

**Variation in use of nouns as nominal premodifiers in advanced student writing
across academic disciplines**

by

Thomas B. Elliott

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Major: Teaching English as a Second Language/Applied Linguistics (Corpus
Computational Linguistics)

Program of Study Committee:
Bethany Gray, Major Professor
John Levis
Jo Mackiewicz

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

Copyright © Thomas B. Elliott, 2019. All rights reserved.

DEDICATION

To Samuel, who graciously waited until my thesis defense was finished before
coming into this world.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES.....	v
ACKNOWLEDGMENTS	vi
ABSTRACT.....	vii
CHAPTER 1. INTRODUCTION	1
CHAPTER 2. LITERATURE REVIEW	8
2.1 Complexity in Academic Writing.....	8
2.2 Disciplinary Variation	10
2.3 Nouns as nominal premodifiers.....	12
2.4 Complexity and Variation in Student writing.....	14
2.5 Research Questions.....	16
CHAPTER 3. METHODS	19
3.1 Corpus.....	19
3.2 Corpus search process	24
3.3 Analysis	27
CHAPTER 4. RESULTS.....	30
4.1 Frequency of Nouns with One or More Nouns as Premodifiers Across Disciplines	30
4.2 Patterns Across Disciplinary Groupings.....	31
4.3 Frequency of Noun Strings by Length	33
4.4 Noun Strings by Becher-Biglan Disciplinary Groupings	35
4.5 Components of Noun + Noun Sequences and their Meaning Relationships.....	37
4.6 Use of Multiple Nouns as Nominal Premodifiers	43
4.7 Analysis of Head and Premodifying Noun Types	47
CHAPTER 5. DISCUSSION.....	54
5.1 Limitations.....	55
5.2 Implications	56
REFERENCES	58
APPENDIX: GROUPING DISCIPLINES BY UNIVERSITY OF MICHIGAN DEPARTMENT AND BECHER-BIGLAN TYPOLOGY	62

LIST OF FIGURES

	Page
Figure 4.1 Frequency of nouns premodified by one or more nouns.....	31
Figure 4.2 Frequency of noun phrases with at least one noun as nominal pre- modifier in the Becher-Biglan typology academic groupings (normalized per 10,000 words).....	32
Figure 4.3 Frequency (per 10,000 words) of noun + noun premodification sequences by length and discipline.....	34
Figure 4.4 Noun string length frequency (per 10,000 words) by Becher-Biglan disciplinary groupings	36
Figure 4.5 Percentage of all nouns premodified by nouns represented by each string length	37

LIST OF TABLES

	Page
Table 3.1 Number of texts and total word count by discipline	20
Table 3.2 Number of texts by discipline and paper type	21
Table 4.1 Frequency (per 10,000 words) of noun + noun premodification sequences by length and discipline, and percentage of total number of nouns with noun premodifiers.....	34
Table 4.2 Frequency (per 10,000 words) of noun + noun premodification strings by length and discipline grouping	35
Table 4.3 Type-token ratios for Becher-Biglan groupings	38
Table 4.4 Twenty most frequent nominal premodification strings for each Becher- Biglan category (normalized per 10,000 words)	39
Table 4.5 Most frequent head nouns premodified by nouns in each Becher-Biglan category (<i>frequencies normalized per 10,000 words</i>)	47

ACKNOWLEDGMENTS

I would like to first thank my committee chair Dr. Bethany Gray for all of your support and patience throughout this project and its many drafts. Thank you as well to my committee members Dr. John Levis and Dr. Jo Mackiewicz for your guidance and valuable input during my defense. Also, thank you to Kimberly Becker and Erin Today at the Center for Communication Excellence for your helpful feedback on my early drafts.

I want to give special thanks to my family. We've been through quite a year, and I appreciate your patience with me while I was busy and distracted with this project. Last but not least, to my wife Adry, thank you for your support, understanding, and kringlas.

ABSTRACT

Academic writing possesses many characteristics that distinguish it from other registers, such as heightened information density, elaborated reference, and an impersonal style (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Biber & Gray, 2016). However, disciplines manifest these language features differently (Biber & Gray, 2013), posing additional challenges to students as they acquire the specialized language of their field. For these reasons, understanding how student writing varies by discipline is essential for both pedagogy and research in writing development.

One of the most important distinguishing features of academic writing is the frequent use of complex noun phrases, which consist of a head noun modified by one or more premodifiers or postmodifiers (Biber et al., 1999; Biber, Grieve, & Iberri-Shea, 2009; Biber & Gray, 2013). Among these noun phrase complexity features, the use of nouns as nominal premodifiers is particularly challenging due to the informational density they enable, the variety of meaning relationships they can express, and their potential ambiguity (Halliday, 1989; Biber & Gray, 2011; Wisniewski, 1996). Understanding how this type of complexity is used by student writers is therefore important in helping students meet the disciplinary communicative norms and expectations which are key to their advancement. To help achieve this goal, this study presents an analysis of the Michigan Corpus of Upper-Level Student Papers (MICUSP) examining the use of nouns premodified by nouns across 16 disciplines and four groupings based on the Becher-Biglan typology for identifying disciplinary families.

Results show that advanced student writing exhibits discipline-based variation in use of nouns with nouns as premodifiers in frequency, number of premodifying nouns, and type. Overall, findings align with previous work on disciplinary variation in other

varieties of academic writing, including research articles and textbooks (Biber & Gray, 2013; Gray, 2015; Jalilifar, White, Malekizadeh, 2017), and lower-level student writers (Musgrave & Parkinson, 2014), showing higher frequencies of nouns premodified by nouns in the hard sciences than in the soft disciplines. Results support work on development of complexity in student writing (Biber, Gray & Poonpon, 2011) and suggest that upper-level student writing may be useful as a more accessible model for less proficient writers as they advance.

Keywords: noun phrase, premodifiers, student writing, disciplinary variation

CHAPTER 1. INTRODUCTION

Acquiring proficiency in academic writing is an essential element of student success at the university level, and specialized linguistic knowledge influenced by situational variables such as discipline, genre, and level is required to achieve that proficiency (Hyland, 2002; Zhu, 2004; Gardner, Nesi & Biber, 2018). Even when taking into consideration these variables, however, writing in the university setting has been shown to possess general characteristics that distinguish it from other registers, such as heightened information density, elaborated reference, and an impersonal style (Biber, Conrad, Reppen, Byrd, & Helt, 2002).

Mastery of academic writing is a skill that takes many years to acquire, and even for students who are native speakers of English, this register can create confusion due to the informational density and structural features that distinguish it from other registers of English (Biber, Gray & Staples, 2016). These features of academic prose are "not acquired naturally, and many native speakers of English rarely (or never) produce language of this type" (Biber, Gray & Poonpon, 2011, p. 29). This specialized language can result in feelings of alienation from students, both native and non-native, as they encounter forms of expression far removed from their familiar day-to-day communication (Martin & Halliday, 1993). Because of the challenges involved in acquiring this essential skill, it is important to understand the linguistic features that characterize academic writing. Effectively describing these features can be useful in guiding both first and second-language learners as they work to master this register in English for Academic Purposes (EAP) contexts, as well as other secondary and post-secondary academic writing instruction environments.

One of the most important features distinguishing the academic register from others is the frequent use of complex noun phrases, which consist of a head noun modified by one or more premodifiers or postmodifiers (Biber, 1989; Biber & Gray, 2010). While spoken

language makes greater use of clausal complexity, including features such as relative clauses and complement clauses (see Ex.1, from the Michigan Corpus of Academic Spoken English), written registers, particularly academic writing, have been shown to feature more use of complex noun phrases with extensive pre- and post-modification of head nouns (Ex. 2, from the Michigan Corpus of Upper-Level Student Papers).

- (1) *Who wants to tell me what they think this is?*
- (2) *The social justice aspect of ICE production and use echoes that of the economic view.*

The present study is focused on premodifiers in particular because in academic writing, nominal complexity is expressed most often by pre-modification, with around 25% of all noun phrases containing a pre-modifier, and 20% a post-modifier. (Biber, et al., 1999, p. 578). Noun premodifier types include attributive adjectives (*medical information*) participial modifiers (*surrounding areas*), and other nouns (*substance abuse*). Among these types of premodification, the use of nouns as nominal premodifiers account for approximately 30% of all premodification in academic prose (Biber et al., 1999), making them an important feature for student writers to acquire and apply in their own writing. However, this form of nominal complexity may be particularly challenging for learners because the resulting noun + noun sequences offer no clear indication of the meaning relationship between the two (or more) elements due to a lack of function words to make the logical connections explicit (Wisniewski, 1996). For example, the noun + noun sequence *steel mill* could mean a mill that produces steel, or a mill that is made of steel. The meaning

relationship between the two words is not explicitly stated and requires contextual knowledge to decode.

The ambiguity in noun premodification relationships increases when multiple nouns are used as premodifiers, which adds additional burden to readers and suggests that occurrence of longer noun premodification strings may be an important consideration when examining development of nominal complexity in student writing. Complexity is added as the modification relationships increase, and it is rare for all premodifiers to modify the head noun directly (Biber et al., 1999, p.597). These complex noun phrases facilitate the dense prose style that characterizes academic writing by allowing more elaborated clausal structures to be compressed into the noun phrase. This can be illustrated by rephrasing a complex noun phrase using more explicit clausal structures. Example 3, taken from a Mechanical Engineering paper in the MICUSP corpus, contains two head nouns premodified by one or more other nouns. To best illustrate the density of information encoded in noun + noun premodification strings, we will focus on the first one (bolded).

(3) *There is no one **material strain energy function** for a nonlinear elastic material. (MEC)*

In this noun group, we find three premodifying nouns preceding the head noun *function*. Correctly identifying the premodifying relationships may require disciplinary knowledge due to the loss of explicitness in this type of structure, but one way of unpacking the string *material strain energy function* might be *the relationship between the amount of strain in a material and the energy required to produce it*. It is clear that using nouns as premodifiers is a more efficient way of expressing the idea. Although this compression of information introduces ambiguity, this presents less of a problem in academic writing, which is generally intended for a specialist audience. Sequences of multiple premodifying nouns

such as this represent an important development in 20th century academic writing as the register became increasingly specialized (Biber & Gray, 2016), which makes them a key feature for students to acquire in their own writing as they work to join an academic community.

Another important characteristic of noun phrase complexity in academic writing is the use of nominalized forms (Biber & Gray, 2013), or nouns derived from other parts of speech such as verbs and adjectives (e.g. *argument*, *beauty*), often using derivational suffixes such as *-ity* and *-ment*. This is an important tool in increasing informational density, as it allows clausal processes to be reduced to a noun, which can then be discussed or characterized directly as the head of a new clause, increasing cohesion and informational density, and allowing the writer to emphasize or de-emphasize different parts of the message (Halliday & Matthiessen, 2014; Jalilifar et al., 2017; Ravelli, 1988). The use of nominalized forms to premodify other nouns has grown considerably in 20th century academic writing, increasing the number and complexity of meaning relationships in noun sequences (Biber & Gray, 2016).

Knowledge about the use of complex noun phrases is particularly important for student writers as they work to acquire the language features of academic writing and their particular discipline of study. In a study of complexity features in professional academic writing, Biber, Gray and Poonpon (2011) hypothesized developmental stages in student writers, moving from conversation-like clausal elaboration to complex noun phrases. Building on that study, Musgrave & Parkinson (2014) found that upper-intermediate level L2 students use nouns as nominal premodifiers less than 'typical' academic writing. The students in their study were "relatively new to academic texts in English", so their writing

showed an expected lower frequency of noun phrase complexity, specifically overuse of attributive adjectives and underuse of nouns as nominal premodifiers.

While much research has been done on the characteristics of academic writing and the type of complexity it contains, it is important to also consider the linguistic variation that exists across the various disciplines under the larger umbrella of academic writing. Studies of student writing have found disciplinary variation in citation practices (Adel & Römer, 2012), lexical bundles (Durrant, 2017), cohesion markers (Crossley, Russel, Kyle & Römer, 2017) and grammatical complexity (Staples, Egbert, Biber & Gray, 2016). Overall, it has been shown that academic disciplines often use very different language features, and discipline influences language characteristics even more than the writer's first language and cultural background (Adel & Römer, 2012). Therefore, student success depends on understanding how writing functions within a discipline and how content is produced, communicated, and critiqued within that academic community (Fang, 2012). Taking into account these disciplinary variations is essential to achieving particular learning objectives (Neuman, Parry & Becher, 2002), and teaching of discipline-specific writing skills has been shown to be a more successful approach than instruction in general academic writing (Hyland, 2002).

The variety of disciplines that exist in the higher educational context have often been grouped and studied according to the Becher-Biglan typology (Neuman, et al., 2002), which categorizes academic fields along the axes of Hard-Soft and Pure-Applied, which are defined by characteristics such as quantitative or qualitative focus, gregariousness or independence, and concern with universals or particulars. By grouping similar disciplines in this analysis rather than examining them individually we may gain insight into broader patterns in academic language use, as it would likely be impractical for each disciplinary subdivision in

a university to develop its approach to writing pedagogy. Although the Becher-Biglan categories are based on characteristics other than language usage, grouping disciplines in this way may offer valuable information when analyzing language use variation patterns because the features used to create these categories may relate to relevant situational variables, such as article type and research methodology, which have been shown to be an important influence on language use patterns (Gray, 2013). In addition, consolidating disciplines in this way offers a more practical source of information for efficiently and effectively teaching writing to students from different academic programs.

Nouns as nominal premodifiers have been explored in detail in registers such as conversation, fiction, news, and academic prose (Biber et al., 1999; Biber & Gray, 2013), as well as variation across academic disciplines (Gray, 2015; Biber & Gray, 2016; Staples et al., 2016). In addition, noun phrase complexity has been examined in the writing of intermediate-level university English learners in EAP (English for Academic Purposes) courses (Musgrave & Parkinson, 2014). However, less attention has been paid to the use of this feature in advanced university student writing, particularly as it varies across disciplinary divisions. Examining nouns as nominal premodifiers in academic writing from advanced L1 students may offer additional useful information, helping to "capture development toward the complex styles of academic writing" (Biber, Gray & Poonpon, 2011, p.32). Their importance in organizing academic discourse and increasing informational density, as well as the challenge they present to novice members of the academic writing community suggest the need for further analysis of how students use this type of nominal complexity in disciplinary contexts. Given the important disciplinary variables shown in previous studies, a more detailed description of frequency, length, and type differences of nouns premodified by

nouns in student writing across disciplines may provide valuable information for both discipline-specific writing research and pedagogy.

For these reasons, the present study seeks to discover how the use of nouns as nominal premodifiers in advanced student writing varies along traditional academic departmental divisions as defined using the Becher-Biglan typology. To examine patterns of variation in frequency and length of noun strings, as well as the particular nouns that occur in these constructions, corpus linguistics methodology was used, allowing specific grammatical features to be targeted across a large collection of texts. The Michigan Corpus of Upper-Level Student Papers (MICUSP) (Römer & O'Donnell, 2011) was chosen because it provides a large, representative sample of student writing across a variety of disciplines.

This paper is structured in five chapters, with the present introduction as Chapter 1. Section 2 reviews the literature to give an overview of previous work relevant to the topic of this study. Section 3 addresses the methodology of the project, with a description of the choice of corpus and the tools used in its analysis. Section 4 presents the quantitative and qualitative findings and discusses their relationship to the research questions and previous work in the area. Finally, Chapter 5 is a discussion of the findings, the limitations of the study, and implications for future research.

CHAPTER 2. LITERATURE REVIEW

The present chapter will give an overview of previous research related to the topic of noun phrase complexity in student writing. The characteristics of complexity in academic writing will be addressed in the first section. Next, differences in complexity features across disciplinary divisions are examined. The third section reviews the use of nouns as nominal premodifiers, and finally complexity in student writing is addressed in the final section.

2.1 Complexity in Academic Writing

Complexity in academic writing has been a frequent topic of previous research in applied linguistics. Important initial steps were taken by researchers who examined the differences between complexity in academic writing and other registers. Halliday (1989) contrasted the types of complexity found in spoken and written language, pointing out the use of noun phrase elaboration to increase informational density in academic writing, as opposed to spoken language's more frequent use of complex clauses. Using multidimensional analysis techniques with large corpora, Biber (1989) provided empirical evidence of these register differences, showing clausal elaboration to be more common in spoken communication, and noun phrase complexity more characteristic of informational writing. This focus on simpler clause structure and more noun phrase modification results in a 'compressed' style of discourse in academic writing, which is efficient for experts but results in less explicit meaning relationships that may present difficulties for less experienced readers (Biber & Gray, 2010). This is especially true in the case of noun phrases with multiple premodifiers, which offer an efficient compression of information, but increase difficulty for readers as premodification relationships become more complex (Biber et al., 1999).

Although the importance of noun phrase complexity as a distinguishing characteristic of academic writing has been demonstrated in previous research, clausal measures such as the T-unit have often been used in assessing complexity of students' academic writing. In an overview of approaches to measuring language development, Wolfe-Quintero et al. (1998) suggested that clausal measures, such as clauses per T-unit and dependent clauses per independent clause, were the most effective means of discriminating between learner levels. That suggestion was reflected in the wide use of clausal measures in a survey of second language instruction research by Ortega (2003), which pointed out the need for more varied complexity measures, including both phrasal and clausal features.

Later research by Biber, Gray and Poonpon (2011) examined grammatical complexity in professional academic writing to determine the best features for evaluating levels of complexity in student writing. The study argues that clausal subordination measures such as length of T-unit fail to distinguish writing proficiency differences. The authors use corpus methodology to provide evidence that clausal complexity is in fact more characteristic of conversation than writing and that complex noun phrases are a more important feature of academic writing. Based on these observations, they suggest a path of development in which writers begin with a more clausally elaborated style characteristic of conversation and gradually use more phrasal complexity as they gain mastery of formal academic writing. This contrast between spoken language and academic writing was explored in more detail by Biber and Gray (2016), which showed the increasing importance of nominal complexity in academic writing in the last century.

Although academic writing as a whole has been shown to possess distinguishing linguistic features such as increased noun phrase complexity, it is also important to consider the variation that exists across the disciplines within the register in order to best prepare

student writers for the communicative needs of their particular fields. The following section will review previous work on the topic of variation in language features across disciplinary lines.

2.2 Disciplinary Variation

The frequent use of complex noun phrases is a key distinguishing feature of academic writing, creating an informationally-dense and specialized type of discourse whose complexity is concentrated in noun phrases rather than the clausal elaboration typical of conversation. However, the degree of nominal complexity that appears in a piece of academic writing has been shown to vary based on the discipline to which it belongs.

In a study of complexity in professional academic writing across both discipline and historical periods, Biber and Gray (2016) indicate several important differences between science and humanities. Especially relevant to the present study is the description of nouns as premodifiers, which are shown to have increased dramatically in the 20th century. This change included the extension of possible meaning relationships between nouns and their premodifiers and increased occurrence of strings of two or more premodifying nouns. Both of these elements increased noun phrase complexity considerably, and this study hypothesizes that both number of nouns as premodifiers and increased scope of meaning relationships might offer insight into student writing development. Comparisons were made across the four categories of Humanities, Popular Science, Social Science, and Specialist Science, and clear differences are shown in the use of nouns as nominal premodifiers, with the lowest frequency occurring in humanities and the highest in Specialist Science writing. Examining academic research articles, Gray (2015) also shows variation in linguistic features, including nouns as nominal premodifiers, across disciplinary lines, and finds correspondences with traditional academic groupings, such as humanities and hard sciences.

An important contribution from both publications is consideration of the complex interactions of situational variables such as paper topic and research methodologies, details that have not often been considered in studies of academic writing as a whole.

Jalilifar, White and Malekizadeh (2017) also compared 'hard' and 'soft' sciences in their use of complex noun phrase features, using Physics and Applied Linguistics textbooks as examples of each category. Their findings showed that noun phrases in Physics were more frequently modified both with pre- and post-modifiers. The Applied Linguistics texts, by contrast, were more likely to post-modify nouns using relative clauses. Their study combines premodifying adjectives and nouns under the term 'classifiers'; however, these two types of premodification have previously been shown to vary differentially across registers (Biber et al., 1999). Future research building on Jalilifar et al. (2017) might find different patterns of variation by looking at noun and adjective nominal premodifiers independently for a more detailed description of disciplinary writing norms.

A variety of language features have been examined in previous studies of complexity in academic writing. To contribute to the existing body of knowledge, it may be beneficial to look in more detail at particular elements of noun phrase complexity. Among these complexity features, the use of nouns as nominal premodifiers has been targeted in this study due to the informational density they enable, the variety of meaning relationships they can express, and their potential ambiguity (Biber & Gray, 2011; Halliday, 1989; Wisniewski, 1996). The following section will explore previous research on the use of nouns as nominal premodifiers and their importance in academic writing.

2.3 Nouns as nominal premodifiers

Nouns as nominal premodifiers constitute an important element of noun phrase complexity in academic writing, representing approximately 30% of premodifiers in that register (Biber et al., 1999). Their frequency in academic writing is related to the importance of conciseness in this register, which is facilitated by the compression of information made possible by nominal premodifiers. For example, a noun + noun (+noun) sequence (Ex.1) rephrased without using nouns as nominal premodifiers results in a much longer string of words (Ex. 2).

(1) *gravel base course design method*

(2) *method of designing the course at the base, which is made of gravel.*

By using multiple nouns as nominal premodifiers in the first example, the writer is able to carry across the message with as few words as possible. This allows efficient transmission of information among specialists in a field; however, attempting to unpack the structure without prior knowledge of the topic is challenging due to the lack of explicitly indicated relationships and potential overlapping of premodifying relationships.

In addition to the increased informational density they enable, noun + noun sequences can represent a wide variety of meaning relationships (Biber et al., 1999; Ravelli, 1998), which places an increased burden on readers due to the lack of structure words to indicate functional relationships. For example, the noun sequence *consumer washing machine* might encode two different meaning relationships based on the premodifying relationships, either a machine for washing intended for consumer use, or a machine that is used for washing consumers. Although the meaning is clear due to the familiarity of reference, there is no

explicitly stated relationship in the structure of the phrase. In cases of more specialized reference, there is potential for confusion among novice members of the knowledge community, which makes this type of noun phrase complexity an important consideration both in writing instruction.

The importance of nouns as nominal premodifiers in academic English has also been shown by Biber and Gray (2016), who described the dramatic increase in their usage over the past three centuries as academic writing developed into an increasingly distinct and specialized register. Along with the more frequent appearance of nouns as premodifiers has come an increase in the types of nouns involved, variety of meaning relationships, and grammatical functions. Relevant to student writing is the loss of explicitness of meaning that accompanies the use of noun + noun premodification sequences, since students must learn to balance conciseness with clarity of expression in their goal of demonstrating knowledge in an academic context.

The use of nouns as nominal premodifiers are an important consideration when looking at student writing for several reasons. They provide a key tool in increasing information density in academic writing, an important skill for student writers as they develop from conversation-like clausal elaboration to a more target-like style of academic discourse. However, the frequency of nouns as nominal premodifiers has also been shown to vary based on discipline and genre (Gray, 2015). These findings suggest that a general approach to academic writing instruction may not be the most effective way to prepare student writers for the expectations of writing in their specific program of study. In addition, the loss of explicitness in noun + noun meaning relationships presents a challenge to novice members of knowledge communities. The need for contextual knowledge for understanding

and producing academic discourse of this type provides support for discipline-specific writing instruction and specific focus on the use of complex noun phrases.

2.4 Complexity and Variation in Student writing

The Michigan Corpus of Upper-Level Student Papers (MICUSP) is a publicly accessible collection of upper-level university student writing created in 2009 (Römer & O'Donnell, 2011). It contains over 2.6 million words in 829 pieces of student writing from senior year of undergraduate and first, second, and third years of graduate study. This corpus provides a valuable resource as a collection of texts across a variety of disciplines within the broader register of academic writing. Previous studies have used this corpus to address a variety of research questions related to complexity in academic writing. In a study of lexical complexity and textual cohesion features, Crossley, Russel, Kyle and Römer (2017) examined disciplinary and sub-disciplinary variation in MICUSP. Their focus was on science and engineering student writing, and the sub-disciplines within each of those categories. Their findings showed differences at both the macro and micro levels and found that year of study (final year undergraduate; 1st, 2nd, 3rd year graduate) was not related to the frequency results. This study also suggested that paper type, which is a variable included in MICUSP, did not interact significantly with the linguistic features they addressed. However, the authors point out that the small sample size prevents a definitive answer to that question.

Variation across both level and discipline in the use of lexical bundles have also been examined in MICUSP (Adel & Römer, 2012). Little variation in frequent lexical bundles was found across student levels from final year undergraduate to third year graduate students, suggesting that students acquire these forms early in their academic development. A second

aspect analyzed was attribution of sources, which was examined in relation to discipline and was found to vary both in form and frequency by discipline.

A wide variety of linguistic features were examined by Hardy and Römer (2013) in a multidimensional analysis of MICUSP, which sought to identify the co-occurring features that coincide with disciplinary variation in the corpus. Four dimensions of variation were identified: involved academic narrative vs. descriptive, informational discourse; expression of opinion and mental processes; non-procedural vs. procedural; and production of possibility statement and argumentation. The four dimensions of co-occurring linguistic variables were found to distinguish most clearly the humanities from the physical sciences, with social sciences falling somewhere between the two. While this study didn't specifically target nouns as nominal premodifiers, certain features associated with them, such as higher level of informational density in the physics and biology, were found to support previous research in disciplinary variation.

The developmental stages proposed by Biber, Gray and Poonpon (2011) were supported by further research (Parkinson & Musgrave, 2014) showing that more advanced L2 students used nouns as premodifiers more often than less experienced students, who relied more on attributive adjectives. Although the variable of paper type wasn't considered in the study, the authors recognized the possible influence on linguistic features. Developmental trends in student writing were also found by Staples, Egbert, Biber and Gray (2016) in a study showing more frequent use of phrasal rather than clausal features as L1 writers advance in their post-secondary studies. Most relevant to the present study is their finding that discipline was an important variable in the frequency of phrasal elaboration, particularly that the arts and humanities students used more clausal features than the sciences. In addition, this study found differences in complexity type associated with paper type, examining four

genres: case studies, critiques, essays, and explanations. Essays, which occurred most commonly in the Arts and Humanities, showed the greatest usage of clausal elaboration, while case studies and explanations, which are more frequent in the Life and Physical sciences, showed generally higher use of phrasal complexity features, such as nouns as nominal premodifiers.

The importance of noun phrase complexity as an indicator of advanced student writing was also explored in a study of L2 student writing by Taguchi, Crawford, and Wetzel (2013). They found that more highly rated student essays featured more frequent use of complex noun phrases rather than clausal subordination. Their results indicate the importance of this feature in meeting norms of academic writing, and the authors suggested that writing instruction could benefit from more explicit work on phrasal complexity.

2.5 Research Questions

Writing is an essential skill in academia, so it is important to understand the characteristics of writing in an academic context in order to best foster student progress. While much work has been done with published academic writing, further analysis and description of student writing may offer a more useful model for less proficient student writers (Gardner et al., 2018). The advanced student writing found in MICUSP may offer a more useful model for lower-level undergraduate writers in English for Academic Purposes (EAP) or discipline-specific writing courses. In analyzing the advanced student writing in the MICUSP corpus, the goal of this study is to gain insight into patterns in advanced students' use of complex noun phrases, particularly nouns as nominal premodifiers.

The frequency of use, length, and the types of meaning relationships encoded by noun + noun sequences are important indicators of student writers' progress toward disciplinary

norms of noun phrase complexity described in previous studies. This is a particularly important element of academic writing for several reasons. First of all, the increased noun phrase complexity and informational density created by the use of one or more noun premodifiers is a key distinction between conversation and academic writing. In addition, among the developmental stages proposed by Biber, Gray and Poonpon (2011) nouns as nominal premodifiers were found in the fourth of five, indicating their later acquisition and relative difficulty. The challenge involved in acquiring this type of complexity makes its usage among students a key consideration for both pedagogy and research of writing development. In addition, the lack of explicit meaning relationships between nouns and the nouns premodifying them requires contextual knowledge, often specific to the discipline in which they appear. Therefore, frequent use of this feature may reflect students' increased topical and disciplinary knowledge as they advance in their studies.

For these reasons, comparing usage of nouns as premodifiers across disciplines in student writing might provide valuable information about whether, and to what extent, advanced student writers are acquiring and following the linguistic norms of their fields. Prior studies have investigated noun phrase complexity in L2 learner writing (Lu, 2011; Parkinson & Musgrave, 2014; Taguchi et al., 2013). Also, complexity features in British university student writing has been analyzed in detail by Gardner, Nesi, and Biber (2018) using the British Academic Written English Corpus, a collection of university student writing across 30 disciplines. Working with MICUSP, Hardy and Römer (2013) used multidimensional analysis to describe variation across disciplines using frequently co-occurring linguistic features. The present study aims to build on their work, giving a more detailed picture of the use of nouns as nominal premodifiers across discipline. In order to accomplish this, this study seeks to answer the following research questions:

1. In advanced student writing, to what extent does the variation in frequency of nouns premodified by one or more nouns correspond with traditional academic departmental divisions?
2. How does the length of noun premodification sequences vary by disciplinary groupings?
3. How does the use of particular head nouns, nominal premodifiers, and the meaning relationships between them vary across disciplinary divisions?

CHAPTER 3. METHODS

Quantitative and qualitative approaches were taken in this analysis in order to address the stated research questions and understand the frequency, string length, and type of nouns used in sequences of nouns premodified by nouns in student writing. Corpus linguistics methodology was used to enable the examination of a broad range of texts and linguistics features, better enabling generalizable findings about disciplinary writing features. This methodology is well suited to answering questions about patterns of variation, as it allows comparison of a large number of texts.

The following chapter addresses the choice of corpus, followed by a discussion of the analytical tools used, and the approach taken to grouping the disciplines into broader categories.

3.1 Corpus

This study is based on the Michigan Corpus of Upper-Level Student Papers (MICUSP), a publicly accessible collection of university student writing created in 2009 (Römer & O'Donnell, 2011). This corpus contains over 2.6 million words taken from a total of 829 pieces of student writing from 16 disciplines (Table 3.1). The texts are drawn from four advanced levels of post-secondary education, including senior year of undergraduate and first, second, and third years of graduate study, levels at which students begin to approach professional disciplinary norms.

Table 3.1 Number of texts and total word count by discipline

Discipline	Number of texts	Total Words
Biology (BIO)	67	168,344
Civil and Environmental Engineering (CEE)	31	95,527
Economics (ECO)	25	70,733
Education (EDU)	46	267,700
English (ENG)	98	148,950
History and Classical Studies (HIS)	40	166,134
Industrial and Operations Engineering (IOE)	42	133,819
Linguistics (LIN)	41	150,536
Mechanical Engineering (MEC)	21	99,823
Natural Resources and Environment (NRE)	62	171,258
Nursing (NUR)	42	157,640
Philosophy (PHI)	44	128,847
Political Science (POL)	62	44,360
Physics (PHY)	21	209,448
Psychology (PSY)	104	317,744
Sociology (SOC)	72	213,654
Total	818	2,544,517

This corpus is particularly suited to the aims of this study, as it provides a large collection representative of student writing in an authentic academic context. Vital to the aims of this study is the inclusion of 16 sub-corpora, representing a variety of academic

disciplines, and enabling comparisons to be made of linguistic features across disciplinary lines to answer the stated research questions. All papers collected for the corpus were part of students' coursework, which provides authentic, representative examples of student writing in a university context. Additionally, because all papers included received high grades, they can be said to meet the standards and expectations of their respective academic communities, thereby making possible an analysis of disciplinary linguistic norms. This collection was chosen in part because it includes only upper-level student writing, final year undergraduate and three years of graduate school, as opposed to the similarly-designed British Academic Written English Corpus (BAWE), which presents student writing across the undergraduate years and is more suited to studying developmental trajectory.

MICUSP includes seven paper types, including argumentative essays, creative writing, critique/evaluation, proposal, report and response paper, representing a variety of academic genres within the broader register of academic writing. However, due to their uneven distribution among the disciplines (Table 3.2), an analysis of the influence of paper type in noun pre-modification was not undertaken. Further discussion of this aspect of the analysis can be found in Chapter 5.

Table 3.2 Number of texts by discipline and paper type

Discipline	Argumentative Essay	Creative Writing	Critique/ Evaluation	Proposal	Report	Research Paper	Response Paper
BIO	3	0	0	5	31	26	2
CEE	1	0	1	0	19	10	0
ECO	0	0	2	2	8	11	1
EDU	4	0	4	4	26	2	6
ENG	65	4	5	0	22	0	2

Table 3.2 (continued)

HIS	16	0	2	0	20	0	2
IOE	1	0	6	5	16	11	3
LIN	4	0	3	4	13	15	2
MEC	0	0	0	3	10	19	0
NRE	7	1	5	2	37	10	0
NUR	6	0	0	5	28	3	0
PHI	20	0	10	0	12	1	1
POL	19	0	8	3	29	3	0
PHY	0	0	1	1	12	7	0
PSY	16	1	10	6	53	14	4
SOC	23	1	4	7	28	8	1

While previous research has described the importance of linguistic variation across sub-disciplines (e.g. Crossley et al., 2017), this study sought to examine variation patterns across broader academic groupings. Combining the 16 sub-corpora into four categories (Table 3.3) also helped to compensate for the limitations of relatively small word counts for each sub-discipline. Results along larger disciplinary groupings were sought in order to give more usable insight into academic language use, as it would likely be impractical for each discipline in a university to develop its own approach to writing pedagogy.

In order to group the disciplines in a meaningful way, a pilot analysis was carried out prior to completing further analysis. When finding frequency counts across disciplinary groupings, calculations were initially made using two different systems of categorization. The first was based on the University of Michigan's academic departments, an approach taken by Römer and O'Donnell (2011) in their work with MICUSP (Table 5.1.1). The

second (Table 5.1.2) was the Becher-Biglan taxonomy (Neuman, et al., 2002), which classifies disciplines based on two axes: hard/soft, and pure/applied. Hard is used to describe the use of quantitative methods and tendency toward collaborative work, while soft disciplines tend to be more qualitative and solitary. Pure disciplines are more holistic in focus and deal with universals, while applied fields make use of pure knowledge to solve particular problems. Using these axes results in four primary groupings: Hard Pure, Hard Applied, Soft Pure, and Soft Applied (Table 3.3). Becher's initial typology also included a third dimension, life/non-life; however, although it may be applicable to the disciplines being examined in this study, it has not often been included in previous studies that make use of the typology, so for the sake of consistency with previous literature, the present study was limited to the two primary axes of hard/soft and pure/applied.

A comparison of the two groupings was undertaken to find out which offered clearer differentiation of disciplinary groupings based on frequency of nouns as nominal premodifiers. When comparing the two (see Appendix A) both indicate a progression in frequency levels. However, of these two options, the Becher-Biglan typology offered a clearer differentiation between the two higher frequency groupings, hard-applied and hard-pure, which were only slightly different in the University of Michigan groupings (Biological & Health Sciences, Physical Sciences). For this reason, the study defined the categories according to the Becher-Biglan typology of academic disciplines (Table 3.3).

Table 3.3 MISCUSP disciplines grouped using Becher-Biglan typology

Category	Discipline	# of texts	# of words
Hard-Pure	Biology (BIO)	67	168,334
	Natural Resources and Environment (NRE)	62	171,258
	Physics (PHY)	21	209,448
	Subtotal	150	590,040

Table 3.3 (continued)

Hard-Applied	Civil and Environmental Engineering (CEE)	32	157,640
	Industrial and Operations Engineering (IOE)	42	133,819
	Mechanical Engineering (MEC)	21	99,823
	Nursing (NUR)	42	95,527
	Subtotal	137	486,809
Soft-Pure	English (ENG)	98	148,950
	History & Classical Studies (HIS)	40	166,134
	Linguistics (LIN)	41	150,356
	Philosophy (PHI)	44	128,847
	Sociology (SOC)	72	213,654
	Subtotal	295	807,941
Soft-Applied	Economics (ECO)	25	70,733
	Education (EDU)	46	267,700
	Political Science (POL)	62	44,360
	Psychology (PSY)	104	317,744
Subtotal	237	700,537	

3.2 Corpus search process

Before analysis could begin, the full-text, downloaded version of MICUSP was tagged for part of speech using CLAWS (Constituent Likelihood Automatic Word-tagging System) (Garside & Smith, 1997). This allowed searches to target nouns as a general class, and using these tagged files, independent searches were carried out within each discipline for noun + noun sequences using AntConc, a "freeware corpus analysis toolkit" (Anthony, 2014).

The CLAWS tagger recognizes a wide variety of noun types, but not specifically nouns as nominal premodifiers. To return the desired results, a regular expression search (see below) was used in AntConc to capture noun + noun (+ noun) sequences.

Regular expression for n + n sequences

$\backslash w + _ NN \backslash w * \backslash s + (? = \backslash w + _ NN \backslash w *$

Search results were exported to Microsoft Excel for manual coding to determine whether the results represented noun + noun premodification relationships. The first step of this process was to eliminate false positives (bolded in the examples below). One type of false positive found was the occurrence of a word tagged as a noun, but which belonged to another part of speech (Ex.1). In the example below *present* was identified as a noun by the tagger, but looking at the context of the sentence, we see that it is in fact a verb.

- (1) *Juvenile offspring of mothers that were undernourished throughout pregnancy **present** hyperphagia.* (BIO)

In addition, the tagging software marked unknown words as nouns, including typos (Ex.2), words from other languages (Ex.3), and words preceded by a single quotation mark (Ex.4).

- (2) *originating in **culturenot biologythat** establishes patterns* (SOC)
 (3) *ipse fortunam **benigno** adloquio*
 (4) *Skinner had admitted in his book '**About** Behaviorism'*

In the next stage, it was necessary to remove results that included noun strings that were tagged correctly, but which did have a premodification relationship. These occurred relatively frequently, and required careful manual checking to avoid miscategorization and inflated frequency counts. In Example 5 we have two nouns that are adjacent, but which belong to separate constituents of the sentence, with a prepositional phrase functioning adverbially (*in the former districts*) preceding the subject noun (*cholera*).

(5) *In these low populated districts, cholera can be seen as a disease with irregular outbreaks, whereas in the former **districts Cholera** exhibits an endemic behavior.* (SOC)

Finally, results were checked for duplications resulting from noun strings of three or more words, which returned separate results for each two-word combination (Ex.6). That is, NPs with more than one noun as a prenominal modifier were considered one noun phrase and counted once. Although these strings contain multiple noun premodification relationships, the decision was made to count them as entire noun strings rather than by individual premodification instances. This choice was made due to the complex nesting of premodification relationships that can occur in noun + noun sequences, where a premodifier does not necessarily modify only the word immediately to its right (Biber et al., 1999). Examining entire strings rather than individual premodifications provides more contextual information for analysis and gives insight into noun phrase complexity levels across disciplines that may be less apparent when examining only two-word units.

(6) *proton relaxation enhancement ability* (BIO)

**relaxation enhancement ability*

**enhancement ability*

After completing the data cleaning and coding, frequencies were calculated in each disciplinary division for 2, 3, 4, and 5-word noun + noun premodification strings. Combining these results, the most frequent noun strings of any length were calculated for each discipline, giving an overall picture of frequency of nouns premodified by one or more

nouns. In addition to these overall frequency counts, calculations were done for each string length to show what percentage of all occurrences of nouns with noun premodifiers each represents. To do this, frequency of occurrence for each string length was divided by the total number of occurrences of nouns premodified by nouns. This measure was undertaken to indicate whether a discipline tends toward longer or shorter noun premodification strings.

. To begin answering the third research question, the type-token ratio was calculated for each academic discipline's noun + noun sequences, dividing the number of forms by the total occurrences. This measure indicates how many among the total noun + noun sequences (of all lengths) were unique occurrences and gives an indication of the diversity of noun + noun sequences that exist in a discipline or disciplinary grouping. In addition, frequencies were calculated for head nouns and premodifying nouns in each discipline to find those that occurred most often. Resulting frequencies for each discipline were normalized per 10,000 words to enable comparison across sub-corpora of different sizes. The results of each measure for individual disciplines were then combined into four categories (see Table 3.3) based on the Becher-Biglan typology (hard-pure, hard-applied, soft-pure, soft-applied) in order to find broader patterns in variation.

3.3 Analysis

After calculating the most frequently occurring noun + noun sequences, the twenty most common for each disciplinary grouping were analyzed qualitatively to identify meaning relationships between the head noun and premodifier, or premodifiers. These meaning relationships are an important aspect of noun phrase complexity to consider, because nouns as nominal premodifiers pack information densely, but lose explicitness in meaning due to the absence of function words. Thus, it might be hypothesized that a wider variety of

meaning relationships will appear as students advance in their mastery of academic writing and gain the disciplinary background knowledge that allows more information to be expressed implicitly.

This analysis was undertaken to answer the third research question, which asked what meaning relationships are found in noun + noun premodifications in student writing across the disciplines. To accomplish this, the analysis of meaning relationships between nominal premodifiers and head nouns was based on the framework presented in Biber et al. (1999, p. 590), which identifies 15 categories of meaning relations between noun + noun sequences (Table 3.4).

Table 3.4 Meaning relationships in noun + noun sequences, adapted from Biber et al. (1999)

Meaning relationship	Explanation	Example
Composition	N ₂ is made from N ₁	<i>glass windows</i>
Purpose	N ₂ is used for N ₁	<i>war fund</i>
Identity	N ₂ has same referent as N ₁ but classified by different attributes	<i>compression process</i>
Content	N ₂ is about N ₁	<i>algebra text</i>
Source	N ₂ comes from N ₁	<i>whale meat</i>
Objective Type 1	N ₁ is the object of N ₂ process	<i>egg production</i>
Objective Type 2	N ₂ is the object of N ₁ process	<i>discharge water</i>
Subjective Type 1	N ₁ is the subject of N ₂ process	<i>eye movement</i>
Subjective Type 2	N ₂ is the subject of N ₁ process	<i>labor force</i>
Time	N ₂ is found at the same time given by N ₁	<i>summer conditions</i>
Location Type 1	N ₂ is found or takes place at the location given by N ₁	<i>thigh injury</i>

Table 3.4 (continued)

Location Type 2	N ₁ is found or takes place at the location given by N ₂	<i>staff room</i>
Institution	N ₂ identifies an institution for N ₁	<i>insurance companies</i>
Partitive	N ₂ identifies part of N ₁	<i>family member</i>
Specialization	N ₁ identifies an area of specialization for N ₂	<i>finance director</i>

CHAPTER 4. RESULTS

Chapter 4 of this thesis presents the results of analysis of the use of nouns premodified by nouns across all disciplines individually and across the four Becher-Biglan groupings. Frequencies are presented for all noun premodification strings and for strings of different lengths. In addition, the frequencies of particular head and premodifying nouns are explored. Findings will then be discussed, addressing the research questions introduced in Chapter 2.

4.1 Frequency of Nouns with One or More Nouns as Premodifiers Across Disciplines

The first research question in this study asked to what extent patterns of variation in the frequency of nouns as nominal premodifiers correspond with traditional academic divisions. In order to begin addressing this question, the analysis started with an examination of the overall frequency of nouns premodified by one or more nouns for all MICUSP disciplines individually (Figure 4.1). A broad range of frequencies were found, with Philosophy containing the lowest number (49.8 per 10,000 words) and Mechanical Engineering the largest (562.9 per 10,000 words). These results clearly demonstrate the importance of discipline as a variable influencing the frequency of use of nouns as nominal premodifiers in student writing.

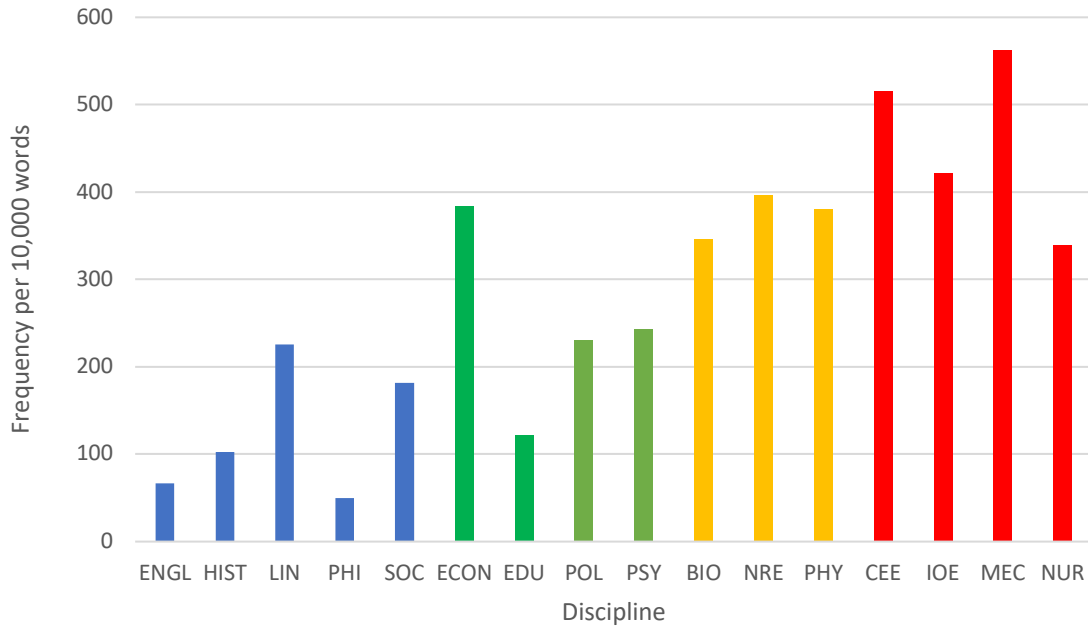


Figure 4.1 Frequency of nouns premodified by one or more nouns.

Blue = Soft-Pure; Green = Soft Applied; Orange = Hard=Pure; Red = Hard-Applied

4.2 Patterns Across Disciplinary Groupings

In comparing the values shown in Figure 4.1 above, some generalizations are suggested, as engineering and science disciplines appear to show more frequent use of noun + noun sequences. In order to begin looking for generalizable findings, the following section will approach the data using wider disciplinary categorizations.

After calculating the frequency variation across all disciplines, the next step in the analysis involved combining the results by disciplinary groupings in order to answer the first research question, which asked how frequency varied across academic divisions. When grouping the disciplines using the Becher-Biglan typology, clear differences in premodification patterns were found across the categories (Figure 4.2).

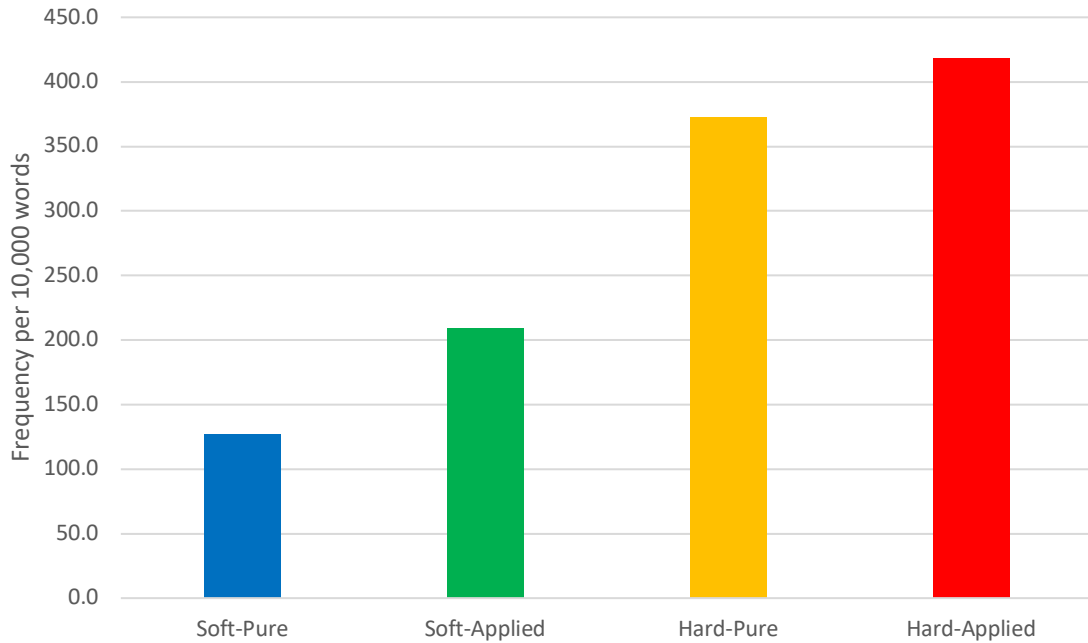


Figure 4.2 Frequency of noun phrases with at least one noun as nominal pre-modifier in the Becher-Biglan typology academic groupings (normalized per 10,000 words)

The Hard-Pure and Hard-Applied groupings showed noticeably higher overall usage of noun + noun premodification strings. The Soft-Pure category showed far fewer occurrences of this feature, with the Soft-Applied falling between the two poles. Though we find less distinction between the two 'hard' categories, patterns of usage appear to correspond with the Becher-Biglan typology's categories. The overall pattern for use of nouns premodified by one or more nouns reveals the 'hard' categories with clearly higher frequencies of usage than the 'soft' disciplines. The dimension of pure/applied also has an interesting influence on the use of this feature as well. Application of hard or soft disciplines' knowledge appears to increase the frequency of noun + noun sequences in disciplinary writing, as we find both Soft-Applied and Hard-Applied showing higher frequencies than their Pure counterparts. Later in this chapter, further exploration of the specific nouns and

meaning relationships used in these disciplinary groupings will attempt to find reasons behind the effect that this variable has on the use of nouns as nominal premodifiers.

4.3 Frequency of Noun Strings by Length

In addition to frequency counts for overall use of nouns premodified by one or more noun, calculations were done for each string length to show what percentage of all occurrences of nouns with noun premodifiers each represents (Table 4.1, Figure 4.3). To do this, frequency of occurrence for each string length was divided by the total number of occurrences of nouns premodified by nouns (see section 4.1). Search results returned 2, 3, 4, 5, and even 6-word noun + noun sequences, but sequences of five or more words occurred very infrequently they were therefore left out of the following discussion.

The results showed that 2-word strings were most common for all disciplines, as expected (Biber et al., 1999, p.597), but clear variation in frequency across disciplines appears when comparing normalized counts at each level. The highest values for each string length were found in engineering fields (Civil & Environmental Engineering for 2- and 4-noun sequences, and Mechanical Engineering for 3-noun sequences). All three string lengths occurred least frequently in the Philosophy sub-corpus.

These variations in frequency may provide some insight into one measure of noun phrase complexity in each discipline. For example, while Education shows one of the lowest frequencies of 2-word strings at 98.5 per 10,000 words (Table 4.1), which might suggest lower levels of nominal complexity in the discipline, the percentage (81%) shows us that the discipline tends toward longer noun strings despite an overall low frequency. The History sub-corpus, by contrast, contained a similar number of 2-word strings (96.4), but those represent 94% of all noun as nominal premodifier strings, meaning writers in that discipline are less likely to use longer, more complex sequences of noun premodification.

Table 4.1 Frequency (per 10,000 words) of noun + noun premodification sequences by length and discipline, and percentage of total number of nouns with noun premodifiers

Disciplines	2-words		3-words		4-words	
	Tokens	% of total	Tokens	% of total	Tokens	% of total
Soft-Pure						
ENG	63.4	95%	2.9	4%	0.2	0.3%
HIS	96.4	94%	5.9	6%	0.5	0.5%
LIN	206.5	91%	17.3	8%	1.9	0.8%
PHI	48.4	97%	1.4	3%	0.1	0.2%
SOC	166.9	92%	13.4	7%	1.5	0.8%
Soft-Applied						
ECO	333.7	87%	46.8	12%	3.2	0.8%
EDU	98.5	81%	20.7	17%	2.8	2.3%
POL	204.4	89%	22.7	10%	2.7	1.2%
PSY	218.2	90%	22.4	9%	2.1	0.9%
Hard-Pure						
BIO	301.6	87%	40.9	12%	3.9	1.1%
NRE	329.5	83%	58.3	15%	7.0	1.8%
PHY	337.9	89%	40.8	11%	2.0	0.5%
Hard-Applied						
CEE	421.2	82%	82.3	16%	10.9	2.1%
IOE	360.0	85%	56.8	13%	3.9	0.9%
MEC	458.8	82%	91.1	16%	10.5	1.9%
NUR	292.0	86%	42.3	12%	5.0	1.5%

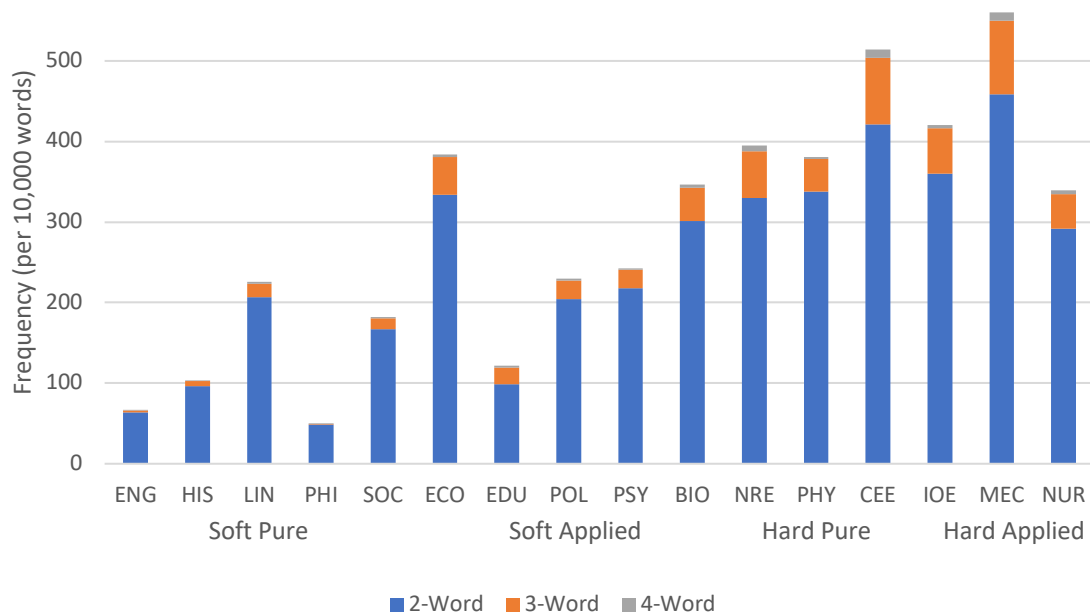


Figure 4.3 Frequency (per 10,000 words) of noun + noun premodification sequences by length and discipline

4.4 Noun Strings by Becher-Biglan Disciplinary Groupings

After comparing frequency of noun strings by length and individual discipline, results were grouped according to the Becher-Biglan typology (Table 4.2, Figure 4.4). Frequency counts for each string length were combined into the Hard-Pure, Hard-Applied, Soft-Pure, and Soft-Applied groupings, and divided by the total number of nouns premodified by nouns occurring in each grouping to find the proportions for each string length. Calculating these frequencies and proportions was done to further address the question posed in the first two research questions by providing additional measures of noun phrase complexity.

Table 4.2 Frequency (per 10,000 words) of noun + noun premodification strings by length and discipline grouping

Becher-Biglan groups	2-word	% total	3-word	% total	4-word	% total	All
Soft-Pure	114.3	90.1%	11.4	9.0%	0.99	0.78%	126.8
Soft-Applied	201.4	96.2%	7.3	3.5%	0.62	0.29%	209.4
Hard-Pure	318.2	85.4%	48.7	13.1%	5.05	1.36%	372.7
Hard-Applied	370.7	88.7%	41.3	9.9%	5.14	1.23%	418.0

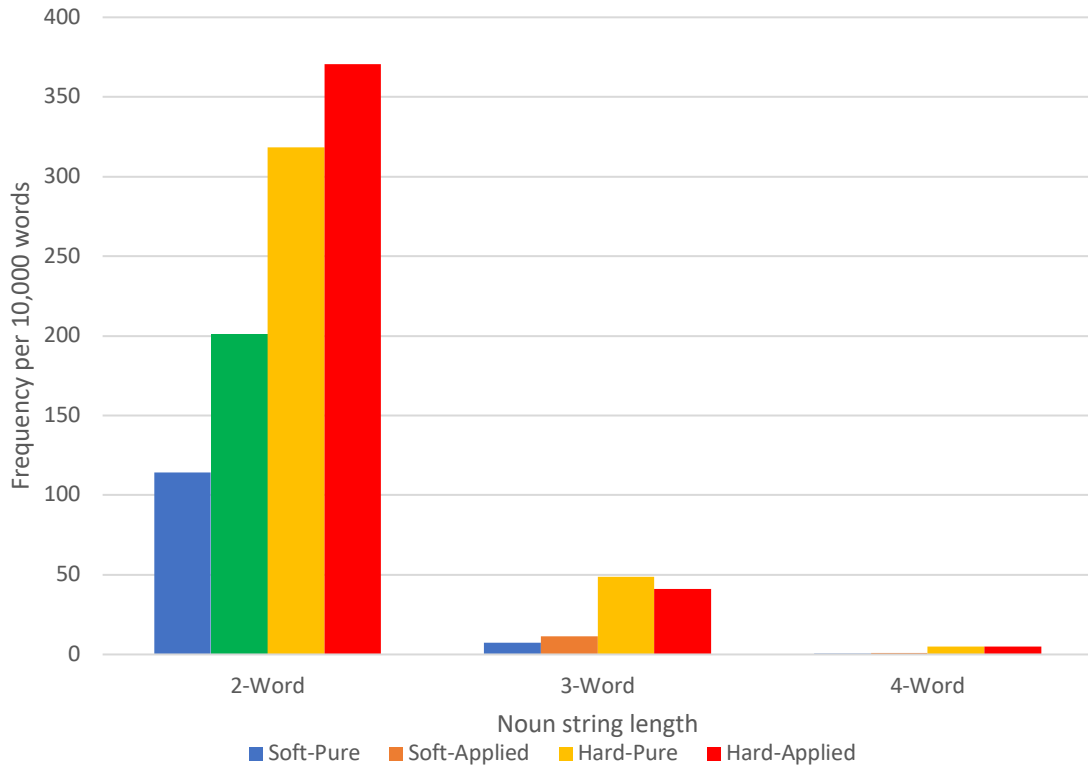


Figure 4.4 Noun string length frequency (per 10,000 words) by Becher-Biglan disciplinary groupings

When examining the frequencies of occurrence for each string length per disciplinary grouping (Table 4.2, Figure 4.4), the Hard-Applied grouping contains the highest frequency of total strings, and the highest frequency for 2 and 4-word strings individually. This finding suggests that this group of disciplines places a high importance on economy of expression with concentration of information into noun phrase structures through the use of nouns as premodifiers. When looking at the Soft-Pure disciplinary grouping, we find a much lower overall frequency of noun + noun premodification strings when compared to both Hard disciplines and the Soft-Applied group.

The patterns are a bit different when taking into account percentages of overall use of nouns with noun premodifiers (Figure 4.5). Among the disciplinary groupings examined,

Hard-Pure showed the highest proportion of 4-word noun strings at 1.36% of all nouns with noun premodifiers, although only slightly higher than Hard-Applied (1.23%) (see Table 4.2). The more interesting finding, however, was that the Soft-Applied grouping uses a higher percentage of shorter (2-word) noun premodification strings than Soft-Pure, despite a much higher overall usage of nouns with nouns as premodifiers. This suggests that although the Soft-Applied fields use nouns as nominal premodifiers much more often than Soft-Pure, the noun + noun sequences are simpler in structure. To better understand the differences in their usage will require a closer look at the specific strings that appear.

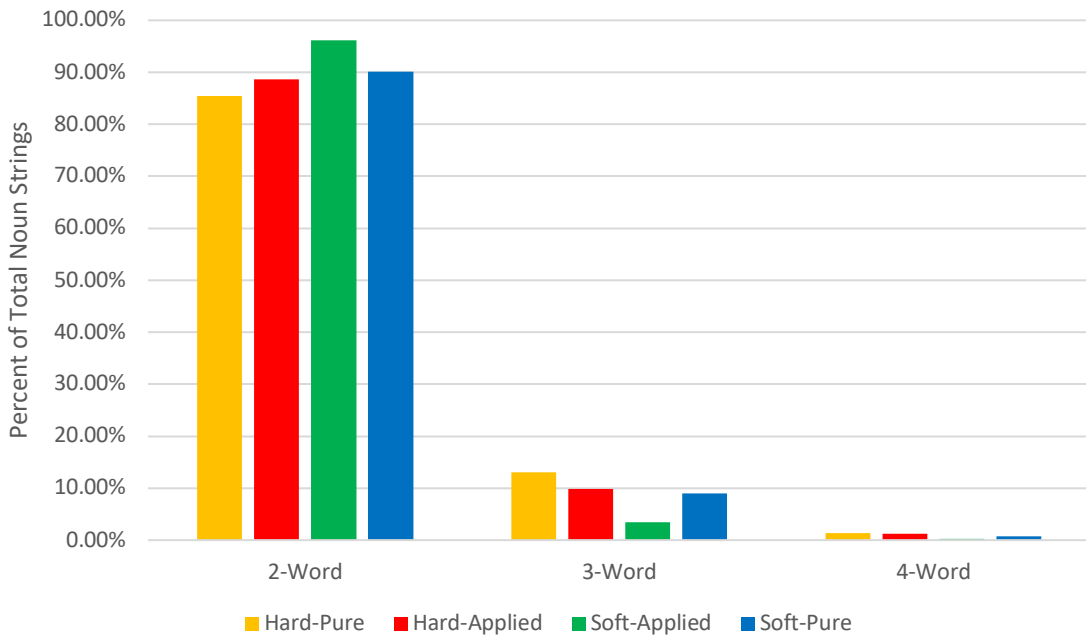


Figure 4.5 Percentage of all nouns premodified by nouns represented by each string length

4.5 Components of Noun + Noun Sequences and their Meaning Relationships

Looking in more detail within the four disciplinary groupings involved first identifying the most frequent noun premodifying strings. To begin answering the third research question about the types of nouns being used in these constructions, the type-token

ratio was calculated for each academic discipline. This measure indicates how many among the total noun + noun sequences (of all lengths) were unique occurrences (Table 4.3) and was taken as an additional indication of noun phrase complexity in that it shows whether a small number or a wider variety of forms are used in a discipline.

Results showed only small variation in type-token ratios for the disciplinary groupings. One generalization that might be made was that Pure disciplines had a slightly higher ratio, meaning they contained a wider variety of nouns premodified by nouns. In a future study it may be interesting to compare these type-token ratios to both lower-level student writers and published academic writing as a measure of developing lexical diversity and increasing mastery of noun premodification.

Table 4.3 Type-token ratios for Becher-Biglan groupings

Disciplinary Grouping	Tokens	Types	Type-Token Ratio
Soft-Pure	11,453	6,935	0.61
Soft-Applied	18,728	10,140	0.54
Hard-Pure	14,313	8,400	0.59
Hard-Applied	20,694	11,707	0.56

In an attempt to characterize the types of noun + noun premodification sequences common to each disciplinary grouping, frequencies were calculated for all occurrences of nouns premodified by one or more nouns. The twenty most frequent strings for each group are included in the table below (Table 4.4) Also included in the table is range, which indicates the number of texts in which a noun + noun sequence appeared. For the purposes of this study, strings with very low ranges offer little insight into broader disciplinary trends, as they may only reflect the idiosyncrasies of a particular writer or topic.

Table 4.4 Twenty most frequent nominal premodification strings for each Becher-Biglan category (normalized per 10,000 words)

Hard-Pure (148 texts)	Frequency	Range
<i>climate change</i>	1.43	14
<i>water quality</i>	1.07	13
<i>concept plan</i>	1.02	1
<i>Forest Service</i>	0.99	3
<i>community members</i>	0.99	6
<i>energy sources</i>	0.70	4
<i>property rights</i>	0.60	2
<i>land use</i>	0.57	19
<i>carbon dioxide</i>	0.55	12
<i>justice movement</i>	0.52	2
<i>fossil fuels</i>	0.49	5
<i>shore impact zone</i>	0.47	1
<i>food web</i>	0.42	4
<i>reforestation project</i>	0.42	1
<i>size structure</i>	0.42	1
<i>reforestation policy</i>	0.36	1
<i>age class</i>	0.34	14
<i>forest dwellers</i>	0.34	1
<i>property owners</i>	0.31	2
<i>action channel</i>	0.31	4
Hard-Applied (143 texts)		
<i>health care</i>	1.31	28
<i>breast cancer</i>	1.26	2
<i>egg donation</i>	1.11	1
<i>heat sink</i>	0.94	2
<i>pressure ulcers</i>	0.90	1
<i>data collection</i>	0.86	18
<i>grip force</i>	0.77	1
<i>solenoid valve</i>	0.73	1
<i>condom use</i>	0.71	5
<i>family members</i>	0.69	8
<i>output variables</i>	0.66	5
<i>Cycle Time</i>	0.62	3
<i>advanced practice Nurses</i>	0.62	3
<i>colon therapy</i>	0.62	1
<i>health care providers</i>	0.62	15
<i>washing machine</i>	0.60	1
<i>Flow Variations</i>	0.60	1
<i>chest pain</i>	0.60	3
<i>patient care</i>	0.60	9
<i>interaction diagram</i>	0.56	1

Table 4.4 (continued)

Soft-Pure (293 texts)		
<i>BLEU Algorithm</i>	0.49	1
<i>Question Operator</i>	0.45	2
<i>truth value</i>	0.45	6
<i>suicide rates</i>	0.41	2
<i>sports literacy</i>	0.35	1
<i>sports rhetoric</i>	0.33	1
<i>sign language</i>	0.32	2
<i>climate risk disclosure</i>	0.31	1
<i>vowel quality</i>	0.29	5
<i>party culture</i>	0.28	1
<i>fairy tales</i>	0.27	2
<i>Adult Input</i>	0.27	2
<i>subject position</i>	0.26	7
<i>gender roles</i>	0.23	10
<i>power relations</i>	0.22	7
<i>starting point</i>	0.22	16
<i>Reunion Islanders</i>	0.22	1
<i>word sense</i>	0.22	1
<i>stop closure</i>	0.20	1
<i>college women</i>	0.20	1
Soft-Applied (235 texts)		
<i>community colleges</i>	1.14	2
<i>community violence</i>	1.08	1
<i>job satisfaction</i>	0.95	5
<i>community violence exposure</i>	0.90	1
<i>party identification</i>	0.71	4
<i>citizenship education</i>	0.63	1
<i>lesson plan</i>	0.58	8
<i>credit card debt</i>	0.58	1
<i>Ultimatum Game</i>	0.52	1
<i>college students</i>	0.50	20
<i>gender differences</i>	0.47	14
<i>stereotype threat</i>	0.47	3
<i>identity formation</i>	0.43	7
<i>Labor Supply</i>	0.40	4
<i>health care</i>	0.39	19
<i>Congresswoman Kaptur</i>	0.39	1
<i>sex education</i>	0.37	1
<i>optimism hypothesis</i>	0.37	1
<i>health insurance</i>	0.36	5
<i>education system</i>	0.35	10

Looking at the overall frequencies and ranges across the four disciplinary groupings, it appears that most noun + noun premodification strings are generally discipline specific, or even specific to the paper topic, based on the generally low ranges for all of the most

frequent strings. Also, none of the 20 most frequent strings are shared by more than one disciplinary grouping, and ranges are generally very low within the groups as well.

Among the most frequent for each grouping we find many discipline-specific terms (*climate change, health care, solenoid valve, etc.*). One difference that arises is a more frequent focus on topics in the physical world in the two 'hard' categories, with nouns such as *climate change, carbon dioxide, grip force, and solenoid valve*. By contrast, we see more abstract nouns in the 'soft' categories, often related to human experience, like *question operator, truth value, job satisfaction, and violence exposure*.

To begin with the Hard-Pure category, the noun strings with a frequency above 0.50 and a range over 10 included *climate change, water quality, land use, and carbon dioxide*. All of these occurred primarily in the NRE (Natural Resources and Environment) discipline. This may indicate that NRE texts tend to focus a smaller set of broader topics, rather than the highly specific subject matter that characterizes other pure sciences. The Hard-Pure category is typically thought to de-emphasize agents and focus primarily on abstracted physical processes. However, the phrases *justice movement, property rights, and community members* are closely related to human experience rather than the physical world. These terms were used commonly in the Natural Resources and Environment, a discipline whose roots are in ecology and geology, but which may be developing more soft-applied characteristics as environmental issues associated with climate change have become a more pressing concern. These results suggest that NRE, despite its ostensible subject matter, may belong in a different disciplinary grouping. In fact, the growing popularity of interdisciplinary programs may necessitate a different approach to academic disciplinary categories.

The Soft-Pure group differs from the others in that its most frequent noun sequences are less than half as frequent as the top results for the other groups. This may suggest greater

diversity of noun + noun premodifications, based also on the higher type-token ratio previously mentioned (Table 4.5.1). Another explanation might relate to the existence of a wider variety of topics in these disciplines, a characteristic of the soft sciences described in a study by Neuman, Parry and Becher (2002). The only two noun strings to appear in more than ten texts were *gender roles* and *starting point*. Both occurred in three or more disciplines within the grouping, suggesting their importance in the Soft-Pure category as a whole. *Starting point* is interesting in that it is used as a means of structuring analyses. It is often used to use a specific example in order to build a larger analysis and make wider generalizations (Ex.1,2,3).

- (1) *He first asks 'why should I want to interfere with other people 's religious practice, provided that they are not able to impose that practice on me? ' This question frames a principal **starting point** for the discussion, the sources of intolerance, and sets the stage for introducing the problems with a basis for tolerance. (PHI)*
- (2) *Gathering an impressive body of evidence, Diner makes a very convincing case for the preeminence of the Lower East Side as a **starting point** for Jewish American narratives and cultural memory. (ENG)*
- (3) *Coleman represents the torture and slaughter of wolves as something incomprehensible, as something so alien to our experiences that it reminds of the differentness of the past while at the same time providing us with a **starting point** from which we can reconstruct past societies and their systems of meaning. (HIS)*

In examining the Hard-Applied grouping, we find *health care*, *data collection*, and *health care providers* occurring both at a frequency over 0.50 and a range over 10. Both *health care* and *health care providers* occur only in the Nursing discipline, and clearly represent discipline-specific vocabulary. *Data collection*, on the other hand, is common to all disciplines within the Hard-Applied but does not appear in the top twenty in any other academic grouping, which came as a surprise. This may reflect a more empirical approach in the Hard-Applied disciplines, as opposed to more theory-focused orientations in others. Looking overall at the twenty most frequent strings, we find that most come from Nursing, which may reflect a more focused range of topics in the field, resulting in more frequent use of a few important noun sequences.

Finally, in the Soft-Applied group we find *college students*, *gender differences*, *health care*, and *education system* occurring in relatively high numbers of texts (range). *College students* occurs only in Education and Psychology, and *gender differences* only in Psychology. *Health care* and *education system* primarily in Political Science, but also appear in the other Soft-Applied disciplines to a lesser degree. These results also suggest that most noun premodification strings are discipline or topic specific, rather than being common to wider disciplinary groupings, although nearly all of the twenty most frequent strings in Soft-Applied concern human experiences rather than physical processes.

4.6 Use of Multiple Nouns as Nominal Premodifiers

Biber and Gray (2016) described the increased frequency of nouns premodified by multiple other nouns, a structure that rarely appeared before the 20th century. In the texts examined in the present study, these longer noun premodification strings occurred most often in engineering texts. Civil and Environmental Engineering in particular showed a relatively

high frequency of longer strings. In addition to more 4-word noun sequences (*Ex. 4,5*), this discipline included some with five nouns (*Ex. 6,7*). In these premodification strings we see multiple meaning relationships represented, including Composition, Content, Purpose, and Objective Type 1 (Biber et al., 1999, p.590), with complex and sometimes ambiguous premodification patterns. To illustrate, we could analyze the first example as (*median {base [(course grain) size]}*), or alternatively as (*median {base [course (grain size)]}*).

- (4) *median base course grain size*
- (5) *tension angle shear force*
- (6) *gravel base course design method*
- (7) *polymer exchange membrane fuel cell*

In their original contexts, these phrases often occur alongside diagrams and graphs, which help to disambiguate the complex interrelationships between the words. Also, mathematical formulas frequently appear, with the nominal phrases functioning as verbal analogues to elements of the formulas. Although it is beyond the scope of the present study, a more detailed analysis of the functional relationships between text and visuals in engineering may offer some interesting insights into the use of language in the engineering disciplines.

A common feature of these longer strings was the use of one or more nominalized processes, often in the penultimate position (*Ex. 8,9,10*).

- (8) *waste heat recovery systems*
- (9) *film transistor fabrication process*
- (10) *driver steering behavior models*

This type of nominal density reflects the primary subject matter of the discipline, which concerns the description of physical properties and processes. Re-wording these constructions using clauses or prepositional phrases would require many more words, violating norms of brevity that often characterize the engineering fields (Becher & Trowler, 2001). For example, the first example could be rephrased as *systems for recovering heat that has been wasted*. Condensing this information into a noun phrases allows it to take the subject or object position in a clause, facilitating the discussion of processes without visible actors. In addition, the specialized nature of the field may alleviate problems of ambiguity that result from the implicit nature of inter-nominal relationships in noun premodification structures.

This concentration of clausal information into noun phrases enables not only more efficient communication, but also increased cohesion. In the following excerpt from a Mechanical Engineering text, the four-noun premodification string (underlined) encapsulates several elements from the previous paragraph (bolded), including both processes (waste, recovery, methods) and properties (energy, heat), creating cohesion between one paragraph and the next.

*The rest of the combustion **energy is wasted**, including 40% that is lost to exhaust gas. **Various methods** have been developed to try to **recover some of this lost energy**, [...]Additionally, some hybrid cars including the Toyota Prius use regenerative braking to recover some of the **kinetic energy** [heat] lost when stopping the vehicle.*

Thermoelectric generators are attractive potential waste heat recovery systems[...]

The discipline with the fewest 3 and 4-word noun strings was Philosophy, which included only one 4-word string (*input output conversion function*), although the relationship between the first two words may be better characterized as coordinated premodification (Biber et al.,1999, p.600). It is interesting to note, however, that the text in which the phrase appeared was a philosophical argumentative essay on the topic of scientific explanation, which may be one reason for the use of a structure more typical of hard sciences. Noun phrase modification in philosophy texts strongly favored 2-word strings, (97%) with some of the most frequent being *memory processes*, *pain/pleasure qualia*, and *truth value*. These include primarily abstract nouns, and following Biber's categorizations (1999, p. 590) show a content between the premodifying and head noun.

Noun + noun premodification strings in the English discipline showed frequencies nearly as low as Philosophy. The longer strings are distinguished from most other disciplines by the appearance of premodified proper nouns (Ex.11-14). However, this classification may be contested, because in many noun phrases containing proper nouns, there can be ambiguity in determining whether premodification, rather than appositives or titles, are being used. (Biber et al.,1999, p.584). This was a more common feature of expository registers in the past but has not shown the same increased frequency as premodification of common nouns (Biber, 2003). Proper noun premodification also appeared in Political Science, showing the importance in both fields of attributing processes and ideas to specific actors, a tendency uncommon to hard sciences (Hardy & Römer, 2013).

(11) *Liquor manufacturer Reb Bratzlawer (ENG)*

(12) *theatre critic Robert Brustein (ENG)*

(13) *genre theorist Carolyn Miller (ENG)*

(14) *field marshal Shamil Busaev (POL)*

To uncover more detail about variation in the types and functions of nominal noun premodification, the next section will examine the nouns that occur most frequently in both modifier and modified positions.

4.7 Analysis of Head and Premodifying Noun Types

The third research question posed in this study asks what kinds of nouns are being used as premodifiers and premodified head nouns. To begin addressing that question, the most frequent head nouns (premodified by other nouns) for all string lengths were calculated for each category (Table 4.5).

Table 4.5 Most frequent head nouns premodified by nouns in each Becher-Biglan category (frequencies normalized per 10,000 words)

Hard-Pure	Freq.	Hard-Applied	Freq.	Soft-Applied	Freq.	Soft-Pure	Freq.
<i>color</i>	6.35	<i>analysis</i>	198	<i>exposure</i>	1.70	<i>system</i>	1.26
<i>genes</i>	5.81	<i>rate(s)</i>	198	<i>study(ies)</i>	1.63	<i>theory</i>	1.26
<i>venation</i>	3.15	<i>system</i>	187	<i>theory</i>	1.62	<i>structure</i>	0.99
<i>species</i>	2.24	<i>care</i>	130	<i>satisfaction</i>	1.55	<i>culture</i>	0.95
<i>system</i>	2.01	<i>process</i>	127	<i>students</i>	1.53	<i>movement</i>	0.95
<i>change</i>	1.90	<i>time</i>	110	<i>development</i>	1.29	<i>position</i>	0.88
<i>area</i>	1.82	<i>model</i>	89	<i>system</i>	1.20	<i>students</i>	0.84
<i>plan</i>	1.74	<i>method</i>	89	<i>violence</i>	1.20	<i>rates</i>	0.82
<i>rates</i>	1.67	<i>data</i>	88	<i>colleges</i>	1.18	<i>process</i>	0.79
<i>water</i>	1.61	<i>variables</i>	85	<i>students</i>	1.08	<i>relations</i>	0.74
<i>members</i>	1.59	<i>weight</i>	80	<i>education</i>	1.03	<i>women</i>	0.70
<i>process</i>	1.56	<i>level</i>	77	<i>disorder</i>	1.00	<i>studies</i>	0.68
<i>service</i>	1.51	<i>members</i>	75	<i>task</i>	0.94	<i>programs</i>	0.64
<i>project</i>	1.35	<i>systems</i>	75	<i>act</i>	0.90	<i>algorithm</i>	0.60
<i>data</i>	1.33	<i>area</i>	74	<i>research</i>	0.88	<i>members</i>	0.58
<i>rate</i>	1.33	<i>design</i>	71	<i>differences</i>	0.88	<i>section</i>	0.58
<i>use</i>	1.28	<i>force</i>	71	<i>group</i>	0.84	<i>education</i>	0.54
<i>structure</i>	1.25	<i>size</i>	69	<i>coverage</i>	0.82	<i>policy</i>	0.53
<i>sources</i>	1.22	<i>test</i>	65	<i>policy</i>	0.82	<i>value</i>	0.53
<i>quality</i>	1.22	<i>plan</i>	65	<i>use</i>	0.78	<i>groups</i>	0.52

Among the most common nouns occurring with nouns as pre-modifiers, we first notice much higher frequencies in the Hard-Pure and Hard-Applied groups. In the Hard-Pure grouping in particular, we find head nouns primarily focused on physical referents which are then classified using concrete nouns as premodifiers, indicative of the 'descriptive, informational production' that characterizes the 'hard' fields (Hardy & Römer, 2013), and showing a variety of meaning relations (Biber et al., 1999) (Ex.15-17).

(15) *mutant head color* - **Location Type 1**

(16) *ballast water* - **Purpose**

(17) *prey species* - **Identity**

The head nouns found in the Hard-Applied category are characterized by abstraction, specifically terms used in description and experimentation, such as *method*, *data*, and *analysis*, which are premodified by a variety of nouns. Very common among first-position (immediately preceding head noun) premodifiers are nouns derived from material processes (Ravelli, 1988), often with a more concrete noun preceding it, demonstrating the complex nesting of noun meaning relationships that sometimes occur in these strings (Ex. 18-22).

(18) *motion analysis*

(19) *body configuration variables*

(20) *energy storage methods*

(21) *Microprocessor chip cooling system*

(22) *facility location planning process*

The clearest difference arising when moving to the Soft-Applied category is a reduction in frequency, which may reflect the less unified topic ranges in the Soft disciplines (Neuman et al., 2002). This may also result from the overall lower use of nouns as nominal premodifiers, as discussed in the previous section. Although there are some characteristics of soft disciplines, the intermediate nature of this category appears, in the types of head nouns found.

Similar to the Hard-Applied category, the most frequent head nouns include a range of abstract nouns used for organizing knowledge, such as *model*, *system* and *theory*. An important difference arises in the way they are premodified, however. Rather than material processes, premodifying nouns relating to human experience are more common, such as *power structures*, *belief system*, *personality development*, *attachment theory*. We see other nouns preceding the first-position premodifier less often, so the meaning relationships lack the complex nesting of meaning relationships present in the Hard fields' description of physical processes and are mostly limited to Content meaning relations.

In addition to the abstract, research-related head nouns, we see a difference arising with the head nouns *education* and *students* among the ten most frequent. Unlike the common head nouns from the previous categories, these relate to human experiences rather than physical processes. These are also modified by a wide variety of premodifying nouns (*high school*, *undergraduate*, *music*, *humanities*, etc.), but generally are limited to a Specialization meaning relationship.

Finally, when examining the nominally premodified head nouns in the Soft-Pure category, we see an even further drop in overall frequencies, suggesting a distinct difference in the linguistic norms of this knowledge community. Most of the frequent head nouns found are abstract nouns used for organizing knowledge, common academic

vocabulary, some of which also appears in the other disciplinary groupings. Like the Soft-Applied category, we find first-position premodifiers primarily related to human experience, such as *justice system*, *trauma theory*, and *kinship structure*. In addition, the head nouns *movement*, *process*, *relations*, and *culture* appear frequently, used to discuss broader systems within human experience, with most premodifying nouns specifying categories of human actors, such as *elite position*, *class relations*, and *suffragist movement*.

One of the most frequent head nouns in the Soft-Pure category that functions somewhat differently from the others is *process*. In this sub-corpus, it is often found with nominalizations (nouns derived from verbs using derivational suffixes) in the premodifying position, such as *standardization process*, *negotiation process*, *gender attribution process*, and *alienation process*. These forms were used to represent behavioral processes, based on Ravelli's categories of grammatical metaphor (1988). *Process* also appeared among the ten most frequently modified head nouns in the Hard-Applied category and was often premodified using nominalizations. However, in this case the processes most often encoded in the nominalization were material, such as *wash process*, *manufacturing process*, *benchmarking process*, and *production process*.

When examining the nouns most commonly used as nominal premodifiers in each disciplinary grouping, we again find distinct differences (Table 4.3). Notably, there are some nouns that occur in more than one column, such as *water* and *energy* in hard-pure and hard-applied, and *family*, *research*, *gender*, and *language* in soft-pure and soft-applied. Those overlaps do not cross the hard/soft divide, with the exception of *community*, which will be further discussed shortly. The overlapping words focus on physical elements in the hard disciplines, and aspects of human social life in the soft fields reflecting the research areas typically addressed in these disciplines.

First, in the Hard-Pure category, we find a preponderance of nouns related to physical entities, such as *water*, *body*, and *food*. The one exception to this is *community*, the majority of which were found in the Natural Resources and Environment category, again reflecting the interdisciplinary nature of that discipline as it addresses human interests more directly (Ex. 25).

(23) *dispersal plays a key role in the development of **community** structure (BIO)*

(24) *biotic interactions may also play a role in shaping **community** structure in these lakes (NRE)*

(25) *programs of education, **community** development, and infrastructural improvement point to its commitment to help the poorest countries (NRE)*

Table 4.3 Most frequent premodifying nouns (any position) in each Becher-Biglan category (*frequencies normalized per 10,000 words*)

Hard-Pure	Freq.	Hard-Applied	Freq.	Soft-Applied	Freq.	Soft-Pure	Freq.
<i>water</i>	6.75	<i>design</i>	5.89	<i>community</i>	4.31	<i>language</i>	2.25
<i>forest</i>	6.62	<i>water</i>	3.83	<i>family</i>	3.84	<i>gender</i>	2.02
<i>energy</i>	6.56	<i>steel</i>	2.59	<i>health</i>	3.66	<i>word</i>	1.89
<i>body</i>	6.17	<i>energy</i>	2.19	<i>research</i>	3.09	<i>class</i>	1.79
<i>community</i>	4.95	<i>data</i>	2.16	<i>college</i>	2.80	<i>family</i>	1.60
<i>eye</i>	4.69	<i>grip</i>	2.01	<i>group</i>	2.73	<i>research</i>	1.27
<i>food</i>	4.56	<i>soil</i>	1.91	<i>school</i>	2.65	<i>education</i>	1.11
<i>mutant</i>	4.53	<i>efficiency</i>	1.80	<i>language</i>	2.61	<i>material</i>	1.10
<i>wing</i>	4.48	<i>surface</i>	1.74	<i>policy</i>	2.49	<i>question</i>	1.08
<i>land</i>	3.72	<i>wall</i>	1.59	<i>gender</i>	2.40	<i>sports</i>	1.06

Among the most common premodifying nouns in the Hard-Applied category, we also find several physical referents, such as *steel*, *water*, and *soil*. While *water* most often premodified *quality* in the Hard-Pure group (an identity relationship), in the Hard-Applied group it most often premodified *utilities*, representing a purpose relationship (Biber et al.,

1999, p.590). These differences suggest the theoretical or descriptive focus in the Hard-Pure disciplines, while the expression of purpose relationships in Hard-Applied reflects the way those fields put theory to use in practical applications. The applied nature of this category is also reflected in the frequently occurring premodifiers *efficiency* and *design* (Ex.26, 27).

- (26) *the importance of including seismic design considerations in all tunnels and underground constructed facilities. (CEE)*
- (27) *This efficiency analysis was performed using an input-oriented non-discretionary (non- controllable) output approach. (IOE)*

In the two 'soft' categories, the most common premodifying nouns relate to human experiences, such as *family*, *education*, and *gender*. Considerable overlap occurs among the most frequent, with four of the ten occurring in both groupings, suggesting a more focused range of topics in these disciplines. Some differences appear in the use of words that occur in both categories. For example, the word *family* modifies *members* most commonly in both groups, but other common head nouns reveal different patterns. In the Soft-Pure disciplines, we find *family history*, *family name*, and *family secrets*, all content relationships, among the most frequent noun + noun sequences (Ex.28). By contrast, in the Soft-Applied group *family conflict* (subject type 1), *family responsibilities* (source), and *family support* (source) were among the most common (Ex.29).

- (28) *Isaac attempts to renounce his inheritance by not accepting the family land, but he cannot disinherit his own family history (ENG)*
- (29) *witnessing and victimization was inversely related to daily family support (PSY)*

Although beyond the scope of the present study, further analysis of meaning relationships in noun + noun sequences using the framework from Biber et al. (1999, p.590) may reveal broader patterns across disciplines. One question for a future study may be whether some disciplines or disciplinary groupings use a wider variety of noun + noun meaning relationships. Subject matter, text type, and other situational factors will also be important to consider as influences on use of complex noun phrases.

CHAPTER 5. DISCUSSION

In the first research question, this study asked to what extent patterns of variation in the frequency and function of nouns as nominal premodifiers correspond with traditional academic departmental divisions as defined using the Becher-Biglan typology. Based on the corpus analysis of MICUSP that was carried out, clear divisions were found in frequency of use both for individual disciplines and the groupings based on Becher's Hard-Soft and Pure-Applied categories. The general tendency uncovered in the data was a greater use of nouns as premodifiers in the Hard-applied category as reflected in two features: overall frequency of noun + noun premodification, and length of noun premodification sequences. By contrast, the Soft-Pure disciplinary grouping showed less use of this feature, with less than half the total normalized rate of noun + noun premodification. The two other groupings fell between the two extremes, with hard-pure disciplines higher and soft-applied lower.

Overall, the results of the MICUSP analysis showed that upper-level student writing aligns with previous findings on the topic of disciplinary variation (Biber & Gray, 2013; Gray, 2015; Jalilifar et al., 2017), showing higher frequencies of the nouns as nominal premodifiers, a 'compression' feature, in the hard sciences than in the 'soft' disciplines. As in Gray's study (2015), intermediate frequency levels appeared in disciplines associated with the social sciences. When comparing these results with previous work on academic research articles, it appears that the high-level student writers represented in MICUSP follow the same tendencies of professional disciplinary norms in use of nouns as nominal premodifiers. However, future studies focused more on writing proficiency development may find interest in comparing the frequencies found here with those in professional writing, as well as lower-level student writing, in more detail.

Because nominalizations are feature often used to increase informational density by reducing clausal processes to a noun and leaving meaning relations implicit (Halliday, 1989), it was expected that they would appear more frequently among modified head nouns and nominal premodifiers in the 'hard' disciplines. However, while they were found in the data, they did not appear at a much higher frequency in 'hard' than in the 'soft' disciplines. This aligns with Biber and Gray's work (2016), which found little variation across academic sub-registers in the use of nominalizations. It may be useful to examine this feature in more detail in future research by subdividing the category of nominalizations, perhaps using Ravelli's (1988) categorizations, which may reveal different frequency patterns within the sub-groupings.

5.1 Limitations

While the academic groupings used proved useful in analyzing the use of nouns as nominal premodifiers, some difficulties should be considered when categorizing disciplines. For example, while Natural Resources and Environment is grouped with the hard-pure category due to its grounding in environmental sciences, the language features found in this study were more characteristic of soft and applied disciplines, perhaps due to changes the field's goals, or possibly the types of topics that are addressed in class assignments for that major. The difficulty of categorization that was found in Natural Resources and Environment can also apply to the field of Linguistics, whose sub-disciplines can vary from 'hard' computational to 'soft' philosophical approaches. This increasing interdisciplinary quality in academia may call into question some aspects of traditional departmental grouping and shows the value of undertaking more empirical analyses of academic writing to better

understand disciplinary patterns and provide students with the linguistic resources needed to succeed in their knowledge communities.

A key limitation in this study was the fact that paper type was not examined as a contributing factor to language variation. The differences between the task types undertaken and their goals within a discipline have been shown to have an important influence on writer's language choices (Gray, 2015). Due in part to uneven distribution of paper types in MICUSP, the scope of the present study did not allow consideration of text type as an influence on use of nouns as premodifiers. However, a future study of MICUSP examining these same language features using Gray's situational framework (2015) could offer valuable information to complement the findings from the present study. In addition to the Subject/Topic category in that framework, other elements such as use of visuals accompanying a text could be valuable analytical tools given the importance of figures and illustrations in engineering and science student writing, for example.

5.2 Implications

The results of the analyses carried out in this paper provide information about noun premodification in advanced student writing, which may be compared to previous work on this feature in professional academic journal articles. Findings show students approaching the register norms for this feature, although the situational variables cannot be considered the same when comparing writing in a classroom context to published journal articles. Despite those differences, understanding noun phrase complexity in advanced student writing, particularly A-graded papers like we see in MICUSP, may offer valuable guidance to lower-level students still working to acquire the linguistic features that characterize their discipline. This may be especially useful in the context of English for Specific Purposes (ESP) and English for Academic Purposes (EAP) courses, as well as discipline-specific writing

programs, which face challenges in addressing the varied needs of student writers (Cooper & Bikowski, 2007).

While informationally-dense noun phrases are not necessarily the most easily comprehensible way of expressing a concept, they are an important characteristic of specialized research writing, and ESP programs must consider the type of writing needed in a field of study rather than aiming at more general conceptions of good or effective writing. In addition to understanding the meanings embedded in these premodification strings, students must also learn to use them in their own writing in order to meet expectations for concise and discipline-appropriate forms of expression.

When examining the particular nouns and noun strings used in premodification constructions in this study, it is important to note the lack of overlap between disciplinary groupings in their most common strings, as well as the small number of texts in which most of them appeared. This suggests that most noun + noun sequences are specific to discipline, and paper type and even topic may also be a strong influence on their use. These findings underline the value of a discipline-specific approach to teaching academic writing, either as a supplement to or a replacement for teaching more generalized academic English. Though it is valuable to understand a common core of academic writing, that alone may not be the most effective way to prepare students for the communication norms of their chosen field as they advance further in their education. Overall, due to the many situational influences on how noun phrase complexity features are used, the best approach may be to emphasize to students the importance of context on language use, avoiding an overarching idea of "good writing" and raising awareness of how language features serve specialized purposes in different knowledge communities.

REFERENCES

- Ädel, A., & Römer, U. (2012). Research on advanced student writing across disciplines and levels: Introducing the Michigan Corpus of Upper-level Student Papers. *International Journal of Corpus Linguistics*, 17(1), 3-34.
- Anthony, L. (2013), *AntCLAWSGUI* (Version 1.1.0w), Tokyo, Japan: Waseda University.
- Anthony, L. (2017), *AntConc* (Version 3.5.0), Tokyo, Japan: Waseda University.
- Becher, T., & Trowler, P. (2001). Academic tribes and territories. Buckingham: SRHE.
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27(1), 3-44.
- Biber, D. (1991). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (2003). 17 Compressed noun-phrase structures in newspaper discourse. *New media language*, 169.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *tesol Quarterly*, 36(1), 9-48.
- Biber, D., Egbert, J., Gray, B., Oppliger, R., & Szmrecsanyi, B. (2016). Variationist versus text-linguistic approaches to grammatical change in English: Nominal modifiers of head nouns. *The Cambridge Handbook of English Historical Linguistics*, 351-375.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2-20.
- Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language & Linguistics*, 15(2), 223-250.
- Biber, D., & Gray, B. (2013). Nominalizing the verb phrase in academic science writing. *The verb phrase in English: Investigating recent language change with corpora*, 99-132.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?. *Tesol Quarterly*, 45(1), 5-35.
- Biber, D., Gray, B., & Staples, S. (2016). Contrasting the grammatical complexities of conversation and academic writing: Implications for EAP writing development and teaching. *Language in Focus*, 2(1), 1-18.

- Biber, D., Grieve, J., & Iberri-Shea, G. (2009). Noun phrase modification. *One language, two grammars*, 182-193.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of written and spoken English. *Harlow: Longman*.
- Cooper, A., & Bikowski, D. (2007). Writing at the graduate level: what tasks do professors actually require?. *Journal of English for Academic Purposes*, 6(3), 206-221.
- Crossley, S., Russell, D., Kyle, K., & Römer, U. (2017). Applying natural language processing tools to a student academic writing corpus: How large are disciplinary differences across science and engineering fields?. *Journal of writing analytics*, 1.
- Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2), 165-193.
- Fang, Z. (2012). Language correlates of disciplinary literacy. *Topics in Language Disorders*, 32(1), 19-34.
- Fang, Z., Schleppegrell, M.J., & Cox, B.E. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research*, 38(3), 247-273.
- Gardner, S., & Nesi, H. (2012). A classification of genre families in university student writing. *Applied linguistics*, 34(1), 25-52.
- Gardner, S., Nesi, H., & Biber, D. (2018). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*.
- Garside, R., and Smith, N. (1997) A hybrid grammatical tagger: CLAWS4, in Garside, R., Leech, G., and McEnery, A. (eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. Longman, London, pp. 102-121.
- Gray, B. (2013). More than discipline: Uncovering multi-dimensional patterns of variation in academic research articles. *Corpora*, 8(2), 153-181.
- Gray, B. (2015). *Linguistic variation in research articles*. Amsterdam.
- Halliday, M. A. K. (1989). *Spoken and Written Language (Language Education)*. USA: Oxford University Press.
- Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 8(2), 183-207.
- Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied linguistics*, 20(3), 341-367.

- Hyland, K. (2002). Specificity revisited: how far should we go now?. *English for specific purposes*, 21(4), 385-395.
- Jalilifar, A., White, P., & Malekizadeh, N. (2017). Exploring nominalization in scientific textbooks: A cross-disciplinary study of hard and soft sciences. *International Journal of English Studies*, 17(2), 1-20.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *Tesol Quarterly*, 45(1), 36-62.
- Martin, J. R., & Halliday, M. A. K. (1993). *Writing science: literacy and discursive power*. Routledge.
- Michigan Corpus of Upper-level Student Papers. 2009. Ann Arbor, Michigan: The Regents of the University of Michigan.
- Musgrave, J., & Parkinson, J. (2014). Getting to grips with noun groups. *Elt Journal*, 68(2), 145-154.
- Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge University Press.
- Neumann, R., Parry, S., & Becher, T. (2002). Teaching and learning in their disciplinary contexts: A conceptual analysis. *Studies in higher education*, 27(4), 405-417.
- O'Donnell, M. B., & Römer, U. (2009). Michigan corpus of upper-level student papers.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied linguistics*, 24(4), 492-518.
- Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, 14, 48-59.
- Ravelli, L. (1988). Grammatical metaphor: an initial analysis. *Pragmatics, discourse and text. Some systemically-inspired approaches*, 133-147.
- Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): the design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 6(2), 159-177.
- Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, 33(2), 149-183.
- Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *Tesol Quarterly*, 47(2), 420-430.

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language*, 35(3), 434-453.

Wolfe-Quintero, K., Inagaki, S., Kim, H. Y., Kim, H. Y., & Inagaki, S. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity* (p. 187). Honolulu: University of Hawai'i Press.

Zhu, W. (2004). Faculty views on the importance of writing, the nature of academic writing, and teaching and responding to writing in the disciplines. *Journal of second language Writing*, 13(1), 29-48.

**APPENDIX: GROUPING DISCIPLINES BY UNIVERSITY OF MICHIGAN
DEPARTMENT AND BECHER-BIGLAN TYPOLOGY**

Table A1 University of Michigan Academic discipline groupings

Academic division	Disciplines	Academic division	Disciplines
Humanities and Arts	English (ENG) History and Classical Studies (HIS) Linguistics (LIN) Philosophy (PHI)	Biological and Health Sciences	Biology (BIO) Natural Resources and Environment (NRE) Nursing (NUR)
Social Sciences	Economics (ECO) Education (EDU) Political Science (POL) Psychology (PSY) Sociology (SOC)	Physical Sciences	Civil and Environmental Engineering (CEE) Industrial and Operations Engineering (IOE) Mechanical Engineering (MEC) Physics (PHY)

Table A2. MISCUSP disciplines grouped by Becher-Biglan typology

Category	Discipline	Category	Discipline
Hard-Pure	Physics (PHY) Biology (BIO) Natural Resources and Environment (NRE)	Soft-Pure	English (ENG) History & Classical Studies (HIS) Sociology (SOC) Linguistics (LIN) Philosophy (PHI)
Hard-Applied	Mechanical Engineering (MEC) Civil and Environmental Engineering (CEE) Nursing (NUR) Industrial and Operations Engineering (IOE)	Soft-Applied	Education (EDU) Economics (ECO) Political Science (POL) Psychology (PSY)

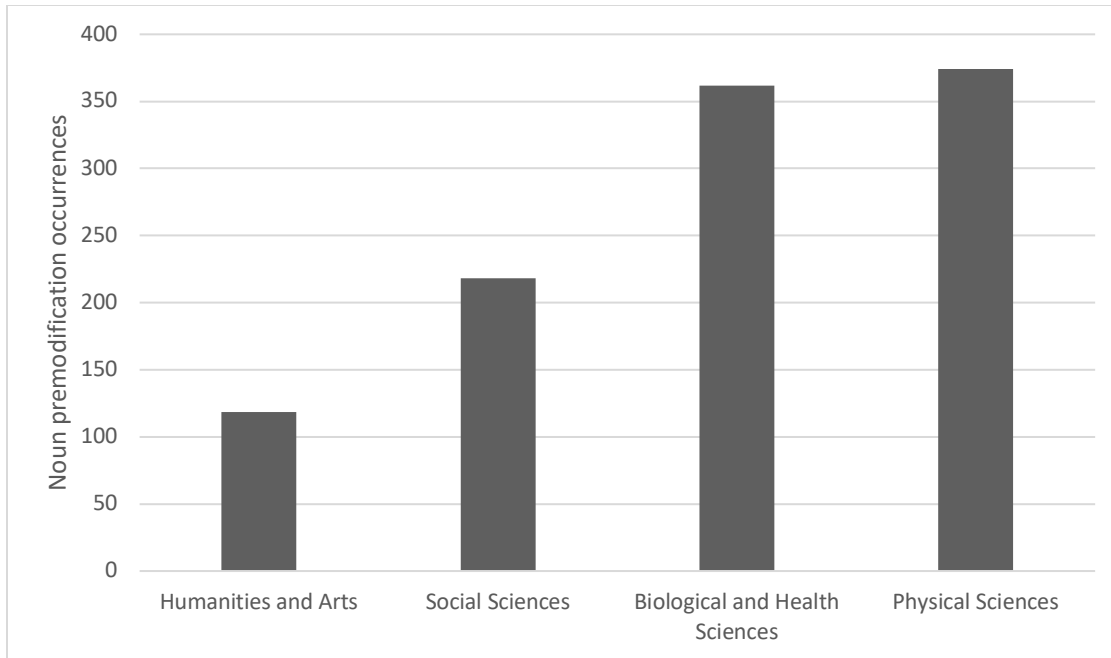


Figure A1. Frequency of nouns with one or more nouns as nominal premodifiers by University of Michigan academic groupings. (normalized per 10,000 words)

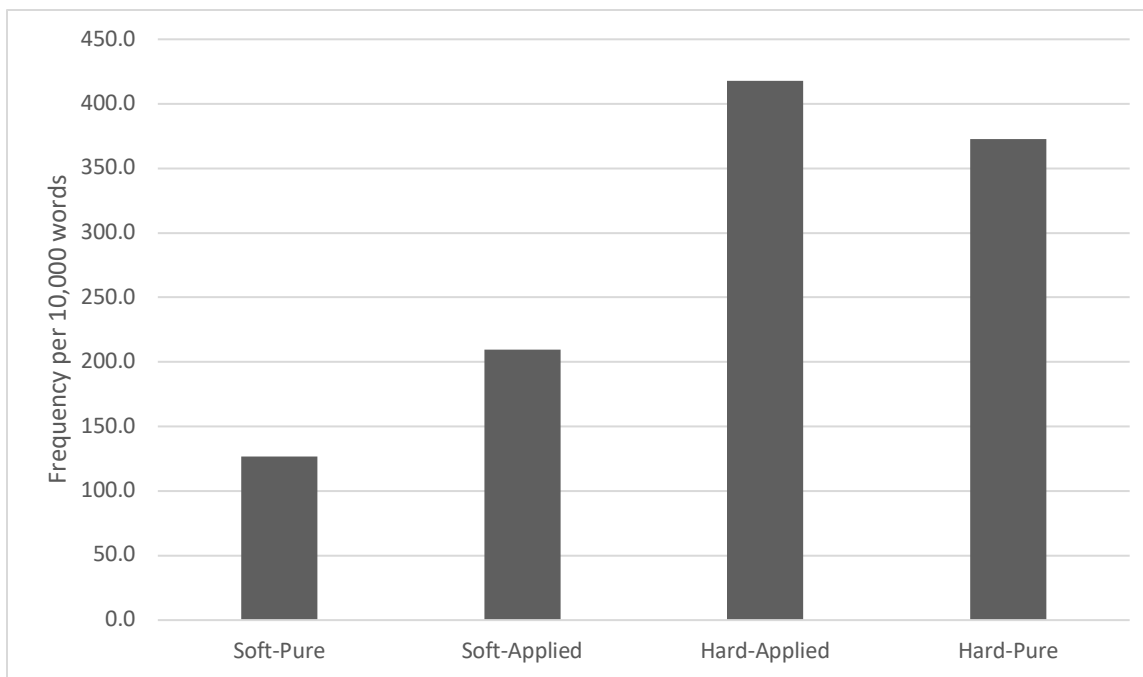


Figure A2. Frequency of nouns with one or more nouns as nominal premodifiers in the Becher-Biglan typology's academic groupings. (normalized per 10,000 words)