

NOMINALIZATION AS HEAD NOUN OR MODIFIER: A CORPUS-BASED STUDY
OF DISCUSSION SECTIONS IN FORESTRY RESEARCH ARTICLES

***Nominalização como núcleo ou modificador do sintagma nominal: um estudo
baseado em corpus de seções de discussão em artigos científicos de Engenharia Florestal***

DOI: 10.14393/LL63-v41S-2026-14

Shirlene Bemfica de Oliveira*

Deise Prina Dutra**

Jasper Vilan Braga***

* Doutorado em Estudos Linguísticos pela Universidade Federal de Minas Gerais. Filiação: Instituto Federal de Minas Gerais – Campus Ouro Preto. ORCID: 0000-0003-1435-1728. E-mail para contato: shirlene.o(AT)ifmg.edu.br.

** Doutorado em Linguística pela Universidade da Flórida. Filiação: Programa de Pós-Graduação em Estudos Linguísticos. Faculdade de Letras da Universidade Federal de Minas Gerais. ORCID: 0000-0001-5799-5174. E-mail para contato: Dpdutra(AT)ufmg.br.

*** Licenciatura em Língua Portuguesa. Faculdade de Letras da Universidade Federal de Minas Gerais. ORCID: 0009-0002-3391-037X. E-mail para contato: kukehkenkun(AT)gmail.com.

RESUMO: A escrita de artigos científicos envolve tomar decisões não apenas quanto ao conteúdo da pesquisa a ser apresentado, mas também às escolhas linguísticas em seções específicas do artigo (Prina Dutra; Berber Sardinha, 2021). Análises aprofundadas de seções são escassas (Biber; Gray, 2013), especialmente no que diz respeito ao uso de nominalizações em seções de discussão (Moreno, 2021). Este estudo visa investigar a estrutura de sintagmas nominais nominalizados nas seções de discussão de artigos de pesquisa em Engenharia Florestal. O objetivo é identificar os sufixos nominalizadores mais comuns e discutir como esses sintagmas nominalizados se comportam em seções de discussão em termos de estrutura e função discursiva. Esta pesquisa de cunho quantitativo e qualitativo compara a seção de discussão com o corpus de estudo como um todo, composto por noventa artigos de alto impacto da área de Engenharia Florestal, totalizando 511.232 palavras. As nominalizações foram extraídas com a ferramenta *CQL* do *Sketch Engine* e analisadas com base na proposta de categorização das nominalizações de Comrie & Thompson (2007). Os resultados revelaram que essas seções são predominantemente escritas com sintagmas nominais nominalizados incorporados, aumentando a impessoalidade do texto (Comrie; Thompson, 2007; Miranda, 2022; Nordquist, 2024). A alta frequência desse padrão indica densidade de informação, abstração e complexidade gramatical (Nguyen; Edwards, 2015; Jalilifar et al., 2017; Biber; Gray, 2016), sugerindo qualidade textual e contribuindo para a aceitação da publicação. Os resultados dessas pesquisas podem informar pesquisadores da área e contribuir para a divulgação internacional da produção científica brasileira em língua inglesa.

PALAVRAS-CHAVE: Língua Inglesa. Escrita acadêmica. Artigos científicos. Nominalizações. Corpus de Ciências Agrárias. Engenharia Florestal.

ABSTRACT: Writing scientific articles involves not only deciding about the content of the research to be presented but also about the linguistic choices in specific sections of the article (Prina Dutra; Berber Sardinha, 2021). In-depth analyses of sections are scarce (Biber; Gray, 2013), especially regarding the use of nominalizations in discussion sections (Moreno, 2021). This study focuses on the structure of nominalized noun phrases in the discussion sections of Forestry research articles. The objective is to identify the most common nominalizing suffixes and discuss how these nominalized phrases behave in discussion sections in terms of structure and discursive function. This quantitative and qualitative research explores nominalizations in discussion sections and compares them with the study corpus, consisting of ninety high-impact Forestry articles with 511,232 words. Nominalizations were extracted using the *CQL* tool of *Sketch Engine* and analyzed based on the proposed categorization of nominalizations from Comrie & Thompson (2007). The results revealed that these sections are predominantly written with embedded nominalized noun phrases, increasing the impersonality of the text (Comrie; Thompson, 2007; Miranda, 2022; Nordquist, 2024). The high frequency of this pattern indicates information density, abstraction, and grammatical complexity (Nguyen; Edwards, 2015; Jalilifar et al., 2017; Biber; Gray, 2016), suggesting textual quality and contributing to the acceptance of the publication. The results of this research can inform researchers in the field and contribute to the international dissemination of Brazilian scientific production in English language.

KEYWORDS: English language. Academic writing. Research articles. Nominalization. Forestry corpus. Forestry.

1 Introduction

“Nominalization allows a process, more obviously realized as a verb, to be realized as a noun and hence to become a participant in a further process” (Bloor and Bloor, 2004, p. 2024)

In English grammar, nominalization is one of the processes of word formation in which a verb, an adjective, or another part of speech is used as / or transformed into a noun. It can be a head noun or a modifier of other head nouns (Comrie; Thompson, 2007; Nordquist, 2024). This process is often pervasive in academic writing, as it can express objectiveness, impersonal tone, and make the text more concise because it packs a great deal of information in a few words (Bruce, 2023; Nordquist, 2024). Technical terms and concepts, quite often nominalizations, are frequent in academic writing, and activities, such as experimenting, measuring, and analyzing, which are abstract conceptual units, are pervasive and help to make the text more impersonal.

Corpus Linguistics (CL) is an empirical approach that uses computer tools to analyze large, authentic collections of written and spoken texts. By identifying statistical patterns and usage, it enhances linguistic description, discourse analysis and language teaching based on real, observed data rather than theoretical statements (Bocorny et al., 2025). CL studies show that nominalization use may be linked to the specific characteristics of research articles (RAs), in which one must inform a lot with a few words (Gebrin, 2003; Biber; Gray, 2016). Moreover, very packed, informative, and concise information in noun phrases can characterize high lexical density.

Nominalizations are often part of such noun phrases, which have the property of encapsulation or condensation, through a realignment of the elements of the message, being extremely important in RAs (Nguyen; Edwards, 2015; Jalilifar, 2017). Since nominalizations in RAs seem to favor discourse impersonalization, information packing leads to the omission of participants, such as the agent, time relations, and the writer's positioning or the indication of modality (Gebrin, 2003; Dutra et. al, 2020). This way, there is a “dehumanization of the process that becomes objectified (reification), that is, it becomes an inanimate and abstract thing, losing its dynamic character. Therefore, nominalizations can also mystify power relations” (Gebrin, 2003, p. 1). Nominalizations are also used to adapt

discourse to the purpose of communication within the discipline and to meet the expectations of scientific discourse in general (Jalilifar, 2017).

Despite all the studies mentioned about nominalizations, there is a paucity of nominalization research in RA, especially in discussion sections, except for Peacock (2002), Parkinson (2011) and Moreno (2021). In addition to that, the nominalized noun phrases preceded by nouns and/or adjectives, the interest of this research, have not been previously explored in Forestry RAs.

Given that discussion sections demand a lot of writers because they need to transform raw data into scientific knowledge, requiring a high level of synthesis and critical analysis (Peacock, 2002; Parkinson, 2011), the major goal of this study is to identify and discuss the nominalized phrases in the discussion section of Forestry articles. This quantitative and qualitative research explores nominalizations in discussion sections and compares them with all the other sections of our Forestry study corpus. This corpus of high-impact Forestry research articles is part of a larger corpus named Corpus of Scientific Research Articles in Agrarian Sciences (CorAgrarian), which was manually compiled and organized in sections (abstract, introduction, methodology, results, discussion and conclusion) and it is fully described in the Methods section (Teixeira et al., 2022; Marques, 2024).

This study, therefore, has two specific objectives. First, identify the most common nominalization suffixes used by the Forestry scholars in their RAs and compare the normalized frequency counts of the whole corpus (except discussion section) and of the discussion sections. Second, investigate how the most frequent nominalized noun phrases in the Forestry corpus may differ in structure and discourse functions as compared to what is used in its discussion sections. For this qualitative part of the study, Comrie and Thompson's categorization of nominalization based on the nominalized noun phrase structure and discursive function was used. The results of this descriptive study will be used to inform the preparation of materials to help improve academic writing (e.g. Almeida et al. 2023) in Forestry higher education. Raising the awareness of nominalization patterns could help novice writers (e.g. graduate students) in the Forestry area to improve their texts by making them more impersonal, concise, and objective.

2 Nominalization in academic writing: shifts in RAs

Understanding RAs features has become more challenging due to evolving stylistic trends that have altered academic language grammatically (Biber; Gray, 2011). These grammatical shifts in different registers and genres across various fields are also a result of the volume of scientific output. Biber and Gray (2016) highlight that grammatical complexity in RAs varies across disciplines. Humanities RAs, for example, tend towards descriptive, qualitative writing with more structural elaboration, leading to longer texts that explore metaphors and abstract concepts. In contrast, hard sciences RAs prioritize structural specification and information compression, resulting in more precise, objective texts focused on experimental results and quantitative data. To enhance clarity and conciseness while minimizing figurative language, hard science RAs often exhibit high lexical density and grammatical complexity, which has been singled out as complexity at the phrasal level by Biber and Gray (2016).

When studying nominalizations, one can observe their structure and their functions, condensing points and allowing researchers to express ideas concisely with high information density (Biber; Gray, 2011; Jalilifar et al., 2017; Dutra et al., 2020). Comrie and Thompson (2007) classified nominalization into semantic categories, which may be productive categories in the corpus of this study. Such categories are shown below with examples from our Forestry corpus:

- state nominalization (*soil **fertility**¹, soil **salinity**, plant **susceptibility***);
- action nominalization (*soil **compaction**, seed **germination**, water **infiltration***);
- agentive nominalization (*the thick litter **layer**, plant **sucker**, fewer leaf **chewer***);
- instrumental nominalization (*size **analyser**, shear **tester**, soil **chamber**, NaOAc **buffer***);
- locative nominalization (*sediment **storage**, carbon **storage**, biomass **storage***);
- manner nominalization (*soil **warming**, tree **seedling**, timber **harvesting**, wood **processing***);
- productive nominalization (***ability**, **availability**, **flowability**, **probability**, **creation***);
- reason nominalization (***decision**, **diversity**, **argument**, **interference***).

The study of the nominalization process is a recent development which is important to understand grammatical complexity in English. Biber and Gray (2011) discuss the historical evolution of nominalization, demonstrating its impact on academic writing,

¹ Examples in this article are in italics and the word(s) on focus is(are) in bold.

particularly in RAs. They note a shift towards a more abstract and especially compressed linguistic style. Their research into noun phrase extension in English, which analyzed grammatical changes, lexical associations, and semantic functions, concluded that these evolving patterns emerged from the communicative demands of written discourse (Biber; Gray, 2011, p. 225). The development in science and the need for conciseness in communication have pressured academic writing to adopt a grammatical style reliant on phrasal modification, often minimizing verbs and clausal modification in favor of nouns and dense phrasal structures (Biber; Gray, 2011, p. 226). Consequently, modern academic writing is characterized by compressed information and a high density of phrasal modifiers, including nominalizations, attributive adjectives, appositive noun phrases, nouns as nominal premodifiers, and prepositional phrases as nominal postmodifiers (Nguyen; Edwards, 2015; Jalilifar, 2017). Biber and Gray (2011, p. 229) present “five grammatical devices that have been important in the development of this compressed discourse style:

- nominalizations (**consumption, comparison, sustenance**)
- attributive adjectives (*gradually **expanding cumulative** effect*);
- nouns as nominal premodifiers (**baggage inspection** procedures);
- prepositional phrases as nominal postmodifiers (*a high incidence of **heavy alcohol consumption amongst patients***);
- appositive noun phrases (*Dallas Salisbury, **CEO of the Employee Benefit Research Institute***)”

Recent literature highlights nominalization as the most studied aspect from these five that brought changes in academic writing. Scholars like Halliday and Martin (1993/1996) have documented the rising prevalence of nominalization in research articles, noting shift, such as the modification of nominalizations by attributive adjectives (**natural regeneration, dominant trees**), verb-to-noun conversions (**canopy clearance**), and the increased use of derivational suffixes, alongside the modern trend of nouns modifying nominalizations (**soil absorption, matrix infiltration**), a rare pattern in the past. Biber and Gray (2011; 2016) credit this evolution to factors including time constraints for writing and review, the breadth of specialized information, scientific and technological advancements, and the peer review process.

Furthermore, Biber and Gray (2016, p. 176) explain that noun-noun nominalization sequences, functioning as both head nouns and premodifiers, emerged in the mid-19th

century. While most nominalizations refer to processes or activities derived from verbs (via morphological derivation or conversion), their expansion through the 20th century also included nominalized and converted forms from verbs used as noun pre-modifiers. The authors also note the use of nominalizations derived from nouns or adjectives, acting as noun pre-modifiers to convey abstract attributes or qualities, rather than just processes. These compressed phrasal sequences are evident in the example sentences below.

- (1) Suppressed ***fine root production*** and biomass have been reported in Norway spruce forest (Majdi, 2004), maple tree seedlings (Wan et al., 2004), and Harvard Forest under soil warming (Zhou et al., 2011).
(CorAgrarian.For.woodtec.22_discussion)
- (2) Therefore, ***higher light availability*** after fire can promote growth and establishment of *E. globulus* trees. In forested habitats, seedling survival was higher in burnt sites, which contrasted with the opposite pattern observed in shrublands. The consumption or scorching of canopies by fire in forested habitats increases light availability, whereas in shrublands seedlings grew in full light both in burnt and unburnt sites.
(CorAgrarian.For.silviculture.26_discussion)

The 20th century also saw an increased use of intangible nouns (abstract nouns) and nominalizations as noun pre-modifiers, leading to multi-word noun pre-modifier sequences (*annual **growth efficiency***). These sequences often convey complex and abstract relationships, unlike earlier, more straightforward modifications (Biber; Gray, 2016). Biber and Gray (2016,) note that nominalized sequences with multiple pre-modifiers became common in noun phrases by the turn of the 20th century. By the latter half of the century, sequences with two or more pre-modifiers, previously rare in academic texts, became more frequent, expanding the range of possible semantic relationships due to the greater variety of pre-modifying nouns, as shown in the examples below:

- (3) To decide what type of restoration and how to restore is a difficult task (Emborg et al., 2012) because making decisions at a landscape scale is largely a ***“top-down strategic management”*** approach where governments set objectives [...]
(CorAgrarian.For.formanag.03_discussion)
- (4) Fertilization and application of other agrochemicals is standard procedure in timber plantations and industrial crops. Therefore, land management around native forest is crucial to maintaining ***woody native vegetation*** and the biodiversity associated with the forest reserves. If we consider the ongoing dramatic land use changes in South America, improving biodiversity conservation in its biomes is a priority (Boit et al., 2016).
(CorAgrarian.Agr.ecology.03_discussion)
- (5) The ***cold steppe vegetation cover*** is sparse, characterized by alpine grasslands mainly comprised of hemicryptophytes (Dvorsky et al., 2015, 2011).
(CorAgrarian.Agr.soil.08_methods)

Biber and Gray (2016) further emphasize that the grammatical use of nominalizations as main nouns introduced new semantic relationships. In these instances, pre-modifier nouns often function as the semantic patient or theme of the process described by the main noun. They note that not all pre-modifying nouns have a singular, clear relationship with the head noun, and even tangible nouns can exhibit varied relationships. This evolution encompassed nominalizations derived from both intransitive and transitive verbs, aligning pre-modifier nouns with their logical subjects or direct objects (Biber; Gray, 2016, p. 179). According to the authors, in the late 19th and 20th centuries, the range of noun relationships broadened.

(...) the individual nouns employed in noun–noun sequences are often not technical, but the meaning relationships underlying those sequences are technical and require specialist background knowledge for understanding. Thus, the historical shift towards an increasing reliance on such forms has resulted in a surprising loss of explicitness in specifying the intended meaning (Biber; Gray, 2016, p. 184).

Ultimately, the phenomenon of multiple nouns modifying each other has become varied and this diversification, primarily observed from the late 19th through the 20th centuries, is directly linked to the expanded range of uses and meanings that noun sequences now convey, besides contributing to the phrasal level packaging of information in discourse (Biber and Gray, 2016). To examine how these structural changes and information packaging tendency influenced within specialized corpora, this study evaluates these patterns empirically. The specific operational frameworks, sampling criteria and linguistic parameters used to analyze these structures are detailed in the Methods section.

3 Methods

This section presents, first, the process of compilation of the corpus used for this research, the Corpus of Agrarian Sciences Scientific Articles (CorAgrarian), from which our study corpus comes. Second, the study corpus of this research (the Forestry Corpus) and the subcorpus (the Discussion Section Corpus) are detailed. Third, the preparation of the corpus and subcorpus, and analysis tools are described, as well as the qualitative and quantitative analysis rationale are explained.

3.1 The Agrarian Sciences corpus

For this research investigation, the Forestry area of the Corpus of Agrarian Sciences Scientific Articles (CorAgrarian) was used (Teixeira et al., 2022; Marques, 2024). CorAgrarian is a specialized corpus of scientific articles in the field of Agrarian Sciences published in English in high-impact journals (Qualis A), compiled manually within the scope of a three-year (2023-2026) project titled “Academic writing in English in the Agrarian Sciences: needs and input for pedagogical applications”, whose members have studied the academic language of the Agrarian Sciences. Since the articles collected were published in peer-reviewed journals, the language use is considered adequate. Consequently, we did not filter if the authors were native or non-native speakers of English. To be representative of this area, the corpus is composed of five subareas, defined based on the courses offered by the Agrarian Sciences Institute of the Federal University of Minas Gerais, namely: Agronomy, Agricultural Engineering, Food Engineering, Forestry and Animal Science. In addition, all compiled articles had the following sections: abstract, introduction, methodology, results, discussion and conclusion and were, therefore, subdivided and saved, manually into sections. Technically, the articles from the subareas of CorAgrarian (e.g. Agronomy, Forestry) are its subcorpora. For the purpose of this paper, the subcorpus of Forestry articles is called the Forestry corpus (F). We have selected to closely study the nominalizations that occur in the discussion sections (DS) of the Forestry RAs. Therefore, DS will be referred to as a subcorpus of the Forestry corpus.

Considering the intention to select representative texts of contemporary academic writing, a temporal criterion was adopted: the articles of CorAgrarian were published between 2014 and 2023. Finally, to enable more in-depth linguistic analyses, the corpus was tagged in Part of Speech (POS) using the Biber Tagger (Biber, 1988), which tags the corpus in more than 130 lexico-grammatical features. The CorAgrarian configuration is presented in Table 1:

Table 1: CorAgrarian by areas

Areas	Nº of articles	Nº of words	Average word per article
Agronomy	87	526639	6053,32
Agriculture Engineering	90	523245	5813,83
Food Engineering	90	459535	5105,94
Forestry	90	505454	5616,16
Animal Science	90	487159	5412,88
Total	447	2,502,032	5,597.39

Fonte: Marques (2024)

As shown in Table 1, CorAgrarian consists of 447 texts totaling approximately 2.5 million words (Teixeira et al., 2022; Marques, 2024). Regarding the corpus balancing, it was defined that each subcorpus (each subarea of Agrarian Sciences) would comprise 90 texts. In the case of the Agronomy, it was necessary to exclude 3 articles due to compilation problems, such as the presence of duplicate texts.

3.2 The study corpora: the Forestry corpus (F) and its Discussion Section sub corpus (DS)

Table 2 presents the Forestry area corpus (F), showing the numbers of words in each of the RA sections. The sub corpus of interest for this study is the discussion section (DS), with 137,006 words, which will be compared to all the other sections of the Forestry corpus, except the discussion section itself. The remaining Forestry corpus, with abstract, introduction, methods, and conclusion sections corresponds to 374,226 words.

Table 2: Number of words in the Forestry Corpus per section and total

Forestry Corpus	Abstracts	Introductions	Methods	Results	Discussions	Conclusions	TOTAL	# RAs
Words	24,708	88,237	142,335	98,362	137,006	20,584	511,232	90

Source: Marques (Thesis in progress)

The Forestry corpus includes RAs from the areas of Forest Management, Silviculture, and Wood Technology (see Table 3), covering a wide range of Forestry-related topics: Ecological Engineering, Forest Ecology and Management, Geomorphology, Agricultural and Forest Meteorology, Biological Conservation, Forest Ecology and Management, International Biodeterioration & Biodegradation, Science of the Total Environment, and Biomass &

Bioenergy. The journals where the RAs were collected from cover the mentioned topics. The number of articles from each journal can be seen in Table 3.

Table 3: Forestry corpus journals

Areas	Journal	# of articles
Forest Management	<i>Ecological Engineering</i>	10
	<i>Forest Ecology and Management</i>	11
	<i>Geomorphology</i>	9
Silviculture	<i>Agricultural and Forest Meteorology</i>	10
	<i>Biological Conservation</i>	7
	<i>Forest Ecology and Management</i>	13
Wood Technology	<i>International Biodeterioration & Biodegradation</i>	10
	<i>Science of the Total Environment</i>	12
	<i>Biomass & Bioenergy</i>	8

Source: Marques (Thesis in progress)

3.3 The preparation of the corpus and analysis tools

First, the Forestry corpus (F) and its sub corpus (DS) were cleaned with the help of Corpus Script^{2,3}. This software can convert all characters to lowercase, remove line breaks, page numbers, special characters, diacritics, Greek and Cyrillic characters, allowing the processing of the corpus analysis to flow smoothly. Second, (F) and (DS) were uploaded into the Sketch Engine, a tool used to explore language through algorithms for text analysis or text mining applications (Kilgarrieff et al., 2014).

Second, all the nominalizations were retrieved by using the query function (CQL). The Corpus Query Language (CQL) feature in Sketch Engine serves as a syntax-driven search mechanism that enables the extraction of complex grammatical and lexical patterns from textual data. By utilizing token-based expressions enclosed in square brackets, CQL allows for the interrogation of corpora beyond simple keyword queries, combining word forms, lemmas, and Part-of-Speech (POS) tags (Kilgarrieff et al., 2014). This functionality facilitates

² CorpuScript (Alves, 2024), a software tool designed for preprocessing files in corpora compilation. The application stands out for its ability to apply both personalized and traditional cleaning parameters across an entire corpus, regardless of its size.

³ Organizer's Note: The software is available as CorpusWright and can be downloaded at <https://github.com/jhlopesalves/CorpusWright>.

the identification of specific nominalizations, multi-word sequences, variable syntactic structures, and distant collocational frameworks (See Appendix 2 for query syntax used).

The choice for 16 (sixteen) nominalization suffixes (*-tion, -ity, -ment, -er, -ing, -is, -ance, -age, -al, -sion, -ness, -ence, -ant, -ism, -dom, and -ury*) (Table 5) come from Biber et al. (1999) and Jalilifar et al. (2017), who also investigated the occurrence of these suffixes in academic writing. A total of 32 tables were generated by the Sketch Engine search (16 tables from Forestry corpus and 16 tables from discussion section subcorpus). Table 4 below illustrates how part of the results was shown. In this sample, the lemmas are the base form of the identified nominalized phrases with its corresponding absolute raw count on frequency column, for example “forest management” appears exactly 90 times in this specific dataset. The *Freq. per million* column corresponds to the normalized frequency. In this example “forest management” appears roughly 572 times. This allows us to compare these results with other corpora of different sizes. The *% of concordance* column presents the percentage of the total search results that this specific phrase represents. Out of all the noun + “-ment” noun combinations found by the query, “forest management” accounts for 30.5% of them.

Table 4: Noun nominalization sequences and their frequency

Query: [pos= “n”] [pos= [“n” & word= “*.ment”];GEDEX300

Lemma	Frequency	Freq. per million	% of concordance
<i>forest management</i>	90	572,63929	30,50847
<i>seedling establishment</i>	18	114,52786	6,10169
<i>seedling recruitment</i>	10	63,62659	3,38983
<i>wood recruitment</i>	8	50,90127	2,71186
<i>seedling development</i>	7	44,53861	2,37288
<i>drought treatment</i>	5	31,81329	1,69492
<i>WSD treatment</i>	5	31,81329	1,69492
<i>SM treatment</i>	4	25,45064	1,355593
<i>addition treatment</i>	4	25,45064	1,355593
<i>CWO treatment</i>	4	25,45064	1,355593
<i>species recruitment</i>	4	25,45064	1,355593

Source: Data from the present study

All the results from the 32 tables were manually checked, and the 16 (sixteen) suffixes were identified in both corpora and coded as pre-modified by nouns (NN) or by adjectives (AN). The raw and the normalized frequency were accessed, as shown in Table 4.

The raw frequency indicates the occurrence of a word or word sequence (e.g. *forest management* – noun + nominalization) in the corpus. The use of normalized frequency is recommended when the purpose is to “compare frequencies in two corpora or subcorpora of different sizes” (Lindquist, 2009, p. 43). To normalize frequency counts, “the raw frequency count should be divided by the total number of words in the text and then multiplied by the chosen norming basis to normalize corpus data”⁴ (Biber et al. 1998, p. 263). Biber et al. (1998) refer to “the number of words in the text” since their unit of observation is the text. In this study, the units of observation are the corpus and its subcorpora. The normalization used, in this study, is per 100,000 words. The number of words in the Forestry corpus, without the discussion section, is 374,226, and the discussion section sub corpus has 137,006 words, as presented in section 3.

Finally, the nominalization bigrams were extracted with Sketch Engine by using the query function (CQL) and data was interpreted based on frequency and productivity of suffixes within nominalized noun phrases from both corpora. In this investigation, the bigrams are sequences of two adjacent words that formed the nominalized phrases and that contain meaning when used together. The analysis of bigrams allowed us to grasp the syntactic and semantic context in which nominalized phrases connect to other nouns, adjectives, verbs, or prepositions, revealing the immediate context, argument structure and meaning that a single isolated term would not provide (Baratta, 2010; Jalilifar et al. 2017). On the other hand, this could be a limitation of this investigation because the frequency calculations and the analysis considered only 2-word grams containing nominalizations, rather than long, embedded and expanded nominalized noun phrases. For instance, in long phrases such as, “an accurate chromosome number determination”, “high-value crop production”, “the generalized likelihood uncertainty estimation”, the Sketch Engine searched only the bigrams “number determination” “crop production” and “uncertainty estimation”. This is a limitation because by extracting only the 2-word grams, the study loses the broader meaning provided by the other embedded words. Working with longer word sequences that carry nominalizations would

⁴ Biber (1988) suggests the normalization per 1,000 words due to the size of the texts in his study. In our case, our investigation deals with comparing a corpus and its sub corpus.

For the qualitative analysis, the researchers followed the model of Comrie and Thompson (2007), which involves presenting the roles that nominalized noun phrases play in academic texts, indicating names of activities and concepts, states, and arguments. This categorization allows capturing both the syntactic and semantic context in which the nouns connected to other nouns, adjectives, verbs, or prepositions, as well as the argument structure with meaning that a single isolated term would not provide.

Section 4 presents the quantitative analysis of nominalizations and the qualitative discussion to illustrate the structure and discursive functions of the nominalizations identified in our study. Examples of the bigrams with nominalizations were used for illustrations and discussion.

4 Data analysis and discussion

To achieve the research objectives, this study combines a quantitative analysis with a qualitative discussion. The quantitative phase establishes the frequency distribution of the data, while the qualitative phase illustrates the specific structures and discursive functions of nominalizations across the (F) corpus without the (DS) subcorpus, and the (DS) subcorpus. To support these insights, examples of extracted nominalized bigrams from the (DS) subcorpus are presented and discussed.

4.1 Nominalized noun phrases frequency: A quantitative analysis

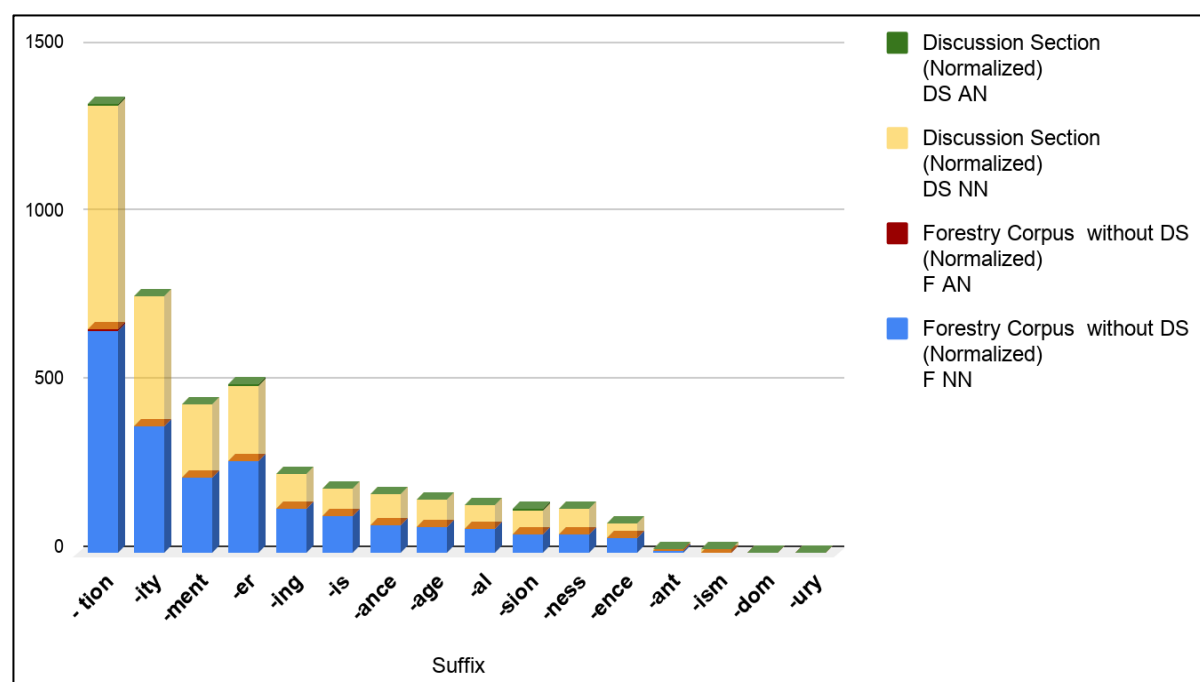
This study identified 16 suffixes via Sketch Engine Queries: *-tion, -ity, -ment, -er, -ing, -is, -ance, -age, -al, -sion, -ness, -ence, -ant, -ism, -dom, and -ury*. Derived nominalizations (from verbs or adjectives) were prevalent in the corpus and specifically in the discussion section (see Table 5). Table 5 shows the suffixes' raw and normalized frequencies in the Forestry corpus (all sections except discussion) compared to the (DS) subcorpus. Notably, NN sequences (noun + nominalization) occur much more frequently than nominalization preceded by adjectives in both (F) without discussion sections and (DS), as shown in Figure 1.

Table 5: Nominalization suffixes pre-modified by nouns (noun-noun, NN) or by adjectives (adjective-noun, AN)

Suffix	Forestry Corpus without DS (Normalized) F NN	Forestry Corpus without DS (Normalized) F AN	Discussion Section (Normalized) DS NN	Discussion Section (Normalized) DS AN
-tion	662.97	2.67	665.66	5.11
-ity	376.24	0.27	386.11	0.00
-ment	226.87	0.53	216.78	0.00
-er	275.23	1.07	224.08	2.92
-ing	133.07	0.00	103.65	0.73
-is	111.96	0.00	81.75	0.73
-ance	84.17	0.80	92.70	1.46
-age	78.83	0.00	79.56	0.00
-al	72.68	0.27	68.61	0.73
-sion	57.72	0.27	72.26	1.46
-ness	55.05	0.00	78.83	0.00
-ence	43.56	0.27	43.06	0.00
-ant	9.09	0.00	4.38	0.00
-ism	2.67	0.00	11.68	0.00
-dom	0.80	0.00	0.73	0.00
-ury	0.80	0.00	0.00	0.00

Source: Data from the present study

Figure 1 - Normalized frequencies of the 16 nominalization suffixes in noun phrases in (F) and (DS)



Source: Data from study

Figure 1 illustrates variations in nominalization across the sixteen suffixes in the Forestry corpus and in the discussion section subcorpus. Visually, in both corpora, nominalizations are much more frequently preceded by nouns than adjectives, showing that nominalization phrases with adjectives as pre-modifiers are not as frequent in RAs in F and DS.

Nominalization contributes to the construction of information packed texts, where lexical density is noticed at the phrase level, more specifically in noun phrases in our study, as well as in previous research, being a characteristic of academic texts (Prasithrathsint, 2014; Gebrin, 2003; Biber; Gray 2011, 2016; Jalilifar et al., 2017; Dutra et al., 2020; Bruce, 2023; Nordquist, 2024) (see Examples 6-9).

The generally higher prevalence of noun-preceded nominalizations in the corpus compared to its discussion section (See Table 5 and Figure 1) might reflect different rhetorical function preferences. According to Comrie and Thompson (2007), this prevalence of noun-preceded nominalizations presented in the results indicates a strong reliance on nominalized concepts to describe phenomena that are important in RAs (e.g. *surface-litter*

decomposition, carbon sequestration, carbon storage). Our (F) corpus may require a denser concentration of technical terms to describe experimental setups, observations, and findings with precision. In contrast, the discussion sections, while still academic, might employ a slightly more interpretative or discursive language, leading to a comparatively lower, though still substantial, reliance on heavily nominalized structures.

The data show an increased occurrence of nominalizations with the suffixes *-tion*, *-ance*, and *-sion* in the discussion sections and this could indicate a move from presenting factual, technical information to a more analytical and argumentative mode of discourse, for example: *chemical reaction* and *wax emulsion* in the following examples: “For future research, it is important to take into account the **chemical composition** of the growth medium, as undesired secondary compounds may be formed”; and “The observed mass loss of paraffin wax emulsion was similar to the samples treated with the UAD. As the paraffin wax emulsion was increased, the mass loss was reduced. This was caused by an increase in **paraffin wax emulsion** from 0.5% to 2.0%”.

In discussion sections, according to Biber and Gray (2013), authors generally interpret the results and formulate generalizations; hence, they focus on process/action, abstraction, and on the argumentative effort to transform experimental processes into theoretical concepts. The differences found between the nominalizations in the Forestry corpus (without DS subcorpus) and its discussion sections distinguish this section from the other RA sections. This phenomenon, where dynamic processes are presented as static concepts, facilitates the concise and abstract discussion of complex phenomena within a specialized field like Forestry. The lower frequencies observed for suffixes, such as *-ant*, *-ism*, *-dom*, and *-ury*, suggest a more peripheral role in the overall nominalization strategy of this discussion section corpus, or that their semantic contributions are less frequently required in the specific contexts of forestry research articles.

5.2 Discourse functions and structural patterns of nominalization: A qualitative approach

A qualitative analysis of nominalization processes involves understanding the roles they play in academic texts that can indicate names of activities, states or arguments (Comrie and Thompson, 2007). In this section, bigrams consisting of noun-noun and

adjective-noun nominalized sequences were retrieved from the (DS) subcorpus to comprehend such roles. The results show that nominalizations function as **head nouns** or **modifiers** in the bigrams. The following examples (6-9) show the nominalized noun phrase expansion, as head nouns or modifiers, with more than two modifiers that could be nouns, adjectives, articles, prepositions, etc.

- (6) results from (**experimental trials testing rotational grazing management**)
(CorAgrarian.AgE.agrmachine.17_discussion)
- (7) the smaller tests size, (**less seminiferous tubule volume density**)
(CorAgrarian.AgE.agrmachine.17_discussion)
- (8) Redominantly exhibited (**a nearly horizontal growth pattern**)
(CorAgrarian.AgE.agrmachine.17_discussion)
- (9) (**the short-term duration of N fertilization**)
(CorAgrarian.For.Woodtec.16.discussion)

To better understand the functions of noun phrases, Comrie and Thompson (2007, p. 338) categorized them into two classes: A and B, where A forms retain certain properties of the verbs or adjectives they are related to, while those in B typically behave syntactically like other nouns in the language, bearing only morphological and (often unpredictable and idiosyncratic) semantic relations to the associated verb or adjective. In this study, the nominalized noun phrases were analyzed based on this categorization and examples from the discussion sections of the (F) corpus were used to illustrate and explain these functions.

5.2.1 Category A: Name of activity or state

Nominalizations may be used to name an activity or a state (Comrie and Thompson, 2007). According to Comrie and Thompson (2007), the first category A is **action or state nouns** that can be formed from a verb phrase consisting of a verb and its object by reversing the order of the verb and the object or state nouns from stative verbs or adjectives, meaning the fact, the act, the quality, or occurrence of that verb or adjective. In the corpus, this strategy is very productive with the suffixes -tion, -ing, -ment, -ance, -ence, -ity, as shown examples (10) to (14).

- (10) By using this multi-scale **restoration** approach, we identified the optimal spatial **arrangement** of the targeted stands that achieve better landscape structural and functional **synergies**, while indicating the **restoration location** and **management** required. (CorAgrarian.For.Woodtec.16.discussion)

- (11) The main reasons for differences between the **recovery** periods from 5 to 25 years were due to some site-related factors such as soil type, texture, and organic matter content, **compaction** level, slope, climate **condition**, and soil biological **activity**. (CorAgrarian.For.formanag.02_discussion)
- (12) Non-forest **vegetation** such as meadows, pastures, and fields also belong to the warmest areas. (CorAgrarian.For.WoodTec.16.discussion)
- (13) The impact of bark beetle **calamity** on the hydrological balance of the Šumava national park **territory** is a widely discussed topic. (CorAgrarian.For.formanag.01_discussion)
- (14) Consistent with our findings, Rab (2004) and Ezzati et al. (2012) reported that soil bulk **density** values did not recover under natural **conditions** and remained 22–68% and 40% greater than the values of undisturbed areas over 10–20 years after logging. (CorAgrarian.For.formanag.02_discussion.txt)

In examples (10) to (14), the nouns “activity, arrangement, condition, recovery, compaction, restoration” come from action verbs “act, arrange, condition, recover, compact, restore”. They maintain properties of the original word and designate a process or a non-process (Comrie; Thompson, 2007). With the same function, the nouns “calamity, vegetation, density, synergy” come from the adjectives “calamitous, vegetative, vegetal, dense and synergistic” that can also be from the verb “synergize”. They are positioned as head nouns (e.g. “bark beetle calamity”) or as modifiers (e.g. “soil bulk density values”).

5.2.2. Category B: Name of an argument

The second category, **B**, or **name of an argument**, can convey agentive, instrumental, manner, locative, objective, reason, productivity, predictability, proposal or refusal ideas. According to Comrie and Thompson (2007, p. 338), “**nominalized agentive nouns** refer to the agent or the person or thing that performs the action described by the verb”. In the Forestry corpus, the agentive and instrumental nominalization plays similar roles with the suffix -er. Most of the occurrences are equipment or developing actions or processes. In the example (15) below, “leaf chewer”, refers to some species that chews leaves and “plant sucker” refers to an organism or a machine that sucks plants or substances made by plants.

- (15) Stand age, i.e. beech stand developmental stage, was of mixed importance for forest herbivore abundance and species richness, with **low leaf chewer species richness** in the canopy of timber-stage stands and an increase in **plant sucker abundance** with increasing stand age. (CorAgrarian.For.formanag.13_discussion)

The nominalized instrumental **nouns**, on the other hand, denote the instrument, tool, or means by which an action is performed. In example (16), the “ring shear tester” is a **device** used for measuring shear strength, whereas the phrase containing “soil chamber” in (17) refers to an apparatus used specifically for measuring soil gas exchange and respiration rates.

(16) In closed systems, such as **the ring shear tester**, the effect of electric charges could thus be mainly neglected. (CorAgrarian.For.woodtec.24_discussion)

(17) The low correlation between the understory ER and **the soil chamber measurement** could also partly be ascribed to a discrepancy in the source area. (CorAgrarian.For.silviculture.05_discussion)

“Arguments can name **nominalized manner nouns** which can refer to both the fact and occurrence of an action and the way it is performed” (Comrie; Thompson, 2007, p. 339).

(18) We found that both light transmittance and **low tree seedling responses** were transitory. (CorAgrarian.For.silviculture.19_discussion)

(19) Rice straw mulch is produced in rice agricultural lands downstream along the Caspian Sea, and sawdust mulch is produced from **wood processing** and wood industries, but from an economic viewpoint, cost estimations regarding these two types of mulch should also be considered at the application area. (CorAgrarian.For.formanag.08_discussion)

(20) Typically, **unmitigated timber harvesting** increases organic matter inputs to streams during and immediately following harvesting, then subsequently reduces terrestrial organic matter during forest regrowth. (CorAgrarian.For.formanag.05_discussion)

In the example (18), the nominalized cluster “low tree seedling responses” refers to the reactions or responses of tree seedlings, which can be interpreted in terms of both the occurrence of these responses and the way in which they, the seedlings, occur. The same type of argument, nominalized manner noun, occurs in “wood processing” (Example 19). “Timber harvesting” in example (20) can refer to the act of harvesting timber or the methods used to harvest timber. The *-ing* nominalization is prevalent in discussion sections with the function of naming an argument.

“Locative nominalizations indicate a place or location which is derived from a verb or adjective” (Comrie; Thompson, 2007, p. 339). In this investigation, in both corpora, the most frequent suffix with this function is *-age*, as in the word “storage” (see Example 21). The item “habitat” in example (22) is not formed by suffixes, but it is a nominalization that comes

from the verb “inhabit or habitare”, from Latin, which means to live or dwell in a place. Therefore, a habitat is the place where a particular species lives, and it is prevalent in the corpus, with 18 occurrences in the whole corpus.

(21) However, enhanced aromaticity of soil organic matter in subsoils (30–45 cm) has been observed decades after fire, highlighting the possibility for rapid translocation or, alternatively, enhanced subsoil pyrogenic matter **storage**. (CorAgrarian.For.woodtec.20_discussion)

(22) At the seedling and sapling stages, **the topographic habitat associations** of tree species are especially variable, for example, 19 of 80 species at Barro Colorado Island (BCI), Panama (Comita et al., 2007), 41 of 60 species at Gutianshan, East China (Lai et al., 2009) and 110 out of 136 of species at Yasuní, Equador (Metz, 2012) had **significant species-topographic habitat associations**. (CorAgrarian.For.silviculture.24_discussion)

In the Forestry corpus, some nominalizations have an affix that forms nouns designating the result, or the typical or ‘cognate’ object of an action: **objective nominalization** (Comrie; Thompson, 2007). Suffixes such as *-tion*, *-ment*, *-ance/-ence*, *-al* and *-ure* often form nouns that represent the result or object of an action.

(23) “In general, these results accord with previous studies indicating that geomorphic impacts of trees and forest **disturbance** are highly contingent on local environmental characteristics, forest **management**, age or stage of vegetation **development**, and timing and frequency of various **disturbance types**”. (CorAgrarian.For.formanag.25_discussion)

In example (23), the words “result”, “impact”, and “development” refer to the product of the described forestry actions. The word “disturbance” reflects the result of disrupting or altering the forest environment and the words “characteristics”, “management”, “development”, “timing”, and “frequency” reflect the result or state of what happens.

According to Comrie and Thompson (2007), the **reason nominalizations** refer to the process of converting verbs or adjectives into nouns that express the cause, reason, or justification for an action or event. Suffixes like *-ation*, *-tion*, *-ment*, *-ance/-ence*, *-ity*, and *-ism* can be used with these functions, as can be seen in examples (24) to (26).

(24) A further possible **explanation** for the high preferential retention of charcoal is a priming effect whereby charcoal enhances the degradation of non-charcoal SOC in the soil. (CorAgrarian.For.woodtec.20_discussion)

(25) The tree crown complementarity, i.e., the partitioning of aboveground space and light, is a universal underlying mechanism for linking positive **biodiversity**

and aboveground biomass in both experimental and natural systems, including forests. (CorAgrarian.For.woodtec.13_discussion)

- (26) In the last few years the intensive system of public support in the European Union has experienced increasing **criticism** for economic, ecological and social reasons and this may influence the corresponding policy. (CorAgrarian.For.woodtec.28_discussion)

In our study, especially in (DS) subcorpus, the suffix form *-tion* (Example 24) is used to indicate the reason or process of explanation which is generally surrounded by the clusters “common, a conclusive, an alternative, a possible, the most likely” and the preposition “for”. The suffix *-ity* (Example 25) forms nouns that denote state or quality, often implying a reason or, in the example, the state of having an ecosystem variety, highlighting the reason for conservation efforts. Moreover, the suffix *-ism* (Example 26) is used to form nouns that indicate a belief or reason.

Another category of argument nominalization implies **predictability or productivity**. Words such as “ability”, “availability”, “flowability”, “probability”, “creation” help to convey important concepts in Forestry, making it easier to discuss and analyze various aspects of forest management and ecology, as shown in examples (27) to (29).

- (27) Considering that even similar thinning intensities could result in different amounts of residues depending on the pre-treatment stand **productivity**, this might also differentiate the microbial biomass response to thinning among the studied forests. (CorAgrarian.For.woodtec.11_discussion)

- (28) The larger the percentage of the finer component, the more the yield locus was shifted towards higher stresses and the smaller the **flowability**. (CorAgrarian.For.woodtec.24_discussion)

- (29) This positive relationship might be explained by bottom-up effects: higher habitat **availability** (suitable dead wood) leads to higher prey density, which in turn supports a larger number of predators and parasitoids. (CorAgrarian.For.formanag.14_discussion)

Comrie and Thompson (2007) state that almost any polysyllabic verb ending in *-ate* will form its action noun by adding *-ion*, as in *create/creation*, *observe/observation* (example 30). Similarly, most adjectives ending in *-able* or *-ible* form nouns in *-ity*: *available/availability* (example 30). But there is no way to predict, for example, that “refuse” will take *-al* while “accuse” will take *-ation* or that “true” will add *-th*. The predictable patterns mentioned contribute to the productivity of language by providing consistent rules

for creating nouns. The exceptions illustrate that while there are productive rules, language also has variability that can affect the ease and consistency of forming nominalizations.

Finally, the DS sub corpus also has **proposal types of nominalizations**. “A proposal may refer either to the fact or activity of proposing or to an actual statement or piece of writing in which an act of proposing is conveyed” (Comrie; Thompson, 2007, p. 342).

(30) The understory vegetation has changed since 1998, during which time the **observations** of Dolman et al. (2004) were conducted, and therefore, their **suggestion** mentioned above does not always occur.
(CorAgrarian.For.silviculture.05_discussion)

(31) An obvious **application** of the results would be to use these models to estimate seedling height distributions for the other NRS-FIA plots with no P2+ measurements (88 percent of all plots).
(CorAgrarian.For.silviculture.27_discussion)

Proposal nominalizations occur mainly in the discussion sections and probably in the conclusions. Words such as “suggestion” and “application” (examples 30 and 31) were used in the discussion section for describing previous methods and procedures and the results in their proposals.

Within the analyzed (DS) subcorpus, this grammatical continuum is exploited by authors to capture dynamic biophysical occurrences. While evaluative terms like “productivity” (Example 27), “flowability” (Example 28), and “availability” (Example 29) are syntactically deployed as standard nouns to fill specific argument slots, they retain a strong, underlying verbal or adjectival force that directly references active physical states and environmental behaviors. This hybrid nature serves a critical discursive role: it allows forestry discourse to stabilize changing ecological processes into fixed nominal concepts, which can then be measured, statistically quantified, and academically debated.

6 Conclusion

This study aimed to investigate the nominalizations in research articles (RAs) written by Forestry scholars by comparing their use in the discussion sections and in the other sections of the articles of the Forestry corpus. Based on quantitative and qualitative approaches, the most common nominalization suffixes used in the RAs in the corpus were: *-tion*, *-ity*, *-ment*, *-er*, *-ing*, *-is*, *-ance*, *-age*, *-al*, *-sion*, *-ness*, *-ence*, *-ant*, *-ism*, *-dom*, and *-ury*.

The nature of nominalized noun phrases was discussed, mainly emphasizing their structure, frequency, and discourse functions. The analysis of the nominalization process in the investigated corpora reveals its pervasive nature, manifesting in expanded nominalized noun phrases mainly preceded by nouns, which lend formality and lexical density to the texts, as noted by Biber, 1988; Biber; Gray, 2013; 2016; Nguyen; Edwards, 2015; Jalilifar et al., 2017; and Moreno, 2021.

A characteristic pattern is the predominance of NPs with nominalization as headers or modifiers over attributive adjectives as modifiers in both the Forestry corpus and its discussion section subcorpus. This trend is accentuated by the frequency with which nouns, often preceded by prepositions, expand the nominalized noun phrases, promoting the grouping of information more concisely through embedded phrases rather than dependent clauses, as observed by Bloor and Bloor (2004) and Gebrin (2003).

Regarding the quantitative analysis, it revealed that *-tion*, *-ity*, and *-ment* are consistently the most common nominalization suffixes employed by Forestry scholars in their RAs, both across the (F) and within its specialized (DS). This finding is evidenced by their highest frequencies. While other suffixes such as *-age*, *-ance*, *-ing*, and *-ness* also contribute to nominalization, their occurrence is significantly lower, with suffixes like *-dom* and *-ury* being particularly rare. The high prevalence of suffixes like *-tion* and *-ment* suggests a strong tendency in Forestry academic writing to transform processes and actions into entities, a characteristic often associated with scientific discourse (Comrie; Thompson, 2007).

Addressing the comparison between Forestry corpus (F) and its discussion section (DS), findings demonstrate that nominalized noun phrases in Forestry RA discussion sections exhibit distinct characteristics in terms of structure, frequency, and discourse functions. Quantitatively, nominalizations are nearly twice as often preceded by nouns than by adjectives in both the main Forestry corpus (F) and in its discussion sections (DS). This preference for noun-preceded nominalizations indicates a strong reliance on nominalized concepts and categories to describe phenomena within the field of Forestry. Qualitatively, the analysis confirmed that nominalization plays a crucial role in academic Forestry research article writing, making texts more formal, dense, and impersonal. The predominance of nominalized noun phrases, especially bigrams compounded by embedded noun-noun and

adjective-noun sequences, contributes significantly to the increased information density and grammatical complexity observed in these texts. This allows for the concise packaging of information, treating actions and processes as abstract entities, and fostering an objective tone by omitting human participants. These structural characteristics align with the demands of scientific discourse, where clarity and conciseness are prioritized as shown in other studies, such as Baratta, 2010; Comrie; Thompson, 2007; Jalilifar et al., 2017.

This study provides quantitative and qualitative evidence regarding nominalization in Forestry research articles. It not only identifies the most prevalent nominalization suffixes, but it also elucidates how these structures contribute to the informational density, formal tone, and specialized communication patterns characteristic of the discussion sections in this specific scientific discipline (Nguyen; Edwards, 2015; Jalilifar, 2017; Moreno, 2021). These insights are valuable for understanding academic writing conventions in Forestry and can inform pedagogical approaches aimed at improving the writing skills of Brazilian researchers in Agricultural Sciences for international dissemination of their research.

Acknowledgment

The authors would like to thank FAPEMIG for funding this project (APQ-01173-22).

References

- ALMEIDA, V. C. E.; ORFANÓ, B. M.; DUTRA, D. P. Is there a better choice? Verb-noun combinations in academic writing. In: Vander Viana. (Org.). **Teaching English with corpora: a resource book**. Oxon: Routledge, 2023, p. 228-232. <http://dx.doi.org/10.4324/b22833-47>
- ALVES, J.H.; BOCORNY, A. E. P.; DUTRA, D. P.; MARQUES, C. G. F.; TEIXEIRA, G. L.; COSTA, D. D. CorpuScript: an automated text-cleaning tool for corpus linguistics. In: XVI Encontro de Linguística de Corpus e XIII Escola Brasileira de Linguística Computacional, 2024, Brasília. **Anais eletrônicos** [...] Associação Brasileira de Linguística de Corpus. 2024, p. 163-167. Disponível em: <https://www.elc-ebralc.net.br/anais>. Acesso em: 21/03/2026.
- BARATTA, A. Nominalization development across an undergraduate academic degree program. **Journal of Pragmatics**. n. 42, p. 1017-1036, 2010. Available at: https://www.researchgate.net/publication/240418319_Nominalization_development_across_an_undergraduate_academic_degree_program. Access on March 13th, 2025.
- BIBER, D. **Variation across Speech and Writing**. Cambridge University Press, Cambridge, 1988.

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus linguistics: Investigating language structure and use**. Cambridge: Cambridge University Press, 1998.

BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. **Longman Grammar of Spoken and Written English**. Harlow: Pearson Education Limited, 1999. xxviii + 1204 pp.

BIBER, D.; GRAY, B. Grammatical change in the noun phrase: the influence of written language use. **English Language and Linguistics**. Cambridge: Cambridge University Press, v. 15, n. 2, p. 223–250, 2011. <http://dx.doi.org/10.1017/S1360674311000025>

BIBER, D.; GRAY, B. Nominalizing the verb phrase in academic science writing. In: AARTS, B.; CLOSE, J.; WALLIS, S. (Eds.), **The verb phrase in English: Investigating recent language change with corpora**. New York, NY: Cambridge University Press, p. 99-132, 2013. <http://dx.doi.org/10.1017/cbo9781139060998.006>

BIBER, D.; GRAY, B. **Grammatical Complexity in Academic English: Linguistic Change in Writing**. Cambridge: Cambridge University Press, 2016.

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus Linguistics: Investigating language structure and use**. Cambridge: Cambridge University Press, 2006.

BOCORNHY, A. E. P.; DUTRA, D. P.; VIANA, V. Corpus Linguistics: Studies and Applications. **Revista Brasileira De Linguística Aplicada**. v. 25, n. 1, e62068, 2025. <https://doi.org/10.1590/1984-6398202562068>

BLOOR, T.; BLOOR, M. **The functional analysis of English: A Hallidayan approach**. London: Arnold. 3rd Edition, 2004.

BRUCE, M. Finding your academic voice: use of nominalizations in academic writing. In: Viana, Vander. **Teaching English with corpora: a resource book**. New York: Routledge, 2023. <https://doi.org/10.4324/b22833>

COMRIE, B.; THOMPSON, S. A. Lexical nominalization. In: Timothy Shopen (ed.). **Language Typology and Syntactic Description**. Grammatical Categories and the Lexicon, Cambridge: Cambridge University Press. v. III, p. 334-381, 2007. <http://dx.doi.org/10.1017/CBO9780511618437.006>

DUTRA, D. P.; QUEIROZ, J. M. S.; MACEDO, L. D.; COSTA, D. D.; MATTOS E. Adjectives as nominal pre-modifiers in chemistry and applied linguistics research articles. In: RÖMER-BARRON, Ü.; CORTES, V.; FRIGINAL, E. (Ed.). **Advances in Corpus-based Research on Academic Writing: Effects of discipline, register, and writer expertise**. Studies in Corpus Linguistics. 2020, n. 95, pp. 206-226. <https://doi.org/10.1075/scl.95.09dut>

GEBRIN, M. Z. D. O uso e a importância da nominalização em resumos científicos. **Estudos Linguísticos**, São Paulo, v. 32, 2003. **CD ROM** <http://www.gel.hospedagemdesites.ws/estudoslinguisticos/volumes/32/htm/comunica/cc057.htm> Access on March 13th, 2025.

HALLIDAY, M. A. K.; MARTIN, J. R. **Writing Science: Literacy and Discursive Power**. Brighton: The Falmer Press, 1993.

JALILIFAR, A.; SALEH, E.; DON, A. Exploring nominalization in the introduction and method sections of Applied Linguistics research articles: a qualitative approach. **Romanian Journal of English Studies**. 2017, v. 14, n. 1, p. 64-80. DOI: <https://doi.org/10.1515/rjes-2017-0009>

KILGARRIFF, A.; BAISA, V.; BUŠTA, J.; JAKUBÍČEK, M.; KOVÁŘ, V.; MICHELFEIT, H.; RYCHLÝ, P.; Suchomel, V. The Sketch Engine: ten years on. **Lexicography**. 2014, v. 1, n. 1, p. 7–36. Available at: <http://www.sketchengine.eu> Access on March 13th, 2025.

MARQUES, C. G. F. CorAgrarian: um corpus de artigos científicos de ciências agrárias. In: **Semana de Eventos da Faculdade de Letras (SEvFALE)**, Universidade Federal de Minas Gerais, Belo Horizonte, 2024.

MARQUES, C. G. F. **Key Feature Analysis applied to the design of online corpus-based discipline specific activities: improving L2 Agrarian Science researchers writing skills in English**. Tese em andamento. FALE / UFMG.

MIRANDA, C. C. Padrões de nominalização em línguas da família Arawák. **Estudos Linguísticos** (São Paulo. 1978), v. 51, n. 3, p. 1113-1136, dez. 2022. Doi: <http://dx.doi.org/10.21165/el.v51i3.3273>

MORENO, A. I. Selling research in RA discussion sections through English and Spanish: An intercultural rhetoric approach. **English for Specific Purposes**, v. 63, p. 1-17, 2021. <https://doi.org/10.1016/j.esp.2021.02.002>

NORDQUIST, R. What Is Nominalization in English Grammar? **ThoughtCo**, Jun. 25, 2024, thoughtco.com/nominalization-in-grammar-1691430 Access on March 13th, 2025.

NGUYEN, T. H. T.; EDWARDS, E. C. An Investigation of Nominalization and Lexical Density in Undergraduate Research Proposals. **Language Education in Asia**, v. 6, n. 1, p. 17-30, 2015. http://dx.doi.org/10.5746/LEiA/15/V6/I1/A3/Nguyen_Edwards Access on March 13th, 2025.

PARKINSON, J. The Discussion section as argument: The language used to prove knowledge claims. **English for Specific Purposes**, v. 30, n. 3, p. 164-175, 2011. <https://doi.org/10.1016/j.esp.2011.03.001>. Available from: <https://www.sciencedirect.com/science/article/pii/S0889490611000196>. Access on March 13th, 2025.

PEACOCK, M. Communicative moves in the discussion section of research articles. **System**. v. 30, n. 4, p. 479-497, 2002. [https://doi.org/10.1016/S0346-251X\(02\)00050-7](https://doi.org/10.1016/S0346-251X(02)00050-7).

PRASITHRATHSINT, A. Nominalization as a Marker of Detachment and Objectivity in Thai Academic Writing. **Manusya: Journal of Humanities**, Special Issue n. 20, 2014. DOI: [10.1163/26659077-01703001](https://doi.org/10.1163/26659077-01703001)

PRINA DUTRA, D., BERBER SARDINHA, T. **A multi-dimensional typology of English research article sections**. Paper presented online at the American Association for Applied Linguistics Conference (AAAL). 2021.

QUEIROZ, J. M. S. **The Grammatical Complexity of English Noun Phrases in Brazilian Learner's Academic Writing: A Corpus-based Study**. MA thesis, Universidade Federal de Minas Gerais. 2019.

Recebido em: 01.07.2025

Aprovado em: 28.04.2026

Appendices

Appendix 1 – Online Tools for Data Analysis

ALVES, J. H. L. **CorpuScript**. Software em desenvolvimento. 2024.

<https://github.com/jhlopesalves/CorpusAid/blob/main/license.txt>

Sketch Engine <https://www.sketchengine.eu>

Appendix 2 - CQL QUERIES

Here are the advanced CQL queries for nominalization pattern:

1) In the CONCORDANCE tool, select the ADVANCED tab.

2) Select the CQL query type.

3) Here you can either

3a) type

[pos = "a"] [pos = "n" & word = ".*tion"] for adjectives + nouns + -tion

[pos = "n"] [pos = "n" & word = ".*tion"] for nouns + nouns + -tion

[pos = "n|a"] [pos = "n" & word = ".*tion"] for nouns or adjectives + nouns + -tion

(The dot and star .* is for "any number of any symbols".)

3b) click the CQL builder, which helps you build CQL queries.

4) Click GO.

For more information about searching with a CQL query, please check

<https://www.sketchengine.eu/news/cql-builder/> and

<https://www.sketchengine.eu/documentation/corpus-querying/>