An exploratory analysis of noun phrases in civil engineering writing
Analyse exploratoire de syntagmes nominaux dans des textes de génie civil

Margaux Guillerit
1. Introduction

In the field of linguistics, numerous studies on academic writing have found that advanced academic writing is syntactically complex, but the nature of this complexity is not consistent across various reports. As explained by Biber & Gray (2010), for a long time, academic writing was, and is still believed to be, clausally complex, with many subordinate clauses. For example, Hyland (2002: 50) claims that academic writing is “structurally elaborate” and “complex” because it resorts to subordination. However, in the last few decades, some studies (Biber 1988; Biber & Gray 2010; Biber & Gray 2011) have also shown that academic writing is actually characterized by phrasal modification rather than clausal elaboration. More specifically, these studies have shown that complex noun phrases are a distinctive feature of academic writing, with numerous premodifiers and postmodifiers such as adjective phrases, nouns or prepositional phrases. Other studies also provide the evidence that noun phrase complexity increases as writers become proficient in their language (Biber, Gray & Poonpon 2011; Parkinson & Musgrave, 2014) and that register plays a role in noun phrase complexity (Lu 2011).

Rather surprisingly, if syntactic complexity has been extensively studied in general academic writing (i.e. without focusing on a particular discipline-related field), few studies have been conducted on field-specific writing (i.e. legal documents, medical texts or civil engineering texts). The problem is that general academic writing and specific purpose writing differ greatly, especially in terms of the types of nouns they use. Specific purpose writing is often fraught with technical vocabulary (i.e. terms), which are mostly nouns (Halliday & Martin 1993). These terms can combine themselves and can also be modified by elements, resulting in long and complex noun phrases,
which may be difficult to understand for non-specialist readers (Halliday & Martin 1993). For example, in civil engineering, phrases such as “the proposed retrofit for the new culvert lining” or “the subsurface exploration geotechnical engineering services” are common and require some disciplinary knowledge in order to understand them quickly.

Very few studies have been conducted on civil engineering writing, even though in civil engineering, writing skills are of the outmost importance, which may be counter-intuitive for people who are not knowledgeable in the field. Engineers do not actually build anything themselves; instead, they communicate information to their customers and partners thanks to different text types. Yet, recent hires are usually not good at writing, resulting in unclear and ambiguous texts, which can cause liability to the firm (Conrad 2017). A handful of studies have shown that this lack of writing efficiency can be seen at different clausal levels (Conrad 2017, 2018; Conrad, Lamb & Pfeiffer 2018), pointing at gaps between professional and student writing. But these studies in civil engineering focus only on clausal complexity; nothing has been written about phrasal complexity and the potential differences between students and practitioners. This is surprising, given that noun phrases are important in English for specific purposes because they encapsulate complex and technical concepts in very few words (Halliday & Martin 1993).

Given this gap in the literature regarding noun phrase complexity in specific purpose writing and more specifically in civil engineering, I propose to investigate noun phrase structure and use in student and practitioner texts (laboratory reports and reports), in order to determine whether there are significant differences between students’ and practitioners’ writing, and between the different types of texts under study.

In this study, I use the terms practitioners to refer to professional engineers, and students to refer to civil engineering students.

2. Literature review

Syntactic complexity in academic writing: an important research topic

Since the late 1980s, research in linguistics has considered syntactic complexity with great attention. In particular, groundbreaking studies such as Biber (1988) and Biber & Gray (2010, 2011) have shown that, contrary to popular belief among the linguistics community, academic writing is characterized by phrasal modification rather than by clausal elaboration. These studies have been able to identify features of academic writing by comparing this register with casual conversation, showing that the features or type of syntactic complexity vary depending on the register. For instance, Biber et al. (1999) showed that noun phrases with a modifier are rare in conversation, but are a distinctive feature of academic prose, with close to 60% of noun phrases having a pre- or postmodifier in this register. Biber & Gray (2010) showed that phrasal modifiers such as premodifying nouns, premodifying adjectives and prepositional phrases as postmodifiers are characteristic of academic writing, while subordination is prominent in conversation. In a diachronic study of noun phrase use, Biber & Gray (2011: 229) also found that over the past three centuries, five non-causal grammatical devices have become increasingly common in academic writing, which suggests that students need
to master their use and structure: attributive adjectives (e.g. “gradually expanding cumulative effect”), premodifying nouns (e.g. “baggage inspection procedures”), postmodifying prepositional phrases (e.g. “a high incidence of heavy alcohol consumption amongst patients”), appositive noun phrases (e.g. “Dallas Salisbury, CEO of the Employee Benefit Research Institute”) and nominalizations (e.g. “consumption”).

In these studies, academic writing is described as “compressed”, which means that it relies heavily on phrases (and especially on prepositional phrases) to add information instead of dependent clauses. This results in numerous compressed noun phrases, i.e. complex embedded noun phrases (Biber & Gray 2010). Consider the following example taken from Biber & Gray (2010: 7):

> From the system perspective, these stages are marked by [the appearance of new systemic mechanisms and corresponding levels of complexity].

This sentence contains several embedded noun phrases (shown in between square brackets), which together form one compressed noun phrase. Note that this noun phrase is quite long. Contrary to what the term “compressed” seems to indicate, nominalized constructions are not necessarily shorter than clausal constructions.

But why are compressed noun phrases used so extensively in academic writing? For Halliday & Martin (1993), one answer is that nominalizations enable scientific writing to represent actions, events and even qualities as if they were objects. Scientific discourse deals with many difficult ideas, so the possibility of having complex concepts being represented as static things makes them easier to examine. Another answer is that with nouns, complex phenomena are summarized and packaged in reduced units of language, which comes in handy in technical domains. Consider these examples given by Halliday & Martin (1993: 87):

1- How quickly cracks in glass grow
2- Glass crack growth rate

These two examples have the same meaning. However, (1) is a finite clause, while (2) is a noun phrase; “how quickly” has been replaced by the single noun “rate”, and the verb “grow” by the noun “growth”. The noun phrase in (2) expresses a complex phenomenon using only nouns, which allows for faster communication, since more information is packaged in few words. Called “grammatical metaphors” because they are the result of the substitution of a grammatical class for another, such noun phrases are characterized by a high lexical density (Halliday & Martin, 1993: 87). For Biber & Gray (2010), compressed noun phrases are favored over clausal phrases because they are more economical and allow for faster and more efficient reading. Fang, Schleppegrell & Cox (2006) note that thanks to the great variety of pre- and postmodifiers, noun phrases can pack a lot of information in a clause.

However, although economical from a syntactic viewpoint, compressed noun phrases may also lead to semantic ambiguity, and may therefore be hard for readers to understand and for writers to use correctly. Halliday & Martin note that such noun phrases (or “grammatical metaphors”) are “the major source of problems facing those who are apprenticed to the discourse of science”, whether they are native speakers of English or not (1993: 57). Indeed, Biber & Gray show that in a compressed noun phrase, it can be hard to determine the exact meaning relationship between the head noun and the phrasal modifiers (2010: 11). As a result, Biber & Gray (2010) explain that the extensive use of phrasal modifiers makes academic writing implicit in terms of meaning and sometimes difficult to understand. Since phrasal modifiers – such as
nouns or prepositional phrases – can express a vast array of relationships while being similar in form, they can be ambiguous. Additionally, Biber & Gray (2011) explained that noun phrases in modern academic writing are more likely to be ambiguous, because the increase of nouns as premodifiers and prepositional phrases as postmodifiers since the 17th century has been accompanied by a major expansion in meaning, leading to difficulties in writing and reading academic texts. Numerous types of nouns can be used as nominal premodifiers and each of them can express a vast range of meaning relationships. For instance, although similar in form, “spring water” can be paraphrased as “an N2 that comes from N1”, while “mineral poisons” can be paraphrased as “an N2 that is composed of N1” (Biber & Gray 2011: 239). Biber & Gray (2011: 238) add that while the last century only permitted a single noun as premodifier, modern academic writing easily allows for two, three or four premodifying nouns, with for instance expressions such as “hill committee report”, “life insurance table”, “plasma concentration time curve”. The situation is similar with prepositional phrases used as postmodifiers: for example, in and on have recently developed abstract meanings, as in “an increase in efficiency”, which does not refer to a physical location (Biber & Gray 2011).

12 If complex noun phrases can be hard to understand and require skills in order for the writer to be able to form them correctly, it may be because they are acquired later in life than clausal expressions. Halliday & Martin (1993) explained that children learn clausal constructions first, and then turn to phrasal structures as they grow up and are more comfortable with their language. Similarly, in their study of academic writing, Biber, Gray & Poonpon (2011) hypothesize the existence of a developmental index which shows that non-causal syntactic complexity goes up at higher proficiency levels. The stages of syntactic complexity develop from finite dependent clauses used as constituents in other clauses to dense phrasal (i.e. non-causal) dependent groups which are used in constituents in noun phrases (Biber, Gray & Poonpon 2011: 28). Parkinson & Musgrave (2014) tested this hypothesis that non-causal noun phrase complexity increases across levels of studies for second language academic writing. They compared two graduate student groups, and found that the more proficient advanced students rely on noun premodifiers and prepositional phrase postmodifiers, while the less advanced writers rely mostly on attributive adjectives as modifiers.

13 However, the level of proficiency does not always account for differences in syntactic complexity; the register plays a role as well. For instance, in a corpus-based study, Lu (2011) analyzed fourteen syntactic complexity measures in texts written by Chinese college students whose major was English, and one of his research questions was to determine if register had an impact on syntactic complexity. Lu found that despite all being written by students of the same level, argumentative essays were syntactically more complex than narrative essays.

14 To sum up, syntactic complexity is a major field of linguistic research in the context of English for Specific Purposes and English for Academic Purposes, and has been studied extensively in academic writing. Overall, researchers have focused on academic writing in general, without differentiating between discipline-specific fields. The problem with that approach is that each specific field is different and has its own characteristic features. Technical domains such as legal English, medical English or civil engineering English have complex terminology and phraseology of their own, and need to be studied individually in order to provide an accurate analysis of their language. In this
paper, I therefore propose to contribute to filling this gap, by studying noun phrase complexity in the specific domain of civil engineering writing.

**Noun phrase complexity in civil engineering writing**

In civil engineering, writing is an invisible but crucial activity: practitioners analyze data, design projects and communicate about them to their customers thanks to different texts. Therefore, practitioners produce written documents on a regular basis, such as technical memoranda and various reports such as design reports (which provide analyses and recommendations for design), investigative results reports (which comprise analyses used by other engineers for design), proposals or specifications. Those texts constitute the very basis of their work and are of the highest importance. They have to be both clear and concise in order to be understood quickly and properly by the readers (Conrad 2017). Effective writing skills are essential in maintaining clients’ satisfaction, executing projects and avoiding unintentional liability for the firm (Conrad & Pfeiffer 2011; Conrad, 2017).

According to senior engineers, recent civil engineering hires in the USA are not competent at writing, even those with high grades in college. For decades, employers have told their new hires that their writing was ineffective or unclear, and recent graduates from prestigious civil engineering programs have expressed their discontent towards the preparation they received for professional writing (Conrad 2017). Only a handful of research articles have looked into civil engineering writing, but recent studies by Conrad (Conrad 2017, 2018; Conrad, Lamb & Pfeiffer 2018) showed that the gap between practitioners and students (or recent graduates) is significant in sentence structure, word choice, genre organization, grammar and punctuation errors and use of passive structures. Conrad’s 2017 study investigated – among other subjects – the difference in sentence structure between student and practitioner writing, and found that students presented more clausal complexity than practitioners in their sentences. Students use subordination more frequently (and therefore embedded clauses) and often express more than one idea per sentence. On the contrary, practitioners use less clausal embedding; their sentences rarely contain more than one dependent clause, because they usually express one idea per sentence, in order to meet their customers’ need for quick reading and to be unambiguous. As explained in another study by Conrad & Pfeiffer (2011), the syntactic complexity of practitioners manifests itself at the noun phrase level, especially thanks to the extensive use of prepositional phrases.

Although they investigate clausal complexity, these recent studies do not discuss whether the heavy use of subordination in student writing is accompanied by less complex noun phrase structure, as would be predicted by the developmental studies on general academic writing. To my knowledge, there has been no study focusing on noun phrase use and structure in student and practitioner civil engineering writing, although in technical subjects noun phrases are important, because they package complex concepts in very few words (Halliday & Martin 1993). In the present study, I will therefore investigate the noun phrase structure in both student and practitioner texts, in order to determine if the gap between practitioner and student writing (Conrad & Pfeiffer 2011; Conrad 2017) is also to be found in noun phrase complexity.

Specifically, the present study addresses the following research questions:
In what way do noun phrases used in student papers differ from noun phrases in practitioner texts, especially with respect to noun phrase frequency, structure and length? Does the register (i.e. reports v. laboratory reports) play a role in the way noun phrases are used in the students’ papers?

3. Methodology

This study is part of a larger project called the Civil Engineering Writing Project (CEWP),² initiated by Susan Conrad at Portland State University (Oregon, USA) in 2009. It was launched to address a deep-rooted issue in engineering education, i.e. the discrepancy between students’ writing skills and the demands of writing in the workplace. In order to gain better understanding of the situation and to find solutions, Conrad’s research team analyzes and compares the writing of students and engineers. The contributors use their findings to design courses and materials which address particular weaknesses found in student papers. The materials are used in engineering classes or for independent analyses. They also study whether students’ writing skills improve over time after using the new materials.

The project is based on a corpus linguistics approach. Different kinds of written documents were collected from undergraduate students and engineers in the United States; those texts altogether form the corpus on which the project is based. For confidentiality reasons, people and company names were anonymized using the template “Companyname1, Companyname2”, etc. and “Lastname1, Lastname2”, etc. Currently, the corpus comprises laboratory (lab) reports, technical memos and reports written by students, as well as reports, field notes and technical memos written by practitioners. Various types of analysis were conducted in order to compare the vocabulary, grammar and organization used by practitioners and students.

For the present study, I used a subset of the corpus gathered for the CEWP project, which comprises manually coded noun phrases taken from practitioner reports, student reports and student laboratory reports. For clarity purposes, I call this corpus the NPcoded corpus (see Table 1). Regarding the nature of the registers represented in the NPcoded corpus, both practitioner and student reports comprise design reports and investigative reports. Both are intended for real or hypothetical clients, even for the student reports. Laboratory reports are assignments in which all students do the same experiment with a given data set or parameters; the audience is the instructor for the class. The practitioners are from private firms or public administration, and the students are third and fourth year students.

These three registers – practitioner reports, student reports and student lab reports – were chosen for different kinds of comparisons. They form a continuum going from academic writing to professional writing. Student reports and lab reports are a form of academic writing, since they are both academic assignments given by the professor teaching the class. Student reports are actually the most advanced kind of writing that students do for their capstone project for their undergraduate degree. That being said, student reports are also very close to professional writing because they are supposed to mimic actual engineers’ reports and are designed for real or imagined clients. Thus, student reports, although they are a form of academic writing, are supposed to resemble practitioner reports, which are written in industry, not academia. In other words, the boundary between academic and professional writing is fuzzy, due to the
complex interactions between the academic and the professional worlds. As ESP specialists, if we want to help prepare students to the professional world, it is necessary to conceptualize learners' needs in their future jobs.

The NPcoded corpus was grammatically annotated by five research assistants using an interactive coding tool specifically created for this project, called NPcoder. The coding process required manual coding of each first order noun phrase (i.e. noun phrases not embedded in other noun phrases) and their pre- and postmodifiers. After the coding was done, the coded data was manually reviewed in order to correct potential coding mistakes. Counts of noun phrases, premodifiers and postmodifiers for each register were obtained using software programs designed for the project; the counts were normed per million words afterwards in order to allow comparisons with previous studies. I used SPSS 25.0 (IBM, 2017) to analyze the data and compare the different groups of texts, using descriptive statistics. Boxplots were used in order to compare the distributions of noun phrases for each group. Then, a concordancer specifically designed for the project was used to find examples.

Table 1 - NPcoded corpus

<table>
<thead>
<tr>
<th>Register</th>
<th>Number of words</th>
<th>Number of texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practitioner reports</td>
<td>37,246</td>
<td>36</td>
</tr>
<tr>
<td>Student reports</td>
<td>29,942</td>
<td>29</td>
</tr>
<tr>
<td>Student laboratory reports</td>
<td>40,452</td>
<td>57</td>
</tr>
<tr>
<td>Total</td>
<td>107,640</td>
<td>121</td>
</tr>
</tbody>
</table>

For the purpose of the present study, a noun phrase was defined as follows: a phrase with a noun or a pronoun as its head, which may include one or more determiners, premodifiers and postmodifiers (Biber et al. 1999). The basic structure of noun phrases is:

Determiner + (premodification) + head noun + (postmodification and complementation). (Biber et al. 1999)

There are several types of noun premodifiers and postmodifiers, which allow for expansion of the noun phrase. For the present study, the premodifiers and postmodifiers were selected following those in Biber et al. (1999). Premodifiers include adjectives, nouns, genitives and numerals, while postmodifiers include prepositional phrases, relative clauses, nonfinite clauses (-ed clauses, -ing clauses, to clauses and bare infinitive clauses), complement clauses and noun phrases as appositives. For both premodifiers and postmodifiers, a category called “other” was added, which included any element that did not fit in the aforementioned categories, such as directional adverbs modifying nouns (e.g. “the area south of Pease Road”) or adjective phrases used as postmodifiers (e.g. “Local water well logs available from the Oregon Department of Water Resources website”). Noun modifiers can be phrasal (adjectives, nouns, prepositional phrases, numerals, genitives, appositives) or clausal (nonfinite clause, relative clauses, complement clauses).
4. Results and discussion

This section begins with quantitative comparisons of noun phrases in the NPcoded corpus (4.1). Noun phrase frequency and structure (4.1.1) are analyzed, as well as length (4.1.2). Then I compare the statistical results to published frequencies (4.2) in Biber et al. (1999). In section 4.3, I specifically answer the research questions.

4.1. Quantitative comparison of noun phrases in the NPcoded corpus

4.1.1. Frequency and structure

Although the differences between registers are not great, Table 2 shows that lab reports have the highest number of noun phrases per million words, then the student reports, and finally the practitioner reports.

Table 2 – Frequencies of first order noun phrases per register (per million words)

<table>
<thead>
<tr>
<th>Register</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student lab reports</td>
<td>160,511</td>
</tr>
<tr>
<td>Student reports</td>
<td>149,989</td>
</tr>
<tr>
<td>Practitioner reports</td>
<td>138,028</td>
</tr>
<tr>
<td>Total</td>
<td>448,528</td>
</tr>
</tbody>
</table>

These frequency patterns are twice as low as those reported in Biber et al. (1999) for academic writing, but it is because we only took into account first order noun phrases (i.e. noun phrases not embedded in other noun phrases). Consider for instance this passage, taken from a student report (first order noun phrases are in between brackets with the head underlined, and embedded ones are in bold):

[Our objective for determining location] was to find [a place on site that could accommodate our structure’s footprint, providing consideration for future road access through the site to the putting green annex and resulting in minimal disturbance of existing features]. [We] have determined from [our site investigation] and after meeting with [Companyname1] and [stakeholders] that [the best possible site for the landscaping bin project] is along [the western perimeter].

If we only look at first order noun phrases, it seems as though the paragraph only has eight noun phrases. But if we take into account the embedded noun phrases, then the count is seventeen noun phrases total. The proportions double, which make the NPcoded corpus close to Biber et al.’s proportions. So when considering embedded noun phrases as well, the NPcoded corpus confirms previous observations (Halliday & Martin 1993) that technical/scientific writing is heavy on nouns in general, although for practical reasons, we only studied first order noun phrases in the present study.

Figure 1 shows that in the NPcoded corpus most noun phrases have pre- or postmodification (or both), for all three registers. The table beneath the stacked bars...
shows the raw counts for each register. About 70% of noun phrases have pre- or postmodification in practitioner and student reports, and about 60% in student lab reports. The following example, taken from a student report, illustrates this tendency:

[Organizationname1] is [a not-for-profit community group in Cityname1, Statename1 that was created in Cityname2, Statename2 and replicated across the country as small community groups with goals to improve the walkability and vitality of their respective neighborhoods].

In this excerpt, there are two first order noun phrases (heads in bold). One of them is not modified, but the second one has extensive modification, resulting in a very long noun phrase. For “group”, premodifiers are underlined and postmodifiers are shown between square brackets:

a not-for-profit community group [in Cityname1, Statename1] [that was created in Cityname2, Statename2 and replicated across the country as small community groups with goals to improve the walkability and vitality of their respective neighborhoods].

Such lengthy and extensively modified noun phrases are very common in the NP coded corpus. As explained by Biber et al. (1999), such patterns result in a high density of information throughout the texts.

Figure 1 - Distribution of noun phrases with premodifiers and postmodifiers

Surprisingly, the proportions in Figure 1 are generally similar for practitioner and student registers. More particularly, Figure 1 shows that for practitioner and student reports, the most common type of modification is premodification only, with about 30% of noun phrases having a premodifier. But for student lab reports, no modification is the most common case, with close to 40% of noun phrases having no modifier. So the factor determining the most common modification case does not seem to be based on the nature of the writer (practitioner v. student), but rather on the text type (reports v. lab reports). This difference in modification patterns may be due to the difference in audience between the two registers; reports are intended for clients, while lab reports are intended for the professor teaching the class. Since the students know (or assume)
that their instructor is familiar with what they write, they may be more likely to use single nouns, instead of being precise and using modifiers. This point is discussed in detail further, in section 4.3.

Regarding premodifier types across registers, adjectives and nouns are by far the most common type, as shown in Figure 2. They represent around 90% of premodifiers in each of the three registers. Numerals, genitives and other premodifiers represent only 10% of premodifier types. So in the three registers, premodifier types are ranked similarly: nouns and adjectives first, and then numerals, genitives and other premodifiers. More specifically, adjectives are the most widely used premodifier for both student reports and student lab reports. For practitioner reports, the most widely used premodifier is nouns. Practitioner reports have the most numerous adjectives and nouns as premodifiers, then student reports, and finally student lab reports.

Figure 2 - Frequency of premodifier types across registers

Figure 3 shows that for all three registers, prepositional phrases are by far the most common postmodifier type. Prepositional phrases represent around 70% of postmodifier types in each register (70.1% in practitioner reports, 69.6% in student reports and 73.8% in student lab reports). For student registers, the frequency of clauses (relative clauses, noun complement clauses and non-finite clauses) is low – 7.4% and 8% - which confirms Biber’s observation regarding the non-clausal nature of student academic writing (Biber & Gray 2010).
4.1.2. Length

As shown in Table 3, noun phrases average between 4 and slightly over 5 words in the corpus.

Table - Descriptive statistics on noun phrase length

<table>
<thead>
<tr>
<th>Register</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practitioner reports</td>
<td>5.08</td>
<td>4</td>
<td>1</td>
<td>46</td>
</tr>
<tr>
<td>Student reports</td>
<td>4.58</td>
<td>3</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>Student lab reports</td>
<td>4.21</td>
<td>3</td>
<td>1</td>
<td>40</td>
</tr>
</tbody>
</table>

Practitioner report noun phrases are on average one word longer than student register noun phrases. The medians for the three registers confirm that most noun phrases are under five words. But despite this relatively short noun phrase average length and the low medians, observations of the minimum and maximum lengths show how wide the range is for noun phrase length in the corpus. The smaller noun phrases are one word long, which means that they are either single nouns (with no determiner) or pronouns. Some examples include: “this”, “vehicles”, “Companyname1”, “Excel” or “ourselves”. But the longer noun phrases are extremely long in all three registers: 46 words for practitioner reports, 36 words for student reports and 40 words for student lab reports. Noun phrases can be particularly long for different reasons. The first reason is that
they can be lists of things (places, regulations, measurements, agencies...), such as in the example below, which is 36 words long (shown in between square brackets, with the head bolded):

(1a) The project would likely require [the following permits from regulatory agencies including: Streambed Alteration Agreement from CDFG Section 404 permit from Army Corps of Engineers Section 401 Water Quality Certification from Regional Water Quality Control Board Consultation with USFWS and NMFS].

Phrase length may also be explained by the fact that they can have many embedded modifiers, which usually are prepositional phrases. For instance, below is an example taken from a student report:

(1b) There must be a separated bay [for each of the three types of materials being stored: topsoil, hemlock mulch, and compost].

The noun phrase in (1b) is in italics and is 18 words long. The head is “bay”, and it has a determiner (“a”), a premodifier (underlined) and one long postmodifier (in between square brackets). The postmodifier is one single prepositional phrase, in which there are five embedded postmodifiers – each of them being at different syntactical levels:

for each...
...of the three types [modifies “each” - so it is embedded postmodifier 1
...of materials being stored: -> modifies “types” - so it is embedded postmodifier 2
...topsoil, hemlock mulch, and compost. -> modifies “stored” - so it is embedded postmodifier 3

Such dense and long postmodification patterns are common in the corpus, which results in complex and long noun phrases. Extreme cases also abound in the three registers (practitioner reports, student reports and student lab reports), as shown in Figure 4.

Figure - Boxplots for noun phrase length

Figure 4 shows that for each register, there is a great number of very long noun phrases, represented by the circles and asterisks above the 4th quartile of the boxplot. The number next to those symbols refers to the case number of the noun phrase. It is worth noting that my analysis of these outliers showed that most of them are objects or appositives, which seems to indicate a strong relation between noun phrase length and the syntactic role a noun phrase fulfils. Example 1a above displays a striking contrast between the subject “The project” (2 words long) and the object of the verb “require” (36 words long). This discrepancy between the length of the subject versus that of the predicate is typical in the NPcoded corpus. Anecdotally, my analysis of randomly chosen noun phrases used in subject role in each register also showed that most of them were between 1 and 5 words. This pattern conforms to the principle of “end weight”, according to which heavy elements (usually object/complements/modifiers) are at the end of a sentence while the subject is kept as short as possible in order to facilitate the comprehension process. Indeed, if we start from example 1a, a sentence such as 1c (see below) would be almost impossible to understand at first, and multiple readings would be necessary.

(1c) The following permits from regulatory agencies including: Streambed Alteration Agreement from CDFG Section 404 permit from Army Corps of Engineers Section 401 Water Quality Certification from Regional Water Quality Control Board Consultation with USFWS and NMFS would likely be required by the project.
Reports and lab reports are supposed to be brief and understood quickly (Conrad 2017: 194), hence the absence of such sentences in the NPcoded corpus and the brevity of subjects. As shown above in example 1b, having very a long subject is likely to prevent the reader from understanding the sentence quickly.

4.2. Quantitative comparison with published frequency patterns

In section 4.1, I focused on the differences in use, frequency and length of noun phrases within the NPcoded corpus, which allowed me to give an overview of how noun phrases were used by each group and in each register. In the present section, I compare the quantitative results (given in percent) with proportions for academic prose reported in Biber et al. (1999), in order to show how noun phrases are used differently in the different registers and groups. This comparison is reported in Tables 4 and 5. I chose to compare the NPcoded corpus with academic writing to conduct a fine-grained analysis of the NPcoded corpus. These two text types being both produced by experts, they are comparable.

Table - NPcoded corpus data for modification patterns compared to published data

<table>
<thead>
<tr>
<th>Features</th>
<th>Practitioner reports</th>
<th>Student reports</th>
<th>Student lab reports</th>
<th>Biber et al. (1999)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Premodifier only</td>
<td>34.7%</td>
<td>31.8%</td>
<td>27.4%</td>
<td>25%</td>
</tr>
<tr>
<td>2. Postmodifier only</td>
<td>17.7%</td>
<td>19.7%</td>
<td>19.8%</td>
<td>20%</td>
</tr>
<tr>
<td>3. No modifier</td>
<td>26%</td>
<td>30.3%</td>
<td>37.8%</td>
<td>41%</td>
</tr>
<tr>
<td>4. Both premodifier and postmodifier</td>
<td>21%</td>
<td>18.1%</td>
<td>14.8%</td>
<td>12%</td>
</tr>
</tbody>
</table>

*These figures are estimations taken from bar graphs.

Regarding noun phrase modification in general (features 1, 2, 3, 4 in Table 4), the NPcoded corpus has more pre- or postmodification (or both) than academic writing. Figure 1 showed that 70% of noun phrases had pre- or postmodification (or both) for practitioner and student reports, and 60% for student lab reports. In academic prose, however, the proportion is only 57%. Furthermore, Figure 1 showed that the most typical pattern was premodification only for student and practitioner reports, while student lab reports had no modifier as their preferred pattern. Table 4 shows that for academic prose, the most common case is also no modifier, as in student lab reports, with a similar proportion for the two registers. In general, student lab reports have much closer proportions to those of academic writing than those of practitioner and student reports. This similarity between lab reports and academic prose may be due to the fact that lab reports are fully academic papers (i.e. written by students for an assignment), while reports – even those written by students – are designed for a professional purpose, whether real or imaginary.
Table - NPcoded corpus data for pre- and postmodifier types compared to published data

<table>
<thead>
<tr>
<th>Features</th>
<th>Practitioner reports</th>
<th>Student reports</th>
<th>Student lab reports</th>
<th>Biber et al. (1999)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Nouns as premodifiers **</td>
<td>49.1%</td>
<td>45.1%</td>
<td>45%</td>
<td>30%</td>
</tr>
<tr>
<td>6. Adjectives as premodifiers **</td>
<td>43.9%</td>
<td>47.2%</td>
<td>46.9%</td>
<td>70%</td>
</tr>
<tr>
<td>7. PPs as postmodifiers ***</td>
<td>70.1%</td>
<td>69.6%</td>
<td>73.8%</td>
<td>77%</td>
</tr>
<tr>
<td>8. Appositives as postmodifiers ***</td>
<td>6.2%</td>
<td>3.6%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>9. Relative clauses as postmodifiers ***</td>
<td>6.1%</td>
<td>8%</td>
<td>6.6%</td>
<td>12%</td>
</tr>
<tr>
<td>10. Nonfinite clauses as postmodifiers ***</td>
<td>11%</td>
<td>10.5%</td>
<td>9.8%</td>
<td>6%</td>
</tr>
</tbody>
</table>

*These figures are estimations taken from bar graphs.
**The percentages are calculated out of the total number of premodifiers for each register.
***The percentages are calculated out of the total number of postmodifiers for each register.

Let’s now have a look at premodifiers, displayed in features 5 and 6 in Table 5. Figure 2 showed that nouns and adjectives were widely used in all three registers, with a slight preference for nouns in practitioner reports and a slight preference for adjectives in the student registers. In Biber et al. (1999), on the contrary, adjectives are by far the most common premodifier, and nouns are less frequent. This obvious difference may be due to the high number of noun sequences in the NPcoded corpus which account for names of tests, programs, regulations, etc., leading to a great number of nouns as premodifiers, as the examples below show.

- The Proctor tests type of test
- The Hec-ras program computer program
- The University Honors College building facility
- The Access Oregon Plan regulation
- The Soil Mitigation Plan regulation

The high frequency of nouns as premodifiers in the NPcoded corpus could also be linked to the need for conciseness and brevity in reports and lab reports. As explained in Biber et al., noun + noun sequences (and by extension multiple noun sequences) bring about extremely densely packaged information and therefore are great when brevity is required, because they can express a vast array of relationship meanings (1999: 589–90). Examples for the NPcoded corpus include:

- funding issues issues to find funding
- cost estimates estimates for the cost of something
- cast iron pipe a pipe made of cast iron

Precisely, reports and lab reports in civil engineering have to be both short and dense in terms of information, so that what is at stake can be understood very quickly by the clients (Conrad 2017: 194). Furthermore, naming exact entities is of the highest
importance, according to practitioners themselves (Conrad 2017: 204). Choosing the accurate word or phrase to refer to something prevents them from causing potential liability for the firm (Conrad 2017: 204). For example, “The Soil Mitigation Plan” is not any plan about soil regulation but a very specific one, hence the need to mention it fully. In the NPcoded corpus, practitioner reports have the highest proportion of nouns compared to student registers, which suggests that noun + noun sequences are favored by engineers over adjectives.

In table 5, postmodifier use (features 7, 8, 9, 10) shows that the NPcoded corpus and academic prose rely heavily on prepositional phrases, making them their preferred postmodifier. This preference might be explained by the fact that prepositional phrases are economical – few words are needed to express a vast array of meaning relationships –, which is very useful in the context of reports and lab reports, since the writers have to be brief (Conrad 2017). Prepositional phrases also make localizations, measurements and descriptions as precise as possible, which is essential in civil engineering documents (Conrad & Pfeiffer 2011) (e.g. “the mouth of the Columbia River Gorge: a total length of 35 feet”; “a pavement section consisting of 3 inches of asphalt over 12 inches of base rock”). The NPcoded corpus and academic prose are also similar in their use of appositives, although practitioner reports have twice as many as the other registers. But for relative clauses, the difference between the NPcoded corpus and academic writing is striking. Academic writing has around twice as many relative clauses as the NPcoded corpus. One possible explanation could be that since civil engineering texts have to be brief (Conrad 2017), a nonfinite clause or a shorter alternative will almost always be preferred to a relative clause (e.g. “the final design plans dated June 28” is a shorter alternative than “the final design plans which were dated June 28”). A relative clause requires a finite verb and most of the time a relativizer, which makes it longer than a nonfinite clause. Indeed, there are more nonfinite clauses in the NPcoded corpus than in academic prose.

4.3. Answers to the research questions

In this section, I sum up the results presented above, answering the research questions.

In what way do noun phrases used in student papers differ from noun phrases in practitioner texts, especially with respect to noun phrase frequency, structure and length?

The analyses showed few differences between student and practitioner texts, but many similarities. In terms of overall frequency, although student registers have the most frequent first order noun phrases, the difference is very slight. In fact, when we break down the data and look at the structure of the noun phrases, students and practitioners use the same patterns to build noun phrases. Both students and practitioners use nouns and adjectives as their favored premodifiers, with similar frequencies – although practitioners tend to use nouns a little more. For postmodifiers, both students and practitioners overwhelmingly use prepositional phrases as their favored postmodifier, with similar frequencies as well. There are slight differences in the way the other postmodifiers are used, but their proportions are so small that more data would be needed in order to draw conclusions. In terms of noun phrase length, the similarities are striking, with means, median and range being almost the same for students and practitioners. They both have about the same amount of extreme cases of long noun phrases, which seems to suggest that civil engineering writing in general resorts to
lengthy and complex noun phrase in order to express technical and complex concepts as briefly as possible. This confirms the findings from previous studies (Biber 1988; Biber & Gray 2010; Biber & Gray 2011) which showed that academic writing relied heavily on complex noun phrases.

The fact that both practitioners and students – whose minimum level of schooling is third year undergraduate – use complex and long noun phrases seems to follow Halliday & Martin’s (1993) idea that phrasal structures are acquired once the writer is comfortable with their mother tongue, and here with the language of civil engineering. In contrast to the observations in Biber, Gray & Poonpon (2011) and Parkinson & Musgrave (2014), it does not seem as if there is an increase in noun phrase complexity as the level of expertise of the writers goes up. This might be due to the fact that the students have reached a form of proficiency in their specialized domain and are high level students. They are third and fourth year undergraduate students, whereas the aforementioned studies were focused on both undergraduate and graduate international students, as well as graduate L2 students. Even though practitioners use slightly more nouns than adjectives as premodifiers (nouns being higher on the developmental index), the difference is too thin to draw conclusions from it. Furthermore, prepositional phrases are by far the most common type of postmodifier in student lab reports (the student lab reports being the register with the largest proportion), which corresponds to the top of the developmental index created by Biber, Gray & Poonpon (2011).

Although earlier studies have shown that students are neither clear nor efficient writers (Conrad 2017, 2018), these resemblances between student and practitioner writing may suggest that students received good and accurate guidance on how to write civil engineering reports and lab reports when it comes to noun phrase structure and length in general. But these similarities are only on the quantitative level; some preliminary qualitative analyses of specific noun phrase examples taken from the NPcoded corpus seem to indicate that the difference between student and practitioner writing lies in the quality of noun phrases. As shown in Table 6, I observed that in student papers, there are errors in noun phrase structure and use that are not present in practitioner papers.

| Table - Examples of types of errors in noun phrase structure and use in student papers |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------|
| Relative clauses too far from the head noun they modify for easy reading. | There were a total of ninety-one crashes in a four-year period (2003 to 2006) at this intersection, which were analyzed. |
| Long noun phrases, after which students seem to lose track of the sentence structure. | OrganizationName4 is also a campus that strives to become as sustainable and “green” as possible; a mission which was recently recognized by the Association for the Advancement of Sustainability in Higher Education (AASHE) in the designation of “Gold Ranking” to the campus sustainability efforts as well as scoring as the third most sustainable campus in [State]. |
One possible explanation for these types of errors could be that students do not master detailed noun phrase modification as well as practitioners because they are not as comfortable with writing as practitioners (Conrad 2017, 2018). It could also be that they are less experienced writers. Even though students master the overall structure of noun phrase modification, they do not quite master the function of highly precise noun phrases yet, hence the presence of these errors in noun phrase construction.

Does the register (i.e. reports v. laboratory reports) play a role in the way noun phrases are used in the students’ papers?

As suggested by Lu (2011), the register under study seems to play a key role, here in the case of general pre- and postmodification structure of noun phrases (i.e. the frequency of premodifier only, postmodifier only, no modifier and both; see Figure 1 and Table 4). Overall, student lab reports have less pre- or postmodification than student reports. For student reports, the most common type of modification is premodification only, while for student lab reports, no modification is the most common case. These differences suggest that depending on the type of text the student is writing, the structure of the noun phrase is different. This situation may be due to the difference in the purpose of the writing task. When writing lab reports, students know that they are writing for their professor in the context of a graded assignment. On the contrary, when writing a report, students are writing with a professional purpose, with hypothetical or even real clients as their audience.

For example, in the lab reports, the single nouns are mostly pronouns, especially “it” (in impersonal structures or as a pronoun replacing a noun phrase), “we” and “I”. Both “we” and “I” are usually used to remind what the students have done before writing the report for the class or what was expected of them (e.g. “In this experiment, we were supposed to explore gradually varied flow”; “[The graphs that I generated […]”). Some single nouns are also software names, such as “excel”, which the students mention in order to explain what they have done in order to get their results (e.g. “The purpose of this experiment was to create an open channel flow profile using excel.”), while they are rarely used in reports. Another possible explanation for the presence of more single nouns in lab reports might be that in lab reports much information is already shared between the writer and the reader. Therefore, instead of being specific about a material or a place by using modifiers, students only use a single noun. Here are some examples of single nouns: “the tests”, “the lab”, “the results”, the material”, “the mean”, “The solutions should be reported in…” etc.

The observation that student lab reports have less pre- or postmodification than student reports is supported by the fact that regarding the general pre- and postmodification structure of noun phrases, student reports are very similar to practitioner reports, while the student lab reports structure is close to the one observed by Biber et al. (1999) when studying academic papers.
I also observed that even though the quantitative analysis did not show significant differences in detailed noun phrase structure between student lab reports and student reports, some preliminary analyses of a few specific examples seem to suggest discrepancies in terms of functions of noun phrases. Consider the following example taken from a student lab report:

Understanding of this concept has allowed for the calculation of important information such as yield strength, which is the maximum amount of stress in the elastic region for the material, and rupture strength, which is the stress experienced by the material at the point of fracture.

In this sentence, some noun phrases are used because they enable the student writer to show that they know what they are acquiring a form of expertise. The two relative clauses (underlined) are used to explain what the referent of the relativizer is: they want to show that they master the concepts they are dealing with, so they define the terms. On the contrary, this use of relative clauses was not observed in the few student reports which were analyzed in detail, because the stakes of writing with a professional purpose are obviously different. Further analysis would be required to investigate this point. To this end, conducting interviews with students could prove useful, asking them why they use noun phrases a certain way depending on the register.

5. Conclusion

This exploratory study has investigated noun phrase structure, frequency and use in both student and practitioner texts, in two different registers (reports and laboratory reports). The purpose was to provide a first quantitative analysis of noun phrase complexity in civil engineering writing, and to determine the differences in noun phrase complexity between the two types of writers (students and practitioners) and the two student registers (reports and laboratory reports), if there were any.

Contrary to what previous studies on clausal complexity have demonstrated, our findings show that the only notable difference between students and practitioners is the overall frequency of first order noun phrases, with students using more first order noun phrases than practitioners, and the difference is very slight. Surprisingly, in terms of noun phrase structure, students and practitioners resort to the same patterns. Despite being clausally more complex – as shown in previous studies –, student writing is apparently not phrasally less complex than practitioner writing. Instead, the difference between students and practitioners seems to lie in the quality of noun phrases (i.e. in specific examples and detailed noun phrase structure), and more investigation on this matter would be required. As predicted by Lu (2011), the register does play a role in noun phrase complexity, with laboratory reports being different from reports in terms of noun phrase modification structure; student lab reports have less pre- or postmodification than student reports, and for student reports, the most common type of modification is premodification only, while for student lab reports, no modification is the most common case. These differences between the two registers appear to be even clearer when specific examples are used, which confirms that further investigation would be needed.

Overall, the findings of the present study are intriguing, because they do not follow the same path as in earlier general academic writing studies. This could be due to the nature of the corpus, which is made of specific purpose texts. Another explanation
could be that there is no significant difference between the noun phrases in student and practitioner writing because the students are at a language level where they master the structure but not the function, which means that they see and they can mimic complex noun phrases, but they may not master the effective functions. Further studies would be necessary in order to understand the situation better, possibly in other specific fields. Furthermore, a larger-scale and deeper analysis would be required in order to confirm our findings given the small size of our corpus.

I would like to express my deepest gratitude and special thanks to Pr. Susan Conrad from Portland State University (USA) for her patient guidance during the course of this research and for giving me access to the Civil Engineering Writing Project resources. I am very grateful to her for her useful comments on the previous versions of this article, which greatly improved its quality. Many thanks to her research assistants as well (Naila Bairamova, Santiago Gustin Lopez, Philippa Otto, Sima Sokolov), who coded most of the corpus I used for my research. I would also like to thank Dominique O’Donnell, who wrote the coding programs I used for this article.

BIBLIOGRAPHY


BIBER, Douglas, Bethany Gray & Kornwipa Poonpon. 2011. “Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?”. *TESOL Quarterly* 45/1, 5–35.


NOTES

1. Throughout the present article, register is used in the sense of Biber & Conrad (2009). It refers to a perspective which consists in the identification of the typical linguistic features in a category of texts, along with the analysis of the situation of use of the category (2009: 2).
3. SPSS is a software package developed by IBM. SPSS can be used to do statistical analyses, including frequencies, descriptive statistics, cross-tabulation, etc.

ABSTRACTS

Recent linguistics studies have shown that academic writing is characterized by complex noun phrases, but few studies have focused on specific-field writing. The few studies conducted on civil engineering texts have shown a gap between practitioner and student writing, with students having a higher clausal complexity than practitioners. However, none of these studies have focused on noun phrases. I investigated in what way noun phrases in student papers differed from noun phrases in practitioner texts – with respect to noun phrase frequency, structure and length – and if the register played a role in noun phrase use. The study relies upon a corpus gathering practitioner and student texts (reports and laboratory reports) and was analyzed using corpus linguistics tools and methods. Surprisingly, results show that noun phrase length and structure are similar in practitioner and student texts. The results also show that register plays a key role in the use of noun phrases in student registers.

Les études récentes en linguistique montrent que les écrits universitaires se caractérisent par des syntagmes nominaux complexes. Cependant, peu d’études concernent la langue de spécialité ; les rares études relatives aux textes en génie civil font état de différences entre la langue des ingénieurs et des étudiants, ces derniers utilisant des phrases plus complexes que les ingénieurs. Toutefois, aucune de ces études ne s’intéresse aux syntagmes nominaux. Dans cet article, nous cherchons donc à déterminer en quoi les syntagmes nominaux dans les textes rédigés par les étudiants diffèrent de ceux écrits par les ingénieurs (en termes de fréquence, de structure et de longueur), et si le registre joue un rôle dans l’utilisation des syntagmes nominaux. Notre article s’appuie sur un corpus de textes écrits par des ingénieurs et des étudiants (comptes rendus et
comptes rendus de laboratoire) qui a été analysé selon les méthodes et outils de la linguistique de corpus. Contre toute attente, les résultats montrent que la longueur et la structure des syntagmes nominaux sont similaires dans les textes des ingénieurs et des étudiants. De plus, le registre joue un rôle déterminant dans l'utilisation des syntagmes nominaux parmi les registres étudiants.

INDEX

Keywords: academic writing, civil engineering, corpus linguistics, noun phrase, technical writing
Mots-clés: anglais universitaire, génie civil, linguistique de corpus, rédaction technique, syntagme nominal

AUTHOR

MARGAUX GUILLERIT

Margaux Guillerit, a graduate from École Normale Supérieure Paris-Saclay, is currently a first-year Ph.D. student in English linguistics and English for Specific Purposes at the University of Paris (former Paris-Diderot University). Her Ph.D. focuses on hierarchical relationships between lawyers and justices during the oral arguments of the Supreme Court of the United States. The present article was written during a year-long research internship in the department of Applied Linguistics at Portland State University, Oregon, USA. mguiller@eila.univ-paris-diderot.fr