CORPUS ANALYSIS OF METADISCOURSE IN UNDERGRADU-ATE ACADEMIC PROJECTS

Amaal Fadhlini Mohamed
Noorzan Mohd Noor

Alam amaalfadhlini@gmail.com
dnoorzan@salam.uitm.edu.my

Abstract:

Writers engage different types of metadiscourse markers in interacting with the readers. They provide indicators in their writing of the contents to help readers comprehend and respond to the text. This paper is a preliminary study to identify types and categories of metadiscourse found in a corpus of undergraduate academic projects (UAP corpus). This study is significant as it identifies types and categories of metadiscourse found in both good and weak undergraduate academic writing. Hence, this study is relevant to further support related research on the use of metadiscourse among tertiary level students to write effectively in academic writing. Since metadiscourse has never been directly taught as a subject to undergraduate students, it may contribute to the lack of awareness on metadiscourse functions in effective writing. In this study, the metadiscourse items in UAP corpus are identified and rated by inter-raters. It is a corpus-based research study that involves collecting, analyzing, and using qualitative and quantitative approaches to identify metadiscourse items in the corpus.

Keywords: corpus analysis, metadiscourse markers, academic project

INTRODUCTION

Metadiscourse is extensively cited as "discourse about discoursing" (Vande Kopple, 1985, p.83; Beauvais, 1989, p.11) or "writing about writing" (Williams, 1985, p.226) which refers to the writer's or speaker's linguistic expression to interact with the readers or listeners.

Over time, the definition of metadiscourse keeps changing and improved upon as the number of researchers increases. According to Thompson (2003), metadiscourse is linguistic expressions in a text which explains on the text itself instead of its propositional content. This definition seems very broad and needs further explanation. Later, Hyland and Tse (2004), Hyland (2004, 2005) came out with a more detailed definition which has been widely cited. They viewed metadiscourse as a "social and communicative process" between writers and readers that writing is a social and communicative process between writers and readers (p.14).

On the whole, metadiscourse is a way of interaction between writers/speakers to readers/listeners and writers to themselves which is not a part of propositional content or idea mentioned in the text (but sometimes it depends on the speaker's meaning) in order to deliver and organize contents or messages effectively. Hence, content is the biggest portion in evaluating an academic writing while language aspect comes second and organization aspect becomes the last.

Literature Review

Jalilifar and Alipour (2008) found that metadiscourse markers organize content and ideas by using connectives and build an interaction between the readers and the writers to become reader-friendly text. The same idea has also been suggested by Vande Kopple (1985). In addition, Hyland (1998) suggested that metadiscourse as a part of text clearly arranges the discourse (content), engages the audience and gives clue on the writer's attitude (on the content). It shows how writers develop themselves into the discussion to in-dicate their attitude toward the contents and the readers (Hyland and Tse, 2004).

According to Kumpf (2000, p.401), writers provide "cues and indicators" in their writing while arranging the con-tents to help readers comprehend and respond to the text. In fact, "cues and indicators" here portray what Hyland (2004, p.142; 2005, p.50) refers as "metadiscourse markers". Although more than 300 markers can be found in a list of metadiscourse items listed by Hyland (2005), there are no exact markers in identifying metadiscourse in a sentence. A possible marker can be a metadiscourse marker in a sentence, but might become a propositional content in another sentence. It depends very much on the semantics of the idea that a particular writer or speaker is trying to deliver.

In most cases, different writers have different approaches in structuring the sentences and putting the words to-gether to convey their intentions to the readers. This means writers interact with their readers in their writing based on their individual style. This is because metadiscourse itself refers to the ways writers or speakers project themselves in the texts to communicate with the readers (Amiryousefi and Rasekh, 2010) and there are no stan-dards of using metadiscourse in writing.

According to Intraprawat and Steffensen (1995), based on students' perspective,the audience is often an examiner who is always assumed as a stickler for grammar and technicalities. They believed that the audience is not some-one primarily interested in the ideas in the text or the development of the essay, and definitely not someone to be engaged in discussion of the text. The students will use better writing styles to interact with their audience every time. They will attempt to give clearer ideas, and pay some attention to whether their potential readers un-derstand their messages (Intraprawat and Steffensen,1995). It is very true that the ability to write effectively has become increasingly significant in the global society, and instruction in writing is now an important field in English as a Second Language (L2). Writers engaged various approaches of writing in delivering the messages.

There are a number of different ways which the features in metadiscourse have been categorized. These were the main theories and concepts that have influenced this research. In this study, the researcher uses Hyland's inter-personal model (2004), as it provides the most comprehensive categories of metadiscourse features and items. The metadiscourse markers are classified into two main categories (interactive and interactional metadiscourse) and sub- categories such as transitions, frame markers, endophoric markers, evidential, code glosses, hedges, boosters, attitude markers, engagements markers and self-mentions.

As Hyland (2004) came out with his classification scheme of metadiscourse features, it is considered as an improved taxonomy version which has been aspired by Vande Kopple's (1985) classification system of metadiscourse and Crismore at. al's (1993) categorizations of metadiscourse. It is believed to be the most comprehensive where meta-discourse is divided to two new terms of category (interactive and interactional metadis-course).

Hyland (1998) improved Crismore et al's taxonomy (1993) by eliminating its formal categories and minimizing the overlapped functions. The interpretive markers category was deleted and the main categories (textual metadiscourse and interpersonal metadiscousre) remained. According to Hyland (1998; p.442), textual metadiscourse markers in the textual metadiscourse category allow the revival of the writer's meaning by building up ideal interpretations of propositions in the text. Meanwhile, interpersonal metadiscourse tells readers about the author's perception towards both the propositional content and the readers themselves. The items in this category are contributing to the "writer-reader relationship and anticipating the subjective negotiability of statements" (Hyland, 1998: p.443 and Hyland and Tse, 2004: p.168).

Corpus

The effective categories of metadiscourse provide researchers with a resourceful instrument for the identification of metadiscourse items in academic text (Hempel and Degand, 2008; Elena, 2009; Heng and Tan, 2010; Abdi et al., 2010). From time to time, more and more studies on metadiscourse in various disciplines are being conducted since the concept of metadiscourse has become one of the essential fields in English Language communication. The evolution of technology has helped the researchers in identifying metadiscourse especially by conducting a corpus-based study. The concordance software helps researchers to identify metadiscourse markers in a corpus faster, especially when the corpus is very large. Thus, these are reasons for using corpus analysis methods for this current study.

The analysis used to study metadiscourse has been highlighted in a variety of corpus-based research such as studies on textbooks (Hyland 1999), L1 and L2 student writing (Crismore et al., 1993; Heng and Tan, 2010), ESL students' writing (Intaraprawat and Steffensen, 1995), L2 postgraduate writing (Hyland, 2004; Lee, 2009), L2 writing (Kuteeva, 2011), slogans and headlines (Fuertes-Olivera, 2001), research articles abstracts (Gillaerts and Van de Velde,2010) and newspaper discourse (Da-fouz-Milne, 2008). Dafouz-Milne (2008) examined these metadiscourse markers in a study characterizing newspaper discourse and identifying the markers that have a more persua-sive function.

2 PURPOSE OF THE STUDY

Previous studies found that good essays have more metadiscourse than weak es-says (Intaraprawat and Steffensen, 1995; Heng and Tan, 2010; Jalilifar and Alipour, 2008). Unfortunately, many of such studies that have been carried out on metadis-course use in L2 writing were conducted in L1 English speaking countries and few stud-ies have examined metadiscourse use in L2 writing in L2 English speaking countries. So, there is a vital need to look at how Malaysian L2 English students use metadiscourse markers in their writing to deliver contents clearly and effectively. To fill the gap, this study investigated metadiscoursal used in Malaysian undergraduate academic projects. It is a pilot study to investigate how metadiscourse markers are utilized in undergraduate academic writing.

Research Questions

The research questions that guided this study are:

1)   What are the common metadiscourse markers used in undergraduate academic projects?

2)   What are the similarities and differences between the use of interactive metadiscourse and interactional metadiscourse in undergraduate academic projects?

3   METHODOLOGY

Sample

Firstly, a corpus of Malaysian undergraduate academic writing (UAP corpus) was col-lected. In the selection of the UAP corpus, writing samples were taken from final year local university students enrolled in a research writing course – Academic Project.

The samples collected are from undergraduate writers with good proficiency in English as they obtained A- and A for this paper, as determined by their respective lecturers. The students had also obtained Band 4 to 6 in the Malaysian University English Test (MUET).

The sample was built from five literature review sections of the academic projects with a total word count of 11,280 words. Only the literature review section was collected as a corpus because it contains the most number of meta-discourse items compared to other sections. This is because undergraduate writers combined and concluded the experts' ideas in the literature review section and organized them by using various metadiscourse markers.

Instrumentation

In analyzing the metadiscourse used in academic writing, Hyland (2005) came out with his classification scheme of metadiscourse as shown in Table 1. It is an improved taxonomy version which has been aspired by Vande Kop-ple's classification system of metadiscourse (1985) and Crismore at. al.'s categorizations of metadiscourse (1993). It is believed to be the most comprehensive and detailed taxonomy in which metadiscourse is divided into two new categories (interactive and interactional metadiscourse). This model was used as a guideline to analyze metadis-course markers found in this study.

The UAP corpus of 11, 280 words was manually analyzed sentence by sentence for identification of metadiscourse markers. For inter-rater reliability, two experienced readers from the field of English Language were involved in the metadiscourse identification.

Metadiscourse markers, features and expressions in the corpus were electronically searched using WordSmith, a basic text analysis and concordance programme. Only sentences that contained metadiscourse markers, features and expressions were thoroughly analyzed and classified into metadiscourse categories. For instance, the transi-tion marker 'and' is classified as a metadiscourse marker when it links two or more ideas/clauses in a sentence. Based on software analysis, occurrence per 11, 280 words of each metadiscourse item found was calculated to find the most common metadiscourse items/categories used in the UAP corpus.

Table 1: A model of metadiscourse in academic texts (Hyland, 2004)

Category	Function	Examples
Interactive resources:	Help to guide reader through the text	
Transitions	Express semantic relation between main clauses	In addition/ but/ thus/ and
Frame markers	Refer to discourse acts, sequence, or text stages	Finally/to conclude/my purpose here is to
Endophoric markers	Refer to information in other parts of the text	Noted above/ see Fig/ in section 2
Evidentials	Refer to source of information from other texts	According to x/ (Y, 1990)/ Z states
Code glosses	Help readers grasp functions of ideational material	Namely/ e.g./ such as/ in other words
Interactional resources:	Involve the reader in the argument	
Hedges	Withhold writer's full commitment to proposition	Might/ perhaps/ possible/ about
Boosters	Emphasize force or writer's certainty in proposition	In fact/ definitely/ it is clear that
Attitude markers	Express writer's attitude to proposition	Unfortunately/ I agree/ surprisingly
Engagement markers	Explicitly refer to or build relationship with reader	Consider/ note that/ you can see that
Self-mentions	Explicit reference to author(s)	I/ we/ my/ our

RESULTS AND DISCUSSION

In sum, the frequency of two main categories of metadiscourse in the UAP corpus was electronically and manually counted. It was found that the total hits of interactive metadiscourse items are 815 while the total hits of interactional metadiscourse items are

330. The frequency of use of interactive and interactional metadiscourse is as displayed below.

Table 2: Frequency of use of interactive and interactional metadiscourse UAP corpus

Table 2: Frequency of use of interactive and interactional metadiscourse UAP corpus
(Total words: 11, 280 words)

Metadiscourse Categories	Total Hits	Occurrence per 10, 000 words	% of Total metadiscourse
Interactive	815	723	71
Interactional	330	293	29

The table shows that interactive metadiscourse has a higher frequency of use (722.5 occurrences per 10, 000 words) as compared to interactional metadiscourse (292.5 occurrences per 10, 000

words). The percentage of difference between the metadiscourse categories is spectacularly huge. At this point, the finding is similar to a few other metadiscourse studies such as Intaraprawat and Steffensen (1995), Hyland (2004), Hyland and Tse (2004), Jalilifar and Alipour (2008) and Heng and Tan, (2010) that showed the use of interactive metadiscourse is higher than interactional metadiscourse. That is to say that in-teractive metadiscourse is often used throughout the text because it helps the readers to understand through transition markers, frame markers, endophoric markers, evidentials and code glosses.

In contrast, the frequency of interactional metadiscourse in the corpus is lower

because it includes ways the writer interacts with the readers by attracting them into the discussion in the text. In this case, the writer would use hedges, boosters, engagement markers, attitude markers and self-mention to attract the readers' participation while reading. By the same token, the use of interactional metadiscourse depends on the writers' writing skill. In a study of metadiscourse in a corpus of leading journal articles from various fields conducted by Hyland (1998), the writers use logical connectives, frame markers, endophoric markers, evidential and code glosses (55.1% of total metadiscourse) and hedges, emphatics, relational markers, attitude markers and person markers (44.9% of total metadiscourse). The percentage gap is rather small. This means that experienced writers write with more interactional metadiscourse than inexperienced writers, especially undergraduates.

The next stage was to analyze the frequency count of metadiscourse use in the UAP corpus based on each sub-cat-egory of metadiscourse. The table below displays the analysis of interactive metadiscourse found.

Table 3: Number of occurrences of the categories of interactive metadiscourse in the UAP Corpus

Metadiscourse Category	Total Hits	UAP Corpus (Total Words: 11, 280)		Examples in students' sentences
		Occurrence per 10, 000 words	% of Total metadiscourse	
1. Interactive				
Evidentials	312	276.6	38.3	According to X e.g.: According to Dennis and Kinney (1998), media richness argues the cotions under when under when each media will be most effective.
Transitions	250	221.6	30.7	Also e.g.: Media richness is also based on language variety.
Frame Markers	173	153.4	21.2	Then| e.g.: The chief focus of this approach, then, has been to show how patterns of interaction between men and women reflect the dominant position of men in society (Lakoff's1975).

Code Glosses	78	69.1	9.5	Such as e.g.: For example, Plumtree Software 2000), defines its corporate portal as a system which is able to bring together in one simple, personalized web page, all the information and productivity tools elevant to corporate users, hosting dynam applications, such as on line reports, e-mail, schedules, calendars, and business services.
Endophoric Markers	2	1.8	0.2	This chapter e.g.: This chapter reviews the studies that have been done by gender experts on conversational styles.
Total		722.5	100.0	

The table shows that evidentials (e.g.: according to, cited, quoted) has the highest frequency count of its use in the UAP corpus constituting almost half of the total metadiscourse (38.3%). Since this corpus consists of the literature review section of the academic projects, the undergraduate writers used evidentials the most in this section as in-text citations to avoid plagiarism. Transitions (e.g.: also, because, but) are recorded with a high percentage which is30.7% with 221.6 occurrences per 10, 000 words. This is similar to the other studies of the same kind (Intaraprawat and Steffensen, 1995; Hyland, 2004; Hyland and Tse, 2004; Hempel and Degand, 2008 and Heng and Tan, 2010) which show marginally high percentages of transitions use. The third highest frequency use is frame markers (e.g.: focus, then, first) which constitute more than 20% of the total metadiscourse (153.4 occurences per 10, 000 words).

On the other hand, code glosses and endophoric markers register low frequency of use. Code glosses (e.g.: such as, for example, defined as) account for only 1.8 occurences per 10,000 words with less than 10% of the total metadiscourse. Meanwhile, endophoric markers (e.g.: this chapter) show the lowest frequency of use with less than 1% of the total metadiscourse. As endophoric markers are meant to refer to a particular section in a text, writers do not use these markers as much as other sub-categories.

Table 4: Number of occurrences of interactional metadiscourse categories in the UAP Corpus

Metadiscourse Category	Total Hits	UAP Corpus (Total Words: 11, 280)		Examples in students' sentences
		Occurrence per 10, 000 words	% of Total metadiscourse	
1. Interactive				
Evidentials	312	276.6	38.3	According to X e.g.: According to Dennis and Kinney (1998), media richness argues the cotions under when under when each media will be most effective.
Transitions	250	221.6	30.7	Also e.g.: Media richness is also based on language variety.
Frame Markers	173	153.4	21.2	Then e.g.: The chief focus of this approach, then, has been to show how patterns of interaction between men and women reflect the dominant position of men in society (Lakoff's1975).
Code Glosses	78	69.1	9.5	Such as e.g.: For example, Plumtree Software 2000), defines its corporate portal as a system which is able to bring together in one simple, personalized web page, all the information and productivity tools elevant to corporate users, hosting dynam applications, such as on line reports, e-mail, schedules, calendars, and business services.
Endophoric Markers	2	1.8	0.2	This chapter e.g.: This chapter reviews the studies that have been done by gender experts on conversational styles.
Total		722.5	100.0	

The above table shows that hedges (e.g.: about, may, tend to) has the highest frequency count of its use in the UAP corpus with more than half of the total metadiscourse, 53.5% with 157 occurences per 10, 000 words. This is similar to a few previous studies of the same kind such as Intaraprawat and Steffensen (1995), Hyland (1998), Hyland and Tse (2004), and Heng and Tan (2010) which recorded hedges as having the highest frequency of use in their analyzed corpora. Boosters (e.g.: found, certain, believe) are recorded at a high percentage which is 30.4% with 89 occurrences per 10, 000 words. In contrast, attitude markers and engagement markers register low frequency of use. Attitude markers (e.g.: important/ly, even X, expected) account for 41 occurences per 10, 000 words with 14% of total metadiscourse, while engagement markers (e.g.: refer to, note that) show the lowest frequency of use, constituting only 2.1% of the total metadiscourse. Self-mention (e.g.: I, we, our, us) is hardly found in this corpus of academic projects (UAP corpus) because academic writing uses pronouns the least. The writer would use 'the researcher/s' instead of using the pronoun 'I/we'. Most of these undergraduate writers were literally reporting the ideas and theories from experts in this section, so self-mention is hardly found in this UAP corpus.

Analysis of Each Metadiscourse Sub-category

The most common five metadiscourse markers in the UAP Corpus were recorded and tabulated based on each main category and sub-category (refer to Table 5 below).

Table 5: List of the most preferred metadiscourse items found in the UAP corpus.

Metadiscourse Category	Total Hits	UAP Corpus (Total Words: 11, 280) Occurrence per 10, 000 words	% of Total metadiscourse	Examples in students' sentences
2. Interactional Hedges	177	157.0	53.5	About e.g.: A study about media choice between Japanese and American workers shows that the concept of low context and high context culture does influence their choice of media.
Boosters	100	89.0	30.4	Found e.g.: These gender experts also found that men tend to be "competitive and dominating" in their conversation.
Attitude Markers	46	41.0	14.0	Important e.g.: It is important for the information to be distributed to the right person as failing to do that will cause failure in a certain work
Engagement Markers	7	6.2	2.10	Note that e.g.: In their study, Shah et al. (2002) stated that it is important to note that the Internet use may not shape individuals' social behaviours regardless towards better or worse.
Self Mention	0	0	0	-
Total		293.2	100.0	

Since this research has been conducted in an academic setting, the results of the study present information of how metadiscourse is used among Malaysian undergraduates in final year academic projects. It shows that undergraduates use more interactive metadiscourse because it helps the readers to understand through the application of transition markers, frame markers, endophoric markers, evidentials and code glosses. In fact, the use of transition markers (but, because, also) and frame markers (first, then, finally) is the easiest way to organize sentences and ideas. .

In comparison, the use of interactional metadiscourse in the corpus is lower because it includes ways the writer interacts with the readers by attracting them into the discussion in the text. In this case, these inexperienced writers (undergraduates) would use fewer hedges (claim, could, tend to), boosters (certain, believe, found), engagement markers (note that, refer), attitude markers (expected, important, usually) and self-mention (I, we, us) to attract the readers' participation while reading. As such, the use of interactional metadiscourse depends on the writers' writing skill which is what most undergraduates are lacking in.

Since the Malaysian government is now implementing a plan to enhance proficiency in English among public university students as a preparation for their working world, this pilot study might be one of ways that researchers can use corpus analysis to start investigating the need for metadiscourse in good writing among undergraduates. Finally, it is hoped that this study would encourage further related research to show the importance of knowledge on metadiscourse among tertiary level students in order to write effectively. Other researchers who are interested in metadiscourse analysis can follow up on this to examine the use of metadiscourse across different cultures and disciplines.

REFERENCES

Amiryousefi,M. & Rasekh, M. (2010). Metadiscourse: Definitions, Issues and Its Implications for English Teachers. 3(4).

Beauvais, P. J. (1989). A speech act theory of metadiscourse. Written Communication, 6 (1), 11-30.

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

Dafouz-Milne, E. (2008). The pragmatic role of textual and interpersonal metadiscourse mark ers in the construction and attainment of persuasion: A cross-linguistic study of news paper discourse. Journal of Pragmatics, 40, 95-113.

Fuertes-Olivera, P. A., Velasco-Sacristan, M., Arribas-Bano, A. & Samaniego-Fernandez, E. (2001). Persuasion and Advertising English: Metadiscourse in slogans and headlines. Journal of Pragmatics, 33, 1291-1307.

Hempel, S. & Degand, L. (2008). Sequencers in different text genres: Academic writing, journalese and fiction. Journal of Pragmatics, 40, 676-693.

Heng, C. H. & Tan, H., (2010). Extracting and comparing the intricacies of metadiscourse of two written persuasive corpora. International Journal of Education and Developement using Information and Communication Technology (IJEDICT), 6 (3), 124-146.

Hyland,K. (2005). Metadiscourse: Exploring interaction in writing, continuum guides to discourse. Continuum: London

Hyland, K. (2004). Disciplinary interactions: metadiscourse in L2 postgraduate writing. Journal of Second Language Writing, 13, 133-151.

Hyland, K. & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. Applied Linguistics, 25(2), 156-177.

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

Ifantidou, E. (2004). The semantics and pragmatics of metadiscourse. Journal of Pragmatics, (37) 1325-1353

Intraprawat, P. & Steffensen, M.S. (1995). The use of metadiscourse in good and poor ESL essays. Journal of Second Language Writing, 4(3), 253-272.

Jalilifar, A. & Alipour, M. (2008). How explicit instruction makes a difference: Metadiscourse markers and EFL learners' reading comprehension skill.

Journal of College Reading and Learning, 38(1). Kumpf, E. P. (2000). Visual metadiscourse: Designing the considerate text. Technical communication quarterly, 9(4), 401-424.

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

Williams, J. M. (1985). Style: Ten lessons in clarity and grace (2nd ed.). Glenview: Scott Foresman