



Characterizing the Metadiscourse of the Pure Math PhD Thesis

Lee Ming Cherk *National University of Singapore*
elclmc@nus.edu.sg

Abstract

Research on postgraduate writing abound, but those that report specifically on thesis writing in the mathematics discipline are rare. The purpose of this research is to systematically study the metadiscourse patterns of PhD theses in the pure math discipline. To this end, six PhD theses in pure mathematics were selected for this study. Investigation was carried out through manual coding and frequency counting using Hyland's metadiscourse model (2005) as the framework of analysis. In addition, materials written about mathematical writing authored by mathematicians themselves were consulted to corroborate and explain the findings. The study reveals that in the theses, great emphasis is placed on using interactive elements to guide readers through logical reasoning. At the same time, there is also a noticeable tendency to use engagement markers to actively involve readers as participants in the epistemological process, dispelling the impression that mathematical writing is highly impersonal.

Key Words: metadiscourse, thesis writing, academic language, mathematical writing language learning, translation, SLA, cognitive linguistics, Relevance

1. Introduction

Scholarly inquiry into research writing is often anchored on the traditional IMRD (Introduction, Method, Results and Discussion) model. While this framework has served as a good starting point for teaching students to organize and apply rhetorical features, in reality, there are disciplinary variations which are already dictated by the epistemological process and expectations of individual academic communities they have to adhere to (Lin and Evans 2011; Paltridge 2002; Posteguillo, 1999; and Yang and Allison, 2004). In addition, the IMRD structure is suited to only empirical research and not to theoretical research (Posteguillo, 1999; Swales, 2004; Yang and Allison, 2004). As such, pure mathematics, given its theoretical nature, cannot be appropriately mapped onto the IMRD structure. As well as that, mathematical writing is very different from other disciplines because of its terse style of writing, in addition to its interweaving of mathematical reasoning and meta-mathematical discussion (Steenrod, 1973), which O'Halloran (2000) describes as a combination of "semiotic resources" consisting of "mathematical symbolism, visual display in the form of graphs and diagrams, and language" (p.360).

Given such distinct features in mathematical writing, this area of research is less well-explored than other disciplines. More recent works on the topic include Kuteeva and McGrath's (2015) study of pure math as a case study on disciplinary practices and discourse and McGrath and Kuteeva's (2012) analysis of stance and engagement in pure mathematics research articles. In the former (Kuteeva and McGrath, 2015), the authors investigated the organizational and rhetorical structures of research articles in the pure math discipline and found that there were considerable variations within the discipline itself. In the latter (McGrath and Kuteeva (2012), the authors applied Hyland's (2005b) Stance and Engagement model to study how the social and disciplinary practices of the pure mathematics community shaped writers' rhetorical choices in research articles. Although the information gleaned from these articles is useful, it may not be directly applicable to PhD theses

since differences in genre features between research articles and PhD theses do exist (Swales, 2004; Kawase, 2015). PhD theses are much lengthier and therefore allow metadiscourse to unfold more extensively. Also, the purpose of writing and readership for both genres are somewhat different. While the PhD thesis serves an educational function and it is written for a highly select and expert audience (i.e. the supervisor and examiner(s)), the research article is written for publication and therefore appeals to a wider audience in the academic community.

With the aim of filling the knowledge gap in PhD mathematical writing, this study sought to answer the following research questions:

1. What do mathematicians say about mathematical writing?
2. What metadiscoursal resources have been employed to navigate readers through mathematical reasoning in PhD theses and allow the writer to express his views in a manner that is acceptable in the academic community?
3. In what way is metadiscourse in PhD pure math theses different from that in other disciplines?

The ultimate goal of this study is to inform the teaching of academic writing to PhD students in the pure mathematics discipline.

2. Method

To answer Question 1:

“What do mathematicians themselves say about mathematical writing?”

Textbooks and articles written on mathematical writing by mathematicians themselves were consulted. From there, we could determine, from the mathematician’s perspective, what constitutes “good mathematical writing”. The information was then used to explain and corroborate the data obtained from the coding and frequency counting of metadiscourse features.

To answer question 2 of this study,

“What metadiscoursal resources have been employed to navigate readers through mathematical reasoning in PhD theses and allow the writer to express his views in a manner that is acceptable in the academic community?”

A total of 6 PhD theses in pure mathematics, which were submitted to the National University of Singapore (NUS) between 2013 and 2015, were retrieved from the NUS Scholar Bank, an online repository of research dissertations and papers. The texts were manually coded using Hyland’s (2005a) metadiscourse model which is subdivided into two main categories: the interactive and the interactional. The interactive category is concerned with how the writer navigates readers through the text so that they are made aware of the writer’s intentions and arguments, whereas the interactional category is concerned with the way the writer brings his authorial “voice” across while actively engaging readers and involving them in the unfolding text.

Manual coding was followed by frequency counts which were then computed into percentages.

Table 1 below shows the main and sub-categories of metadiscourse and descriptions of their functions as given in Hyland (2005a). The examples, however, are taken from our present research.

Category	Function*	Examples
Interactive	Help to guide the reader through the text	
transitions	Express relations between the main clauses	Hence; if and only if; moreover; then; since; given that X is..., then y.... ; thus; therefore; however; moreover; to this end; In addition; first, next, then; similarly; on the other hand
Frame markers	Refer to discourse acts; sequences or stages	In this part, we will prove that... ; Now we are ready to combine the respective estimates for each of the five factors; The next section contains some applications and some miscellaneous remarks.
Endophoric markers	Refer to information in other parts of the text	Recall equation X; see figure Y; We looked at some applications of the theory developed in the last section; Recall in chapter 3, we studied...
Evidentials	Refer to information on other texts	A precise lower bound is obtained by P. Yang and M. Zhu [28] using a symmetrization technique. As in [8], [11] and [25], we consider the following set of...
Code gloss	Elaborate propositional meaning	In other words; In fact; which implies ... such that... ; that is; equation (3.6) immediately implies...; We deduce from (4.10) that there exists a sequence...
Interactional	Involve the reader in the text	
Hedges	Withhold commitment and open dialogue	One <i>may</i> need to develop new a priori estimates
Boosters	Emphasize certainty or close dialogue	This is clearly ...; one can easily see that; indeed
Attitude markers	Express writer's attitude to proposition	The situation is trivial; what is important to us is that...; It is interesting to try a flow approach
Self-mentions	Explicit reference to author(s)	One finds; one sees; one obtains;
Engagement markers	Explicitly build relationship with reader	We are going to show that; we need to; let us assume; we have the following; observe that; let; suppose; define; write; check; fix; compute; identity; calculate; show; state; note

Table 1. Examples of Metadiscourse Features in Mathematical Writing Based on Hyland's (2005a) Metadiscourse Model

To answer Question 3:

“In what ways is metadiscourse in PhD pure math theses different from that in other disciplines?”

The results from Hyland and Tse’s (2004) investigation of metadiscourse in doctoral theses across six disciplines were used as a basis for comparison with our findings. These six disciplines were: applied linguistics, public administration, business studies, computer science, electronic engineering and biology. Since the presentation of our results was different from that of Hyland and Tse (2004), we derived a rank order from their results which could then be conveniently compared against the rank order which was derived in our present study.

The rank order derived from the frequency count (reported as number per 10,000 words) in Hyland and Tse (2004) is shown in Table 2 below:

Dimension	Category	Frequency (Per 10,000 words)	Ranking *
Interactive	Transitions	95.6	1
	Frame markers	76.2	4
	Endophoric markers	40.6	5
	Evidentials	30.3	2
	Code gloss	24.0	3
	(total)	266.7	
Interactional	Hedges	95.6	1
	Boosters	51.9	3
	Attitude markers	35.3	4
	Self-mentions	18.5	5
	Engagement markers	40.2	2
	(total)	241.5	

Table 2. Rank order of Metadiscourse categories in PhD theses Across 6 Disciplines Derived from Hyland and Tse (2004, p. 170)

* This column was not in Hyland and Tse’s (2004) original work. The ranking is derived to enable ease of comparison.

3. Findings

3.1 What mathematicians say about mathematical writing

A mathematician, Reiter (1995) describes the work of the mathematical writer thus:

A successful mathematical writer will lay out for her [sic] readers two logical maps, one which displays the connections between her [sic] own work and the wide world of mathematics and another which reveals the internal logical structure of her work.

Clearly, not only does the mathematical writer need to solve a problem but he or she should also ensure that his or her readers can see how the present work builds on the existing literature and can follow his or her own intricate logical reasoning. As such, mathematicians advocate mindfulness of the audience and clear and concise writing, so that readers may quickly grasp the gist of the paper, identify the main results, and understand how the arguments proceed (Krantz, 1997). Also, since readers are often less interested in the details of a proof than in the outline and the key idea (so that they could learn a technique or principle that can be applied in other situations), much attention should be given to emphasizing the structure of a proof, commenting on the ease or difficulty of each step and highlighting the key ideas that make the proof work (Higham, 1998).

On the whole, since mathematical writing strives for clarity and conciseness, the general rules for spelling, grammar and syntax, clear organization and academic style prevail here as they do in other disciplines.

3.2 Results of the Corpus Investigation

Table 3 shows the results of the present research based on the categories found in Hyland's (2005a) metadiscourse model. The first column shows the two main metadiscourse categories, i.e. the interactive and interactional dimensions, while the second column gives the various categories under each dimension. The third column gives the frequency of occurrences in the 6 PhD theses which were studied (expressed in percentages), and the fourth column shows the rank order of the categories in terms of their frequency of occurrences in the theses. The subtotal percentage of occurrences for each dimension is also given.

Dimension	Category	Average %	Ranking
Interactive	Transitions	34.23	1
	Frame markers	8.68	4
	Endophoric markers	4.178	6
	Evidentials	13.04	3
	Code gloss	7.64	5
	(Subtotal)	67.78	
Interactional	Hedges	1.03	8
	Boosters	0.94	9
	Attitude markers	2.04	7
	Self-mentions	0.81	10
	Engagement markers	27.4	2
	(Subtotal)	32.22	
	(Total)	100	

Table 3. Analysis of Corpus Based on Hyland's (2005a) Metadiscourse Model

From this table, it can be seen that collectively, the interactive dimension features more significantly (67.78%) than the interactional dimension (32.22%) in the six theses which were analyzed. However, in terms of rank order, engagement markers (27.4%), which comes under the interactional dimension and which explicitly focuses readers' attention or includes them as discourse participants in the text, ranks second only to transition markers (34.23%), which comes under the interactive dimension and which clearly spells out the pragmatic connections between stages in an argument. The next three most frequently occurring categories also come under the interactive dimension. Evidentials comes third in rank order (13.04%). Such elements are found in ideas from other sources and are used to guide readers' interpretation, affirm the author's command of the subject, and persuade readers to the author's views. Frame makers, which highlights the stages of an argument, is fourth in rank (8.68%). This is closely followed by code glosses, which is employed to rephrase, explain or elaborate on what has been said, and is fifth in rank (7.64%). Endopohric markers (4.178%), expressions which refers to other parts of the text, is sixth in rank. The other metadiscourse features under the interactional dimension appear much less frequently in the theses, and are therefore of lower ranks: attitude markers (ranked seventh: 2.04%), hedges (ranked eighth: 1.03%), boosters (ranked ninth: 0.94%) and self-mention (ranked tenth: 0.81).

3.3 Comparison of Metadiscourse Studies

We then compared the rank order of metadiscourse categories in our present study with the same categories derived from Hyland and Tse (2004).

This is presented in table 4 below, where the interactive categories and the interactional categories are compared separately.

Ranking for Interactive Categories	Hyland and Tse (2004)	Present study	Ranking for Interactional Categories	Hyland and Tse (2004)	Present study
1	Transitions	Transitions	1	Hedges	Engagement markers
2	Evidentials	Evidentials	2	Engagement markers	Attitude markers
3	Code glosses	Frame markers	3	Boosters	Hedges
4	Frame markers	Code glosses	4	Attitude markers	Boosters
5	Endorphics	Endorphics	5	Self-mention	Self-mention

Table 4. Rank Ordering of Interactive and Interactional Categories

3.3.1 Interactive Categories

Under the interactive domain, transitions are the most frequently found elements both in Hyland and Tse's study and in our present research. This may be unsurprising as doctoral research work involves examining and breaking information into smaller components, determining how these parts relate to each other, identifying causes and effects, making inferences and finding evidence to support generalizations. The types of transitions commonly used mirror these epistemological processes and they mainly serve these functions: combining ideas, highlighting implications, giving explanations (cause and effect, comparison), modifying or setting restrictions, setting conditions, enumerating, highlighting processes, and showing emphasis.

Next to transition, evidentials is ranked second in both Hyland and Tse's (2004) study and this present research. In our present study of mathematical writing in PhD theses, elements that are found to fall under the 'evidentials' category are mainly citations and attributions to works, as well as formulas and techniques put forward by scholars.

Citations are given either by name and year, by name and number, and by number only. Here are some examples:

Example of references given by name, year and number:

In 1983, Paneitz [22] discovered ...

Examples of references given by name and number:

Branson [2] found that

We adopt the approach from Chang-Gursky-Yang [7]

We follow the approach as in Section 7 of Chang-Gursky-Yang [7] and Section 5 of Wei-Xu [25]

Theorem 3.1 (McShane [29]).

Examples of references given by number only:

Theorem 1.1 of [31]

See for instance [9, 20, 31,42]

... related topics have continued to be widely studied from different viewpoints (cf. [10,11,12,20,21,23,24,25])

In the theses analysed, it was observed that there was a tendency to cite by name and year, particularly in the introduction, as it "give(s) an enlivening effect because of the human interest it introduces", as suggested by Higham (1998, p. 95).

In addition to citations, it is common for methods, formulas, equations, techniques, theories etc. to be attributed to individuals. For example:

By Taylor expansion, we have the following estimates

Now from the Kazdan-Warner identity, we have

We use the Lyapunov-Schimidt's reduction method to solve the equation

This thesis is mainly focused on identities motivated by McShane's identity.

The rank ordering of other metadiscourse categories under the interactive dimensions in Hyland and Tse (2004) and our study were only slightly different. In the case of Hyland's study, the rank order was code glosses (third), frame markers (fourth) and endorphoric (fifth), whereas in the present study, the rank order was frame markers (third), code glosses (fourth) and endorphoric (fifth).

From these comparisons, we could infer that despite the visual distinctiveness in mathematical writing, caused by the interweaving of symbols, equations, graphs and language, the interactive domain of metadiscourse in PhD mathematical writing is not very different from that of other disciplines.

3.3.2 Interactional Categories

While the way in which interactive metadiscourse is employed in the math theses is similar to that in other disciplines (as seen in Hyland and Tse (2004)), the use of interactional metadiscourse features is somewhat different. Whereas the categories in Hyland and Tse's study were ranked in this order: hedges (first), engagement markers (second), boosters (third), attitude markers (fourth) and self-mention (fifth), the categories in our present study were ranked thus: engagement markers (first), attitude markers (second), hedges (third), boosters (fourth), self-mention (fifth).

What emerges from this data is that engagement markers ranks highly, i.e. second in rank in Hyland and Tse (2004), and first in rank in our present study. A possible explanation for this is that mathematical writing is like a conversation taking place between two mathematicians strolling through the woods (Krantz, 1997). In this process, one mathematician communicates, persuades and convinces the other in a step-by-step manner and through logical reasoning (Lee, n.d). This epistemological process requires a high level of audience awareness (Fountain, 2009; Higham, 1998; Krantz, 1997; Lee, n.d.; Shelton, n.d.; University of New South Wales, 2014), and so "one also must have the confidence to know that I am writing for X, and not be 'writing primarily to convince himself that his theorem is correct.'" (Krantz, 1997, p.2). Therefore, engagement markers, which are used to involve readers as participants in the reasoning process, are frequently employed. The most used engagement markers are the pronoun "we" and directives (mainly imperatives) that guide readers to perform textual, physical and cognitive acts (Hyland, 2005b). Also, the custom is to use the first person pronoun "we" in modern mathematics (Krantz, 1997), possibly to create a sense of solidarity between the writer and his or her readers (McGrath and Kutteva, 2012).

Another point of note is that while hedges takes first place in the interactional metadiscourse in Hyland and Tse's (2004) study of PhD theses across six disciplines, it ranks only third in our present study. This could be due to disciplinary convention which dictates that mathematics has to be written with a high level of certainty. As hedging is used for withholding commitment and open dialogue (Hyland 2005a), it is used sparingly.

In addition, while attitude markers ranks quite low (i.e. fourth) in the Hyland and Tse's (2004) study, it ranks second in our present study. It is common practice in mathematics for remarks to be made about the aesthetic quality, the simplicity of an equation, or the ease in performing a calculation. According to McGrath and Kutteva (2012), attitude markers can also contribute to engagement as it "allows the author to briefly connect with the reader, without having to withdraw from the main business of the article."

4. Discussion and Conclusion

Discipline variation in use of metadiscourse is to be expected as writers need to conform to the expectations of the academic community. In our present study, we have found that there is more frequent use of the interactive metadiscourse categories than the interactional metadiscourse categories, while the frequency of using interactive and interactional metadiscourse in the theses reported in Hyland and Tse (2004) is about the same. Our finding corroborates with readings about mathematical writing which emphasize the need to highlight the structure of a proof and the key ideas that make the proof work so that the reasoning can be easily understood - a need that is addressed through the use of interactive metadiscourse strategies.

Also, we could conclude from the results of the interactional metadiscourse categories that, contrary to the impression that math writing is impersonal, it actually attempts to engage readers through various features. For example, the greater use of name-year citations and attribution of work, formulas and techniques etc. to individuals gives the works a “human touch”. The common use of the pronoun “we”, directives given in the imperatives, and attitude markers all point to a high level of audience awareness and an eagerness to involve readers in the epistemological process. This finding is consistent with the advice given in readings that audience awareness and engagement of readers are important.

Taking such features into account, the language teacher could highlight the use of transitions and other metadiscourse strategies to engage the reader, particularly in the introduction chapter, and in subsequent introductory and concluding segments of each chapter where the text contains more language than symbols and equations.

While this study has demonstrated the unique way in which metadiscourse is employed in texts where language weaves through symbols, equations and diagrams, the scale of this work is small. Future studies could build on it and employ concordancing software so that a much larger quantity of texts can be analysed and the results may be generalizable. The information can then inform curriculum planning and materials preparation for a postgraduate writing course.

References

- Fountain, J. (2009). *Notes on Writing a Thesis*. Department of Mathematics, University of York. <http://maths.york.ac.uk/www/ResearchWriting>
- Guidelines for Writing a Thesis* (2014). Department of Mathematics and Statistics. University of New South Wales. <https://www.maths.unsw.edu.au/currentstudents/guidelines-writing-thesis>
- Higham, N. J. (1998). *Handbook of Writing for the Mathematical Sciences*. Philadelphia, PA: SIAM (Society for Industrial and Applied Mathematics).
- Hyland, K. (2005a). *Metadiscourse*. London: Continuum.
- Hyland, K. (2005b). Stance and engagement: a model of interaction in academic discourse. *Discourse studies*. 7(2), 173-192.
- Hyland, K., & Tse, P. (2004). Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics*. 25(2), 156-177.
- Kawase, T. (2015). Metadiscourse in the introductions of PhD theses and research articles. *Journal of English for Academic Purposes*. 20(2015), 114-124.
- Krantz, S.G. (1997). *A Primer of Mathematical Writing*. Providence, RI: American Mathematical Society.
- Kuteeva, M., & MacGrath, L. (2015). The theoretical research article as a reflection of disciplinary practices: the case of pure mathematics. *Applied Linguistics*, 36 (2), 215-213.
- Kwan, B.S.C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, 25, 30-55.
- Lee, K. P. (n.d). *A Guide to Writing Mathematics*. University Of California, Davis. <http://web.cs.ucdavis.edu/~amenta/w10/writingman.pdf>
- Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross disciplinary study. *English for Specific Purposes*, 31(3), 150-160.
- MacGrath, L., & M. Kuteeva (2012). Stance and engagement in pure mathematics research articles: linking discourse features to disciplinary practices. *English for Specific Purposes*, 31 (3), 161-173.
- O' Halloran, K. (2000). Classroom Discourse in Mathematics: a multisemiotic analysis. *Linguistics and Education*. 10(3). 359-388.
- Posteguillo, S. (1999). The schematic structure of computer science research articles. *English for Specific Purposes*, 18(2), 139-60.
- Paltridge, B. (2002). [Thesis and dissertation writing: an examination of published advice and actual practice](#). *English for Specific Purposes*, 21(2), 125-143.
- Reiter, A. (1995). *Writing a Research Paper in Mathematics*. <http://web.mit.edu/jrickert/www/mathadvice.html>

Shelton, T. (n.d.) *Guide for Writing in Mathematics: About Writing in Mathematics*. Southwestern University. <http://www.southwestern.edu/live/files/4175-guide-for-writing-in-mathematicspdf>

Steenrod, N. (1973). "Introduction" in N. Steenrod, P. Halmos, M. Schiffer and J. Dieudonne (Eds.): *How to write Mathematics*. American Mathematical Society, 1-18.

Swales, J.M. (2004). *Research Genres: Exploration and Application*. NY: CUP.

Yang, R. Y., & D. Allison (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes*, 22(4), 365-385.

Yang, R. Y., & D. Allison (2004). Research articles in applied linguistics: Structures from a functional perspective. *English for Specific Purposes*, 23(3), 264-279.

About the Author



Lee Ming Cherk teaches at the Centre for English Communication, National University of Singapore. Her academic interests include second language writing, grammar awareness, language policy, quality and change management, peer conferencing and ICT in learning, and more recently, metadiscourse.