THE ASIAN ESP JOURNAL

The Language of Mathematics: A Corpus-based Analysis of Research Article Writing in a Neglected Field

Laurence Anthony and Mark Bowen

Waseda University, Japan

Biodata

Laurence ANTHONY is Professor of Educational Technology and Applied Linguistics at Waseda University, Tokyo, Japan. His main interests include tools development for corpus linguistics research and applications, and English for Specific Purposes (ESP) program design and methodologies. He is the developer of several corpus tools including *AntConc*, *AntWordProfiler*, and *AntMover*. E-mail: anthony@waseda.jp

Mark BOWEN is Associate Professor of Mathematics at Waseda University, Tokyo, Japan. His main research interests are high-order nonlinear partial differential equations, and exploiting a combination of analytical and numerical techniques in their study. He is also interested in developing teaching approaches and materials for native and non-native speakers of English in the sciences and mathematics. E-mail: mbowen@aoni.waseda.jp

Abstract

Research article writing has received a great deal of attention from ESP researchers. Analyses of the general structure of Introduction-Method-Results-Discussion (IMRD) articles, as well as detailed analyses of individual sections, including introductions, results, and discussions sections, have dominated the ESP literature, especially following Swales' pioneering work on introductions in the early 1980s. Surprisingly, however, the writing of mathematics research articles has been almost completely neglected to date. A few reasons for this can be speculated, including a) the assumption that mathematics writing is similar to that in the well-covered areas

of science and engineering, b) the difficulty in analyzing mathematics research articles due to their often extremely specialized content, and c) the difficulty in locating expert mathematicians who would be willing to serve as specialist informants. In this paper, we present an overview of mathematics writing based on a corpus-based analysis of 410 refereed journal articles covering one complete year of publications in a high-impact mathematics journal. The two-million word corpus was divided into sections, and then analyzed using various corpus tools. Next, the analysis was interpreted by the authors, both of whom have a background in mathematics and one of whom is an active and well-published researcher in mathematics. Results of the study reveal that some macro-level aspects of mathematics writing, such as the basic structuring of titles and introductions, can resemble writing in the fields of science and engineering. On the other hand, many features of mathematics writing diverge greatly from the established norms. We offer reasons for these differences and suggest strategies for teaching writing to a mixed group of science and engineers that may include mathematics majors.
Keywords: ESP, mathematics, research article writing, corpus-based analysis

1. Introduction

Research article writing has received a great deal of attention by ESP researchers, especially in the 1990s and early 2000s following Swales' (1981) seminal study of research article introductions (Swales, 1981) and his follow up study that introduced the CARS model (Swales, 1990). To date, studies of research article writing have focused almost exclusively on articles that follow a standard Introduction-Methods-Results-Discussion (IMRD) structure, which has been described extensively in the literature on report writing (e.g., Day, 1979; Swales & Feak, 2004; Robinson et al., 2008). A review of journal articles published in *English for Specific Purposes*, for example, reveals two articles on research article titles, eleven on introductions, one on the methods section, four on the results section, three on discussions, and one covering the conclusion section. The interest in research article writing is highlighted further by noting that five of the top ten most cited papers listed in 2012 in *English for Specific Purposes* are related to research article writing (e.g., Ozturk 2007, Matsuda & Tardy 2007).

Despite the strong interest in research article writing among ESP researchers, there has been a huge variation in the extent to which the genre has been investigated across different fields and disciplines. Figure 1 shows the results of a keyword search for various field and discipline names in the titles of published articles in three of the top international ESP journals, i.e., *English for Specific Purposes, Asian ESP Journal,* and *ESP World.* Clearly, research on business, science, and medical English has dominated the ESP literature. Also, studies on legal English, engineering English, and economics English (possibly a sub-category of business) have been featured although less prominently. What is interesting to observe from Figure 1, however, is the neglect of another major field of study, i.e., Mathematics. To date, only one research paper focusing specifically on mathematics has been published in *English for Specific Purposes*, and no papers have been featured in the other two journals under study.

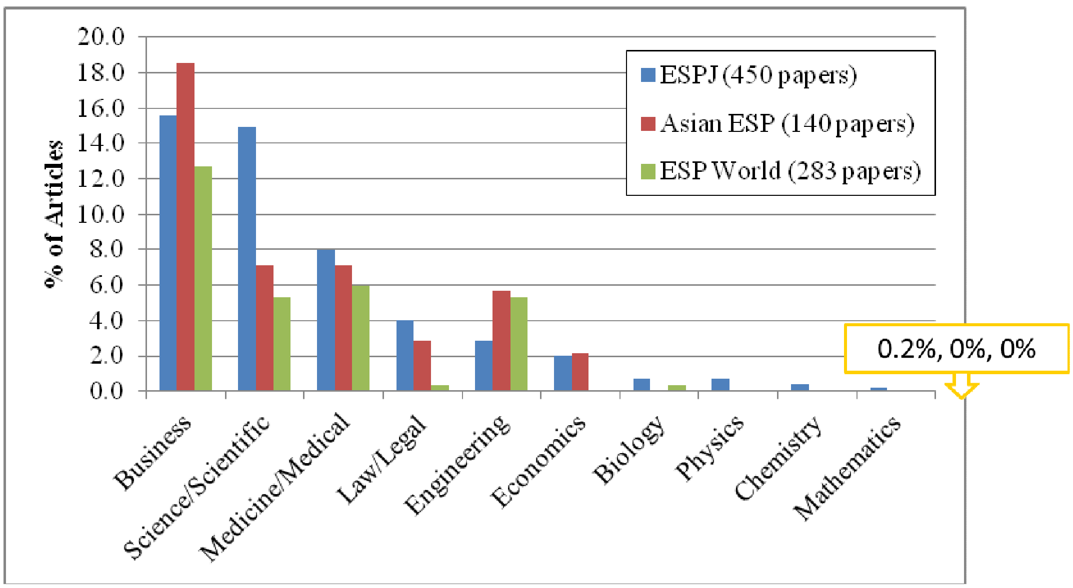

Figure 1: Search terms appearing in titles of research articles in three ESP Journal

It is difficult to ascertain the reasons why mathematics has been neglected as an area of ESP research to date. However, McGrath & Kuteeva (2012), the authors of the only published ESP paper on mathematics, offer three possibilities. First, they argue that researchers may assume that the language of mathematics is similar to that of hard sciences, such as physics, chemistry, and biology, or theoretical disciplines, such as astrophysics, biostatistics and theoretical physics. Indeed, Swales & Feak (2004) follow the latter view. Second, it is possible that ESP researchers consider mathematics to deviate too much from the 'norm' of other sciences to be included in a research study on cross discipline differences, such as that by Hyland (2006).

Third, McGrath & Kuteeva suggest that researchers may consider mathematics discourse to be a type of standardized code that requires little analysis. There is also a fourth possibility. To do a detailed study of mathematics writing inevitably requires a specialist informant from the target field. However, a myth continues that mathematicians are misunderstood, antisocial loners (Devlin 1996). Therefore, ESP researchers may consider that finding an expert mathematician that is willing to serve as a specialist informant is difficult. In fact, the opposite is generally the case. Mathematicians are increasingly providing specialist knowledge of mathematics as part of cross-discipline teams in order to solve complex real-world and theoretical problems. The point is highlighted by the fact that this study is also a cross-discipline endeavor with specialist knowledge of mathematics provided by a practicing mathematician.

Although the study of mathematics English has been largely ignored in the ESP literature, it has featured in other areas of research. For example, there has been a great deal of research that looks at how mathematics is taught in the classroom (e.g., Rowland 1995, 1999; Artemeva & Fox 2011; Street 2005; Leung 2005; Barwell 2005; Morgan 2005; Pimm 1984). Researchers have also looked at how language (not English per se) is used to explain mathematical concepts (e.g., Huang & Normadia 2007; Borasi & Rose 1989; Connolly & Vilardi 1989; Buerk 1990; MacGregor 1990; Countryman 1992; Johanning 2000). In the area of English language, some rare studies on mathematics language can be found, such as an analysis of imperatives (Swales et al. 1988), language and symbolism (O'Halloran 2005), authorial identity (Burton & Morgan 2000), and the above study by McGrath & Kuteeva (2012) on stance and engagement. Clearly, educationalists are interested in the language of mathematics and would find great value in the results of ESP studies on mathematics. Not only would the results help to improve classroom teaching of mathematics, they would also help to build better metacognitive views of mathematics, and of course, help ESP teachers meet the needs of mathematicians in academic and technical reading/writing classes.

In this paper, we report on a study of the writing of research articles in mathematics at both the macro and micro level. At the macro level, we investigate the presence and positioning of the major sections of the research article and compare the results with those for a more traditional science/engineering field, namely mechanical engineering. As a result of this analysis, we hope to establish to what extent mathematics writing (and indeed mechanical engineering writing) follows an IMRD structure.

At the micro level, we look at the writing style of mathematics research articles and identify in what ways the style of mathematics writing resembles or differs from that in a more traditional science/engineering field science and engineering (again, mechanical engineering). Previous literature on research article writing in science and engineering (e.g. Robinson et al. (2008), Sales (2006) and Swales & Feak (2004)) has suggested that a formal style in predominantly used by authors. A formal style manifests itself in many ways, including the *infrequent* use of imprecise, general conversation words and expressions such as "stuff," "things," "bunch," and "whole lot of," the *infrequent* use of phrasal verbs (e.g. "figure out," " make up", "go down"), and the *infrequent* use of connectives such "and," "so," and "but" that even tools such as *Microsoft Word* mark as being informal (Swales & Feak, 2004: 17-24). Here, we investigate the features of these three styles in the hope of establishing whether mathematics writing adopts a formal or informal style. We address the following two research questions:

1. Does mathematics research article writing diverge from the 'norm' of science and engineering research article writing in terms of macro-level structuring, i.e., the presence and positioning of the IMRD ("Title", "Abstract," "Introduction," "Methods," "Results," and "Discussion") sections? If yes, in what way does it diverge from the 'norm'?

2. Does mathematics research article writing diverge from the 'norm' of science and engineering research article writing in terms of style (i.e. does it deviate from a formal writing style)? If yes, in what way does it diverge from the 'norm'?

For the analysis, we use a large corpus of 410 refereed journal articles comprising one complete year of published works in a high-impact mathematics journal. First, we create a structural model of mathematics research article writing based on a qualitative analysis of the texts by the two authors, both of whom have degrees in mathematics and one of whom is an active and well-published researcher in the field. Next, we conduct a quantitative corpus-based analysis of the texts to confirm or reject the structural model and to identify characteristic features of style. We then compare these corpus-based results with those of a comparable reference corpus of texts from a more traditional science/engineering field. Finally, we discuss the results and offer implications for ESP teaching in mathematics.

2. Methodology

2.1 Corpus design

For a corpus-based analysis of writing, it is necessary for the target corpus to be both balanced and representative of the target language (Biber, 1993). It is also important that the corpus has extrinsic validity, i.e., users of the results of the study can understand the relevance and applicability of the results. To ensure a reasonably balanced selection of mathematics topics, an ideal corpus would contain articles from a wide range of research journals covering both pure and applied mathematics. Also, to ensure that each journal was represented accurately in the corpus, a large number of articles from each journal would need to be selected. However, maintaining balance and representativeness would necessitate including many articles from less-prestigious journals. Thus, the validity of the study may be questioned, especially considering that the aim of the study is to provide useful results to teachers of ESP.

For this study, we chose to relegate the importance of balance and focus instead on creating a representative and valid corpus comprised of all articles published in one year of a single, high-impact mathematics journal. To ensure that the corpus was valid, we employed the following journal selection criteria:

- The journal should be ranked in the top 10 highest impact factor journals in the area of applied mathematics, according to the Thomson Reuters (formerly ISI) Web of Knowledge (http://wokinfo.com/).
- The journal should not be a review article journal.
- The journal should cover a broad range of mathematics domains.
- The journal should appeal to a broad audience of both pure and applied mathematicians.

Based on the above criteria, we selected *Nonlinear Analysis: Real World Applications* (hereafter NARWA) for the analysis and collected all 410 articles (approx. 1.9 million words) from Volume 11 (year 2010) of the journal. Details of the target corpus are given in Table 1.

Table 1: Target corpus details

JOURNAL TITLE:	Nonlinear Analysis: Real World Applications (NARWA)
PUBLISHER:	Elsevier
IMPACT FACTOR (2012):	2.043
DATES:	February 2010 - December 2010 (Volume 11: Issues 1-6)
SAMPLING:	Whole population approach (410 articles: 1 entire year)
CORPUS SIZE:	1,917,422 tokens; 30,700 types

It was also necessary to create a corpus of research articles from a more traditional science/engineering field to serve as a comparison. For this study, we chose to analyze articles from the field of mechanical engineering. This is a traditional applied engineering field, which in some ways can be considered to be the exact opposite of mathematics. The following selection criteria were employed:

- The journal should be ranked in the top 10 highest impact factor journals in the area of mechanical engineering, according to the Thomson Reuters (formerly ISI) Web of Knowledge (http://wokinfo.com/).
- The journal should not be a review article journal.
- The journal should cover a broad range of mechanical engineering domains.
- The journal should appeal to a broad audience of mechanical engineers.

Based on the above criteria, we selected the *Journal of Engineering Materials and Technology* (JEMT) for the analysis and collected all 318 articles (approx. 1.6 million words) from Volume 122 (year 2000) of the journal. Details of the target corpus are given in Table 2.

Table 2: Target corpus details

JOURNAL TITLE:	Journal of Engineering Materials and Technology (JEMT)
PUBLISHER:	American Society of Mechanical Engineers (ASME)
IMPACT FACTOR (2012):	0.56
DATES:	January 2000 – December 2000 (Volume 122: Issues 1-4)
SAMPLING:	Whole population approach (318 articles: 1 entire year)
CORPUS SIZE:	1,643,576 tokens; 24,637 types

2.2 Software tools

The analysis of the corpus data was carried out using the *AntConc* 3.3.5 concordancer analysis toolkit (Anthony, 2012), and specially written Python scripts developed by one of the authors (Anthony).

2.3 Qualitative analysis

In order to create an intuitive model of mathematical papers that was not influenced by the corpus data or later quantitative analysis, we (the authors) first discussed at length our own experiences of writing and reviewing mathematics papers. Next, we formulated a preliminary model and then critiqued this model, clarifying, simplifying, or expanding each of its steps to arrive at a final model that the results of the quantitative analysis could be compared against. On completion of the quantitative analysis, we also reviewed the results and discussed the interpretations and implications of the findings with a view to formulating recommendations for future ESP teaching in mathematics.

3. Results

3.1 Macro-level structure of mathematics papers – Intuitive model

Our intuitive model of mathematical research articles is shown in Figure 2. First, it is important to note that mathematics papers can generally be classified on a cline between analytical papers (pure mathematics) and application papers (applied mathematics). We considered the general structure of the two to be equivalent, i.e., they would both include a title, an abstract, and an introduction, and they would discuss the background to the research, explain the methods and results, and optionally finish with some kind of discussion/conclusion section. However, we also agreed that the purpose and details included in each section would reflect the nature of the study and thus vary greatly. For example, in an analytical paper, the purpose is often to prove some mathematical result (stated in the background) and thus the result 'section' would simply offer a closure to the study and may consist of just a single sentence. On the other hand, in an application paper, the aim of the research would be to propose a number of mathematical relationships relating to some real-world applications. Thus, the results of the study could be of three different types: a) details of the relationships, b) experimental results that are used to validate the mathematical relationships in a real-world setting, and/or c) computer simulations that verify the validity of the relationships and perhaps offer further insights into the relationships. Also, we anticipated that the nature of the study would also dictate the actual section headings used in the research articles. We anticipated few would use "Background," "Methods," "Results," or "Discussion/Conclusion" as heading labels.

- **Analytical papers (theory)**
  - Title
  - Abstract
  - Introduction
    - general literature review
  - Background
    - background to the method and/or statement of results
  - Methods
    - analytical tools/methods/lemmas
  - Results
    - study closure
  - Discussion/Conclusion (optional)
    - restatement of the points of the introduction
    - allusion to applications
- **Application Papers**
  - Title
  - Abstract
  - Introduction
    - general literature review
  - Background
    - historical perspective on application
    - introduction of model
    - useful concepts
  - Methods
    - analytical tools/methods/lemmas
  - Results
    - [**Analytical**] math a set of relationships
    - [**Experimental**] experiment to check the relationships in the real-world
    - [**Computational**] simulation to check the relationships
  - Discussion/Conclusion (optional)
    - discussion of related properties
    - impact of results on the application

Figure 2: Intuitive model of mathematics research article structure

3.2 Macro-level structure of mathematics papers – Corpus-based model

To confirm or reject the intuitive model presented in section 3.1, we first wrote a Python script that would extract and count the number of section headings from each article in the target and reference corpora. Figures 3 and 4 show the results of the analysis.

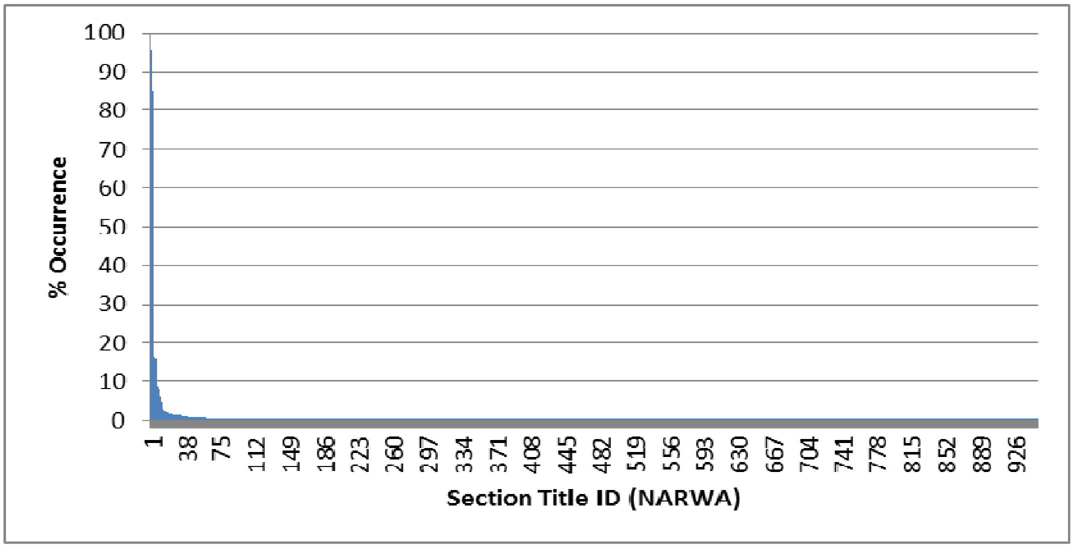

Figure 3: Percentage Occurrence of Section Headings in NARWA

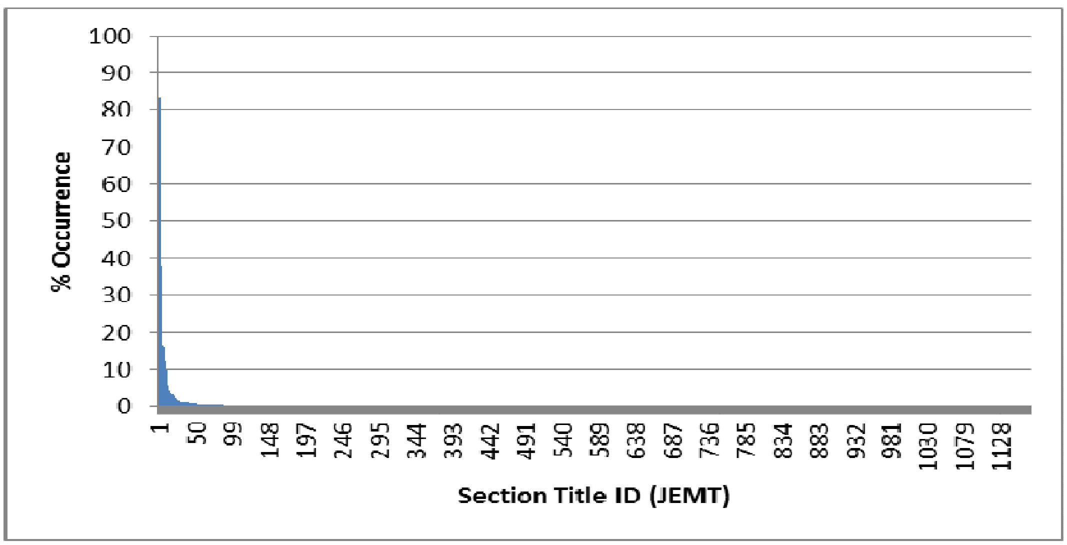

Figure 4: Percentage Occurrence of Section Headings in JEMT

Figures 3 and 4 show the percentage occurrence of different labels in the mathematics (NARWA) and mechanical engineering (JEMT) articles, respectively. In both figures, individual headings in the articles are given an identification (ID) number and arranged in order from the most frequent to the least frequent one. In both figures, the most frequent heading (ID = 1) is "Introduction". Figures 3 and 4 highlight that research articles in both mathematics (NARWA) and mechanical engineering (JEMT) show a great variation in the headings used. In NARWA, 951 different main headings were used (ave. 6.0 per article) with 832 (87%) of them occurring just once. Examples of single occurrence headings are "Exponential convergence," "Blow up phenomenon," and "Classical global solutions." In JEMT, 1167 different headings were used (ave. 7.1 per article) with 1081 (93%) of them occurring just once. Examples of single occurrence headings are "Density measurement," "Previous work," and "Thermal and mechanical analyses."

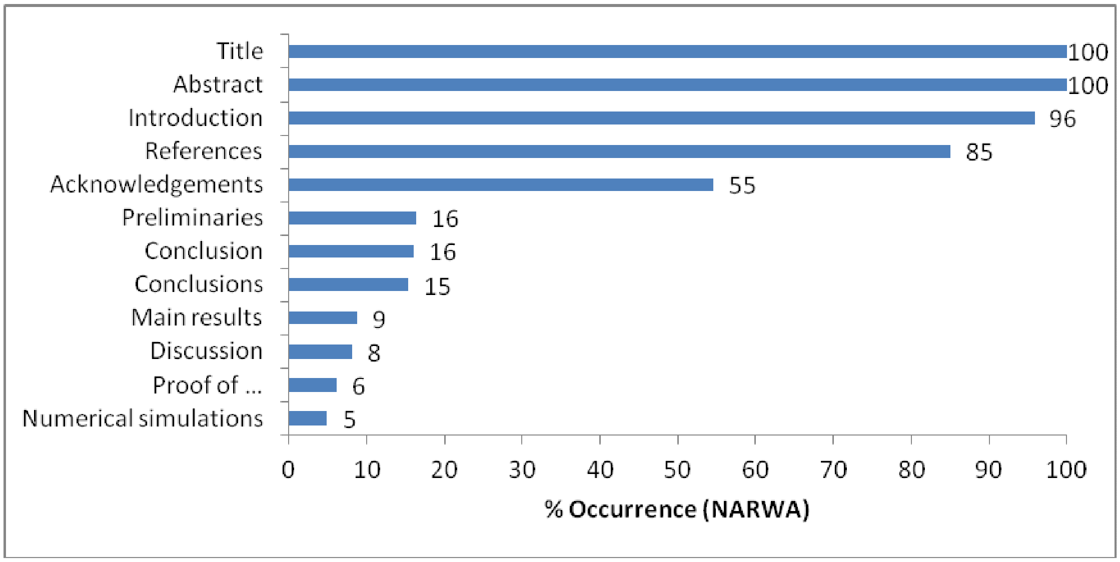

Figure 5: Percentage Occurrence of Title, Abstract and Top 10 Sections in NARWA

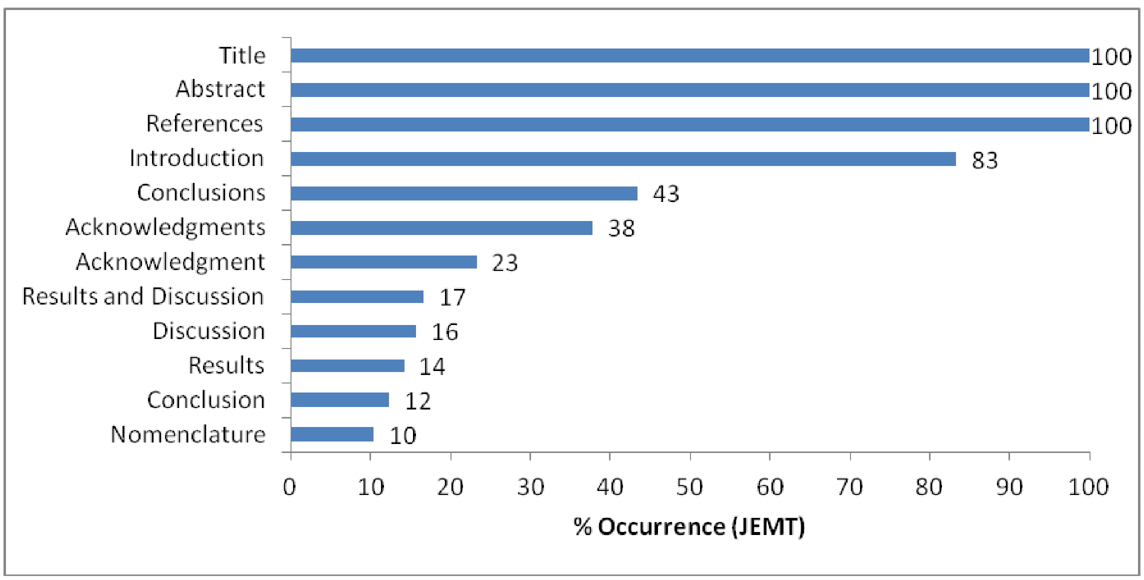

Figure 6: Percentage Occurrence of Title, Abstract and Top 10 Sections in JEMT

Figures 5 and 6 show the frequency of occurrence of the title, abstract, and top ten most commonly used section headings in the NARWA and JEMT corpora. Not surprisingly, most articles in both corpora include an introduction and references. Just over half the articles in both corpora have acknowledgments. However, beyond these very typical sections, few other generalizations can be made. Clearly, mathematics articles do not typically include "Methods," "Results," or "Discussion" sections; a fact that was predicted from the intuitive model. However, surprising to these authors was the fact that mechanical engineering papers also did not show the typical IMRD pattern.

3.3 N-gram analysis of NARWA section titles

The results in Figures 3 and 5 show that mathematics research articles in the NARWA corpus do not show a general IMRD labeling of section headings. This was predicted by the intuitive model given in Figure 2. However, in order to confirm or reject the basic structuring of mathematics article content predicted by the model, it was necessary to clarify whether the variously named sections of the NARWA corpus articles showed any general patterns in terms of content, and if so, establish what these were.

In view of the fact that most section headings occurred only once in the corpus, a complete analysis would require all the articles to be read in full and subsequently analyzed for content structuring. Not only would this be extremely time consuming, it would also introduce the danger of researcher bias, i.e., seeing patterns in the data that did not exist. To counter this danger, we decided to adopt a quantitative analysis of the section structuring based on counts of N-grams (contiguous word sequences of length N) of varying lengths in the different sections. We could then relegate the qualitative analysis to only interpretations of the N-gram frequency tables.

To carry out the analysis, first, the section headings of each article in the NARWA corpus were grouped according to their order of appearance in the article as a whole. In this way, the opening sections of the articles, with headings such as "Introduction," "Introduction and Preliminaries," and "Introduction and Main Results" were grouped together. Similarly, the headings used for the second sections of the articles were grouped together and so on until all headings were accounted for. Next, common patterns of structuring in each group were established by counting N-grams of size one to six using the *AntConc* 3.3.5 (Anthony, 2012) N-gram tool, and ranking them according to frequency. The maximum size of N-gram was chosen based on the frequency distribution of section headings in the corpus, where 374 (91%) of the articles contained six or less sections and just 36 (9%) of corpus articles included seven sections or more. Following this approach, the highest ranked N-grams would represent the most commonly used patterns in each section.

Due to space restrictions, Table 3 shows the results of the analysis for N-grams of size one to three for the first six sections of the corpus articles. Raw frequencies are given for each N-gram in the column adjacent to the N-gram. The N-gram results show several very commonly used headings, For example, Section 1 of the articles reveals a very strong preference for the single word heading "Introduction" although minor variations are possible, such as "Introduction and Preliminaries".

Table 3: Top 10 N-Grams of size 1-3 for each section of the NARWA corpus.

Section 1					
introduction	401	introduction and	11	introduction and *main*	7
and	13	*and main*	7	*and main results*	4
main	7	*and preliminaries*	4	*and main result*	3
preliminaries	5	*main results*	4	introduction and *preliminaries*	3
results	4	*main result*	3	**a kinetic model**	1
result	3	and background	2	**basic physical concepts**	1
the	3	*the problem*	2	introduction and background	1
background	2	**a kinetic**	1	introduction notations and	1
model	2	**basic physical**	1	**kinetic model q**	1
problem	2	*introduction notations*	1	notations and background	1
Section 2					
the	110	of the	42	*of the problem*	14
of	108	*existence of*	17	*formulation of the*	10
and	78	**the model**	17	*existence and uniqueness*	6
preliminaries	76	*main results*	15	the mathml source	6[#]
model	65	*the problem*	14	view the mathml	6[#]
formulation	42	*preliminary results*	13	**of the model**	5
problem	35	*problem formulation*	11	*positive periodic solutions*	5
results	31	*formulation of*	10	*and existence of*	4
existence	28	**governing equations**	10	*and uniqueness of*	4
mathematical	27	**mathematical model**	10	*existence of hopf*	4
Section 3					
of	141	of the	49	*of the main*	8
the	113	*main results*	32	*proof of the*	7
and	68	*existence of*	22	*existence and uniqueness*	6
main	45	*stability of*	16	*the existence of*	6
results	44	*proof of*	15	**analysis of the**	5
existence	40	**periodic solutions**	11	*existence of the*	5
stability	39	*main result*	10	**of the problem**	5
solutions	33	*existence and*	9	solution of the	5
solution	28	*hopf bifurcation*	9	**a priori estimates**	4

analysis	26	*the existence*	9	*and uniqueness of*	4
Section 4					
of	95	of the	30	*and uniqueness of*	5
the	65	*proof of*	16	*existence and uniqueness*	5
and	48	**numerical simulations**	12	**analysis and discussion**	3
numerical	43	*existence of*	10	*and stability of*	3
results	22	and discussion	9	*convergence of the*	3
stability	21	**numerical results**	8	*direction and stability*	3
analysis	20	**periodic solutions**	7	*existence of the*	3
discussion	19	**stability of**	7	**numerical results and**	3
conclusions	18	*existence and*	6	*of limit cycle*	3
existence	17	**analysis and**	5	*of the solution*	3
Section 5					
of	52	of the	15	**results** and discussion	6
conclusions	36	and discussion	8	**an illustrative example**	3
the	36	**numerical simulations**	8	*asymptotic behavior of*	2
and	33	concluding remarks	7	*behavior of the*	2
numerical	28	*proof of*	7	discussion and conclusions	2
conclusion	26	**results and**	7	*nonconstant positive solution*	2
discussion	23	**numerical results**	6	*nonexistence of nonconstant*	2
results	18	discussion and	5	**numerical results and**	2
simulations	11	**illustrative example**	4	*of nonconstant positive*	2
a	10	**numerical simulation**	4	*proof of the*	2
Section 6					
conclusion	21	concluding remarks	7	results and discussion	3
of	21	of the	7	*existence of nonconstant*	2
the	17	**numerical results**	4	*nonconstant positive solutions*	2
discussion	14	results and	4	*of nonconstant positive*	2
conclusions	13	and discussion	3	*of the positive*	2
and	9	**numerical examples**	3	*stability of the*	2
numerical	9	*proof of*	3	*the existence of*	2
remarks	9	*existence of*	2	**a convective condition**	1
concluding	7	final remarks	2	**a critical point**	1
results	7	*nonconstant positive*	2	**a heat flux**	1

*Numbers in the adjacent columns are the raw frequency counts. N-grams in italics are likely to reflect sections from analytical papers. N-grams in bold are likely to reflect sections from application papers. The classifications of the N-grams were determined by the subjective judgments of this paper's authors based on their experience reading and writing mathematics papers.

#These entries relate to the markup of equations in the body of the text. They should be considered as noise.

However, the main result from Table 3 is that the majority of sections in mathematics papers show a complex pattern of section structuring. To understand these patterns more clearly, we went through the N-gram list and attempted to classify each N-gram into one of three categories; a) those that are likely to relate to analytical work (shown in italics in Table 3), b) those that are likely to relate to applications (shown in bold in Table 3), and c) those that are generic in nature (unmarked in Table 3). A summary of this analysis is shown in Table 4 with characteristic N-gram patterns for each type of article appearing below a general description of the section purpose.

Table 4: Summary of N-gram categorization of NARWA section headings.

Analytical Papers	Application Papers
Section 1 (introduction)	
Analytical Papers	Application Papers
introduction introduction and main results introduction and preliminaries	
Section 2 (background/methods)	
Analytical Papers	Application Papers
main results preliminary results	the model mathematical model problem formulation model formulation governing equations mathematical formulations
Section 3 (methods/results)	
Analytical Papers	Application Papers
proof of the main result solution of the problem main result and its applications linear stability and Hopf bifurcation analysis direction and stability of the Hopf bifurcation existence and uniqueness of theorem/solutions	analysis of the data/model/problem a priori estimates of positive solutions existence of a positive periodic solution the main result/results
Section 4 (results/application)	
Analytical Papers	Application Papers
existence and uniqueness of equilibrium point existence and uniqueness of limit cycle direction and stability of the Hopf bifurcation convergence of the series solutions solution and equilibrium points	application equation for the pressure control design for the RTAC system results and discussion
Section 5 (results/application)	
Analytical Papers	Application Papers
proof of the principal theorem proof of the main results nonexistence of nonconstant positive solution periodic solutions	numerical results and discussion numerical results numerical simulations simulation results experimental results results and discussion illustrative example an example
Section 6 (conclusions)	
Analytical Papers	Application Papers
concluding remarks final remarks	concluding remarks final remarks

results and discussion proof of ...	results and discussion numerical examples numerical results

The analysis in Table 4 is largely consistent with the intuitive model presented in Figure 2. Clearly, there is a general ordering of information in terms of introduction, background, methods, results, and discussion/conclusion. However, this ordering may not be immediately apparent unless the reader is well trained in the theories and practices of mathematics. Also, it is clear from Table 4 that the choice of section structuring differs greatly depending on whether the mathematics paper has a focus on analytical methods or applications.

3.4 Comparison of writing styles in the NARWA and JEMT corpora

A preliminary analysis of the NARWA corpus articles revealed multiple occurrences of imprecise, general conversation words and expressions, phrasal verbs, and the connectives "and," "so," and "but," that were described by Swales & Feak (2004) as indicative of informal language. Several examples are shown below, where the informal word or expression is highlighted in bold.

- *It is **easy** to verify that U and F satisfy the operator equation.*
- *There **you can see** the precise conditions...*
- *Here, **you note** that the condition (%%%) above is satisfied.*
- *We* **say** *that, the problem (%%%) and (%%%) is maximal regular...*
- ***Anyway**, we have that EQT is bounded ...*
- *It should be **pointed out** that discrete-time neural networks become more important...*
- *Similarly the second-order solution **works out** to be EQT...*
- *It **turns out** that depending on the locations...*
- ***So** the stability of neural networks has been one of the most active areas of research.*
- ***But** for the Lotka–Volterra predator–prey systems, it is more difficult to discuss.*
- ***And** we arrive at the purpose of the present article.*

In mathematics writing, the researcher is often taking the reader on a journey through various theorems and lemmas to arrive at a proof or new model. In this exposition, a commonly held view among mathematicians is that formality can be sacrificed in exchange for clarity (Halmos, 1973). To investigate if this phenomenon is unique to mathematics writing, we looked in both the NARWA and JEMT corpora at the frequency of occurrence of various imprecise words and expressions, the frequency of occurrence of phrasal verbs, and the frequency of occurrence of the informal connectives "and," "so," and "but." The results are shown in Table 5.

Table 5: Frequency of occurrence of informal features in NARWA and JEMT

Informal Language Feature	% Occurrence	
	NARWA	JEMT
use of adjective "easy"	552 hits (0.46% of all adj.)	52 hits (0.04% of all adj.)
phrasal verbs	4697 hits (18% of all verbs)	7776 hits (16% of all verbs)
informal connectives	714 hits (7.6% of all connectives) and (155 hits: 1.6%) so (372 hits: 4.0%) but (187 hits: 1.9%)	182 hits (2.4% of all connectives) and (91 hits: 0.61%) so (48 hits: 0.97%) but (43 hits: 0.84%)

The results in Table 5 suggest that mathematics articles do indeed show a greater tendency to use informal expressions than articles from mechanical engineering. In particular, the adjective "easy" was used almost ten times as often in NARWA than in JEMT, and the word "so" was used four times as often in the mathematics corpus. However, all the informal expressions investigated also appeared in the mechanical engineering corpus. For this study, we did not calculate if the differences in occurrence of informal expressions were significant. However, there was a large variation in occurrence of informal expressions between different articles, and so we anticipate that the differences are not significant.

4. Discussion

The first research question asked if mathematics research article writing diverges from the 'norm' of science and engineering research article writing in terms of macro-level structuring. Although our results revealed that mathematics articles are structured in widely varying forms and consistently break the traditional IMRD model of Introduction-Methods-Results-Discussion, we also discovered that this is also the case for mechanical engineering research papers. This was a surprising result as we anticipated that mechanical engineering research papers would reveal a more consistent pattern in view of its status as a well-established and traditional engineering field. This result has profound implications for ESP teachers of writing in science and engineering. Many textbooks focus on the IMRD structure of research papers. However, this assumed 'norm' of writing may be less 'normal' than previously assumed. In a real-world scenario, rather than following the IMRD structure, students may be well advised to write their research articles following a less rigid format. In mathematics, the best advice to give to students may be to let the research determine the flow of the research article. For example, if the writer anticipates that the reader will need some preliminary knowledge before understanding the proposed model, then a preliminary knowledge section should be included. Similarly, if the results naturally lead to some interesting applications, then the writer should feel able to include an additional section describing these applications even if it comes between the results and the discussion.

The second research question asked if mathematics research article writing diverged from the 'norm' of science and engineering research article writing in terms of style. Again, the results were surprising. Although expert researchers in mathematics did include informal expressions in their writing, this phenomenon was not unique to their field. In fact, the same informal expressions were observed in mechanical engineering writing, although to a lesser extent. Writers in both mathematics and mechanical engineering used vague terms such as "easy", wrote using phrasal verbs, and linked ideas together using the connectives "and," "but, and "so". All these features are traditionally considered to be inappropriate for a formal academic writing style (see Swales & Feak, 2004) and are even explicitly signaled as inappropriate by style checking tools such as that in *Microsoft Word*. Clearly, ESP teachers need to be aware that informal expressions can be used in advanced technical writing within some disciplines, and they need to inform students of this fact in the writing classroom. Although it may be useful in beginner level classes for teachers to encourage students to follow traditional models of writing style, as students advance in their writing, perhaps a more relaxed view of style is necessary. One way to achieve this is to expose students to corpora in the classroom and allow them to investigate patterns in writing using a data-driven learning approach.

5. Conclusion

In this paper, we have investigated the structure of research article writing in the field of mathematics and compared it to that in mechanical engineering. The results show that the structuring of mathematics papers varies considerably from article to article and that few consistent patterns in the choice of sectioning can be found. However, the same result was found for mechanical engineering papers, suggesting that the traditional IMRD model is not as prevalent in the research literature as it is depicted in textbooks on writing. At a deeper level, however, mathematics papers do reflect the research process itself, starting with an introduction and review of background work, continuing with a description of methods and results, and finishing with a discussion or conclusion. This basic pattern is followed regardless of whether the research is focused more on analytical methods or applications of mathematical models. In terms of style, mathematics articles include various informal expressions that are often considered inappropriate in writing textbooks. However, these same expressions were also found to occur in mechanical engineering articles although to a lesser degree. Again, these findings question the advice given in textbooks on writing and suggest that ESP teachers should encourage learners to be more flexible in their writing and become more aware of actual patterns of writing in professional journals in their field, perhaps through a data-driven learning approach utilizing specialized corpora in the classroom.

Acknowledgements

The authors express their gratitude to two anonymous reviewers for comments on an earlier version of this paper.

References

Anthony, L. (2012). *AntConc* (Version 3.3.5) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/.

Artemeva, N.,& Fox, J. (2011). The writing's on the board: The global and local in teaching undergraduate mathematics through chalk talk. *Written Communication*, *28*(4), 345–379.

Barwell, R. (2005). Language in the Mathematics classroom. *Language and Education*, *19*(2), 96-101.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243-57.

Borasi, R., & Rose B. (1989) Journal writing and mathematics instruction. *Educational Studies in Mathematics, 20*, 347–65.

Buerk, D. (1990). Writing in mathematics: A vehicle for development and empowerment. In A. Sterrett (Ed.), MAA Notes, Vol. 16. *Using writing to teach mathematics* (pp. 78–84). Washington, D.C.: Mathematical Association of America.

Burton, L., & Morgan, C. (2000). Mathematicians Writing. *Journal for Research in Mathematics Education*, *31*, 429-453.

Connolly, P., & Vilardi T. (eds.) (1989). *Writing to learn mathematics and science*. New York: Teachers College Press.

Countryman, J. (1992). *Writing to learn mathematics: Strategies that work*. Portsmouth, NH: Heinemann.

Day, R. A. (1979). *How to write, and publish a scientific paper*. Westport, CT: Oryz.

Devlin, K. (1996). Of men, Mathematics, and myths. *Devlin's Angle*. Retrieved April 18, 2004 from http://www.maa.org/devlin/devlin_aug.html

Halmos, P. R. (1973). How to write mathematics. In N. E. Steenrod, P. R. Halmos, M. M. Schiffer, & J. R. Dieudonne (Eds.), *How to write mathematics* (pp. 19-48) US: American Mathematical Society.

Huang, H., & Normadia, B. (2007). Learning the language of mathematics: A study of student writing. *International Journal of Applied Linguistics, 17*(3), 294-318.

Hyland, K. (2006). Disciplinary differences: Language variation in academic discourses. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 17–45). Frankfurt: Peter Lang.

Leung, C. (2005). Mathematical vocabulary: Fixers of knowledge or points of exploration? *Language and Education*, *19*(2), 126-134.

Johanning, D. I. (2000). An analysis of writing and postwriting group collaboration in middle school pre-algebra. *School Science and Mathematics, 100*(3), 151–61.

MacGregor, M. (1990). Reading and writing in mathematics. In J. Bickmore-Brand (Ed.), *Language in mathematics* (pp. 100–8). Portsmouth, NH: Heinemann.

Matsuda, P. K., & Tardy, C. M. (2008). Continuing the conversation about voice in academic writing. *English for Specific Purposes, 27*(1), 100-105.

McGrath, L., & Kuteeva, M. (2012). Stance and engagement in pure mathematics research articles: Linking discourse features to disciplinary practices. *English for Specific Purposes*, *31*(3), 161-173.

Morgan, C. (2005). Words, definitions and concepts in discourses of mathematics, teaching and learning. *Language and Education, 19*(2): 103–17.

O'Halloran, K. L. (2005). *Mathematical discourse: Language, symbolism and visual images*. London and New York: Continuum.

Ozturk, I. (2007). The textual organization of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes*, *26*(1), 25-38.

Pimm, D. (1984). Who is ‘‘we’’? *Mathematics Teaching*, *107*, 39–42.

Robinson, M. S., Stoller, F. L., Costanza-Robinson, & M. S., Jones, J. K. (2008). *Write like a chemist*. Oxford, UK: Oxford University Press.

Rowland, T. (1995). Hedges in mathematics talk. *Education Studies in Mathematics*, *29*, 327–353.

Rowland, T. (1999). Pronouns in mathematics talk: Power, vagueness and generalisation. *For the Learning of Mathematics*, *19*(2), 19–26.

Sales, H. E. (2006). *Professional communication in Engineering*. New York, US: Palgrave.

Street, B. (2005) The hidden dimensions of mathematical language and literacy. *Language and Education*, *19*(2): 136–41.

Swales, J. M. (1981). *Aspects of article introductions*. Birmingham, U.K.: University of Aston, Language Studies Unit.

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

Swales, J. M., Ahmed, U., Chang, Y., Chavez, D., Dressen, D., & Seymour, R. (1998). Consider this: The role of imperatives in scholarly writing. *Applied Linguistics*, *19*(1), 97–121.

Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students*. Ann Arbor, MI: Univ. Michigan Press.