6

Corpus-Driven Study of Information Systems Project Reports

Ryan T. Miller and Silvia Pessoa

1 Introduction

In recent years, corpus methods have increasingly been applied to the study of disciplinary genres (e.g. Cortes 2004; Flowerdew 2015; Hyland 2008; Nesi and Gardner 2012; Swales 2014). These studies are important because university students must learn to understand and produce disciplinary genres in order to become a full-fledged member of their discipline of study (Johns 1997). However, many faculty members in the disciplines lack explicit knowledge of the rhetorical and linguistic features of disciplinary genres, limiting their ability to effectively teach these genres. Through study of disciplinary genres, applied linguists and ESP tutors can help faculty in the disciplines better scaffold student writing.

In this chapter, we use DocuScope, a corpus-based tool for analysis of rhetorical functions in writing, to study the features of a key genre in the field of information systems (IS), project reports. Through quantitative and quantitatively informed qualitative analysis of model texts and learner writing, we identify rhetorical features that typify this genre, and the extent to which students include these features in their writing.

1.1 Disciplinary genres

In university-level education, students are expected to write a number of genres in a variety of disciplines. Each of these genres reflects the epistemology, culture and discourse that are valued in each discipline (Canagarajah 2002; Duff 2001; Johns 1997). However, disciplinary genres pose a significant challenge to novice

writers because the communication norms and conventions that are unique to each disciplinary genre are often implicit to disciplinary faculty, and opaque to students.

Disciplinary literacy practices can be challenging for students as these practices require an understanding of not only the discipline's content, but also the expectations and demands of the various genres within the discipline. Therefore, there has been a noticeable shift in recent years towards looking at writing within the context of its disciplinary community (Christie and Maton 2012; Hyland 2004). This work emphasizes the fact that it is not enough for students to know about a subject; they also need to be able to use this knowledge in formats that accord with expectations of their discipline. In the present study, we investigate writing in the discipline of IS.

1.2 Writing in information systems

As a discipline, IS focuses on using computers and information technology tools to help businesses achieve their objectives effectively and efficiently. Writing is important in professional IS work, and although writing skills are some of the most requested by employers, a gap still exists between employer expectations and IS graduates' skills (Liu and Murphy 2012). Previous research has recommended that IS faculty take responsibility for helping students improve their written communication skills (Merhout and Etter 2005). Thus, it is vital for IS students to learn to write IS genres, and IS faculty to have explicit genre knowledge in order to more effectively teach these genres (see Miller and Pessoa (2016) for another study of IS genres).

One of the main writing tasks of IS professionals is documentation of IS software development, which occurs via the genre of the project report. A project report is written to document work in each phase of the development of an IS solution. Like other types of software development, IS development often occurs in teams, and occurs incrementally through a series of phases (see Larman and Basili 2003). At the end of each phase, the team documents their progress by writing a project report. Investigating project reports in computer science, Kaneko, Rozycki and Orr (2009) found that although this genre is common among professionals, computer science education lacked instruction in how to write this genre.

The project report itself is a macro-genre (Martin 1992), or a longer text containing multiple subgenres, with each subgenre reflecting the overarching

purposes of the macro-genre. Overarching purposes of an IS project report are to document and report a team's progress in developing an IS software solution, and to show how a team addresses problems they have encountered. Reflecting the importance of documentation in professional IS contexts, project reports are also a part of IS education. In this context, project reports are what Nesi and Gardner (2012: 171) call a genre 'which prepares [students] for professional practice' by discussing and seeking 'solutions to practical problems', and which may differ from more typical academic writing tasks such as argumentative essays or research reports.

1.3 Theory of genre

The basis of our investigation is a theory of genre as representational compositional choices developed by David Kaufer and colleagues (e.g. Ishizaki and Kaufer 2012; Kaufer and Butler 1996, 2000; Kaufer et al. 2005; Klebanov et al. 2016). Kaufer's theory arises from the rhetorical tradition of genre study, which emphasizes pragmatic function over structure of a text. Although rhetorical theories of genre typically rely on 'impressionist, selective, and non-operationalized construals of genre features' (Klebanov et al. 2016: 170), Kaufer's theory seeks to operationalize pragmatic functions of texts by analysing and categorizing language which instantiates these functions. Kaufer's theory views genre as recurrent combinations of micro-rhetorical elements that occur in certain stable proportions, and which prime the reader to have a certain experience or understanding of the text, such as acknowledgement of alternative viewpoints, confidence or objectivity. In Kaufer's theory, these micro-rhetorical elements are organized hierarchically under a number of rhetorical categories, and it is the relative distribution of the various categories which sets genres apart (Klebanov et al. 2016).

Kaufer's theory is congruent with that of Biber (1989, cited in Klebanov et al. 2016) in that both identify genres based on multidimensional analysis of covariation among variables. Biber's system captures functional distinctions through analysis of grammatical categories, such as parts of speech and types of syntactic phrases, and semantic categories, such as communication verbs or certainty adverbs. Rather than linguistic categories, Kaufer's system directly targets functional categories (i.e. the experience that is created in the reader when a phrase is used) (Klebanov et al. 2016). Similar to Hoey's (2005) concept of lexical priming, Kaufer argues that 'words in use prime an audience's

experience and different words prime different experiences' (Kaufer et al. 2004: xvii). Take, for example, the word 'smear': whereas the expression 'to smear a politician' contributes to a negative expression, 'smearing soap' contributes to an expression of everyday motion (Ishizaki and Kaufer 2012: 277).

1.4 Corpus-driven study of academic writing

The study of academic writing development in a second language (L2) has been greatly benefited in recent years by the development of corpora of learner writing, and the increased use of corpus tools in the study of writing development (e.g. Granger, Gilquin and Meunier 2013; Lee and Chen 2009; Wulff and Gries 2011). However, much of this focus has been on the frequency of lexical and grammatical forms, rather than on the functions of these forms as they instantiate rhetorical meanings in discourse (Flowerdew 2009; Upton and Connor 2001). Furthermore, most corpus-based research on writing has focused on academic essays and research reports (Nesi and Gardner 2012), with less attention paid to the writing of professional disciplinary genres.

Corpus-driven research of disciplinary genres can shed light onto not only the linguistic forms used in specialized texts, but also the relationships between lexico-grammatical choices and rhetorical functions. Thus, in this study, we make use of a corpus-based tool (DocuScope) based on Kaufer's theory of genre to investigate rhetorical functions in a specific disciplinary genre, IS project reports, seeking to answer two questions: (i) What are rhetorical features of IS project reports? (ii) To what extent do students adopt these rhetorical features in their writing?

2 Methods

2.1 Data source

The study was conducted in an undergraduate IS programme at a branch campus of an American university in the Middle East. Major components of this IS programme are two hands-on, semester-length team projects, one of which students complete in their third year (termed Junior Project) and one in their fourth year (Senior Project). According to the Junior Project syllabus, the projects are 'team-based project course[s] in team-based software development' in which students collaboratively 'design and build an information or decision

support system, following a disciplined software project lifecycle approach'. As part of this project lifecycle approach, students produce project reports throughout their development process. The major difference between the two projects is that in the Junior Project, the students design their software solution for an imagined client, whereas in the Senior Project the students work with a real-life client. Instructional materials and student writing were collected.

In line with the expectations of the IS discipline, enhancing students' oral and written communication skills for future professional use is a goal of the IS programme at this university. Reflecting this goal, the Senior Project syllabus states that one of the objectives is for students to develop 'professional communications through a structured, guided, hands-on process'. Because each phase had a different focus, the sections and content varied from one phase to the next. However, each project report began with an executive summary which described the team's progress in that specific phase. Because this was the only section which appeared in all of the project reports, we focus our analysis on the executive summary. Instructional materials described the executive summary as one to two-pages that 'tell the reader, in an abbreviated, accurate, and highly readable form, what is in your report'. Also, it 'should communicate to a busy reader all important information … the reader needs to know about your team's project and progress during the phase'. Although sample project reports with executive summaries were given to students as models, there was little instruction about how to write a project report or executive summary, and no focus on language used in writing such reports.

2.2 Texts

Our analysis included three relatively small sets of texts. The first set was sample project report executive summaries provided by IS professors (n = 12; mean length = 381 words, SD = 170.9), which were used by IS professors as models in the Junior Project and Senior Project courses. These sample project reports were written by advanced graduate students at the university's main campus in the United States. Because these texts were used as model texts by the IS faculty, we took the samples as representative of professors' expectations of IS project reports.

The second and third sets of texts were Junior Project (n = 17; mean length = 498, SD = 153.7) and Senior Project (n = 18; mean length = 375, SD = 109.4) executive summaries respectively, written by undergraduate students at the international branch campus. All students were non-native speakers of English,

though they were highly proficient (mean TOEFL iBT score = 97, SD = 12.4). Students came from linguistically and culturally diverse backgrounds, mainly from the Gulf region, the greater Middle East, India, Pakistan and Bangladesh. All project reports were written by teams of three to five students. The Junior Project reports and Senior Project reports were collected in sequential academic years, and thus many of the students were the same in both classes. However, none of the groups consisted of the same members.

2.3 Analysis tool

As described earlier, our analysis was conducted using DocuScope (Ishizaki and Kaufer 2012), a dictionary-based text tagging and visualization tool for identifying instances of language reflecting rhetorical functions. The DocuScope software is a string-matching tool that contains over 45 million unique words and phrases divided into 112 micro-rhetorical function categories (see Ishizaki and Kaufer (2012) and Kaufer et al. (2004) for more information), based on Kaufer and Butler's (1996, 2000) representational theory. For example, DocuScope tags the pronoun 'I' as a first person function, 'if' as associated with a contingency function and 'might happen' as an uncertainty function. The 112 micro-rhetorical functions are grouped into fifteen clusters (see Table 6.1). For example, the narrative cluster contains linguistic elements for past tense verbs, indication of background narrative (e.g. 'by the way') and expressions of time shift (e.g. 'at that moment') and time duration (e.g. 'over the last month'). The analysis in the present study focused on these fifteen rhetorical function clusters.

The DocuScope dictionaries have been found to successfully differentiate patterns associated with genres found in the Brown and Freiburg-Brown (Frown) corpora in ways that comport with the intuitions of human classifiers (Collins 2003). In previous research, DocuScope has been used to analyse rhetorical functions in non-native speaker writing, including argumentation in academic writing (Pessoa, Miller and Kaufer 2014), pragmatic functions in academic writing (Zhao and Kaufer 2013), and features of higher- and lower-graded placement essays (Ishizaki and Wetzel 2008). Although DocuScope has often been used for genre analysis, it should be distinguished from tools whose goal is to classify texts into genres (e.g. Argamon et al. 2007; Stein and Eissen 2008). DocuScope, in contrast, is designed to deepen the understanding of how genre is enacted on the textual surface through language choices (Ishizaki and Kaufer 2012).

Table 6.1 Rhetorical function categories and representative subcategories in the DocuScope system

Academic	**Character**	**Description**	**Elaboration**	**Emotion**
Abstract Concepts: phenomenology	Personal Pronouns: he, she	Sensory Property: bright red	Generalization: all indicate that	Positivity: is beneficial
Citation: points out that	Property of Person: employer, client	Space Relation: the back of the	Example: for example	Anger: livid about
Authoritative Citation: has conclusively reported	Dialogue Cue: ,' he said	Scene Shift: go back to a place	Exceptions: with the exception of	Fear: is afraid of
Contested Citation: maintains that	Oral Cue: uh, hey	Motions: place it behind	Specifiers: and, more specifically	Sadness: lose all hope
Quotation: said '…'			Definition: is defined as	Reluctance: have to admit
Metadiscourse: As this chapter has demonstrated				Apology: apologetically
Future	**Insisting**	**Institutional**	**Interaction**	**Narrative**
Project Ahead: hope to	Immediacy: at this juncture	In Media: on TV	Curiosity: the challenge as we see it is	Narrative Verb: saw the
Predicted Future: there will be	Insisting: it is imperative that	Common Authorities: it is often said that	Question: Did he…?	Time Shift: next week
	Prohibition: you must not	Responsibility: under the supervision of	Open Query: Is there any way that…?	Time Duration: over the last month
		Public Standards: fairness, injustice	Direct Address: I urge you to	Biographical Time: was earlier known as
			Request: We respectfully ask that	
Past	**Personal Relations**	**Reasoning**	**Reporting**	**Subjectivity**
Project back: used to be	Positive Relations: give much credit to	Reason Forward: therefore	Report States: is made of	1st Person: I feel that
Future in past: was to be	Promise: pledged to	Reason back: because	Report Events: arrived at the agreement	Autobio: I have always
	Reassure: find solace in	Support: is evidence for	Report Recurring Events: again	Private Thinking: believe that
	Inclusivity: we all	Contingency: on the condition of	Transformation: changed the nature of	Subjective Time: take our time
	Negative Relations: condemn	Deny: not the case that		Confidence: is definite
		Concessive: it must be acknowledged		Uncertainty: perhaps

DocuScope offers a number of advantages over other tools. First, the size of the DocuScope dictionaries is larger than those of other dictionary-based tools. While many such tools use fewer than 10,000 entries (Klein 2013), the DocuScope dictionaries contain more than 45 million unique words and phrases. A further advantage of DocuScope is its ability to classify strings of varying lengths. DocuScope is able to recognize single words and longer strings (up to seventeen words in length); for example, while 'I' is coded as a first person function, the five-word phrase 'in this paper, I will' is coded as a metadiscourse function in the academic rhetorical function cluster to orient readers at the beginning of an essay.

2.4 Analysis procedure

Our analysis examined rhetorical functions in each of the sets of texts (samples, Junior Project reports and Senior Project reports). First, we used DocuScope to analyse the rhetorical functions in the sample project report executive summaries. We did this by comparing the samples with a reference corpus, the Freiburg-Brown (Frown) corpus, which contains 500 texts of approximately 2,000 words each (a total of approximately a million words). This corpus contains a wide variety of sources and genres, including news (88 texts), general prose (206 texts), scientific texts (80 texts) and fiction (126 texts), and is designed to be representative of 1990s written American English (Biber 1993). This corpus was chosen because previous research using the DocuScope tool has used it as a reference corpus (e.g. Marcellino 2014; Witmore and Hope 2007) and this corpus was used in the validation study noted earlier (Collins 2003). By examining differences between the samples and the reference corpus, we can identify the rhetorical functions which typify the samples. Previous studies using the DocuScope tool have also used this technique (e.g. Ishizaki and Kaufer 2007; Ishizaki and Wetzel 2008; Kaufer et al. 2005).

Following the analysis of the samples, we subsequently analysed the Junior Projects and Senior Projects, and compared these with the samples to determine whether students were using the same rhetorical and micro-rhetorical functions as the samples, and whether they were using these to similar degrees as the samples.

Quantitative comparisons were made using multiple techniques. We used the two-sample Kolmogorov-Smirnov test, a non-parametric method for comparing the distributions of two data sets, and Welch's t-test. We chose these methods because they are robust to unequal sample size and variance, and because they have been used in previous research using DocuScope (e.g. Airoldi et al. 2006;

Kaufer and Hariman 2008; Marcellino 2014). In both cases, we corrected for multiple comparisons using the Benjamini-Hochberg correction set with a false discovery rate of 0.05, and all reported p-values are FDR-adjusted Benjamini-Hochberg p-values. Because such a quantitative analysis reveals only the presence or absence of the rhetorical functions, we also conducted subsequent qualitative analyses to better understand how the rhetorical functions were used in the project report executive summaries.

3 Results

3.1 Sample project report executive summaries

The analysis of the twelve sample project report executive summaries revealed that some rhetorical functions occurred significantly more frequently in the samples than in the reference corpus (higher-occurrence rhetorical functions) while others occurred significantly less frequently (lower-occurrence rhetorical functions). Below, we describe each of these rhetorical functions, with illustrations of usage in the sample executive summaries.

3.1.1 Higher-occurrence rhetorical functions

There were three higher-occurrence rhetorical functions: Personal Relations, Reporting and Future (see Table 6.2).

Personal Relations includes language indicating relationships among people. The qualitative analysis showed that this rhetorical function was used in the samples to show that the authors were working together as a team, and writing the project reports as a team. This included extensive use of first person plural personal pronouns and possessive determiners, such as 'We completed 90% of our tasks on time, and for our project we have completed 5 out of 10 use cases' (Sample 7). It also included use of inclusive noun phrases such as *our team*, *our group* and *collaboration*, as in 'Our team has altered its approach' (Sample 3). Although this is likely a result of the team-based nature of the projects, there was no information in the assignment description indicating that reports should be written in this way. In addition, the use of inclusive 'we' in the project reports is somewhat different from how 'we' has been found to be used in other academic writing, such as to publicize the writer and their work (Harwood 2005).

Table 6.2 Rhetorical functions occurring more frequently in samples than in the reference corpus

	Samples		Frown Corpus		
	M	**SD**	**M**	**SD**	**t (df)**
Personal Relations	3.61	1.44	0.55	0.39	7.49 (11.0)***
Reporting	15.72	3.76	7.26	1.62	7.77 (11.1)***
Future	2.13	0.90	0.83	0.42	4.99 (11.1)***

Note: *** $p < 0.001$; Means indicate the mean number of patterns per 100 words of running text.

The Reporting function includes words for reporting information, particularly verbs. This aligns with the overarching function of the project report in general, and the executive summary in particular, of reporting the team's progress in the development of their software solution. This is different from many other types of writing that students do, such as argumentative writing. Within Reporting, the samples used verbs to report on the processes of their project's development, such as 'The lifecycle of our project began with the introduction of the hospital discharge data. It was then followed by the creation of the database' (Sample 8). Students also used verbs to report on actions they had taken, such as 'We have already introduced the application to our clients' (Sample 4) and 'We identified the most important non-functional requirements' (Sample 5).

The Future rhetorical function, which includes a variety of forward-looking language, was also significantly more frequent in the samples. In the qualitative analysis, we found that this was used to connect the current state of development with plans for future development. For example, the samples used *will*: 'Our proposed system will be based on PHP language in a MVC framework' (Sample 1). In addition, the teams used language that indicates actions taken in order to solve a problem, such as in 'We reduced the complexity of the data model … in order to increase the performance of the system' (Sample 6) or 'Integration will be difficult with [the] existing database … . We will need to build a connector between their old and our new database' (Sample 1).

Overall, the higher-occurrence rhetorical functions suggest that the executive summaries are written from the perspective of a team rather than each individual, they are written to report on the process of development of the project and the actions that were taken during that process and that they use forward-looking language to describe future development of the project or how actions the team has taken will accomplish goals or solve problems.

3.1.2 Lower-occurrence rhetorical functions

There were six rhetorical functions that occurred significantly less frequently in the samples than in the reference corpus (see Table 6.3). Different from the higher-occurrence rhetorical functions, whose presence typifies the genre, lower-occurrence rhetorical functions are notable because of their absence.

The Subjectivity rhetorical function allows authors to express their own involvement in a text (Kaufer and Butler 2000). This function includes use of first person singular pronouns (e.g. I), and because the project reports are written from the perspective of a team (as described earlier), first person singular pronouns were used very little. In addition, this function includes subjective interpretations and evaluations showing confidence (e.g. 'Task assignments were done easily', Sample 4) or uncertainty ('We discovered… our lack of knowledge of the technologies we were using', Sample 8), suggesting that the executive summaries are typically written in a more objective, factual style, with less subjective evaluation.

The Character rhetorical function includes language for naming and describing individuals or entities, and is typical of narrative writing (Kaufer and Butler 2000). Within this category, we saw in particular that the samples had significantly fewer personal pronouns (he, she, they) and few instances of naming individuals or entities according to their role (e.g. 'There is no strong developer in the group', or 'The client may be unable to specify their needs'; Sample 1).

The sample executive summaries were also significantly lower in Description and Emotion. Here, description refers to language which appeals to the senses,

Table 6.3 Rhetorical functions occurring less in samples than in the reference corpus

	Samples		Frown Corpus		
	M	**SD**	**M**	**SD**	**t (df)**
Subjectivity	2.55	0.57	4.78	1.79	12.17 (16.8)***
Character	1.51	0.75	3.88	1.90	10.16 (14.6)***
Description	3.18	0.94	8.06	4.27	14.70 (24.3)***
Reasoning	1.48	1.11	2.90	0.90	2.73 (11.3)*
Emotion	1.37	0.62	2.31	1.02	5.08 (12.5)**
Interaction	0.34	0.34	1.16	0.87	7.94 (14.8)***

Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; Means indicate the mean number of patterns per 100 words of running text.

such as colours, sounds, tastes and so on (Kaufer and Butler 2000). Language appealing to positive and negative emotions was almost completely absent, with most of the matched language being descriptions of defects in the system or problems encountered during system development (e.g. 'We underestimated the time we needed to complete the project', Sample 8), which were not especially emotional in this context. The lack of descriptive and emotional language again reflects the purpose of the project reports as a whole, and the executive summary in particular, as reporting information in an objective, factual manner.

The Reasoning rhetorical function also occurred significantly less in the sample executive summaries. This includes language indicating forward reasoning (e.g. thus, therefore), backward reasoning (e.g. because, owing to the fact) and oppositional reasoning (e.g. it is not the case that) and indicative of logical reasoning in constructing an argument (Pessoa, Miller and Kaufer 2014). There were some instances of these in the samples (e.g. forward: 'Therefore, the team has altered its approach', Sample 3; backward: 'We chose to keep our list … because we do not want to create too large a scope …' Sample 5; oppositional: 'Despite the time, we were able to deliver …' Sample 8), however they were relatively rare. Reflecting the function of the executive summaries as a report, an argumentative rhetorical mode is unnecessary.

Almost completely absent was Interaction, which includes language indicating interaction between author and reader, such as direct requests or questions directed to the reader (Kaufer and Butler 2000).

Overall, the lower-occurrence rhetorical functions suggest that the sample executive summaries present information in a factual, objective manner and without extensive description, evaluation or indication of emotion.

3.2 Junior Projects

After we identified the rhetorical functions in the samples, we examined the student writing, beginning with the seventeen Junior Project reports. First, we compared the Junior Project executive summaries with the reference corpus to determine whether the rhetorical functions that were salient in the samples were also salient in the Junior Projects. Next, we compared the Junior Project reports with the sample reports to determine whether the rhetorical functions occurred to similar degrees. Table 6.4 shows the means for each dataset and the results of the comparisons.

All three of the higher-occurrence rhetorical functions (Personal Relations, Reporting and Future) occurred more frequently in the Junior Project executive

Table 6.4 Frequency of rhetorical functions in the Junior Project executive summaries and comparisons with the Frown corpus and samples

	Junior Project		Frown Corpus			Samples		
Rhetorical function	**M**	**SD**	**M**	**SD**	**t (df)**	**M**	**SD**	**t (df)**
Higher occurrence								
Personal Relations	2.84	1.04	0.55	0.39	9.08 (16.1)***	3.61	1.44	1.72 (18.8)
Reporting	13.46	1.91	7.26	1.62	13.21 (16.8)***	15.72	3.76	1.92 (15.0)
Future	1.76	0.67	0.83	0.42	5.66 (16.4)***	2.13	0.90	1.21 (19.3)
Lower occurrence								
Subjective	3.18	1.15	4.78	1.79	5.52 (18.8)***	2.55	0.57	1.96 (24.8)
Character	1.28	0.79	3.88	1.90	12.35 (22.8)***	1.51	0.75	0.78 (24.6)
Descriptive	3.26	1.50	8.06	4.27	11.69 (25.9)***	3.18	0.94	0.17 (26.7)
Reasoning	1.75	0.83	2.90	0.90	5.60 (17.3)***	1.48	1.11	0.68 (19.2)
Emotional	0.90	0.43	2.31	1.02	12.41 (22.7)***	1.37	0.62	2.25 (18.2)
Interactive	0.53	0.36	1.16	0.87	6.59 (22.7)***	0.34	0.34	1.45 (25.0)

Note: *** $p < 0.001$; Means indicate the mean number of patterns per 100 words of running text.

summaries than in the reference corpus, suggesting that these were also higher-occurrence rhetorical functions in the Junior Projects. In addition, none differed significantly from the samples, suggesting that students used these rhetorical functions at rates appropriate for the genre.

Similarly, all six lower-occurrence rhetorical functions (Subjectivity, Character, Description, Reasoning, Emotion and Interaction) also appeared significantly less frequently in the Junior Project reports than in the reference corpus, and there were no significant differences between the Junior Projects and the samples, further suggesting that the students incorporated these rhetorical functions to an appropriate degree.

However, among the other six rhetorical functions, the Elaboration rhetorical function was found to be higher in the Junior Projects ($M = 4.47$, $SD = 0.95$) than either the reference corpus ($M = 3.54$, $SD = 1.23$; $t(17.9) = 3.94$, $p < 0.001$) or the samples ($M = 3.48$, $SD = 1.08$; $t(21.8) = 2.53$, $p = 0.019$), suggesting an overuse of this function. The Junior Projects tended to include more elaborated explanation, such as in, 'We paid a great attention to this section, as it plays a vital role in fixing our system and meeting users' needs' (Junior Project 13), or 'The technical manual is targeted towards the developers of the system, as it will explain [to] the developers the technical requirements, coding techniques, and design components of the system' (Junior Project 9). This could be a function of the document's status as a classroom assignment, as such explanations tell the professor the students' reasoning for their actions or display their understanding of concepts, which would likely not be necessary in a professional context. The lower amount of elaboration in the samples also reflects the professor's description of the executive summary as 'abbreviated' and presenting only 'the highlights' (Junior Project syllabus).

3.3 Senior Projects

Our analysis of the eighteen Senior Project executive summaries proceeded similarly to our analysis of the Junior Project executive summaries. Table 6.5 summarizes the means for each dataset, and the results of the comparisons.

Among the three higher-occurrence rhetorical functions, two (Personal Relations and Future) were both significantly higher than the reference corpus and not significantly different from the samples, suggesting that students used these functions to an appropriate degree. However, the Reporting function occurred significantly less in the Senior Projects than in the samples, though it was still significantly greater than the reference corpus, indicating that the Senior Projects were in the right direction compared to general English, but were not

Table 6.5 Frequency of rhetorical functions in the Senior Project executive summaries and comparisons with the Frown corpus and samples

	Senior Projects		Frown Corpus			Samples		
Rhetorical function	**M**	**SD**	**M**	**SD**	**t (df)**	**M**	**SD**	**t (df)**
Higher occurrence								
Personal Relations	2.30	1.63	0.55	0.39	4.55 (17.1)***	3.61	1.44	2.42 (25.7)
Reporting	11.66	2.29	7.26	1.62	8.09 (17.6)***	15.72	3.76	3.35 (16.4)*
Future	2.03	0.75	0.83	0.42	6.79 (17.4)***	2.13	0.90	0.30 (20.7)
Lower occurrence								
Subjective	2.81	0.90	4.78	1.79	8.71 (22.2)***	2.55	0.57	0.97 (28.0)
Character	1.72	1.02	3.88	1.90	8.45 (21.5)***	1.51	0.75	0.64 (27.6)
Descriptive	3.58	1.32	8.06	4.27	12.32 (32.2)***	3.18	0.94	0.98 (27.8)
Reasoning	1.65	1.12	2.90	0.90	4.69 (17.8)***	1.48	1.11	0.85 (23.6)
Emotional	0.88	0.43	2.31	1.02	12.90 (24.5)***	1.37	0.62	2.36 (17.9)
Interactive	0.64	0.49	1.16	0.87	4.27 (18.9)**	0.34	0.34	2.67 (25.9)

Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; Means indicate the mean number of patterns per 100 words of running text.

at the level that may be expected in this genre. In particular, the Senior Projects had significantly less reporting of processes than the samples. Closer inspection of the Senior Projects found that they tended to include more description of the product that was being built (e.g. 'The app also has an admin view', 'It is designed to be simple', Senior Project 8), rather than description of the team's process in constructing the product (e.g. 'This development phase started with the successful implementation of several basic CRUD operations', Senior Project 2).

Similar to the Junior Projects, the Senior Projects were significantly lower than the reference corpus and not significantly different from the samples in all six lower-occurrence rhetorical functions. This suggests that the students were able to include these rhetorical functions in their writing in amounts that were appropriate for the genre.

Of the other variables, the Institutional rhetorical function ($M = 6.03$, $SD = 1.88$) occurred more frequently than either the samples ($M = 3.25$, $SD = 1.97$; $t(22.6) = 3.79$, $p = 0.014$) or the reference corpus ($M = 3.71$, $SD = 2.25$; $t(18.8) = 5.09$, $p < 0.001$). The Institutional rhetorical function includes writing that relies on authorized external sources of information rather than the author. From the qualitative analysis, we saw that the Senior Projects tended to refer more to consultations with the client organization than the samples did, for example, 'After carrying out different meetings with our client to gather and record as much information about the project requirements, the team crafted and developed good understanding of the client's expectation of the system' (Senior Project 4). This is in contrast to the sample executive summaries, which also mentioned client requirements, but focused more on the development team's identification of the requirements rather than consulting with the client: 'We identified that the most important non-functional requirements for the client are performance and security' (Sample 5). This may have been a function of the authors of the sample reports being advanced graduate students, who may be more confident in their independent assessment of client needs, without needing as much guidance by the client.

4 Discussion

Using corpus-assisted methods, the present study examined rhetorical functions in a specific disciplinary genre in the field of IS, project report executive summaries. The examination of sample executive summaries revealed a significantly greater presence of three rhetorical functions as compared to a reference corpus. These

suggest that traits of project report executive summaries include increased use of language that indicates the team-based nature of the system development. Following the overarching function of reporting, executive summaries report on the processes of the project development and actions that had been completed. We also saw a significantly greater amount of future-looking language to indicate actions that the team will take, especially in response to problems. In addition, the analysis of the samples found six low-occurrence rhetorical functions. The samples showed a lack of first person personal pronouns, which supports that the executive summaries are written from the perspective of a team. In addition, the executive summary genre seems to include writing which is direct, factual and objective, corresponding to the overarching goal of reporting, rather than descriptive, evaluative or subjective.

The present study found that even without explicit instruction of rhetorical features of the executive summary genre, students nonetheless included most of the same rhetorical functions as the samples, and did so at similar rates as the samples. However, there were some rhetorical functions in which the student writing differed from the samples, including overuse of elaboration and reliance on external sources of information, and underuse of reporting. These findings may reflect differences between academic essay genres, in which students have received explicit language instruction in English courses, and professional genres such as project reports, in which students received no instruction on language use. Nesi and Gardner (2012) point out that when students are asked to write professional genres, there is often a tension as the reader of the text is not a client or colleague, but rather a teacher who evaluates the text. This may cause student writers to feel a need to explain their actions and choices, provide evidence of their understanding or rely more on information from external sources rather than their own interpretations. The result of this is what Wardle (2009: 774) calls a 'mutt genre' with conflicting audiences and purposes.

5 Pedagogical applications

The present study has a number of pedagogical implications. First, the findings highlight the importance of using model texts in teaching writing. Although there was no explicit instruction of rhetorical features for writing project report executive summaries, students were nonetheless able to include most of the rhetorical functions in their writing with only the sample texts as input. Although some researchers have argued that using model texts represents

genres too narrowly and inhibits students' creativity as writers (e.g. Elbow 1999; Smagorinski 1998), the model texts here seem to have allowed students to (perhaps implicitly) pick up micro-rhetorical elements of the sample texts and integrate these into their own writing. We would, however, suggest that instructors go beyond simply supplying model texts by helping students develop explicit knowledge of rhetorical features by making these features (such as those found in the present study) explicit for students. After understanding rhetorical functions through an analysis such as in the present study, instructors can then show students the language through which these functions are implemented in the model texts.

In addition to model texts, instructors can create exercises using extracts from project reports. For example, the present study found that an important feature of IS project reports was reporting, specifically about processes that the team undertakes in developing their product, how steps in the process built on each other and how current problems link with future solutions. This was also an area that students showed difficulty in. The Senior Projects showed significantly less reporting of processes than the samples, and, while many project reports included language that connected the current status of the project (or current problems with the project) with future development plans, some focused entirely on the past, such as one that described how the 'project suffered from major deviation from the original proposal' (Junior Project 10) without describing how the team planned to address this as they moved forward. Thus, this aspect of project report writing might be a suitable target for explicit classroom instruction of genre features. For example, in the classroom, a teacher could provide students with a list of events that may happen during development of a project, such as those reported in Senior Project 16 (which almost entirely lacked descriptions of processes and linkages between events): 'A needs analysis was conducted at (popular tourist destination)', 'We identified different types of visitors', 'The main disadvantages were…', 'We aligned the problems and the needs of the visitors', 'We saw an opportunity for an innovative approach' and 'This web application maps together the two most important things in (this tourist destination).' The teacher could then ask students to, using these events, collaboratively write a report that puts the events into a series or process with linkages between steps.

Last, we would suggest that assignment descriptions and rubrics take into account rhetorical functions and the language through which they are realized. For example, in the courses from which our data were drawn, the assignment descriptions included instructions such as that the reports should 'communicate to a busy reader all important information' and should 'highlight important

conclusions, recommendations, and problems encountered or anticipated'; however, it could be helpful to have more explicit instructions stating that, for example, there should be little elaboration or extended description, and instruction could include examples of language that enacts these functions. Assessment rubrics could also reflect linguistic choices as they construe rhetorical functions.

6 Conclusion

Using a combination of quantitative and quantitatively informed qualitative analyses, the present study identified rhetorical features of IS project report executive summaries, finding that this genre is written from a group perspective, connects current problems with future plans and is direct, factual and objective. Although there was no explicit instruction of these rhetorical functions, students were nonetheless able to include most of them in their writing through exposure to model texts. However, deviances from the model texts were observed, which may have resulted from conflicting exigencies due to the genre's status as both a professional genre and a classroom assignment.

Acknowledgement

This publication was made possible by NPRP grant #5-1320-6-040 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

Airoldi, E.M., Anderson, A.G., Fienberg, S.E. and Skinner, K. K. (2006), 'Who wrote Ronald Reagan's radio addresses?', *Bayesian Analysis*, 2: 289–320.

Argamon, S., Whitelaw, C., Chase, P. and Hota, S. R. (2007), 'Stylistic text classification using functional lexical features', *Journal of the American Society for Information Science and Technology*, 58: 802–22.

Biber, D. (1989), 'A typology of English texts', *Linguistics*, 27: 3–43.

Biber, D. (1993), 'Representativeness in corpus design', *Literary and Linguistic Computing*, 8: 243–57.

Canagarajah, A. S. (2002), *A Geopolitics of Academic Writing*, Pittsburgh: University of Pittsburgh Press.

Christie, F. and Maton, K. (2012), *Disciplinarity: Functional Linguistic and Sociological Perspectives,* London: Bloomsbury Academic.

Collins, J. (2003), 'Variations in written English: Characterizing the rhetorical language choices in the Brown Corpus of Texts', Unpublished doctoral dissertation, Carnegie Mellon University.

Cortes, V. (2004), 'Lexical bundles in published and student disciplinary writing: Examples from history and biology', *English for Specific Purposes*, 23: 397–423.

Duff, P. A. (2001), 'Learning English for academic and occupational purposes', *TESOL Quarterly*, 35: 606–7.

Elbow, P. (1999), 'Individualism and the teaching of writing: A response to Vai Ramanathan and Dwight Atkinson', *Journal of Second Language Writing*, 8: 327–38.

Flowerdew, L. (2009), 'Applying corpus linguistics to pedagogy: A critical evaluation', *International Journal of Corpus Linguistics*, 14: 393–417.

Flowerdew, L. (2015), 'Using corpus-based research and online academic corpora to inform writing of the discussion section of a thesis', *Journal of English for Academic Purposes*, 20: 58–68.

Granger, S., Gilquin, G. and Meunier, F. (eds) (2013), Twenty Years of Learner Corpus Research: Looking Back, Moving Ahead, Louvain, Belgium: Presses Universitaires de Louvain.

Harwood, N. (2005), "Nowhere has anyone attempted… In this article I aim to do just that': A corpus-based study of self-promotional I and we in academic writing across four disciplines'. *Journal of Pragmatics*, 37: 1207–31.

Hoey, M. (2005), *Lexical Priming: A New Theory of Words and Language*, London: Routledge.

Hyland, K. (2004), *Disciplinary Discourses: Social Interactions in Academic Writing*, Ann Arbor, MI: University of Michigan Press.

Hyland, K. (2008), 'As can be seen: Lexical bundles and disciplinary variation'. *English for Specific Purposes*, 27: 4–21.

Ishizaki, S. and Kaufer, D.S. (2007), 'A model of rhetorical design strategies'. Paper presented at the conference of the American Association for Applied Linguistics, Costa Mesa, AZ, 21–24 April.

Ishizaki, S. and Kaufer, D.S. (2012), 'Computer-aided rhetorical analysis', in P. McCarthy and C. Boonithum-Denecke (eds), *Applied Natural Language Processing: Identification, Investigation, and Resolution*, 276–96, Hershey, PA: Information Science Reference.

Ishizaki, S. and Wetzel, D. (2008), 'Computerized rhetorical analysis of L2 freshman placement essays'. Paper presented at the conference of the American Association for Applied Linguistics, Washington, DC, 29 March to 2 April.

Johns, A. M. (1997), *Text, Role, and Context*, London: Cambridge University Press.

Kaneko, E., Rozycki, W. and Orr, T. (2009), 'Survey of workplace English needs among computer science graduates'. Paper presented at IEEE International Professional Communication Conference, Waikiki, HI, 19–22 July.

Kaufer, D.S. and Butler, B. S. (1996), *Rhetoric and The Arts of Design*, Mahwah, NJ: Lawrence Erlbaum.

Kaufer, D.S. and Butler, B.S. (2000), *Designing Interactive Worlds with Words: Principles of Writing as Representational Composition*, Mahwah, NJ: Lawrence Erlbaum.

Kaufer, D.S., Geisler, C., Ishizaki, S. and Vlachos, P. (2005), 'Computer-support for genre analysis and discovery', In Y. Cai (ed.), *Ambient Intelligence for Scientific Discovery: Foundations, Theories, and Systems*, 129–51. New York: Springer.

Kaufer, D. S. and Hariman, R. (2008), 'A corpus analysis evaluating Hariman's theory of political style', *Text & Talk*, 28: 475–500.

Kaufer, D. S., Ishizaki, S., Butler, B. S. and Collins, J. (2004), *The Power of Words: Unveiling the Speaker and Writer's Hidden Craft*, Mahwah, NJ: Lawrence Erlbaum.

Klebanov, B.B., Kaufer, D., Yeoh, P., Ishizaki, S. and Holtzman, S. (2016), 'Argumentative writing in assessment and instruction: A comparative perspective', in N. Stukker, W. Spooren and G. Steen (eds), *Genre in Language, Discourse and Cognition*, 167–92, Boston: de Gruyter.

Klein, H. (2013). Text analysis Info: Category Systems. Available at: http://www.textanalysis.info (accessed 17 January 2017).

Larman, C. and Basili, V. R. (2003), 'Iterative and incremental development: A brief history', *Computer*, 36: 47–56.

Lee, D. Y. W. and Chen, S. X. (2009), 'Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners', *Journal of Second Language Writing*, 18: 149–65.

Liu, M. X. and Murphy, D. (2012), 'Fusing communication and writing skills in the 21st century's IT/IS curricula', *Information Systems Education Journal*, 10: 48–54.

Marcellino, W. M. (2014), 'Talk like a marine: USMC linguistic acculturation and civil-military argument', *Discourse Studies*, 16: 385–405.

Martin, J. R. (1992), *English Text: System and Structure*, Amsterdam: John Benjamins.

Merhout, J. W. and Etter, S. J. (2005), 'Integrating writing into IT/MIS courses', *International Journal of Information and Communication Technology Education*, 1: 74–84.

Miller, R. T. and Pessoa, S. (2016), 'Role and genre expectations in undergraduate case analysis in information systems', *English for Specific Purposes*, 44: 43–56.

Nesi, H. and Gardner, S. (2012), *Genres across the Disciplines: Student Writing in Higher Education*, New York: Cambridge University Press.

Pessoa, S., Miller, R.T. and Kaufer, D. (2014), 'Students' challenges and development in the transition to academic writing at an English-medium university in Qatar', *International Review of Applied Linguistics in Language Teaching*, 52: 127–56.

Smagorinski, P. (1998), 'How reading model essays affects writers', in J. Irwin and M. A. Doyle (eds), *Reading/writing Connections: Learning from Research*, 160–76, Newark, DE: International Reading Association.

Stein, B. and Eissen, M. (2008), 'Retrieval models for genre classification', *Scandinavian Journal of Information Systems*, 20: 93–119.

Swales, J. (2014), 'Variation in citational practice in a corpus of student biology papers: From parenthetical plonking to intertextual storytelling', *Written Communication*, 31: 118–41.

Upton, T. A. and Connor, U. (2001), 'Using computerized corpus analysis to investigate the textlinguistic discourse moves of a genre', *English for Specific Purposes*, 20: 313–29.

Wardle, E. (2009), '"Mutt genres" and the goal of FYC: Can we help students write the genres of the university?', *College Composition and Communication*, 60: 765–89.

Witmore, M. and Hope, J. (2007), 'Shakespeare by the numbers: On the linguistic texture of the late plays', in S. Mukherji and R. Lyne (eds), *Early Modern Tragicomedy*, 133–53, Cambridge: Brewer.

Wulff, S. and Gries, S. (2011), 'Corpus-driven methods for assessing accuracy in learner production', in P. Robinson (ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*, 61–88, Philadelphia: John Benjamins.

Zhao, H. and Kaufer, D. (2013), 'DocuScope for genre analysis: Potential for assessing pragmatic functions in second language writing', in N. Taguchi and J. M. Sykes (eds), *Technology in Interlanguage Pragmatics*, 235–59, Amsterdam: John Benjamins.