Focus on ELT Journal
Vol 5, Issue 2, 2023
ISSN: 2687-5381

Comparing research articles in pulmonology and other disciplines

[a]**Ashleigh Cox** [b]**Eric Friginal**

[a] PhD Student, Georgia State University, Atlanta, GA, acox27@gsu.edu

[b] Prof. Dr., The Hong Kong Polytechnic University, Hong Kong, eric.friginal@poyu.edu.hk

APA Citation: Cox, A., & Friginal, E. (2023). Comparing research articles in pulmonology and other disciplines. *Focus on ELT Journal*, *5*(2), 21-41. https://doi.org/10.14744/felt.2023.5.2.2

ABSTRACT

There have been many arguments for research-informed pedagogy in English for Academic Purposes (EAP) classes, including medical English classes. While there has been a wealth of research on writing in English for medical purposes, there has been little research on academic research articles (RAs) specifically in the pulmonology domain with very little empirical information about linguistic patterning useful in materials design for writing instruction. The extent to which writing in that branch of medicine is similar to writing in hard sciences in general is unknown. The present study analyzes the linguistic features in a specialized corpus of pulmonology research articles and a comparison corpus of research articles from other scientific disciplines using three of the functional dimensions established by Biber (1988) that are relevant to academic writing—the first, third, and fifth dimension. Results indicate that in comparison to research articles in other hard sciences, pulmonology research articles have more densely packed information and less abstract information. Pedagogical implications for academic writing teachers of respiratory therapy students in EFL and ESL contexts and future research directions are discussed.

Keywords
corpus linguistics, academic writing, English for Academic Purposes

Article History
Received : 14.08.2023
Revised : 10.12.2023
Accepted : 10.12.2023
Published : 30.12.2023

Type
Research Article

Introduction

Discipline-specific academic writing has been a widely studied area of applied linguistics research (see for example, Ferris, 1994; Grant & Ginther, 2000; Gray, 2013), and corpus-based research has been an effective way to find characteristic features and trends in academic writing that have useful pedagogical implications (Conrad, 1999; Cortes, 2013; Römer et al., 2020). Various approaches to exploring academic writing corpora have been used over the years, some analyzing the overall organization of texts, and others examining smaller units within holistic texts (Biber et al., 2007). There have been many corpus studies of academic writing with important findings and pedagogical implications in areas such as the differences between written academic registers and other domains (e.g., Biber, 1988), common rhetorical patterns in academic research articles (RAs) (e.g., Gray et al., 2020; Swales, 1990), and features that are associated with high proficiency in essays (e.g., Friginal et al., 2014; Grant & Ginther, 2000).

A popular area of academic writing research is the analysis of academic RAs (e.g., Friginal & Mustafa, 2017; Gray, 2013; Swales, 1990). Many studies have examined specific linguistic

✉**Corresponding Author:** acox27@gsu.edu

features in academic RAs in particular disciplines with the goal of drawing pedagogical implications for writing teachers (e.g., Crosthwaite et al., 2017). While studies on individual linguistic features can have practical implications, to describe a register more broadly, it is also helpful to examine how multiple linguistic features are used together, as linguistic features do not appear in isolation (Biber, 1988; Durrant et al., 2021; Ervin-Tripp, 1972; Halliday, 1988), so that writing teachers can be provided with information about the environments that each feature occurs in and the functions that co-occurring features perform together.

Although general tendencies in academic writing have been found, many studies have examined writing in specific disciplines, since there are different communicative needs in each field. Some disciplinary differences in the linguistic features used in research articles have been identified (Conrad, 2001; Gray, 2013). Nonetheless, many writing researchers group academic disciplines into general categories like "hard sciences" and "soft sciences" because of similarities that have been noted in the writing of related disciplines (e.g., Bazerman, 1981; Bruce, 2008). For teachers, it can be helpful to know what is shared among hard and soft sciences and what is unique to specific disciplines, whether they in are academic writing classroom contexts that have students from many disciplines together in one classroom or contexts designed to target more specific student populations. Currently, there is a lack of research exploring the unique features of pulmonology (a medical specialty that focuses on the respiratory system) writing in relation to other hard sciences, even though many pulmonology scholars using English as an additional language learn about academic writing with the goal of publishing their research in English-medium journals. This study analyzes the co-occurrences of linguistic features in academic RAs in the field of pulmonology using Biber's (1988) dimensions of register variation and compares pulmonology research articles to research articles in other science disciplines to explore how similar pulmonology writing is to writing in "hard sciences" in general, with the goal of providing insight for academic writing teachers.

Why should academic writing researchers care about English in pulmonology?

English is used as a lingua franca around the world for many domains of study, and it is especially prominent in medical fields. Using a shared language internationally is crucial to ensure that important findings can be shared among healthcare workers with different language backgrounds. Medical English has attracted the attention of many ESP researchers because of the high international demand for pedagogical materials. Not surprisingly, differences in the language needs for different branches and specializations within the medicine umbrella have been identified (see for example, Hsu's work on medical vocabulary using corpora separated into different branches of medicine). Since topic has a role in the communicative needs of writers, and the communicative needs drive the choice of words and linguistic features, it is reasonable to hypothesize that each branch of medicine would develop its own disciplinary nuances. It would not be expected, for example, that discussing research on knee replacement surgery would trigger similar combinations of lexico-grammatical features as discussing research on pneumonia, assuming that grammatical choices are driven by communicative functions. ESP researchers have argued for decades that developing field-specific English courses and materials is more efficient than traditional general English alternatives (which Abbott [1980] referred to as "English for no obvious reason"), and while ESP researchers have had a sustained interest in Medical English, the efficiency of Medical English pedagogy could be greatly improved if there was more attention on the language use within sub-specializations of healthcare. In pulmonology literature specifically, there are many articles that discuss topics related to physics, such as air pressure, that scholars in

other medical subfields would not be likely to write about. Describing pulmonology-specific procedures, equipment, and treatments is likely to lead to the use of unique combinations language structures. Academic writing researchers have focused on some specific branches of medicine, such as dentistry (e.g., Crosthwaite, 2016; Crosthwaite et al., 2017), but pulmonology has not caught their attention. As pulmonologists around the world are no less connected through English medium journals than specialists in other medical branches (see for example Souza and de Carvalho's [2014] arguments that publishing English versions of pulmonology research articles originally written in Portuguese or Spanish can increase international readership), it is important for academic writing researchers to consider finding ways to develop materials that can efficiently target pulmonologists' English language needs. Such materials could be useful for the following settings, both in countries where English is widely used as a first language with universities that attract international medical students, and in countries where English is not widely used as a first language where medical students are focused on English for research sharing purposes:

a. Teachers of writing support classes in pulmonology departments
b. Teachers of medical English who want to group students based on specialization for writing activities
c. Teachers of general academic English (e.g., in ESL courses of universities with pulmonology programs that bring international students) who want to group students based on specialization for writing activities
d. Teachers in programs for only specific subfields of medicine (such as language support programs funded by grants specifically for international tuberculosis research, for example)

Although the history of research on Medical English in general has led to many informative findings, as the development of technology and corpus approaches has made it possible for linguists to analyze more specific registers, adding to the body of literature on sub-fields of medicine could take the Medical English research a step further and improve the efficiency of ESP materials development. For exploratory studies on writing in pulmonology, comparing pulmonology to more general scientific writing that EAP teachers are likely to be familiar with could be a helpful starting point.

Literature Review

Previous research on RAs in medical domains

Although there is limited research on pulmonology writing, the research that has been done on medical English writing in general can inform studies on more specific sub disciplines related to medicine. There have been some notable studies with a broader medical English focus. Atkinson (1992) provided an overview of the historical evolution of medical RAs from 1735 to 1985 that could be a useful starting point for understanding the background of this register. He explored how the conventions of writing research articles evolved using linguistic and rhetorical analysis of texts from the *Edinburgh Medical Journal*, which he identified as the oldest medical journal that is still operating. The changes over time that he found reflected changes in how science was conceptualized, growing more objective and evidence based. Early medical research articles were written as letters starting with "Dear Sir," and often reported on only one case study presented chronologically from the perspective of the researcher's observations with discussion of their

thought processes (p. 347). Atkinson (1992) described that over time, medical research articles progressed from single case studies to multiple case studies, then to an integration of many loosely related case studies, and eventually to a presentation of statistical data, reaching the IMRD (Introduction, Methods, Results, Discussion) format by 1945. He found that as the organizational format changed, the writing style of RAs also became less narrative and more technical.

Other studies have sought to provide descriptions of the modern IMRD medical research article genre, analyzing the linguistic features within organizational units and the rhetorical structure. Focusing on abstracts, Salager-Meyer (1992) analyzed the rhetorical moves and verb tense and modality in medical RA abstracts and found associations between tense and aspect choices and rhetorical functions. For example, the present tense was used in conclusions, recommendations, and data syntheses, the present perfect was used to discuss research gaps by showing disagreement with previous studies, and past tense was frequently used throughout multiple sections of abstracts (in the introduction, methods, results, and "case presentation" sections). Biber and Finegan (2001) later examined the distribution of linguistic features in different sections of medical research articles and compared their frequencies in each section to other registers of English. They found that in the introduction sections, *that* verb complement clauses, first person pronouns, possibility modals, and present tense were frequent in comparison to other sections. In the methods sections, past tense and agentless passives were more prominent than in other sections. The results sections had relatively more present and past tense, and the discussion sections had comparatively higher frequencies of *that* verb complement clauses, possibility modals, downtoners, and present tense. In comparison with a reference corpus of other genres, *that* verb complement clauses and agentless passives were more common in the introduction section, past tense and agentless passives were more common in the methods section, downtowners (to a small degree) and agentless passives were more common in the results section, and *that* verb complement clauses, possibility modals, downtoners, and agentless passives were more common in discussion sections. They also found that introductions had more elaborated references and methods sections were more informational, non-narrative, and impersonal than the other sections.

In a study analyzing the broader rhetorical organization of complete RAs, Nwogu (1997) described that medical RAs presented background information about their topic, reviewed previous research and its limitations, presented the goal of the paper and procedures used, described the data collection process, provided details on the experimental procedures and materials, described the data analysis, shared and accounted for the observations made, stated and explained the overall outcomes of the research, and concluded with implications, in this order. These studies are just a few examples of the type of work that has been done on medical RA analysis. Based on these studies, the following conclusions can be formed:

a. Medical research articles change over time (Atkinson, 1992), and ongoing research on this register is warranted.
b. The use of grammatical features in medical research articles is related to communicative functions that they perform (Biber & Finegan, 2001; Salager-Meyer, 1992), and it is thus reasonable to hypothesize that writers would need to be familiar with features used to describe the topics and procedures discussed in their specific subdiscipline of medicine.
c. The organization of medical research articles has some differences from what has been described for research articles in general, and these differences elicit the use of unique linguistic features (Nwogu, 1997), but the differences are small enough that medical

> research articles and research articles from other fields can still be considered comparable.

While what is known about how medical research articles are written is likely to be helpful for teaching pulmonologists, it may not fully capture the nuances of research articles whose readership targets the more specific pulmonology discourse community. The current body of knowledge can help researchers form hypotheses about what is likely to be found in pulmonology research articles, but research more specifically on articles in that domain is needed to verify whether there are unique qualities that differ from other domains related to medicine or health care. To start to explore this topic, studies with broader goals than those listed above may be useful to later inform studies with narrower foci.

What is currently known about pulmonology research articles?

Although scholars for disciplines that require pulmonology research have suggested that language skills are important for success in English medium programs (e.g., AbuNurah et al., 2020), there is not a strong enough research base describing the domain of pulmonology RAs for academic writing teachers to be able to develop activities specifically tailored to the nuances of that register. There is very little research published on "English for pulmonologists," and there is limited research on pulmonology RAs. Hsu's (2013) medical vocabulary list with sub-categories for specific branches of medicine that included pulmonary/respiratory medicine could be a helpful starting point for ESL or EFL writing teachers teaching students in careers or academic programs related to pulmonology. Djamaa (2013) offered practical pedagogical suggestions (supported by survey and interview data reporting a small sample of students' positive perceptions of her approaches) for teaching English for science, and she included charts of the respiratory tract given to her by a pulmonologist in activities, but her focus was not specifically on pulmonology. Among the studies that are focused on language use within pulmonology-related disciplines, Watson and Low (1988) and Watson and Scully (1988) found that the reading level of respiratory therapy textbooks is higher than the reading level of the average respiratory therapy student in the U.S., but it has not been established whether this is still an ongoing issue today. The pulmonology writing research is very sparse, but there is at least one study on pulmonology-related RA writing available.

Hagan (1996) conducted a case study on peer and instructor writing conferences in an undergraduate level respiratory care seminar that required students to write a research paper. Based on an analysis of students' revisions, Hagan considered peer and instructor revision conferences successful in helping students improve their research papers. Although some students only made language related edits, other students made unprompted changes during the revision process. Hagan also found that the literature review section was particularly challenging for students. While this information is pedagogically useful and adds to the body of mixed findings on the effectiveness of peer review in writing teaching, it does not provide any information describing how pulmonology RAs are written.

Based on the limited information available from these studies, academic writing teachers could consider teaching students in majors related to pulmonology useful morphemes (Hsu, 2013), using authentic realia (Djamaa, 2013), targeting reading skills (Watson & Low, 1988; Watson & Scully, 1988) and literature review writing (Hagan, 1996), using peer review and writing conferences (Hagan, 1996), and reading about more general medical RA studies, but there is not

any corpus-based empirical research describing the genre of pulmonology research articles to inform their writing instruction.

Multi-dimensional analysis

Biber (1988) argues that studies on register variation need to address both form and function and linguistic and nonlinguistic features, and this can be achieved using his multi-dimensional analysis (MDA) approach, which is a method of analyzing the co-occurrences of linguistic features and their functions that could be useful in exploratory studies on under-researched registers. The approach involves a factor analysis of the linguistic features in a corpus followed by a functional interpretation of the co-occurrences of linguistic features in each factor. Each factor can provide a different description of a register, but a single factor by itself is not enough to fully characterize the style; multiple descriptions are needed.

Table 1. *Biber's (1988) dimensions*

Dimension	Linguistic Features with Positive Loadings Stronger than .35	Linguistic Features with Negative Loadings Stronger than .35
Dimension 1: Involved versus Informational Production	Private verbs, *that* deletion, contractions, present tense verbs, 2nd person pronouns, *do* as pro-verb, analytic negation, demonstrative pronouns, general emphatics, 1st person pronouns, pronoun *it*, *be* as a main verb, causative subordination, discourse particles, indefinite pronouns, general hedges, amplifiers, sentence relatives, *wh* questions, possibility modals, non-phrasal coordination, *wh* clauses, final prepositions, (adverbs)	Nouns, word length, prepositions, type/token ratio, attributive adjectives, (place adverbials), (agentless passives), (past participial WHIZ deletions)
Dimension 2: Narrative versus Non-narrative Concerns	Past tense verbs, third person pronouns, perfect aspect verbs, public verbs, synthetic negation, present participial clauses	(Present tense) (attributive adjectives)
Dimension 3: Explicit versus Situation-Dependent Reference	*wh* relative clauses on object positions, pied piping constructions, *wh* relative clauses on subject position, phrasal coordination, nominalizations	Time adverbials, place adverbials, adverbs
Dimension 4: Overt Expression of Persuasion	Infinitives, prediction modals, suasive verbs, conditional subordination, necessity modals, split auxiliaries, (possibility modals)	No negative features
Dimension 5: Abstract Non-Abstract Information	Conjuncts, agentless passives, past participial clauses, *by*-passives, past participial WHIZ deletions, other adverbial subordinators	No negative features with loading of at least .35
Dimension 6: On-line Informational Elaboration	*That* clauses as verb complements, demonstratives, *that* relative clauses on object position, *that* clauses as adjective complements	No negative features with loading of .35 or higher

Note. Adapted from Biber (1988, pp. 102-103). Only features with loadings stronger than .35 are displayed. Features in parentheses were not included in Biber's calculation of factor scores, for reasons such as having a heavy loading in more than one factor. Features with stronger loadings are listed before features with weaker loadings.

The steps are detailed in Biber's (1988) *Variation across speech and writing*. The linguistic features in a corpus are tagged and counted, and the frequency counts are normalized. A factor analysis is conducted to determine how the linguistic features cluster together. Each factor has a complementary distribution of features that often occur together and features that rarely occur with

them. The *z*-scores of the linguistic features that have strong loadings with a factor can be added to give a text a "dimension score," and the average dimension score of all the texts in a corpus can be calculated for each factor. Biber believes that linguistic features that have a statistically strong co-occurrence are functionally related—there are communicative reasons that they frequently appear together. Texts that have a very high dimension score for a given factor can be analyzed manually to try to determine the functional reasons behind the clusters that are found. By examining the features used together in context and the communicative functions that they perform, consulting corpus-based grammar books about the functions of each feature when needed, each factor can be assigned a dimension name reflecting the communicative functions behind the co-occurrence of linguistic features. Biber (1988) used this approach to establish patterns in different registers such as conversation, broadcasts, prepared speeches, personal letters, general fiction, press reportage, academic prose, and official documents. His dimensions for register description and the linguistic features with strong loadings (≥.35) in each factor are displayed in Table 1.

Register variation studies using scores on Biber's (1988) original dimensions

A common trend in academic register analysis studies is to describe texts in light of the dimensions that Biber (1988) established by calculating a corpus's factor scores for Biber's dimensions (e.g., Atkinson, 1992; Biber & Finegan, 2001). The features in Biber's (1988) study are tagged in new corpora, and the frequencies are counted, normalized, and standardized based on the means and standard deviations in Biber's reference corpus so that the texts' factor scores for Biber's (1988) dimensions can be calculated. There have been important findings from these studies, and a notable benefit of using Biber's dimensions is that the results can be easily compared to other registers that have been previously analyzed using the same method, like the registers in Biber's (1988) original study and the wide array of registers that have been analyzed using his dimensions since 1988. For example, Conrad (1996) used Biber's (1988) MDA approach to compare a corpus of ecology textbooks, which are the primary form of discipline specific writing that ecology students are exposed to, and ecology research articles, which are the type of writing that ecology students are expected to learn how to produce. Because she used Biber's dimensions, she was able to compare her results with previous MDA studies on general fiction and popular nonfiction texts, which she argues are similar to what students would be exposed to in composition courses. She concluded that there were notable differences between the language used in the texts that students are exposed to in ecology and composition classes and the language that they are expected to produce in research articles, arguing that students would need exposure to examples of the type of RA texts that they needed to write. She later provided a comparison of the ecology texts with history textbooks and research articles, also utilizing Biber's (1988) dimensions (Conrad, 2001). If she had not used Biber's dimensions, she would not have been able to make such direct comparisons between her work and previous work and provide readers with information that they can compare to other studies.

Using Biber's (1988) dimensions in comparative register analysis studies has led to other important findings, such as ways to improve augmentative and alternative communication in occupational settings (Friginal et al., 2013), evidence of improvement in dentistry students' writing before and after instruction (Crosthwaite, 2016), differences in British and American spoken English styles (Helt, 2001), and extreme density of text on official university documents that students need to read (Biber et al., 2002). For initial studies on writing in pulmonology, it may be

helpful to limit the comparison of dimension scores to writing in pulmonology and other disciplines.

The present study

Although there is an established body of literature on medical research articles, studies on pulmonology RAs are still relatively absent from the research available on disciplinary variation. Previous research on writing across disciplines has found different styles and structures across discourse communities, so there is reason to hypothesize that pulmonology RAs would have unique characteristics that differ from other disciplines. Since studies on other registers have successfully provided broad descriptions of multiple characteristics of texts by following Biber's traditions, and Biber's (1988) framework allows for comparison of registers that have previously been studied, exploring pulmonology RAs using Biber's (1988) dimensions is a promising starting point for understanding the unique qualities of this register and providing English teachers with contextualized information about relevant lexico-grammatical features and their functions. The present study uses three (out of the original six) of Biber's (1988) dimensions that are most relevant to academic writing to provide a description of the characteristics of a corpus of pulmonology research articles. A more general corpus of science RAs from different disciplines is also analyzed for comparison purposes to determine how distinct pulmonology writing is from more general scientific writing, with which writing teachers would be likely to be familiar. Although it would also be interesting to compare pulmonology articles to other medical texts, the choice to use a hard sciences corpus as comparison option was driven by the widespread prevalence of general academic English courses and the generalizations in previous literature about writing in hard sciences.

Methods

To explore discipline-specific variation in research articles in the pulmonology domain, scores on Biber's (1988) dimensions were extracted and compared in a corpus of pulmonology research articles and a more general corpus of science research articles. First, the corpora were compiled using Antcorgen (Anthony, 2017). The linguistic features in each text were tagged using Nini's (2014a) Multidimensional Analysis Tagger (MAT), which was also used to calculate the scores for Biber's (1988) dimensions. To determine whether there were significant differences in the two corpora's scores for only three of Biber's dimensions, Mann Whitney tests were conducted using SPSS (Version 28.0). The three dimensions (1 – Informational versus Involved Production, 3 – Explicit versus Situation-Dependent Reference, and 5 – Abstract Non-Abstract Information) were selected for inclusion in this study based on Biber's (1988) descriptions of their relevance to academic writing. The following sections provide details of each step in the procedures used.

Corpora

Two corpora were compiled, one domain specific corpus comprised of pulmonology research articles (i.e., empirical papers), and one general science corpus comprised of research articles from other science fields. Antcorgen (Anthony, 2017), a program that collects research articles in specific disciplines from interdisciplinary Public Library of Science journals (2003-present) makes it feasible to compile very large corpora. A limitation in this study is that the corpora only include research articles published in Public Library of Science journals, and other prominent journals are not represented. However, the limited scope of the corpora could also be seen as a benefit for this

study because the articles in the domain-specific corpus and the general science corpus are from the same publisher, and in many cases the same journals, and they are thus very comparable. Both corpora contained only the bodies of the texts; author information, reference lists, and other text that is not a part of the article body was eliminated using Antcorgen's automatic body extraction feature.

The pulmonology corpus had 8,238 texts (38,092,373 words), and the general science comparison corpus had 7,056 texts (43,379,491 words). The pulmonology articles extracted for the domain specific corpus were from 8 journals: *PLOS Biology*, *PLOS Computational Biology*, *PLOS Clinical Trials*, *PLOS Genetics*, *PLOS Medicine*, *PLOS Neglected Tropical Diseases*, *PLOS ONE*, and *PLOS Pathogens*. The articles in the comparison corpus were from the fields of conservation science, mechanical engineering, analytical chemistry, acoustics, materials physics, and plant genetics. This array of disciplines was chosen to ensure that the corpus would encompass a broad spectrum of fields that had enough Public Library of Science articles available to have a similar size to the pulmonology corpus. There were 1176 texts from each discipline (conservation science, mechanical engineering, analytical chemistry, acoustics, materials physics, and plant genetics) in the comparison corpus, and they were extracted from the ***PLOS Biology, PLOS Computational Biology, PLOS Genetics,*** *PLOS Global Public Health,* ***PLOS Medicine, PLOS Neglected Tropical Diseases, PLOS ONE,*** and ***PLOS Pathogens*** journals (bolded journals are journals that overlap with the journals selected for the pulmonology corpus). To avoid having duplicate articles that fit into more than one discipline or overlap with the pulmonology corpus (such as studies on drugs that treat lung diseases that were categorized under both analytical chemistry and pulmonology disciplines), more than 1176 texts from each discipline were initially collected, and duplicate articles and articles overlapping with pulmonology were eliminated to exclude them from the final corpus.

POS tagging & dimension scores

The linguistic features in each corpus were tagged for part-of-speech using the Multi-Dimensional Analysis Tagger (MAT) (Nini, 2014a), a program that tags all the features used in Biber's (1988) study with 97% accuracy for written texts (Nini, 2014b). It is modeled after the Biber Tagger and follows Biber's (1988) criteria for differentiating between linguistic features. In addition to tagging the linguistic features, MAT also calculates the texts' dimension scores for each of Biber's (1988) dimensions by summing the z-scores of the normalized frequency counts of the linguistic features that have salient loadings for each dimension. The z-scores are calculated using the means and standard deviations from Biber's (1988) work, which has been used as a measure of how a feature's frequency in a text or corpus compares to its frequency in other registers in the English language (Biber & Finegan, 2001). The dimension scores for individual texts and the average dimension scores for the corpus were calculated automatically by MAT, and the z-scores generated by MAT for the calculation of the dimension scores were examined manually to check if any particular features contributed heavily to the dimension scores. For more information about MAT and its limitations, Nini's *Multidimensional Analysis Tagger Manual* with details about tagging and calculations is publicly available online.

Results and discussion

Both corpora most closely matched Biber's (1988) "learned exposition" text type, which describes formal, informational texts with low scores on Dimension 1 and high scores on Dimension 3 and

5. Table 2 summarizes the means and standard deviations of the dimension scores of both corpora. The scores on each dimension and the results of the Mann-Whitney tests are discussed with examples from the corpora in the segments below.

Table 2. *Dimension scores of the Pulmonology and General Science Corpora*

Dimension	Pulmonology Corpus Average Score	Pulmonology Corpus Standard Deviation	General Science Corpus Average Score	General Science Corpus Standard Deviation
1- Informational versus Involved Production	-22.72	2.87	-20.94	3.32
3- Explicit versus Situation-Dependent Reference	6.79	1.64	6.53	1.82
5- Abstract Non-Abstract Information	3.86	1.80	4.66	1.99

Dimension 1: Informational versus involved production

Dimension 1 reflects the purpose of the text (informational or interactive) and the production circumstances (planned or spontaneous) (Biber, 1988). Aligning with the trends reported in other studies on written academic registers (e.g., Biber et al., 2002), the Dimension 1 scores of both corpora are large negative values. The average Dimension 1 score of the texts in the pulmonology corpus was -22.72, and the average Dimension 1 score of the texts in the general science comparison corpus was -20.94. Figure 1 displays the range and mean Dimension 1 scores of the pulmonology and general science corpus in comparison to Biber's (1988) scores for other registers. Biber (1988) explains that the negative features in Dimension 1 (nouns, word length, prepositions, type/token ratio, attributive adjectives) are used to describe information and integrate it into texts, whereas positive features (such as second person pronouns, private verbs, pro-verb *do*) are used for interactive, uncertain, unplanned production.

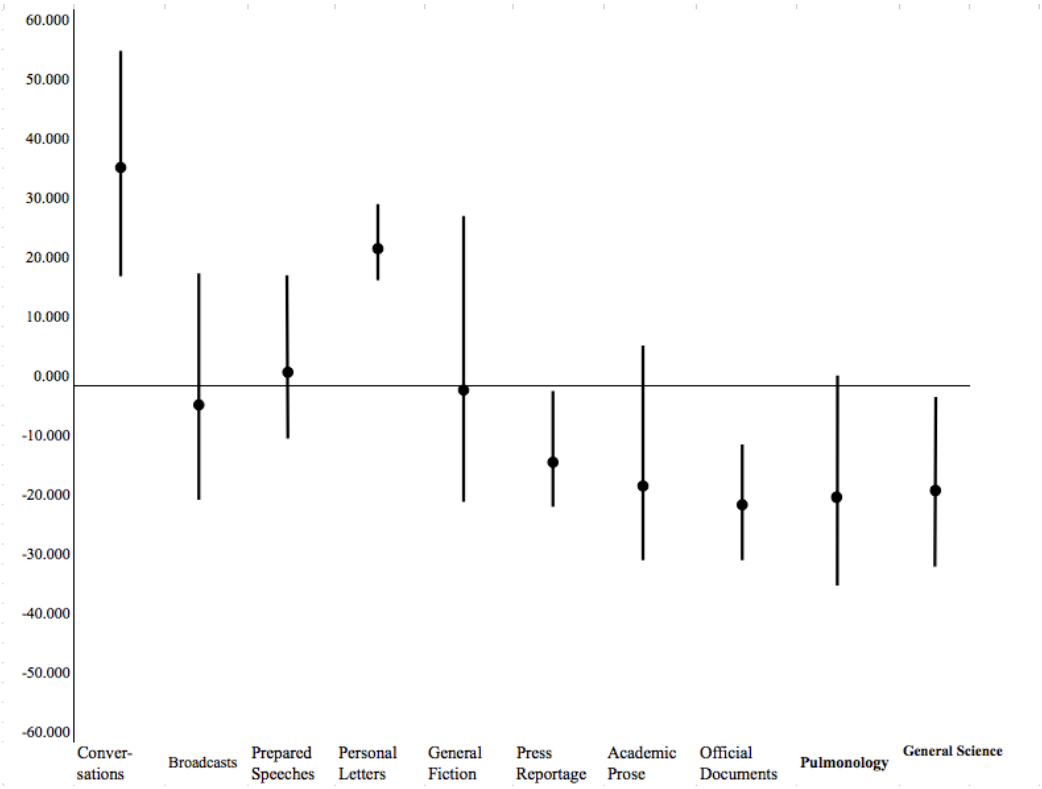

Figure 1. Dimension 1: Pulmonology and General Science Corpus in comparison to other registers

Since nouns carry much of the lexical information in texts, texts with higher frequencies of nouns are informationally denser, and prepositional phrases and attributive adjectives can also provide concise and precise information and descriptions (Biber, 1988, p. 105). High type/token ratios can be expected in texts that are planned with intentional varied vocabulary choices (Biber, 1988, p. 104), and it is reasonable to hypothesize that authors reporting research results would want to choose their words carefully. To determine which linguistic features were driving the large negative dimension scores, the standardized scores (calculated based on the means and standard deviations of Biber's 1988 reference corpus, indicating how the frequencies compare with a range of other registers) used in the automated calculation of the dimension scores were examined. Both corpora had negative standardized scores on all the features that load positively with Dimension 1 except for sentence relatives and non-phrasal coordination, and they had positive standardized scores on all the features that load negatively with Dimension 1. The relatively infrequent use (in comparison to Biber's other registers) of present tense verbs and the *it* pronoun, frequent use of nouns and attributive adjectives, and high average word length (measured in number of letters) in comparison to other registers (Biber, 1988) were especially notable.

The examples below illustrate the dense information packaging in texts from each corpus (numbers marking footnotes for citations in the original texts are eliminated in all the samples), with examples from different sections of the papers. The words contributing to the counts of word classes with negative loadings on Dimension 1 are bolded, with the exception of average word length, as it is not a categorical variable (note that nominalizations are counted separately from "other nouns" in Biber's analyses and in the analyses of this study, and the nouns are marked as "nouns" or "nominalizations" based on Nini's definition of nominalization).

Excerpt A (from a text with a dimension 1 score of -33.19):

"**For** [*preposition*] ***M. tb*** [*noun*], **antimycobacterial** [*attributive adjective*] activity **against** [*preposition*] **strain** [*noun*] **H37Rv** [*noun*] **in** [*preposition*] **vitro** [*noun*] was previously reported **for** [*preposition*] **quinazoline [noun] 2-carboxylate** [*attributive adjective*] **derivatives** [*noun*], more precisely, **thiazoloquinazoline** [*attributive adjective*] **carboxylates** [*noun*]." (Pulmonology Corpus)

In this excerpt from the introduction of an article about drugs for Mycobacterium tuberculosis, the features with negative loadings on Dimension 1 are used together to convey specific detailed information about what chemicals were in the Mycobacterium tuberculosis drug tested, the specific type of tuberculosis it treats, and the type of activity against the strain of tuberculosis that was reported in one succinct sentence. Excerpt B, from the same article introduction, uses features that load negatively with Dimension 1 to give readers detailed information framing the importance of the study discussed.

Excerpt B (from a text with a dimension 1 score of -33.19):

"**Despite** [*preposition*] the **high** [*attributive adjective*] **efficacy** [*noun*] **of** [*preposition*] the **DS-TB** [*noun*] treatment, 490,000 **new** [*attributive adjective*] **cases** [*noun*] **of** [*preposition*] **TB** [*noun*] were reported **in** [*preposition*] 2016 to be resistant **to** [*preposition*] both **RIF** [*noun*] and **INH** [*noun*] and therefore classified as **multidrug-resistant** [*attributive adjective*] (**MDR-TB** [*noun*])" (Pulmonology Corpus; note that MAT does not count "as" as a preposition based on Biber's recommendation to exclude prepositions that have alternate uses falling into other word classes; it is tagged as "preposition or subordinating conjunction")

The details that are concisely packed into this sentence using negative features of Dimension 1 give readers information about the scope of the issue that the study targeted, which could be a strategy to get readers' attention or convince them of the importance of the topic. Excerpt C, taken from an article on pulmonary problems in people working in an environment with a history of water damage, uses the negative Dimension 1 features to concisely summarize information about key findings from the authors' previous publications from their data set.

Excerpt C (from a text with a dimension 1 score of -35.42):

"**In** [*preposition*] **previous** publications using **initial** [*attributive adjective*] **survey** [*noun*] **data** [*noun*], we reported 67 **cases** [*noun*] **of** [*preposition*] **post-occupancy-onset** [*attributive adjective*] **asthma** [*noun*], eight **of** [*preposition*] hypersensitivity **pneumonitis** [*noun*], and six **of** [*preposition*] **sarcoidosis** [*noun*]**."** (Pulmonology Corpus)

In this example, the authors pack the most important information from their previous research into one sentence. Dense sentences were also found in the comparison corpus. The following general science comparison corpus excerpt is from the methods section of a paper on an e-coli outbreak.

Excerpt D (from a text with a dimension 1 score of -30.86):

"A **marker** [*noun*] **peak** [*noun*] based **strategy** [*noun*] **for** [*preposition*] **MALDI-TOF** [*noun*] **MS** [*noun*] **strain** [*noun*] **typing** [*noun*] was evaluated **during** [*preposition*] a **large** [*attributive adjective*] **STEC** [*noun*] **outbreak** [*noun*] **in** [*preposition*] **spring/summer** [*noun*] 2011. **Outbreak** [*noun*] **strain** [*noun*] **specific** [*attributive adjective*] **spectral** [*attributive adjective*] **biomarkers** [*noun*] were discovered by comparison **of** [*preposition*] **reference** [*noun*] **spectra** [*noun*] **from** [*preposition*] **STEC** [*noun*] **outbreak** [*noun*] isolate **TY-2482** [*noun*] (**ATCC** [*noun*] **BAA-2326** [*noun*], **NCBI** [*noun*] **Taxonomy** [*noun*] **ID** [*noun*] 1038844, **BioProject** [*noun*] **accession** [*noun*] **PRJNA67657** [*noun*]) **to** [*preposition*] a **random** [*attributive adjective*] selection **of** [*preposition*] **archived** [*attributive adjective*] **pre-outbreak** [*attributive adjective*] **spectra** [*noun*], which had previously been acquired **for** [*preposition*] **routine** [*attributive adjective*] **MALDI-TOF** [*noun*] **MS** [*noun*]based **species-level** [*attributive adjective*] identification **in** [*preposition*] our **clinical** [*attributive adjective*] **microbiology** [*noun*] **laboratory** [*noun*]." (General Science Corpus)

As seen in this example, explaining procedures in research requires detailed information, and the use of nouns, attributive adjectives, and prepositional phrases allows authors to explain their methods precisely. Excerpt E below, from the general science comparison corpus, is from the discussion section of an article exploring fatigue with vision loss. The authors cite previous studies to relate the results to their own.

Excerpt E (from a text with a dimension 1 score of -33.59):

"A **recent** [*attributive adjective*] **meta-analysis** [*noun*] indicates that **comorbid** [*attributive adjective*] **sleep** [*noun*] **disorders** [*noun*] **in patients** [*noun*] **with** [*preposition*] **medical** [*attributive adjective*] or **psychiatric** [*attributive adjective*] **disorders** [*noun*] can be effectively treated **with** [*preposition*] **CBT** [*noun*]" (General Science Corpus)

This example expresses the main finding of a metanalysis into a single sentence by using the negative features of Dimension 1. While it is not surprising that scientific research articles have constructions that function to express dense information, it may be surprising that the pulmonology texts include more dense text features than research articles in other disciplines. The pulmonology texts (M = -22.72, Mdn = -22.83, SD = 2.87) have significantly lower Dimension 1 scores than the texts in the general science corpus (M = -20.94, Mdn = -21.25, SD = 3.32), $U(N_{\text{Pulmonology}} = 8{,}238, N_{\text{General science}} = 7{,}056) = 19{,}930{,}199.00$, $p < .001$, with a moderate effect size (r = .3). All of the features that have negative loadings with Dimension 1 were more frequent in the pulmonology corpus than in the general science corpus, and sentence relatives, present tense verbs, possibility modals, private verbs, *be* as a main verb, amplifiers, the pronoun *it*, causative subordination, emphatics, *wh* clauses, demonstrative pronouns, *wh* questions, *that* deletion, first person pronouns, final prepositions, non-phrasal coordination, general hedges, pro-verb *do*, and discourse particles, which have positive loadings with Dimension 1, were less frequent in the pulmonology corpus than in the general science corpus. Excerpt F below is a longer sample from a pulmonology text with a very large negative score.

Excerpt F (from a text with a dimension 1 score of -36.14):

"**Pulmonary** [*attributive adjective*] **arterial** [*attributive adjective*] **hypertension** [*noun*] **(PAH)** [*noun*], a **vascular** [*attributive adjective*] **disease** [*noun*] characterized **by** [*preposition*] **persistent** [*attributive adjective*] **pulmonary** [*attributive adjective*] **hypertension** [*noun*] **(PH)** [*noun*] could lead **to** [*preposition*] **progressive** [*attributive adjective*] **right** [*attributive adjective*] **heart** [*noun*] **failure** [*noun*] and **premature** [*attributive adjective*] **death** [*noun*]. **Recent** [*attributive adjective*] **evidence** [*noun*] shows that **abnormal** [*attributive adjective*] **metabolic** [*attributive adjective*] **pathways** [*noun*] may play a **significant** [*attributive adjective*] **role** [*noun*] **in** [*preposition*] the development and **progression** [*noun*] **of** [*preposition*] **PAH** [*noun*]. ... Because **mitochondrial** [*attributive adjective*] **oxidative** [*attributive adjective*] phosphorylation (**with** [*preposition*] **glucose** [*noun*] **uptake** [*noun*] and utilization) has been shown to occur **in** [*preposition*] the **pulmonary**[*attributive adjective*] **arterial** [*attributive adjective*] **endothelium** [*noun*] **of** [*preposition*] **PAH** [*noun*] **patients** [*noun*], **metabolic** [*attributive adjective*] alterations **in** [*preposition*] **PAECs** [*noun*] are more likely to be **representative** [*noun*] **of** [*preposition*] **disease** [*noun*] development." (Pulmonology Corpus)

In this example and in previous examples, the need for specificity appears to drive the heavy use of nouns and attributive adjectives. Pulmonary diseases and their symptoms can be subdivided into smaller categories, and it is clear in the texts in the pulmonology corpus that authors are very specific about what exactly their studies address. That is not to say that the general science corpus did not also have precise reporting, but the pulmonology corpus had long strings of discipline-specific terms with negative dimension 1 features, such as "pulmonary arterial endothelium of PAH patients" which has six words in a row loading negatively with Dimension 1. The texts in

both corpora seem to fit Biber's (1988) idea of informationally dense text very well, and pulmonology texts follow the patterns Biber (1988) observed even more than other scientific research articles. Future studies could examine whether texts in other domains related to medicine display the same trends.

Dimension 3: Explicit versus situation-dependent reference

As shown in the means and ranges displayed in Figure 2, both corpora had positive scores on Dimension 3 (6.79 for the pulmonology corpus and 6.53 for the general science corpus), which is characteristic of texts with context-independent, explicit references. Features that load positively on Dimension 3 include *wh-* clauses, which Biber describes as being used "to specify the identity of referents within a text in an explicit and elaborated manner, so that the addressee will have no doubt as to the intended referent" (1988, p.110). Other positive features are phrasal coordination, nominalizations, and pied piping constructions. Like the negative score on Dimension 1, a positive score on Dimension 3 indicates dense information packaging (Biber et al., 2002).

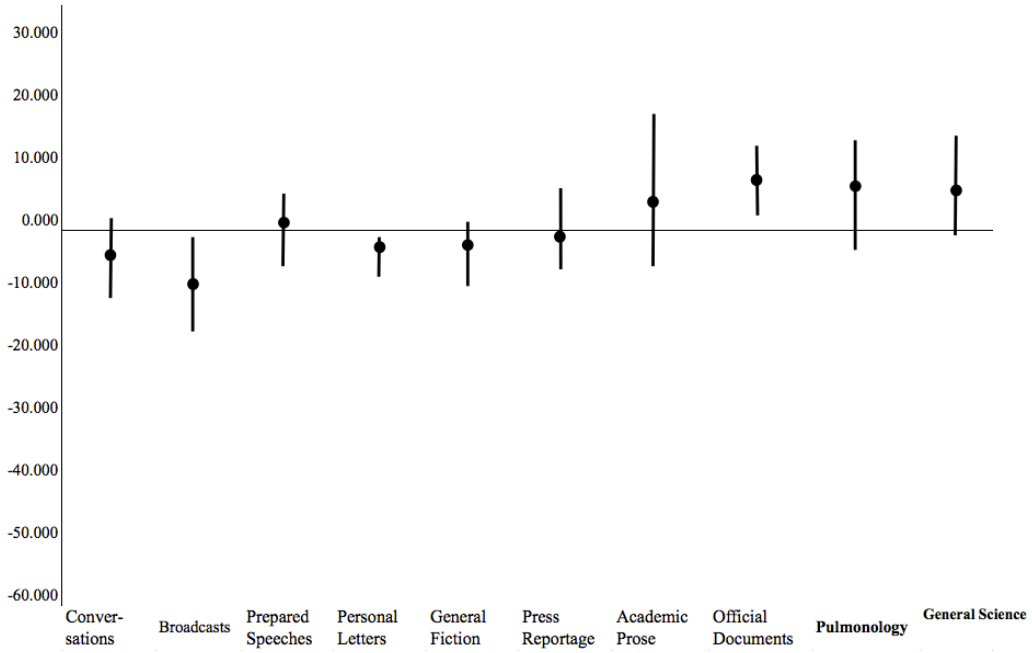

Figure 2. Dimension 3: Pulmonology and General Science Corpus in comparison to other registers

The features driving the positive dimension 3 scores in both corpora were the relatively frequent phrasal coordination and nominalizations and the relatively infrequent use of all of the features with negative loadings. *Wh* relative clauses on object position, pied piping constructions, and *wh* relative clauses on subject position were not used more frequently on average in the pulmonology and general science corpora than in other registers included in Biber's (1988) analysis. Although Dimension 3 is named for the use of explicit or situation dependent references, the name draws on the use of *wh-* clauses, which were actually not very heavily used in the pulmonology and general science corpus. Biber's (1988) description of dense information packing with phrasal coordination and nominalizations in Dimension 3 seems to fit the data in this study more closely. Examples of the phrasal coordination and nominalizations in context can be seen in the excerpts below.

Excerpt G (from a text with a dimension 3 score of 13.82):

> "Furthermore, we recently demonstrated that lung epithelial-specific *Igf1r* deficiency in mice and chronically HDM-challenged IGF1R-depleted mice showed delayed club cell **differentiation** [*nominalization*] **which** [wh-*clause in subject position*] could result in decreased goblet cell hyperplasia **and** [*phrasal coordination*] mucus **production** [*nominalization*]" (Pulmonology Corpus)

In this example from an article about targeting insulin-like growth factor 1 receptor (IGF1R) in mice with asthma that were exposed to house dust mites, the *wh* clause gives specific information, as Biber (1988) describes, about the "differentiation," which is a nominalization, and it contains both a phrasal coordinator and another nominalization inside the clause that add lexical information, contributing to the information density. The following example (Excerpt H) from the general science comparison corpus also packs information into the sentences using nominalizations.

Excerpt H (from a text with a dimension 3 score of 6.39):

> "Recent advances of **instrumentation** [*nominalization*] **and** [*phrasal coordination*] **computation** [*nominalization*] has enabled the simultaneous analysis of a large number of metabolites. Gas chromatography coupled with mass spectrometry (GC-MS) has proven to be an effective **combination** [*nominalization*] for metabolites **identifications** [*nominalization*] **and** [*phrasal coordination*] **quantifications** [*nominalization*] in mammalian cell lines due to its excellent **resolution** [*nominalization*] **and** [*phrasal coordination*] **sensitivity** [*nominalization*]." (General science corpus)

The high frequency of nominalizations, reflected in the Dimension 3 score, and nouns, reflected in the Dimension 1 score, indicate that the texts are rich in lexical information. The Mann Whitney test revealed a significant difference in the two corpora's scores on Dimension 3, $U(N_{\text{Pulmonology}} = 8{,}238, N_{\text{General science}} = 7{,}056) = 26{,}280{,}622.50$, $p < .001$, but the effect size was small ($r = .1$). The texts in the pulmonology corpus ($M = 6.79$, $Mdn = 6.68$) had a higher score than the texts in the general science corpus ($M = 6.53$, $Mdn = 6.41$), with fewer place adverbials and adverbs and more *wh* relative clauses on subject position and phrasal coordination. Excerpt I from an article on the role of respiratory viruses in emergency room and hospital crowding further illustrates how the Dimension 3 features are combined in dense descriptions.

Excerpt I (from a text with a dimension 3 score of 14.33):

> "First, this analysis was done on 2 separate databases **which** [wh-*clause in subject position*] are unlinked and hence a positive virus **detection** [*nominalization*] is not linked with the individual attending the ED. However, the accuracy **and** [*phrasal coordination*] **validity** [*nominalization*] of the database can be corroborated by the peak in ED visits **and** [*phrasal coordination*] **hospitalizations** [*nominalization*] with RTIs but not COPD seen in 2009 due to the H1N1 epidemic." (Pulmonology Corpus)

The *wh* relative clause in this example allowed the authors to elaborate on the details of the databases they were referring to without starting a new sentence. Each sentence contains dense information strung together with phrasal connectors and *wh* clauses.

Dimension 5: Abstract versus non-abstract information

Both corpora had a positive score on Dimension 5, which characterizes abstract, formal, technical discourse like academic prose (Biber, 1988, p. 113). As displayed in the means and ranges charted in Figure 3, the average Dimension 5 score of the pulmonology texts was 3.86 and the average Dimension 5 score of the general science texts was 4.66. The features that load positively on this

dimension are conjuncts, agentless passives, past participial clauses, *by*-passives, past participial WHIZ deletions, and adverbial subordinators. Passives can be used to place focus on the patient and di-emphasize the agent, and as they are common in descriptions of procedures (Biber, 1988, p. 112), it is not surprising that they would be used in scientific research articles, especially in combination with other features used in technical discourse. However, in both corpora, first person pronouns were also common across different journals, which might decrease the need for passive forms for descriptions of procedures. However, passive voice was still common even in articles using first person pronouns (or their corresponding possessive adjectives, which are tagged with the same tag), as in the example, "Subsequently, chromatograms of **our** sequences **were inspected** visually for these nucleotides to check if the chromatogram peak and deduced bases match," (Pulmonology corpus) but active voice was not as uncommon in descriptions of procedures as expected based on Biber's (1988) findings on academic prose.

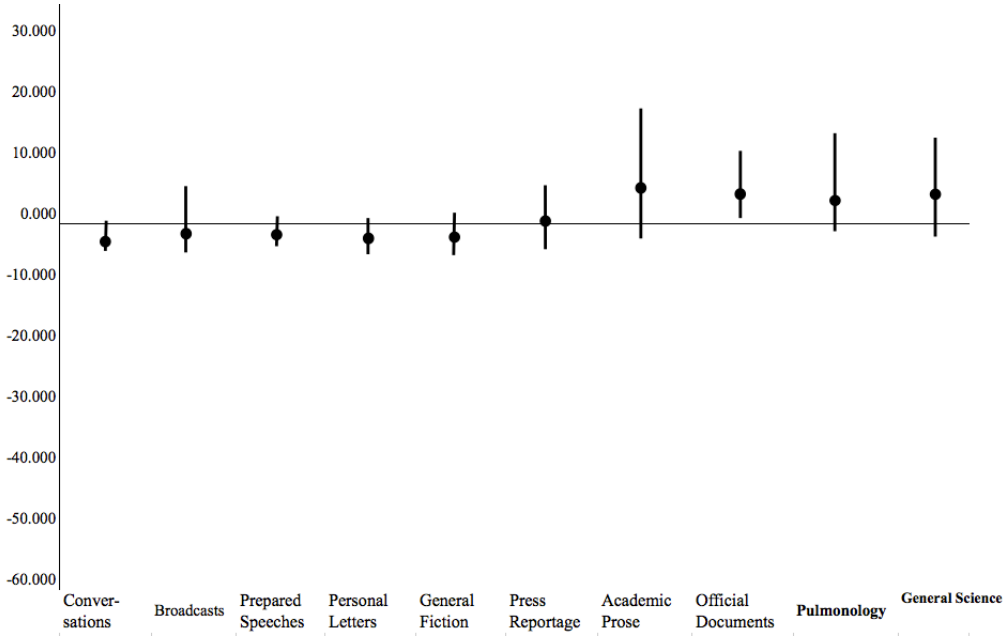

Figure 3. Dimension 5: Pulmonology and General Science Corpus in comparison to other registers

In addition to passive voice, Biber (1988, p. 112) observed that conjuncts and adverbial subordinators can be used together to express complex logical relationships, which are needed in technical discourse. All of the features that load positively with Dimension 5 were more frequent in both corpora than the averages established by Biber (1988) in his corpus of mixed registers, especially conjuncts and past participial clauses. The following excerpts demonstrate how conjuncts, passive forms, past participials, past participial WHIZ deletions, and adverbial subordinators are combined.

Excerpt J (from a text with a dimension 5 score of 15.09):

"In fact, the majority of the A(H1N1)pdm09 cases **were detected** [*agentless passive*] in April, **while** [*adverbial subordinator*] the majority of the A(H3N2) cases **were detected** [*agentless passive*] in March. Fisher's Exact Test for count data **was performed** [*agentless passive*] to test whether **being infected by** [by

> *passive*] A(H1N1)pdm09 and A(H3N2) **is associated** [*agentless passive*] with the month of sampling." (Pulmonology corpus)
>
> **Excerpt K (from a text with a dimension 5 score of 13.79):**
>
> "**As** [*conjunct*] a consequence, the accuracy of segmentation process can **be affected by** [by *passive*] these head motions. **Similarly** [*conjunct*], these movements also affect the statistical properties of signals. **Therefore** [*conjunct*], there is a growing need for the automatic removal of these low frequency components **associated** [past participial WHZ deletion] with head motions from swallowing accelerometry signals." (General science comparison corpus)

In these examples, the conjuncts and adverbial subordinators help to form transitions between the ideas presented, and the passive voice forms relay the information in a technical manner. Although both corpora had positive Dimension 5 scores, there was a significant difference between the Dimension 5 scores in the pulmonology corpus and the general science comparison corpus, $U(N_{\text{Pulmonology}} = 8{,}238, N_{\text{General science}} = 7{,}056) = 22{,}350{,}774.00$, $p < .001$; $r = .2$. The general science comparison corpus ($M = 4.66$, $Mdn = 4.55$) had a significantly higher Dimension 5 score than the pulmonology corpus ($M = 3.86$, $Mdn = 3.78$), indicating that pulmonology research articles are less abstract than the research articles in other scientific domains. The difference in scores was mostly driven by the higher frequencies of conjuncts and adverbial subordinators in the general science corpus.

Comparison to medical research articles

It may also be interesting to note that in comparison to what is known about medical journal articles in general, pulmonology research articles present some unique variation. The dimension 1 score of the pulmonology corpus seems to follow the path that could be predicted from Atkinson's (1992) work. Atkinson had found that medical texts were becoming more informational over time. The latest texts in his corpus were from 1985, and they had an informational score of 18.63 (using the older flipped poles for dimension 1, with positive scores for informational texts and negative scores for "involved" texts, so his 18.63 score equates to a -18.63 score on the scale used in this study and other recent studies with negative scores for informational texts and positive scores for "involved" production). The pulmonology corpus, with recently written texts, had a score of -22.72, following the increasing information-density trend. In terms of dimension 2, there was not a stable pattern over time in Atkinson's data. His score for 1985 medical texts was 2.90, while the score for the pulmonology corpus was 6.79, indicating that the pulmonology corpus has more context-independent, explicit references. In terms of dimension 5, the pulmonology corpus goes against the trend that would have been predicted from Atkinson's work. Dimension 5 scores in Atkinson's corpus were increasing from 1864-1985, with a score of 5.03 in 1985. However, the pulmonology corpus had a lower score of 3.86, indicating that the texts are less abstract than the medical articles in Atkinson's corpus. Although there were significant differences between the scores for the pulmonology and general science corpus in this study, the pulmonology corpus was more similar to the general science corpus than Atkinson's medical research article corpus, perhaps because of the changes in writing conventions over time.

Pedagogical implications

Although more research is needed to have a better understanding of the pulmonology RA domain, some pedagogical implications can be drawn from the results of this study. As the pulmonology corpus's scores on each dimension were closer to the general science corpus's scores than they were to other registers that Biber (1988) reported on, general science writing classes have the potential to move pulmonology students' writing style in the right direction, but the differences between the dimension scores of the pulmonology and general science RA corpora suggest that the nuances of pulmonology writing are unique from other scientific disciplines. Significant differences were found between the pulmonology and general science corpus's scores on Dimensions 1, 3, and 5, and the effect size was medium for Dimension 1. Pulmonology has denser information packaging, and although the texts had abstract information, it was to a lesser degree than other science disciplines. These differences suggest that English for Academic Purposes materials that have been very well designed to reflect scientific writing in general may not necessarily be as reflective of the style of writing that students in fields that do pulmonology research would need to learn.

Academic style may seem like an abstract concept, but knowing concrete groups of features that are used for shared communicative purposes can help teachers understand the writing style in pulmonology research articles. Biber's dimensions provide an operationalization of the communicative functions needed for the register, which can be useful for materials development. The pulmonology research articles were found to be informationally dense based on their Dimension 1 and 3 scores, and frequently occurring features that were key contributors to the dense information packing were nouns (Dimension 1), phrasal coordination, and nominalizations (Dimension 3), which were used more on average in the pulmonology corpus than they are in other registers that Biber (1988) analyzed. With this information, academic writing teachers can try to teach pulmonology students nouns that are common in pulmonology texts, perhaps drawing from word families from the pulmonology/respiratory medicine section of Hsu's (2013) medical vocabulary list, which may also be helpful for finding common nominalization morphemes used in discipline-specific vocabulary. Instructors can also prioritize teaching students how to connect phrases with coordinators correctly. Since the pulmonology texts were on the abstract side of the Dimension 5 continuum, with relatively high frequencies of past participial clauses and conjuncts, writing curricula for pulmonology students could include the grammar of forming past participial clauses and using conjuncts, as well as vocabulary options for conjuncts.

Limitations and future directions

The results suggest that the pulmonology domain has unique styles that merit further investigation. There are some limitations in this study that could be addressed in future research. One limitation is the scope of the corpus. Only including articles from the Public Library of Science Journals in this study allowed for maximum comparability between the pulmonology corpus and the general science corpus. However, to continue to study the domain of pulmonology RAs, future studies could analyze corpora of articles from other journals and publishers and determine whether similar trends are found outside of Public Library of Science journals. A manually compiled corpus rather than an automatically compiled corpus could be useful in future studies because researchers can intentionally select articles to include, but manually compiled corpora have size limitations. However, with size limitations, tagger accuracy limitations can be more feasibly addressed. The 97% tagging accuracy of the MAT (Nini, 2014a; 2014b) was accepted for this study, but in studies with smaller corpus sizes, some of the possible tagging errors could be corrected.

It would also be beneficial to build off of the findings of this study by looking further into the structures that were found to be frequent. For example, conjuncts were more frequent in pulmonology RAs than in other registers, and it could be helpful to compile a list of the commonly used conjuncts for academic writing teachers and materials developers to reference. Future exploratory studies on pulmonology RAs could also use a new factor analysis instead of using Biber's (1988) dimensions to add another perspective of describing the register. Using Biber's (1988) dimensions had the advantage of allowing comparisons to be made to other registers of English but conducting a new factor analysis could reveal new patterns of co-occurrences. Another option for future studies could be to apply Gray's (2013) dimensions rather than Biber's (1988) to see how pulmonology research articles compare to the research articles in her study. Subcorpora of different types of empirical studies (Gray, 2013) could also be compared in future analyses. It has also been recommended to triangulate MDA with other methods, such as move analysis, to find variation that might be overlooked in an MDA (Atkinson, 1992; Conrad, 2014). Future studies could apply other methods of studying register variation to add different types of descriptions of pulmonology RAs to the literature.

Conclusion

By calculating scores for Biber's (1988) dimensions, it was determined that in comparison to other registers, pulmonology research articles have dense information packing and a technical writing style with frequent nouns, phrasal coordination and conjuncts, and past participial clauses. Each individual dimension does not provide adequate information to describe the pulmonology RA domain, but Dimensions 1, 3, and 5 together reflect multiple facets of the RA writing style in the pulmonology research discourse community. In comparison to science RAs from other disciplines, pulmonology RAs were denser and less abstract. The results suggest that pedagogical materials developed specifically for the pulmonology domain could be worthwhile. In addition, the differences between the pulmonology RA corpus and the general science RA corpus suggest that further investigation of pulmonology RAs is merited. Future research could explore the unique qualities of pulmonology RAs further using other approaches to provide a fuller description of the pulmonology RA register.

Disclosure Statement

No potential conflict of interest was reported by the authors.

References

Abbott, G. (1980). ESP and TENOR. *ELT Documents*, 107, 122–4.

AbuNurah, H. Y., Zimmerman, R. D., Murray, R. B., & Gardenhire, D. S. (2020). The impact of respiratory therapy international education in the United States: Saudi international students' perceptions. *Respiratory Care*, *65*(7), 961-965. https://doi.org/10.4187/respcare.06961

Anthony, L. (2017). AntCorGen (Version 1.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp

Atkinson, D. (1992). The evolution of medical research writing from 1735 to 1985: The case of the *Edinburgh Medical Journal*. *Applied Linguistics*, *13*, 337–374. https://doi.org/10.1093/applin/13.4.337

Bazerman, C. (1981). What written knowledge does: Three examples of academic discourse. *Philosophy of the Social Sciences*, *11*, 361-388. https://doi.org/10.1177/004839318101100305

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Biber, D., Connor, U., & Upton, T. A. (2007). Discourse on the move: Using corpus analysis to describe discourse structure. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/scl.28

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly*, *36*, 9-48. https://doi.org/10.2307/3588359

Biber, D., & Finegan, E. (2001). Intra-textual variation within medical research articles. In S. Conrad & D. Biber (Eds.) *Variation in English: Multi-dimensional studies* (pp. 108-123). Longman.

Bruce, I. (2008). Cognitive genre structures in methods sections of research articles: A corpus study. *Journal of English for Academic Purposes*, *7*, 38-54. https://doi.org/10.1016/j.jeap.2007.12.001

Conrad, S. (1996). Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education*, *8*, 299–326. https://doi.org/10.1016/S0898-5898(96)90025-X

Conrad, S. (1999). The importance of corpus-based research for language teachers. *System*, *27*, 1–18. https://doi.org/10.1016/S0346-251X(98)00046-3

Conrad, S. (2001). Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 94–107). Pearson Education.

Conrad, S. (2014). Expanding multi-dimensional analysis with qualitative techniques. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Mult-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 297-316). John Benjamins Publishing.

Cortes, V. (2013). Corpora in language for specific purposes research. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-7). Wiley Blackwell. https://doi.org/10.1002/9781405198431.wbeal0225

Crosthwaite, P. (2016). A longitudinal multidimensional analysis of EAP writing: Determining EAP course effectiveness. *Journal of English for Academic Purposes*, *22*, 166–178. https://doi.org/10.1016/j.jeap.2016.04.005

Crosthwaite, P., Cheung, L., & Jiang, F. (2017). Writing with attitude: Stance expression in learner and professional dentistry research reports. *English for Specific Purposes*, *46*, 107–123. https://doi.org/https://doi.org/10.1016/j.esp.2017.02.001

Djamaa, S. (2013). Scientific English in the EFL classroom: Rethinking our pedagogies. *Journal of Language Teaching and Research*, *4*(5), 939-952. https://doi.org/10.4304/jltr.4.5.939-952

Durrant, P., Brenchley, M., & McCallum, L. (2021). *Understanding development and proficiency in writing. Quantitative corpus linguistic approaches*. Cambridge University Press. https://doi.org/10.1017/9781108770101

Ervin-Tripp, S. (1972). On sociolinguistic rules: Alternation and co-occurrence. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 213–250). Holt, Rinehart & Winston.

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, *28*(2), 414–420. https://doi.org/10.2307/3587446

Friginal, E., Pearson, P., Di Ferrante, L., Pickering, L., & Bruce, C. (2013). Linguistic characteristics of AAC discourse in the workplace. *Discourse Studies*, *15*(3) 279–298. https://doi.org/10.1177/1461445613480586

Friginal, E., Li, M., & Weigle, S. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing*, *23*(2), 1–16. https://doi.org/10.1016/j.jslw.2013.10.001

Friginal, E., & Mustafa, S. (2017). A comparison of U.S.-based and Iraqi English research article abstracts using corpora. *Journal of English for Academic Purposes*, *25*, 45-57. https://doi.org/10.1016/j.jeap.2016.11.004

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, *9*(2), 123–145. https://doi.org/10.1016/S1060-3743(00)00019-9

Gray, B. (2013). More than discipline: Uncovering multi-dimensional patterns of variation in academic research articles. *Corpora*, *8*(2), 153-181. https://doi.org/10.3366/cor.2013.0039

Gray, B., Cotos, E., & Smith, J. (2020). Combining rhetorical move analysis with multi-dimensional analysis: Research writing across disciplines. In Roemer, U., Cortes, V., & Friginal, E. (Eds.), *Advances in corpus-based research in academic writing: Effects of discipline, register, and writer expertise* (pp. 137–168). John Benjamins Publishing Company. https://doi.org/10.1075/scl.95.06gra

Hagan, G. D. (1996). Revision pedagogy in the technical writing of respiratory care students. *Respiratory Care Education Annual*, *5*, 19-30.

Halliday, M. A. K. (1988). On the language of physical science. In M. Ghadessy (Ed.), *Registers of written English* (pp. 162–178). Pinter.

Helt, M. (2001). A multi-dimensional comparison of British and American spoken English. In S. Conrad, & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 171-183). Longman.

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, *17*(4), 454–484. https://doi.org/10.1177/1362168813494121

Nini, A. (2014a). Multidimensional Analysis Tagger (Version 1.3.1) [Software]. Available from: http://sites.google.com/site/multidimensionaltagger.

Nini, A. (2014b). The Multi-Dimensional Analysis Tagger. In T. Berber Sardinha & M. Veirano Pinto (Eds), *Multi-Dimensional Analysis: Research Methods and Current Issues (pp.* 67-94). Bloomsbury Academic. https://doi.org/10.5040/9781350023857.0012

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, *16*(2), 119-138. https://doi.org/10.1016/s0889-4906(97)85388-4

Römer, U., Cortes, V., & Friginal, E. (2020). *Advances in corpus-based research in academic writing: Effects of discipline, register, and writer expertise*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.95

Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in Medical English abstracts. *English for Specific Purposes*, *11*, 93-113. https://doi.org/10.1016/S0889-4906(05)80002-X

Souza, R., & de Carvalho, C. R. R. (2014). Brazilian Journal of Pulmonology and Portuguese Journal of Pulmonology: Strengthening ties in respiratory science. *Portuguese Journal of Pulmonology*, *20*(6), 285-286. https://doi.org//10.1016/j.rppneu.2014.11.001

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

Watson, M. E., & Low, G. G. J. (1988). Reading-level evaluation of respiratory physiology textbooks. *Respiratory Care*, *33*(10), 937-938.

Watson, M. E., & Scully, S. D. (1988). A survey of reading-level assessment methods and determination of reading level of selected respiratory therapy textbooks. *Respiratory Care*, *33*(11), 1020-1026.

Copyrights