be addressed:

Use of the interactional engagement marker 'allow'

Instances in UNNCELE are few (five instances, two of which were propositional). Audience engagement, therefore, is minimal. In contrast, the BAWE essays display a far higher degree of audience engagement.

UNNCELE: Acknowledgement of, or counter to, anticipated problems

1. Apart from this, current solar systems only allow more than 20% sunshine being used (Masters 2004)

2. … affordable technology of household solar water heating allow Brazil to decrease the energy expenses by 25% (Wuppertal Institute, 2007).

Compare and contrast

1. In contrast, the new media allow the users react immediately such as adding comments as soon as the news co

BAWE: Acknowledgement of, or counter to, anticipated problems

2. A way around this might be to simplify rules initially in order to allow for DECPRO, and then introduce more complexity at a later stage.

3. … indeed, many have questioned whether it was ethical to allow Rashbrook to purchase such treatment given her age, and the potential …

4. … "the Hobbesian scheme has no place for the notion of significance. It will allow only for purely quantitative judgements."

5. Learners may lose a valuable source of learning which is available from learning in a group, as centres only allow group work to be undertaken by prior arrangement, most learning is done in isolation …

6. He was an inflexible Eurocentric. A 'scientific' historian would not allow preference for political and European history to eclipse the need to …

BAWE: Offering further supporting evidence/strengthening arguments

1. They further elaborate that 'focused instruction will allow learners to notice the target features in subsequent input and interaction …

2. … once again democracy's inability to allow science to operate freely is illustrated with the public's attempts to …

3. Levi's worked to sustain its brand image successfully. For example it did not allow Tesco from selling 501 Original Jeans since it undermines its brand image …

4. … will have more significant effects. Whilst a revisable budget will allow the company a larger degree of flexibility …

Use of the interactional engagement marker 'see'

Occurrences in UNNCELE account for only 0.34 percent of the total metadiscourse markers for this category, compared to 4.61 percent in BAWE.

CELE essays: Commenting on evidence from the literature

1. .. for the advanced energy-efficient materials, techniques and products but cannot see the long-term benefit that saving money from energy bills at the moment.

2. … which is economical and environmental, becomes an inevitable tendency. To see if they are alternatives to the use of fossil fuels…

CELE essays: Drawing a conclusion

1. In conclusion, we can see that ranching, logging and resettlement are of great significance in the pro

BAWE: evaluations made on propositional content

From these examples, it is evident that the authors of the BAWE essays use the engagement marker 'see' to make critical evaluations of propositional content. This is often achieved through the use of the inclusive 'we', which in fact, collocates 61 times in the L2 position with, as shown below, the highest frequency rating. This shows a high degree of awareness of the need to draw the reader into the argument as a way of making claims that much more persuasive.

N	Cluster	Freq.	Length
1	WE CAN SEE	43	3
2	TO SEE THE	33	3
3	CAN SEE THAT	22	3
4	TO SEE HOW	18	3
5	SEE THAT THE	18	3
6	SEE FIGURE 1	12	3
7	WE SEE THAT	9	3
8	TO SEE A	9	3
9	TO SEE IF	9	3

2. We therefore cannot see tense as a relationship with time in this way. These examples may leave us a …

3. … we can see that the patient does indeed seem to be able to hold a conversation successfully …
4. When looking at these design features with reference to chimpanzees, we can see that the sign language they are using does indeed confer to the rules of …

5. These people also struggled to see how such linguistic theories could be used in therapy to help patients with …

6. Traditionalists see tense as having a very close relationship with time and as more of a temporal …

7. … whether consciously or not, we tailor our speech to fit the occasion. We can see an example of this by the fact that we notice when people don't do this and

Appendix 11: hedges

Hedges (overall results don't look highly significant, given the corpus sizes, but individual features are significant. Does this align with the literature?)					
	UNNCELE			BAWE sub-corpus	
	Frequency	**%age of category markers**	**%age of total discourse markers**	**%age of token**	**%age of token**
About	91 (+58 propositional)	5.11		80 (+1,065 propositional)	0.67
Almost	21	1.18		164	1.36
Apparent	2 (+ 3 boosters)	0.11		44 (+71 boosters)	0.37
Apparently	0	0		33	0.27
Appear	8 (+3 propositional)	0.45		115 (+51 propositional)	0.96
Appeared	1 (+2 propositional)	0.06		26 (+20 propositional)	0.22
Appears	20 (+2 propositional)	1.12		169 (+36 propositional)	1.41
Approximately	64	3.6		61	0.51
Argue	21	1.18		260	2.16
Argued	48	2.7		390	3.24
Argues	8	0.45		351	2.92
Around	31 (+32 propositional)	1.74		89 (+195 propositional)	0.74
Assume	2	0.11		83	0.69
Assumed	3	0.17		95	0.8
Broadly	0	0		24	0.2
Certain amount	3	0.17		4	0.03
Certain level	(2 propositional)	0		7 (+3 propositional)	0.06
Claim	11	0.62		206	1.71
Claimed	26	1.46		90	0.75
Could	93 (+10 propositional)	5.22		946 (+413 propositional)	7.87
Couldn't/could not	1 (+5 propositional)	0.06		(119 propositional)	0
Doubt	(9 boosters)	0		5 (+64 boosters)	0.04
Doubtful	2	0.11		7	0.06
Essentially	3	0.17		94	0.78
Estimate	4	0.22		10 (+21 propositional)	0.08
Estimated	83	4.66		58	0.48
Fairly	5	0.28		56	0.47
Feel	2	0.11		46 (+134 propositional)	0.38
Feels	0	0		4 (+19 propositional)	0.03
Felt	(2 propositional)	0		36 (+93 propositional)	0.3
Frequently	12	0.67		88	0.73
From my perspective	0	0		0	0
From our perspective	0	0		0	0
From this perspective	0	0		4	0.03

Generally	51	2.87		209	1.74
Guess	0	0		5	0.04
Indicate	11	0.62		91	0.76
Indicated	14	0.79		79	0.66
Indicates	13	0.73		123	1.02
In general	8	0.45		74	0.62
In most cases	2	0.11		8	0.07
In most instances	0	0		0	0
In my opinion	0	0		23	0.19
In my view	0	0		5	0.04
In this view	0	0		4	0.03
In our opinion	0	0		1	0.008
In our view	0	0		0	0
Largely	7	0.39		164	1.36
Likely	30	1.69		410	3.41
Mainly	83 (+5 propositional)	4.66		119	0.99
May	254	14.27		1,579 (+76 propositional)	13.13
Maybe /may be	67	3.76		465	3.87
Might	29	1.63		353 (+3 propositional)	2.94
Mostly	7	0.39		52	0.43
Often	75	4.21		668	5.56
On the whole	0	0		9	0.07
Ought	3	0.17		36	0.3
Perhaps	4	0.22		418	3.48
Plausible	0	0		22	0.18
Plausibly	0	0		2	0.02
Possible	22	1.24		475	3.95
Possibly	9	0.51		65	0.54
Postulate	0	0		5	0.04
Postulated	0	0		11	0.09
Postulates	0	0		5	0.04
Presumable	0	0		1	0.008
Presumably	0	0		15	0.12
Probable	6	0.34		15	0.12
Probably	33	1.85		120	1
Quite	24	1.35		176	1.46
Rather x	4 (+12 propositional)	0.22		63 (+580 propositional)	0.52
Relatively	71	3.99		195	1.62
Roughly	4	0.22		19	0.16
Seems	70	3.93		492	4.09
Should	115	6.46		20 (+905 propositional) – criteria used: should as a hedge shows a possible result/conditional sentence	0.17
Sometimes	20	1.12		115	0.96
Somewhat	1	0.06		69	0.57
Suggest	4	0.22		236	1.96
Suggested	12	0.67		204	1.7
Suggests	8	0.45		364	3.03
Suppose	0	0		7	0.06

Supposed	2	0.11		58 (+1 propositional)	0.48
Supposes	0	0		2	0.02
Suspect	0	0		6	0.05
Suspects	0	0		3	0.02
Tend to	10	0.56		137	1.14
Tended to	1	0.06		53	0.44
Tends to	15	0.84		52	0.43
To my knowledge	0	0		2	0.02
Typical	12	0.67		68	0.57
Typically	2	0.11		45	0.37
Uncertain	4	0.22		24	0.2
Uncertainly	0	0		1	0.008
Unclear	2	0.11		15	0.12
Unclearly	0	0		0	0
Unlikely	3	0.17		61	0.51
Usually	42	2.36		138	1.15
Would	5 (+102 propositional)	0.28		118 (+1887 propositional)	0.98
Wouldn't / would not	(4 propositional)			5 (+134 propositional)	0.04
	Total 1780	**Total**		**Total 12024**	**Total**
	Range 253			**Range 1578**	
	SD: 34.77596			**SD:** 216.4239	

CELE students use the hedges they have been taught. This adds weight to the argument that their writing is lacking because of their limited knowledge of linguistic devices (Hood, 2004).

Hedging devices not listed by Hyland (2005) yet used in CELE essays. Is this because they are taken from Seitz?

- What is more, global warming tends to be the most catastrophic impact which is in part true that developing nations would suffer more owing to ….

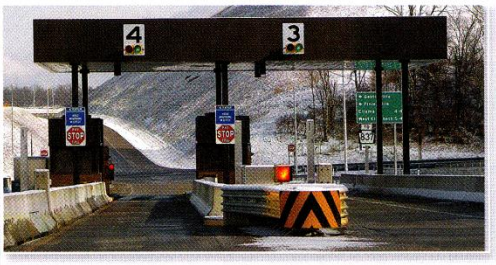
- … population growth, plus less ability in coping with the greenhouse gas emissions in part as a consequence of limited economic capability. Furthermore, the Intern

'Before' used as sequencing markers and not listed by Hyland (2005)

- Before August 2007, 439 nuclear reactors were put into operation in 30 countries
- .. is often not treated before being discharged into the water

INTELLIGENT TRANSPORTATION SYSTEMS

1 Intelligent transportation systems—or ITS—use current communication technologies to improve the efficiency of existing transportation systems. Electronic devices such as road sensors, smart traffic signals, and variable message signs can provide valuable information to drivers, dispatchers, and police.



2 Efficiency is a constant challenge for transportation engineers because transportation demand varies by time of day. Almost all daily congestion occurs during peak periods; however, in most cities, it lasts only a few hours. Therefore, it is difficult to argue that building more roads is the most cost-effective way to eliminate traffic problems, especially if something else can be done. Three alternatives are possible:

a. Do nothing. Some people believe that traffic control in cities is now effectively accomplished by the congestion. Adding more vehicle capacity would only encourage more traffic.

b. Increase the cost of using the highway during peak periods. City governments could collect tolls or increase downtown parking prices to discourage use of the highway.

c. Use all the technology possible to encourage drivers to use the highway when it is less crowded. Communications technologies can help drivers make better decisions about when and where to travel.

3 The development of intelligent transportation systems (ITS) is a result of the third alternative. ITS are designed to improve existing road systems. This helps transportation engineers overcome the efficiency, cost, environmental, and safety challenges[1] they face.

TECHNOLOGY USED IN ITS

4 Road sensors are key technology for ITS. Buried under the roads, they sense heavy traffic flow at road intersections. The road sensors send a signal to "smart" traffic lights, which automatically change the signal timing to help move the vehicles more quickly through an intersection. This allows express buses to travel quickly through crowded downtowns during rush hours because all the traffic lights are green for them. If the smart traffic lights are connected to a traffic operations center, a dispatcher at the center can also control the traffic signals, allowing ambulances and police to travel quickly in an emergency situation. In Shanghai, for instance, transportation engineers have already installed road sensors and smart traffic lights near on-ramps to elevated roads. The smart light turns red shortly after an accident. This prevents other vehicles from entering the elevated roads and reduces traffic congestion.

5 Freeway management systems are another example of ITS. There are different types of freeway management systems. Some are used to collect tolls electronically. Traditionally, collecting tolls has meant that a person in a toll booth collects money from each driver, which increases the travel time of the trip. Freeway management systems that use ITS can collect tolls electronically without increasing total travel time. For example, in Toronto, Canada, many drivers who use the freeway install transponders in their vehicles, which signal a central billing system every time the driver uses the freeway. For vehicles that don't have transponders, cameras take photos of the vehicles' license plates. This identifies the drivers who should be billed. This freeway management system increases the cost of using the freeway without increasing total travel time. This reduces traffic congestion on the freeway.

[1] **overcome . . . challenge** *exp* to successfully solve a problem in the face of difficulties, challenges, or opposition

References

Abdi, R. (2002). Interpersonal metadiscourse: an indicator of interaction and identity. *Discourse Studies*, *4*(2), 139-145.

Aijmer, K. (2002). Modality in advanced Swedish learner's written interlanguage. In S. Garner, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55-76). Amsterdam: John Benjamins Publishing Co. Ltd.

BAWE (2004-2007). The data in this study come from the British Academic Written English (BAWE) corpus, which was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800).

Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics - Special issue on using large corpora: II*, *19*(2), 219-241.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Biggs, J. (1996). Western misperceptions of the Confucian Heritage learning culture. In D. Watkins & J. Biggs (Eds.), *The Chinese learner: cultural, psychological and contextual influences*. Hong Kong: Comparative Education Research Centre/Australian Council for Educational Research.

Brown, V. (1993). Decanonizing discourses: Textual analysis and the history of economic thought. In T. Dudley-Evans, W. Henderson, & R. Backhouse (Eds.), *Economics and language*. London: Routledge.

Canagarajah, S. (2002). *Critical academic writing and multilingual students.* Michigan: Ann Arbor: University of Michigan Press.

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, *6*(4), 289-302.

Cook, G. (1998). The uses of reality: a reply to Ronald Carter. *ELT Journal*, *52*(1), 57-63.

Crismore, A., Markkanen, R., & Steffensen, M.S. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication*, *10*(1), 39-71.

Currie, G. (2005). *"Beyond our imagination": The voice of international students on the MBA.* University of Nottingham.

Ede, L., & Lunsford, A. (1984). Audience awareness/audience invoked: The role of audience in composition theory and pedagogy. *College Composition and Communication*, *35*, 155-171.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, *6*(1), 3-17.

Flowerdew, L. (2001). The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry, & R.L. Roseberry (Eds.), *Small corpus studies and ELT: theory and practice* (pp. 363-380). Amsterdam: John Benjamins.

Fox, H. (1994). *Listening to the world: Cultural issues in academic writing.* Urbana: National Council of Teachers of English.

Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, *6*(4), 319-335.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam: Benjamins.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. New York: Longman.

Hinds, J. (2001). Reader versus writer responsibility: a new typology. In T.J. Silva & P. Matsuda, Kei (Eds.), *Landmark essays on ESL writing*. New Jersey: Lawrence Erlbaum Associates.

Hood, S. (2004). *Appraising research: taking a stance in academic writing*. PhD Thesis, University of Technology, Sydney.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hyland, K. (1998a). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, *30*, 437-455.
Hyland, K. (1998b). Talking to students: Metadiscourse in introductory coursebooks. *English for specific purposes*, *18*(1), 3-26.
Hyland, K. (2002a). Directives: argument and engagement in academic writing. *Applied Linguistics*, *23*(2), 215-239.
Hyland, K. (2002b). Specificity revisited: how far should we go now? *English for specific purposes*, *21*, 385-395.
Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, *12*, 17-29.
Hyland, K. (2004a). *Genre and second language writing*. Michigan: University of Michigan Press.
Hyland, K. (2004b). Patterns of engagement: dialogic features and L2 undergraduate writing. In L.J. Ravelli & R.A. Ellis (Eds.), *Analysing Academic Writing* (pp. 5-23). London: Continuum.
Hyland, K. (2005). *Metadiscourse: exploring interaction in writing*. London: Continuum.
Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, *27*, 4-21.
Hyland, K. (2009). Writing in the disciplines: Research evidence for specificity. *Taiwan International ESP Journal*, *1*(1), 5-22.
Hyland, K., & Hamp-Lyons, L. (2002). EAP: issues and directions. *Journal of English for Academic Purposes*, *1*, 1-12.
Johns, A.M. (1993). Written argumentation for real audiences: suggestions for teacher research and classroom practice. *TESOL Quarterly*, *27*(1), 75-90.
Johns, A.M. (1997). *Text, role and context: developing academic literacies.* Cambridge: Cambridge University Press.
Johns, T. (2002, 19-24 July, 2000.). *Data-driven Learning: The Perpetual Challenge*. Paper presented at the The Fourth International Conference on Teaching and Language Corpora, Graz.
Kirkpatrick, A. (1997). Traditional Chinese text structures and their influence on the writing in Chinese and English of contemporary mainland Chinese students. *Journal of Second Language Writing*, *6*(6), 223-244.
Krishnamurthy, R., & Kosem, I. (2007). Issues in creating a corpus for EAP pedagogy and research. *Journal of English for Academic Purposes*, *6*(4), 356-373.
Kuteeva, M. (2011). Wikis and academic writing: Changing the writer-reader relationship. *English for specific purposes*, *30*, 44-57.
Liu, J. (2008). The Generic and Rhetorical Structures of Expositions in English by Chinese Ethnic Minorities: A Perspective from Intracultural Contrastive Rhetoric. *Language and Intercultural Communication*, *1*(1), 2-20.
Locker, K.O. (1989). *Business and administration communication*. Boston: Irwin.
Luzon, M.J. (2009). The use of *we* in a learner corpus of reports written by EFL Engineering students. *Journal of English for Academic Purposes*, *8*(3), 192-206.
Maton, K. (2009). Cumulative and segmented learning: exploring the role of curriculum structures in knowledge-building. *British Journal of Sociology of Education*, *30*(1), 43-57.
Maton, K. (forthcoming). *Knowledge and knowers: Towards a realist sociology of education*. London: Routledge.
McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. London: Routledge.
Mei, W.S. (2007). The use of engagement resources in high- and low-rated undergraduate essays. *Journal of English for Academic Purposes*, *6*(3), 254-271.
Mohan, B., & Lo, W.A.-Y. (1985). Academic writing and Chinese students: Transfer and developmental factors. *TESOL Quarterly*, *19*(4), 515-534.
Mukherjee, J. (n.d.). Corpus linguistics and language pedagogy: The state of the art – and beyond. *Justus Liebig University*. Giessen: http://www.uni-giessen.de/anglistik/ling/Staff/mukherjee/pdfs/Corpus%20Linguistics%20and%20Language%20Pedagogy.pdf Accessed: 12 April 2011.

Myers, G.A. (1992). Textbooks and the sociology of scientific knowledge. *English for specific purposes*, *11*, 3-17.
Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 125-152). Amsterdam: John Benjamins.
Park, D.B. (1986). Analyzing audiences. *College Composition and Communication*, *37*(4), 478-488.
Paxton, M. (2007). Tensions between textbook pedagogy and the literary practices of the disciplinary community: a study of writing in first year economics. *Journal of English for Academic Purposes*, *6*(2), 109-125.
Scardamalia, M., & Beretier, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (Vol. 2: reading, writing and language learning, pp. 142-175). Cambridge: Cambridge University Press.
Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
Scollon, R., & Scollon, S.W. (2001). *Intercultural Communication*. (2nd ed.). Oxford: Blackwell.
Scott, M. (2009). WordSmith Tools version 5.0. Liverpool: Lexical Analysis Software.
Searle, J.R. (1976). A classification of illocutionary acts. *Language in Society*, *5*(1), 1-23.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Swales, J.M. (1995). The role of the textbook in EAP writing research. *English for specific purposes*, *14*(1), 3-18.
Swales, J.M., Ahamed, U.K., Chang, Y.-Y., Chavez, D., Dressen, D.F., & Seymour, R. (1998). Consider this: the role of imperatives in scholarly writing. *Applied Linguistics*, *19*(1), 97-121.
Thompson, G. (2001). Interaction in academic writing: learning to argue with the reader. *Applied Linguistics*, *22*(1), 58-78.
Vande Kopple, W.J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, *36*(1), 82-93.
Weber, J.-J. (2001). A concordance- and genre-informed approach to ESP essay writing. *ELT Journal*, *55*(1), 14-20.
Williams, J., & Hill, D. (2010). *Academic Connections 3*. New York: Pearson Education.
Wilson, D., & Sperber, D. (2004). Relevance theory. In G. Ward & L.R. Horn (Eds.), *Handbook of Pragmatics* (pp. 607-632). Oxford: Blackwell.
Wollman-Bonilla, J.E. (2001). Can first-grade writers demonstrate audience awareness? *Reading Research Quarterly*, *36*(2), 184-201.
Wu, S.M. (2007). The use of engagement resources in high- and low-rated undergraduate geography essays. *Journal of English for Academic Purposes*, *6*(3), 254-271.
Zarei, G.R., & Mansoori, S. (2007). Metadiscourse in academic prose: a contrastive analysis of English and Persian research articles. *The Asian ESP Journal*, *3*(2), 24-40.