Macrothink Institute™

International Journal of Linguistics
ISSN 1948-5425
2018, Vol. 10, No. 2

A Cross-disciplinary Investigation of Textual Metadiscourse Markers in Academic Writing

Ebtisam Saleh Aluthman, PhD

Department of Applied Linguistics

Princess Nourah bint Abdulrahman University

P.O. Box: 84428, 11671. Riyadh, Saudi Arabia

E-mail: esaluthman@pnu.edu.sa

Received: Dec. 16, 2017 Accepted: December 24, 2017 Published: March 30, 2018

doi:10.5296/ijl.v10i2.12916 URL: https://doi.org/10.5296/ijl.v10i2.12916

Abstract

This paper is a cross-disciplinary investigation of textual metadiscourse markers (TMDMs) in the British Academic Written English (BAWE) corpus. To this end, corpus analysis tools are utilized to investigate the density of each category of TMDMs and the distribution of these categories across BAWE academic divisions. The premise under the investigation in this study is that metadiscourse is a social interaction that facilitates knowledge communication between writers and their targeted readers. Within the framework of Hyland' s (2000, 2005) claim of disciplinary discourse, this study shows a correlational relation between the use and frequency of different TMDMs across academic disciplines. The quantitative results indicate that the use of transitions outnumbered all other TMDMs in the four BAWE academic divisions, accounting for almost 76.76% of the total occurrences of TMDMs. The results also indicate a greater tendency to use endophoric markers in the physical sciences whereas the social sciences and arts and humanities exhibit larger quantities of evidentials and code glosses. The results are interpreted in terms of the epistemological nature of the academic divisions as well as the target readership.

Keywords: Academic writing, Metadiscourse, Textual metadiscourse markers, Academic disciplines

1. Introduction

1.1 Preliminaries

There has been a focus on academic writing within the field of English for academic purposes (EAP) over the last few decades (Römer, 2012). Biber and Conrad (2009) argued that the practices of academic writing can differ in variables such as the intended audience, communicative purposes, and academic discipline. In their analysis of genres across the disciplines, Nesi and Gardener (2012) identified 13 genres of academic writing in higher education, namely case study, critique, design specification, empathy writing, essay, exercises, explanation, literature survey, methodology recount, narrative recount, problem question, proposal, and research report. Despite the numerous studies on the process and practices of effective academic writing, few have focused on the discourse-coherent aspects of these practices (Hyland, 2009). Moreover, few studies have informed EAP practitioners' decisions regarding what should be included and how to teach coherence in academic writing for the various academic disciplines.

Recently, there has been a gradual but significant shift in the field of discourse analysis (DA), with more of a focus being placed on the language learner's authentic production of language in academic discourse (e.g., Biber, 2007; Hyland & Tse, 2004, Swales, 2004). Elements of DA research and corpus tools have been integrated into what has been coined the corpus-informed discourse analysis (Hyland, 2009), which is a more enriched and insightful analysis of discourse in language use. Within the context of this framework, corpus tools have proven to be particularly significant, featuring significant evidence of language use among a particular community and indicating the typical patterning of language use in a particular context. Approaching a text as a package of linguistic features, corpus tools bring evidence into the research of language variations across different disciplines of written academic discourse (Hyland & Tse, 2004, 2009).

Utilizing BAWE and a heuristic approach of corpus-informed discourse analysis, the current study reports on a cross-discipline investigation of the frequency and use of TMDMs across four main academic divisions: the arts and humanities (AH), the social sciences (SS), the life sciences (LS), and the physical sciences (PS). Drawing on Hyland and Tse's (2004) modified version of Crismore, Markkanen, and Steffensen's (1993) taxonomy of MDMs and using the tools available in the Sketch Engine platform (see https://www.sketchengine.co.uk/) such as "Frequency" and "Relative Frequency" (Nesi & Thompson, 2014), the current study examines the occurrences of TMDMs in academic writing. The results are interpreted within the context of well-documented epistemologies and knowledge underlying the academic discipline under investigation.

The signficance of the present study is that identifying the most frequent TMDMs in the four main academic disciplines can have enormous potentials for a number of fields. First, the current study contributes to the body of research in the field of disciplinary discourse and discipline-specific knowledge-making practices (Hyland, 2000, p. xii). Second, the current study will increase students' awareness of the expert's use of TMDMs in academic writing. Third, it will have a significant impact in the field of EAP because it will guide the

language-teaching approaches to TMDMs in terms of preparing pedagogical materials and developing textbooks in different academic disciplines. Implications drawn from the results of the current study may contribute to the design of EAP academic writing courses for various academic disciplines.

1.2 Review of Related Literature: Disciplinary Discourse and Metadiscoursal Markers

This section briefly reviews the most significant issues of previous studies and their results regarding the topic at hand. First, the functions of MDMs are presented, including their relation to Halliday's (1973) theory of macro-meta functions of language. Second, the literature on disciplinary variations are reviewed, specifically the studies with references related to MDMs across disciplines.

The scholarly research on MDMs is theoretically grounded in Halliday's (1973) theory of macro-meta functions of language. Writers and speakers refer to a variety of language resources used to organize the content of a discourse in a coherent way that is acceptable by the readership. According to Halliday, language use has three different, yet independent, functions. The first macro-function is the ideational one related to the expressing of facts, experiences, concepts, and ideas. It is mainly concerned with the ideational consent of the language. The second macro-function is the interpersonal one that relates to the establishment of the writer's or speaker's feelings and attitudes by means of linguistic resources. The third macro-function is the textual one that is concerned with the coherence and cohesion of the spoken/written language. Metadiscourse scholars (such as Crismore et al., 1993; Hyland, 2000) have collectively claimed that metadiscourse achieves both textual and interpersonal functions but not ideational ones.

The concept of "metadiscourse" dates back to 1959 when the term was coined by Zellig Harris in describing the writer's or speaker's linguistic resources to guide a reader's or hearer's perception of a text (Hyland, 2005). It is simply defined as the discourse about discourse. According to metadiscourse theorists, communication is not only about the exchange of messages, but it is also the way people monitor, articulate, and construct this exchange. Hyland (2005, p. 4) maintained that metadiscourse provides the speaker or writer with a variety of resources "to convey his personality, credibility, audience-sensitivity and relationship to the message."

Given the diversity of metadiscourse functions realized by the theorists of MDMs, a number of different classifications have been developed. Most are based on the classification of MDMs proposed by Vande Kopple (1985), who classified MDMs into textual and interpersonal (see Hyland [2005] for a detailed description of all the classifications of MDMs). Hyland (1998) also categorized MDMs into textual and interpersonal markers. According to Hyland (1998), textual metadiscourse markers (TMDMs) refer to the resources used by the writer for the sake of establishing the clarity and coherence of a given text for a particular readership. Interpersonal metadiscourse markers (IMDMs), on the other hand, are the resources used to construct the writer's perspectives and evaluation toward ideational information. Table 1 shows Hyland 's (2000) modified version of MDMs, the classification adopted in the present study, which is based on Crismore et al.'s (1993) classification.

Table 1. Functions of metadiscourse in academic texts (Hyland, 2000, p. 111)

Category	Function	Examples
	Textual metadiscourse markers	
Logical connectives "Transition markers"	Express the semantic relation between main clusters	In addition, but, thus, and
Frame markers	Explicitly refer to discourse acts or text stages	Finally, to repeat, here we try to
Endophoric markers	Refer to information in other parts of the text	Noted above, see Figure, in section 2
Evidentials	Refer to sources of information from other texts	According to X/ Z states
Code glosses	Help readers grasp the meaning of ideational material	Namely, e.g., such as, i.e.
	Interpersonal metadiscourse markers	
Hedges	Withhold the writer's full commitment to the statement	Might, perhaps, possible, about
Boosters	Emphasize force or the writer's certainty of the message	In fact, definitely, it is clear
Attitude Markers	Express a writer's attitude about the propositional content	Unfortunately, I agree, X claims
Relational markers	Explicitly refer to or build the relationship with the reader	Frankly, note that, you can see
Person	Explicit reference to author(s)	I, we, my, mine, our

The written discourse has attracted an increasing amount of attention in different academic fields with a clear acknowledgment that knowing a discipline implies understanding its written discourse (Bazerman, 1993). Bazerman stated:

> …everything that bears on the professions bears on professional writing. Indeed, within the professions, writing draws on all the professional resources, wends its way among the many constrains, structures, and dynamics that define the professional realm and instantiates professional work. (1993, p. vii)

The "professional resources" in Bazerman's statement refer to the linguistic and discursive resources by which knowledge is constructed, negotiated, and communicated within a particular readership. According to Hyland (2000) and Swales (1990), it is "how" writers of different disciplines write rather than "what" they write that creates the differences between disciplinary discourses. This means that the writers of a particular discipline display a professional competence in their discipline-approved practices. Actually, "writers in different disciplines go about producing knowledge, epistemic belief, and institutional structures of academic communities" (Hyland, 2000, p. 2).

Metadiscourse in academic writing research is based on this well-acknowledged claim among metadiscourse theorists that academic texts manifest discipline-specific knowledge-making practices (Hyland, 2000, p. xii). Three decades ago, the concept of metadiscourse was investigated in a variety of contexts, including school textbooks (Crismore, 1989), undergraduate textbooks (Hyland, 2000), casual conversation (Schifrin, 1980), oral narratives (Norrick, 2001), and postgraduate dissertations (Bunton, 1999; Hyland, 2000; Hyland & Tse,

2004; Swales, 1990). However, Hyland's (2000) *Disciplinary Discourse: Social Interactions in Academic Writing* and Hyland's (2005) *Metadiscourse* both initiated corpus-informed discourse analysis (Hyland, 2009) in metadiscourse research and have been considered significant references in the field. Hyland (2000, 2005) investigated a corpus of 1.5 million words of 1,400 texts from five genres in eight disciplines in a cross-sectional investigation of the regularities and repetitions of some literacy practices. Hyland's data are mainly composed of research articles (RAs), book reviews, scientific letters, abstracts, and textbook chapters. In his analysis of metadiscourse practices, Hyland investigated seven textbooks in eight disciplines, illustrating some of the significant ways that writers in these disciplines "represent themselves, their views, and their audience in constructing both a perspective of a disciplinary consensus and an identity as an expert" (Hyland, 2000, p. xiiii). Hyland (2000) also investigate the use of MDMs in university textbooks; the data consist of 56 textbooks chapters (481,000 words) and display 32,543 instances of MDMs, which equals an average of 580 MDMs per chapter (refer to Hyland [2000, p. 114, 116] for a detailed clarification of the accounts of MDMs in academic textbooks per 1,000 words and the proportion of textual metadiscourse across disciplines in percentages).

Among Hyland's most significant results is that the textual frame markers account for about 60% of the overall MDMs. This reflects the significance of MDMs as communicative strategies that guide the reading process and emphasize the instructional aspects of the genre. On the other hand, interpersonal MDMs such as boosters and hedges are only one-fifth of the overall MDMs and half of the interpersonal MDMs. According to Hyland (2000), this shows that textbook authors are concerned about expressing their argumentation explicitly in a coherent and organized way.

The investigation of metadiscourse markers in academic writing has received noticeably more attention recently, especially in the genre of RAs. Cao and Hu (2014) conducted a contrastive investigation of interactive metadiscourse in a corpus of 120 qualitative and quantitative RAs in three different fields: education, applied linguistics, and psychology. Their analysis revealed both cross-disciplinary and cross-paradigmatic differences that are related to the use of some MDMs, such as comparative transitions and linear transitions, after the RAs' method sections. Rashidi and Alihosseini (2012) explored the use and frequency of MDMs in the abstract parts of 20 RAs across two different disciplines: engineering and sociology. They found no significant variations in the use and frequency of MDMs.

Because both RAs and postgraduate theses display similar rhetorical and stylistic characteristics, scholars researching MDMs have been drawn to this area of inquiry. Both Swales (1990) and Bunton (1998) pointed out that PhD dissertations display a greater use of MDMs than RAs. Partridge and Starfield (2007, p. 89) justified this by noting that PhD theses have "much lengthier" textual characteristics. In this context, Kawase (2015) compared the use of MDMs in the introduction parts of PhD theses and RAs by the same authors and observed a greater usage of MDMs in the introduction of RAs compared with theses. Hyland (2000, 2005) extensively examined metadiscourse across disciplines in academic RAs, including popular science articles and introductory textbooks. The genre of PhD dissertations has also been investigated in terms of the density of MDMs compared to other academic

genres. However, scholarly research is lacking in the investigation of undergraduate's assignments and texts, which is one reason behind the present study.

Understanding the density and distribution of MDMs in the genre of undergraduate-produced texts and assignments can contribute to the understanding of the distance between academics and professionals in different academic disciplines and their counterpart audiences. Because of the limited scope of the current study, it is limited only to the investigation of TMDMs. Hyland (2000) indicated a greater use of TMDMs than IMDMs (60% of the overall MDMs) in his data. The investigation conducted in the present study resonates with Hyland's (2000) claim of disciplinary variation in university textbooks. The analysis is guided by two main questions: (1) What are the most and less frequent TMDMs categories and subcategories in the four BAWE main academic divisions listed above, and (2) Are there any frequency-based significant variations of TMDMs across BAWE academic divisions. The investigation is guided by the hypothesis that there is a correlation between cross-disciplinary differences and the use of some TMDMs. From a methodological perspective, a number of methodological tools have been used to answer the research questions. First, a descriptive analysis of the count, sum, average, and variance of TMDMs has been used to show their distribution among different disciplines. Second, Sketch Engine corpus tools such as “Frequency” and “Relative Frequency" are used to obtain a detailed description of the distribution of TMDMs in the data.

2. Corpus Material and Analysis Procedures

In the current study, BAWE data are used. BAWE is a representative corpus of British Academic English writing and is equally distributed among the main academic disciplines (AH, SS, LS, and PS) and among levels. Appendix A shows a BAWE corpus summary (Nesi & Gardener, 2012, p. 8). The BAWE corpus contains a total of 2,761 samples of proficient evaluated undergraduate students' academic writings that are about 500 words to 5000 words length. BAWE is compiled by Paul Thompson and Alois Heuboeck at the University of Reading, Tagged by with POS CLAWS v7 and semantic category with Matrix. BAWE can be obtained through the Sketch Engine platform, where it can be investigated by a number of corpus tools such as word sketch, thesaurus, word list, n-gram, and concordance. Table 2 summarizes the metadata of BAWE.

Table 2. Metadata of BAWE (Source: The Sketch Engine Platform)

General Information			
Language	English		
Encoding	UTF-8		
Compiled	11/08/2016 05:39:17		
Tagset	POS CLAWS v7		
Counts			
Tokens	8,336,262		
Words	6,968,089		
Sentences	293,113		
Paragraphs	127,401		
Documents	2,761		
Sub corpus Statistics			

Sub corpus	Tokens	Words	%
AH_Disciplines-	2,243,330	1,875,147	30.83
All_except_English_as_author_first_language	2,570,523	2,148,640	30.83
Discipline_Biological_Sciences	455,734	380,937	5.46
Discipline_English	329,853	275,716	3.95
English_as_author_first_language	5,765,739	4,819,448	69.16
Genre_Essays	37,768	31,569	0.45
LS_Disciplines-	1,754,545	1,466,583	21.04
PS_Disciplines-	1,611,261	1,346,815	19.32
SS_Disciplines-	2,727,126	2,279,541	32.71
Tag	358		
Lemma	137,598		
Lempos	157,008		
Sem	2,751		
Textpart	14		
lc	148,879		
lemma_lc	137,598		

3. Results

The current study uses Hyland's (2005, p. 218–224) taxonomy of metadiscourse markers .The distribution of each category of TMDMs in the four BAWE main disciplines are presented first. Then, frequency-based significant variations of these categories across the four BAWE divisions are illustrated by the variance value. The results illustrate the number of times each BAWE category appears in per million words of BAWE. Significant occurrences of some TMDMs are illustrated per million words.

3.1 Textual Metadiscourse Markers in BAWE

Figure 1 shows the distribution of TMDMs in BAWE. The total number of TMDMs occurrences in BAWE is (n =426,984). Evidentials are the least used TMDMs in the whole BAWE (n=4,326, 1.01%), followed by endophoric markers (n=18,446, 4.32%), then code glosses (n= 20,960, 4.91%), and frame markers (n=55,472, 13%). Transition markers account for the most frequent TMDMs in the data (n=327,357, 76.76%), indicating the writers' tendencies to frame the material they are presenting in their texts and illuminate prepositional connections for readers. Figure 1 shows a clear category-related variation in the use of TMDMs. The distribution of TMDMs categories in the four divisions are illustrated in Table 3.

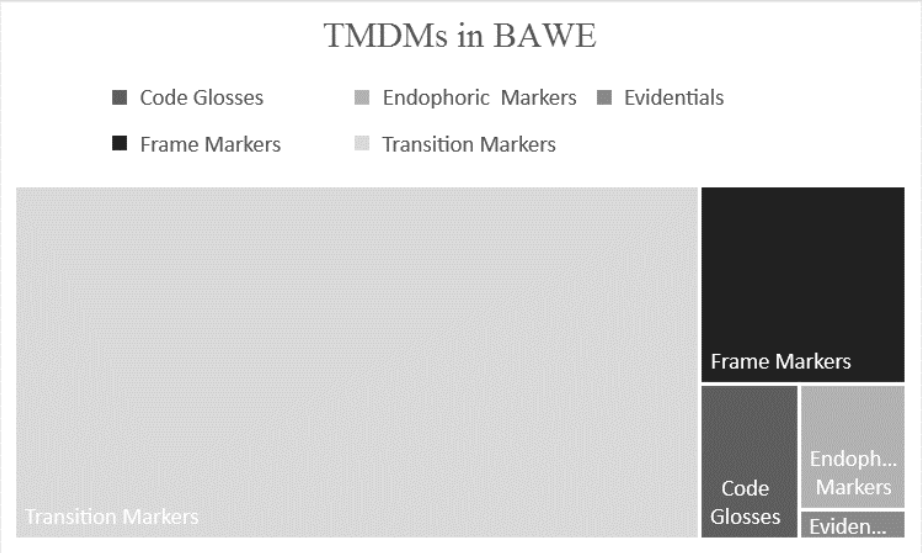

Figure 1. The distribution of TMDMs in BAWE

Table 3. The distribution of TMDMs categories in the four divisions (per million words)

Category	Total no.	SS	LS	AH	PS
Transition markers	327,357 76.76%	110,380 33.71%	72,128 22.03%	87,338 26.68%	57,511 17.58%
Frame markers	55,472 13%	16,463 29.68%	10,081 18.17%	15,016 27.07%	13,912 25.08%
Endophoric markers	18,446 4.32%	4,379 22.98%	4,017 21.08%	3,177 16.67%	7,483 39.27%
Evidentials	43,26 1.01%	1,852 42.81%	984 22.75%	1,026 23.72%	464 10.72%
Code glosses	20,960 4.91%	7,477 30.91%	5,088 21.03%	7,206 29.79	4,421 18.27
Total	422,473				

One of the most significant results obtained is that the transitional marker "and" is significantly the most frequent of the TMDMs (n=208,696, 25,034.70 per million) with an equal distribution among the four main academic divisions, as illustrated in Figure 2.

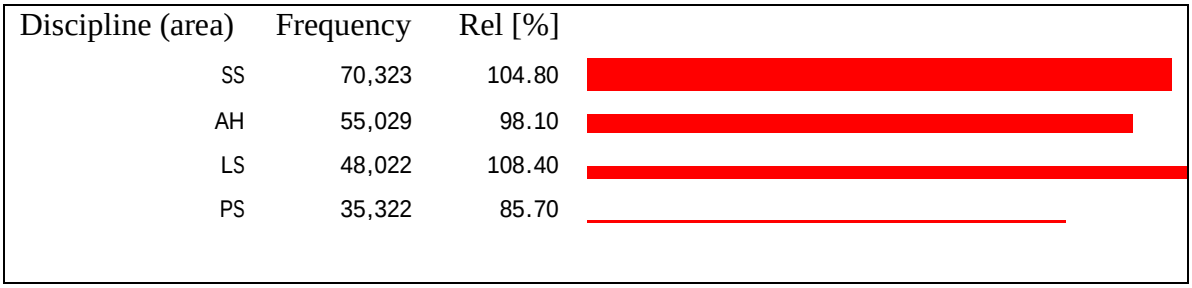

Figure 2. The distribution of "and" in the four BAWE main academic divisions

The corpus quantitative analysis also shows that the transitional markers "also" and "but" are the next most frequently TMDMs used in the BAWE (also, n= 16,740, 2,008.10 per million)

and (but, 15,262, 1,830.80 per million). The distribution of "also" and "but" among the four main academic divisions is illustrated in Figure 3 and Figure 4.

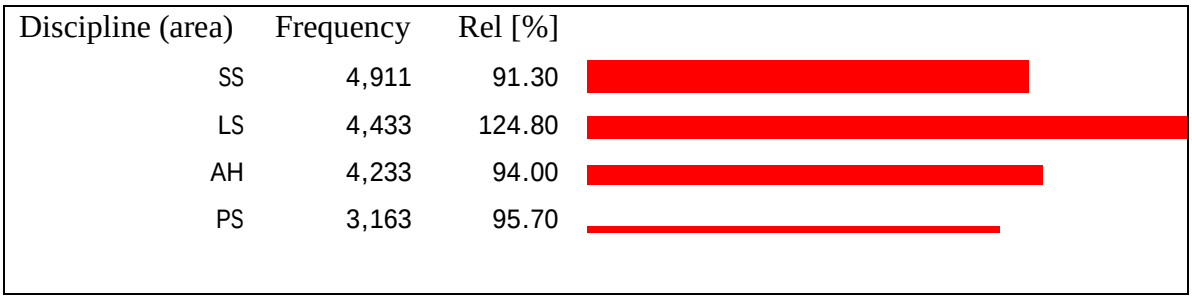

Figure 3. The distribution of "also" in the four BAWE main academic divisions

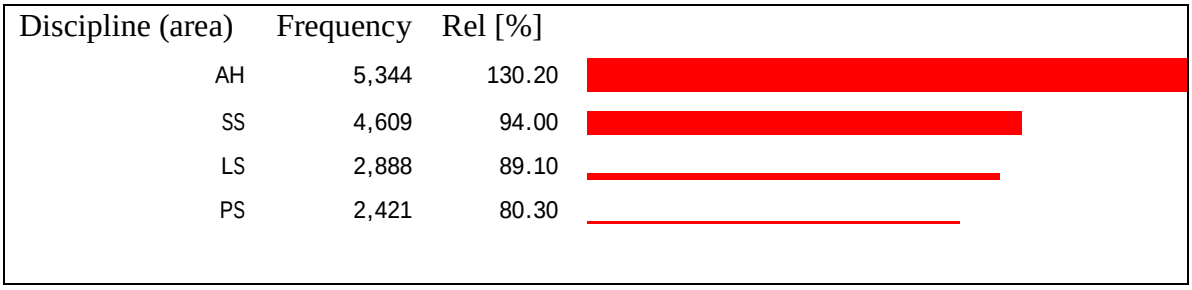

Figure 4. The distribution of "but" in the four BAWE main academic divisions

The code gloss TMDM "as a matter of fact" occurred the least in the BAWE data (n=13, 1.56 per million) and was mostly in the social sciences category, as illustrated in Figure 5.

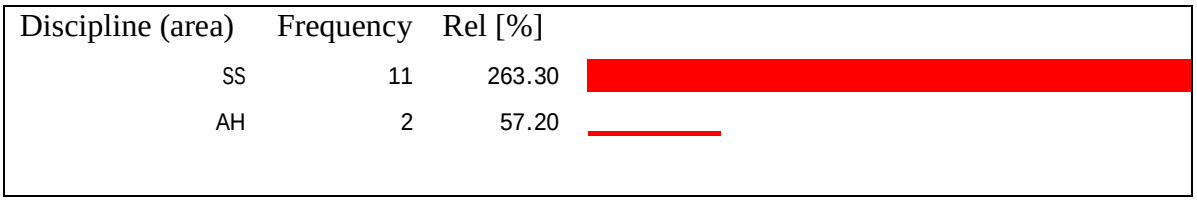

Figure 5. The distribution of "as a matter of fact" in the BAWE four main academic divisions

3.2 Transitional Metadiscourse Markers in BAWE

Forty-seven transitional TMDMs were searched for via Sketch Engine "frequency" and "relative frequency" tools. The occurrences of each transitional TMDMs per million words was also calculated with the same tools.). Figure 6 shows the distribution of transitional TMDMs in BAWE.

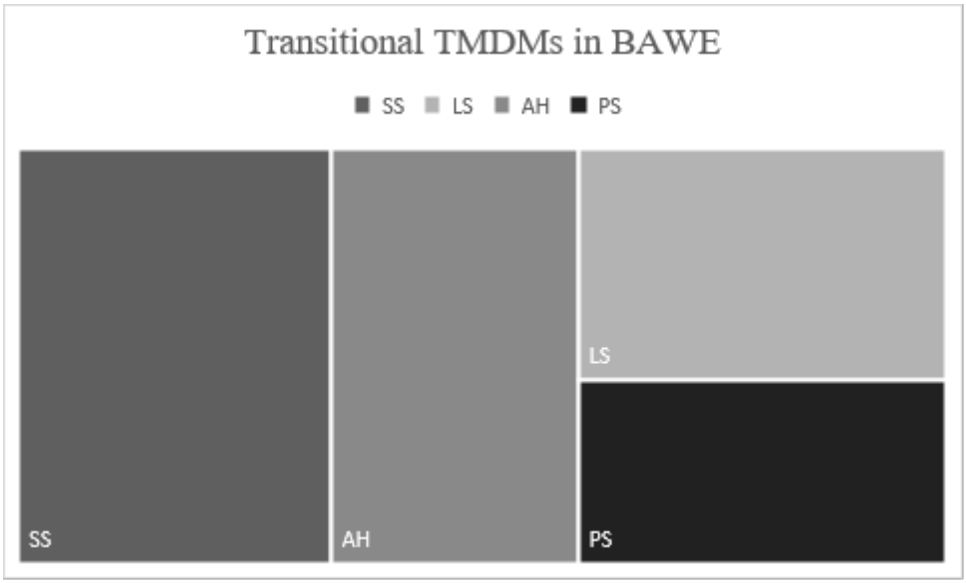

Figure 6. The distribution of transitional TMDMs in BAWE

The use of transitional TMDMs accounts for 76.76% of the total number of TMDMs used in the data. As illustrated in Table 3, transitional TMDMs occurred the most in the social sciences (n=110,380, 33.71%) and the least in the physical sciences (n=57,511, 17.58%). The sum, average, and variance of transitional TMDMs in BAWE are given in Table 4.

Table 4. Descriptive statistics of transitional TMDMs across BAWE divisions.

Division	Count	Sum	Average	Variance
SS	47	110,380	2348.511	198,040,008
LS	47	72,128	1534.638	48,774,776
AH	47	87,338	1858.255	64,043,045
PS	47	57,511	1223.638	26,421,322

3.3 Frame Metadiscourse Markers in BAWE

The occurrences of 57 frame TMDMs have been calculated. The number of frame markers in each category is as follows: sequencing TMDMs (n= 23), labelling TMDMs (n=17), announcing goals TMDMs (n=10), and shifting topic TMDMs (n=7). The total number of frame TMDMs that occurred in BAWE is (n=55,472), constituting only 13% of the total TMDMs found in the data. Figure 7 and Table 5 show that they are almost distributed equally in the four BAWE divisions.

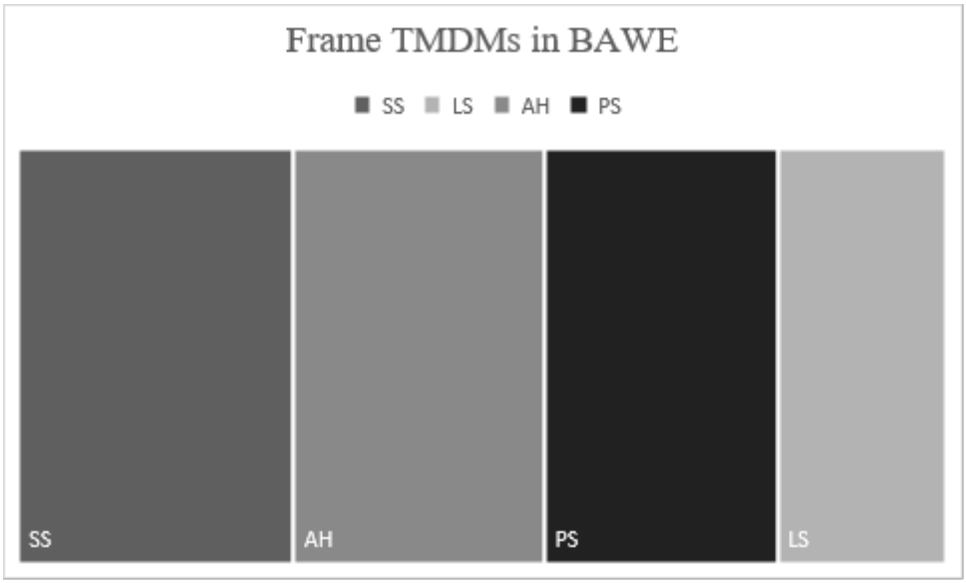

Figure 7. The distribution of frame TMDMs in BAWE

Table 5. Descriptive statistics of frame TMDMs across BAWE divisions

Frame TMDMs-Sequencing				
Division	Count	Sum	Average	Variance
SS	23	6,797	295.5217	186,681.1
LS	23	4,770	207.3913	140,737.9
AH	23	7,147	310.7391	303,986.5
PS	23	7,277	316.3913	423,728.2
Frame TMDMs-Labelling stages				
Division	Count	Sum	Average	Variance
SS	17	1,948	114.5882	54,123.76
LS	17	1,067	62.76471	14,171.32
AH	17	1,570	92.35294	36,695.74
PS	17	1,358	79.88235	31,171.86
Frame TMDMs- Announcing Goals				
Division	Count	Sum	Average	Variance
17	10	3,854	385.4	69,312.04
17	10	1,660	166	27,413.78
17	10	2,431	243.1	29,294.77
17	10	1,455	145.5	17,356.72
Frame TMDMs- Shifting Topics				
Division	Count	Sum	Average	Variance
SS	7	3,864	552	892,874.3
LS	7	2,584	369.1429	510,771.5
AH	7	3,868	552.5714	1,051,834
PS	7	3,822	546	1,292,054

3.4 Endophoric Textual Metadiscourse Markers in BAWE

The occurrences of 14 endophoric TMDMs were searched and calculated. Table 3 shows that the occurrence of endophoric TMDMS comes out to 18,446 times, constituting 4.32% of the total number of TMDMs in the data. As illustrated in Figure 8 and Table 6, endophoric

markers occurred 7,483 times in physical sciences and only 3,177 times in arts and humanities.

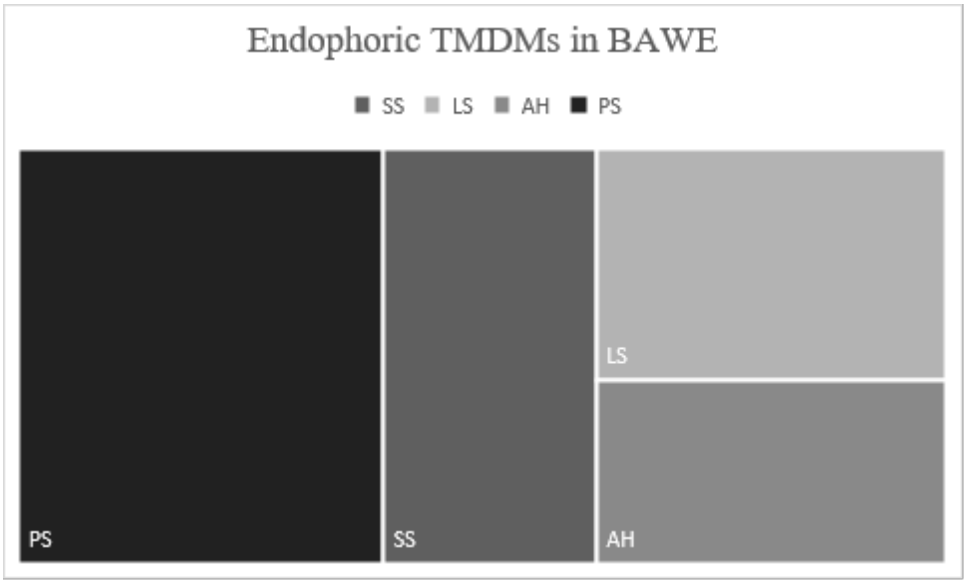

Figure 8. The distribution of endophoric TMDMs in BAWE

Table 6. Descriptive statistics of endophoric TMDMs across BAWE divisions

Division	Count	Sum	Average	Variance
SS	14	4,379	312.7857	120,390.3
LS	14	4,017	286.9286	160,144.5
AH	14	3,177	226.9286	52,611.61
PS	14	7,483	534.5	756,977.8

3.5 Evidential Metadiscourse Markers in BAWE

Three evidential TMDMs have been investigated in the studied BAWE disciplines. Evidential TMDMs occurred the least in the BAW data (n=4,326, 1.01%). The distribution of evidentials in the four BAWE divisions are illustrated below.

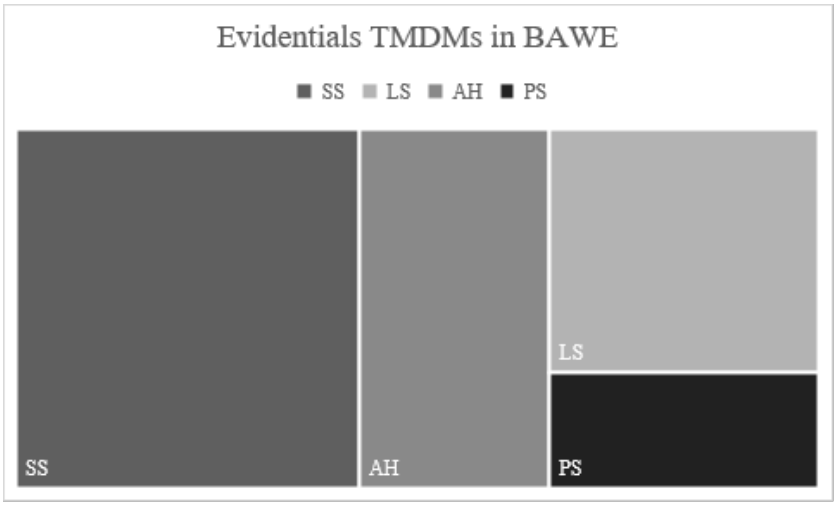

Figure 9. The distribution of evidential TMDMs in BAWE

Table 7. Descriptive statistics of evidential TMDMs across BAWE divisions

Division	Count	Sum	Average	Variance
SS	3	1,852	617.3333	187,922.3
LS	3	984	328	77,161
AH	3	1,026	342	57,232
PS	3	464	154.6667	37,830.33

3.6 Code Gloss Metadiscourse Markers in BAWE

Twenty-one code glosses were searched across the BAWE data, calculated, and tabulated. Code gloss occurred 20,960 times in the BAWE data, constituting 4.91% of the total number of TMDMs. The distribution of code gloss TMDMs in the four divisions are illustrated in Figure 10 and Table 8 below.

Figure 10. The distribution of code glosses TMDMs in BAWE

Table 8. Descriptive statistics of code glosses TMDMs across BAWE divisions

Division	Count	Sum	Average	Variance
SS	21	7,477	356.0476	225,466.7
LS	21	5,088	242.2857	157,096.8
AH	21	7,206	343.1429	251,695.5
PS	21	4,421	210.5238	68,290.86

4. Discussion

The present study used a cross-sectional design to conduct a cross-disciplinary investigation of TMDMs. The data under investigation was the four BAWE academic divisions, and to analyze these divisions, corpus quantitative tools were utilized. The investigation was limited to TMDMs are used by writers to "organize prepositional information in ways that a perceived audience is likely to find coherent and convincing" (Hyland, 2000, p. 112).The quantitative results obtained in the current study are to be interpreted in the light of Hyland's (2000, 2005) main findings. Below is a plausible interpretation of each of the main findings.

The most important finding in the current study is that MDMs are of significant value in providing a framework for facilitating communication, supporting a claim, maintaining reliability, and building a relationship with an audience. This is indicated by the high density of TMDMs in the data (n=4,269, 84 per million words). This is in line with Hyland's (2000, 2005) findings, where TMDMs accounted for 59.4% of the total MDMs in academic textbooks and 55.1% in RAs. The predominance of TMDMs in all genres investigated stresses the significance of guiding the reader through the text by expounding discourse organization, connections, and meanings. This overwhelming predominance emphasizes that MDMs are significant resources in guiding the reading process, showing readers where they are and where they should be going.

The comparative quantitative analysis displays a number of differences and similarities regarding the distribution of TMDMs across BAWE academic divisions. Results show considerable variations that account for the overall higher use of transitional TMDMs -conjunctions and adverbial phrases that guide the reader to recognize pragmatic associations between ideas by indicating additive, relatives, and contrastive associations—in all academic divisions. The use of transitional TMDMs accounts for almost three quarters of the TMDMs markers used (n=327,357, 76.76%). This is consistent with Hyland's (2000, 2005) findings where transitions account for 19.7% in RAs and are the second most used MDMs after hedges which constitutes 23.3% of the total. In academic textbooks, Hyland (2000) found that transitions account for 37.0% of the total MDMs.

The average and variance values in Table 4 indicate that the social sciences and arts and humanities exhibit greater use of transitional markers but with noticeable variations. The social sciences corpus was striking in its heavy use of transitional TMDMs, containing almost twice as many as the physical sciences corpus. This variation indicates that disciplinary differences seem to have a correlational influence on the use of transitions, with the physical sciences and life sciences being less concerned with guiding the reader to recognize

pragmatic associations between ideas. Because transitions are the resources used to guide the readers through the text, it is predictable that the soft disciplines of social sciences and arts and humanities would display the greatest number of transitions. In these soft sciences, the writer's exposition and claims count more than proofs of facts, figures, and statistics.

Second to transitions, the use of frame markers constitutes 13% of the total TMDMs in the data. Frame markers refer to text boundaries that indicate schematic text structuring; these include sequencing, labeling stages, announcing goals, and shifting topics. Unlike transitional TMDMs, no significant variations are found regarding frame markers across the BAWE divisions (see Figure 7 and Table 5). Students in all academic divisions demonstrate a considerable care when it comes to framing their assignments and texts. Thus, disciplinary differences have no influence on the use of frame markers across academic divisions. However, it is worth mentioning here that the frame marker "so" occurs the most in the data (n=10,362) with almost an equal distribution across the four academic divisions. The sequencing frame markers "first," "second," and "then" outnumbered the other frame markers.

Endophoric TMDMs constitute only 4.32% of the total number of TMDMs in the BAWE data. The analysis reveals a significant number of variations, with physical sciences exhibiting most endophoric TMDMs, followed by the social sciences. Unlike the case with transitional TMDMs, endophoric markers occur the least in the arts and humanities (see Figure 8 and Table 6). This also is in line with Hyland's (2000, 2005) findings where 85% of the endophoric markers identified in textbooks were found in scientific books. In dissertations, Hyland (2005) found that the hard sciences, such as electronic engineering, computer sciences, and biology, displayed the most endophoric markers per 10,000 words (43.1% in electronic engineering, 25.9% in computer sciences, 23.0% in biology). Endophoric refers to the markers used to point to tables, figures, and other sections in the text to relate visual concepts with verbal expositions. The high density of endophoric TMDMs in the physical sciences demonstrates their significance to scientific discourse, which combines scientific verbal discourse with visual representation of facts, such as tables, images, or graphs. Disciplines within the physical sciences, such as architecture, chemistry, computer science, electronic engineering, mathematics, and physics, convey scientific concepts and facts. They are, as Lemke (1998, p. 8, as cited in Hyland, 2000) pointed out, "semiotic hybrids, simultaneously and essentially verbal, mathematical, visual graphical, and action operational." It is the function of endophoric makers (such as above, below, this chapter, this figure, figure below, this part, in this section, etc.) to scaffold meaning through the discursive combination of text and visual graphical semiotics, making such discursive devices central to scientific discourse.

Evidentials account for only 1.01% of the total TMDMs occurring in the data. This is also in accordance with Hyland's (2000) investigation of academic textbooks (2.7%) and of RAs (10.7%). However, in his investigation of the genre of dissertations, evidentials accounted for 47.8% of the all metadiscourse markers. Evidentials are defined as "metalinguistics representations of an idea from another source" (Thomas & Hawes, 1994, p. 129). They are discursive resources by which writers establish an authorial voice, maintain credibility, and

show command of the related literature. Items such as according to, reported, quoted, and cited are used as integral citation strategies to establish credibility and integrity when locating text related to the literature. The low percentage of evidentials in the data might be justified by the fact that unlike RAs and dissertations, undergraduate texts and assignments do not require the same amount of reviews of the related literature. References to other related works are anticipated but do not come with lengthy arguments that locate the text into the context of the related literature as in the case of RAs and dissertations.

There is also a noticeable variation in the use of evidentials across the academic divisions. The social sciences have the greatest density of evidential TMDMs, followed by arts and humanities (See Figure 9 and Table 7), indicating a significant variation across disciplines regarding the use of endophoric TMDMs. In Hyland's work (2005), sociology constituted the largest amount of evidentials, followed by marketing. In academic textbooks, sociology displayed the greatest amount of evidentials at 4.9 per 1, 000 (16.2%), with mechanical engineering and electronic engineering (0.1 per 1,000 words) being the lowest. Evidentials play a significant role in locating the text within the wider disciplinary framework of knowledge. Apparently, this role is more evident in sociology and the humanities, which contain twice as many evidentials as in the physical sciences and life sciences. This may be because claims in the hard sciences are driven by facts and observations; thus, new discoveries emerge from an existing knowledge base. The scientific discourse in academic research is characterized by the cumulative knowledge of theories, facts, and technical lexicon, so the scientist will employ a highly standardized code of communication with the academic community (Kuhn, 1970; Bazerman, 1988). This is not the case with sociology and humanities, where knowledge is communicated in a less clear-cut manner. Establishing claims in soft disciplines requires greater attention to locating the research claim within the larger disciplinary context of knowledge by greater use of evidentials.

Code glosses account for 4.91% of the total number of TMDMs, with 20,960 instances distributed with a noticeable variation across the academic divisions (See Figure 10 and Table 8). Similar to the case of evidentials, the social sciences have the greatest use of code glosses (n=7,477 per million words, 30.91%) followed by the arts and humanities (n=7,206 per million words, 29.79%). The physical sciences displayed the lowest amount of code glosses (n=4,421 per million words, 18.27%). Code glosses are discursive resources by which writers can supply additional information or explain existing information. Examples of code glosses are as a matter of fact, for example, for instance, this is defined, this is called, known as, in other words, that means, such as, indeed, namely, and i.e. Writers usually rely on such discursive resources to relate to the reader's knowledge-base, thus ensuring greater clarity and readability. According to Hyland (2000, p. 118), they are used also to decrease the cognitive load of prepositional new knowledge. These discursive resources are not only used for clarifying meanings. Rather, they are used as instructional tools that instruct the reader while adding new information or providing a definition in making sure that readers comprehend knowledge, claims, and information the way the writer intended. The clear-cut nature of scientific discourse justifies the dearth of code glosses in the physical sciences, whereas the argumentative nature of discourse in the social sciences and arts and humanities, where

claims are framed through extensive utilization of discursive resources, explains the density of code glosses.

In summary, the results indicate the effect of discipline-related differences on the density and variations of TMDMs across BAWE academic writing. Related literature has shown that MDMs represent the social purposes of writers as social actions rather than a consequence of language items. Thus, academic writers seek to utilize MDMs that best evoke a social standardized communicative code among a particular audience. Thus, discipline-related differences have been shown as influencing the factor-determining types and patterns of MDMs across academic disciplines. In the current study, undergraduate texts and assignments were found to be consistent with prior scholarly research regarding how they demonstrate the standardized communicative code that is used among each academic community.

5. Concluding Remarks

The present study was particularly useful in lending support to Hyland's (2000) assumption of disciplinary discourse. It also added to the scarce literature on the investigation of MDMs across academic disciplines by showing the distribution of TMDMs in undergraduate's academic writing.

An awareness of the use of TMDMs across disciplines is strongly related to the practices of teaching and learning academic writing. The investigation of the use of TMDMs in academic writing leads both academic readers and academic writers in identifying the cognitive load that an academic text has and the best communicative discursive practices in communicating this load. This awareness-improving process is best maintained through both direct and indirect teaching instructions through which authentic academic texts are presented to the learners. An awareness of the use of TMDMs across disciplines can also guide academic writers into participating in the related academic community and facilitating the process of gaining a " membership in the academic and professional discourse community" (Hyland, 1994, p.244).

Other benefits of using TMDMs proficiently include aiding in the persuasiveness, credibility, as well as coherence of their written texts. Learners of different academic disciplines should be strongly advised to analyze the occurrences of TMDMs in different genres of writing and in different contexts. This will help in aiding them to maintain organization of their texts as well as guidance of their readers.

The availability of a representative of undergraduate's academic writing in different disciplines provide practitioners in the field of ESL with a facility in raising this awareness of TMDMs, thus ensuring proficient use of them. With the corpus searching tools, learners can easily scan texts of a specific discipline and specific genre to identify the most frequent TMDMs used. This also raises the cognitive comprehension of the different relationships that are expressed between academic writers and their targeted audiences. Corpus searching tools, such as frequency, concordance, and sketching, facilitate examining them into their authentic contexts providing good examples to follow.

Acknowledgments

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 (formerly of the Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES000-23-0800).

References

Bazerman, C. (1984). The writing of scientific non-fiction. *Pre/Text*, 5, 4-23.

Bazerman, C. (1993). Foreword. In N. Blyler & C. Thralls (Eds.), *Professional communication: the social perspectives*. (pp. ix-x). Newbury Park, CA: Sage.

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

Biber, D., & Conrad, S. (2009). *Register, Genre and Style. Cambridge*: Cambridge University Press.

Bunton, D. (1998*). Linguistic and textual problems* (Doctoral dissertation). University of Hong Kong, Hong Kong. Retrieved from http://hup.hku.hk/handle/10722/39580

Bunton, D. (1999). The use of higher level metatext in PhD theses. *English for Specific Purposes, 18*, 41-56.

Cao, F., & Hu, G. (2014). Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics, 66*, 15-31.

Crismore, A. (1989). *Talking with readers: metadiscourse as rhetorical act*. New York: Peter Lang.

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

Halliday, M. (1973). *Explorations in the functions of language*. London: Edward Arnold.

Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics, 30*, 437-455.

Hyland, K. (2000). *Disciplinary discourse.* University of Michigan, Michigan.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Hyland, K. (2009). Corpus informed discourse analysis. In M. Charles, S. Hunston, & D. Pecorari (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 110-128). London: Continuum.

Hyland, K., & Tse, P. (2004). Metadiscouse in academic writing: A reappraisal. *Applied Linguistics, 25*(2), 156-157. https://doi.org/10.1093/applin/25.2.156

Kawase, T. (2015). Metadiscourse in the introductions of PhD thesis and research articles. *Journal of English for Academic Purposes, 20,* 114-124.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. Martin, & R. Veel (Eds.), *Reading science: Critical and functional perspectives on discourses of science* (pp. 87-113). New York, NY: Routledge.

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education.* Cambridge, MA: Cambridge University Press.

Nesi, H., & Thomson, P. (2014). *Using Sketch Engine with BAWE (Version III).* Retrieved from http://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/search-the-bawe-corpus/

Norrick, N. R. (2001). Discourse markers in oral narratives. *Journal of Pragmatics, 33,* 389-78.

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

Rashidi, N., & Alihosseini, F. (2012). A contrastive study of metadiscourse markers in research articles abstracts across disciplines. *Philology and Cultural Studies, 5*(54).

Römer, U. (2012). Corpora and teaching academic writing: Exploring the pedagogical potential of MICUSP. In J. Thomas, & A. Boulton (Eds.) *Input, process and product: Developments in teaching and language corpora* (pp. 70-82). Brno: Masaryk University Press.

Schifrin, D. (1980). Metatalk: Organizational and evaluative brackets in discourse. *Sociological Inquiry: Language and Social Interaction, 50,* 199-239.

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

Thomas, S., & Hawes, T. (1994). Reporting verbs in medical journal articles. *English for Specific Purposes, 13,* 129-148.

Vande Kopple, W. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication, 36,* 82-93.

Appendix A

BAWE corpus summary

Table 1. BAWE corpus summary (Source: Nesi & Gardener, 2012, p. 8)

		Level 1	Level 2	Level 3	Level 4	Total
Arts and Humanities (AH) Archeology; Classics; Comparative American studies; English; History; Linguistics/English Language studies; Philosophy; others	assignments	239	228	160	78	705
	texts	225	229	160	80	724
	words	468,353	583,617	427,942	234,206	1.714,118
Life Sciences (LS) Agriculture; Biological science; Food science; Health; Medicine; Psychology	assignments	180	193	113	197	683
	texts	188	206	120	205	719
	words	299,370	408,070	263,668	441,283	1.412,391
Physical Sciences (PS) Architecture; Chemistry; Computer science; Cybernetics/ Electronic engineering; Mathematics; Meteorology; Physics; Planning	assignments	181	149	156	110	596
	texts	181	154	156	133	624
	words	299,370	314,331	426,431	339,605	1,381356
Social Sciences (SS) Anthropology; Business; Economics; Hospitality, leisure, and tourism; Management; Law; Politics; Publishing; Sociology	assignments	207	197	166	207	777
	texts	216	198	170	207	791
	words	371,473	475,668	447,950	704,039	1,999,130
Total students		333	302	235	169	1039
Total assignments		807	767	595	592	2761
Total texts		840	787	606	625	2858
Total words		1,440,185	1,781,686	1,565,991	1,719,133	6,506,995