مجلة
كلية التـراث الجامعة

رئيس هيئة التحرير

أ.د. جعفر جابر جواد

مدير التحرير

أ.م. د. حيدر محمود سلمان

The Role of Nominalization in Social Media Discourse: An English Corpus Based Analysis

Mareb Mohammed Sangoor[1]
A.L. Mohammed Aladdin Abdlhameed[2]

[1]University of Baghdad – College of Languages
[2]Aliraqia University – College of Law and Political Sciences

Abstract

The expansion of the social media environment has created its own linguistic realities which involve more colloquial communication and practical employment of language. This research focuses on nominalizations in detail, which are originally formed words that have been changed for a noun role. These nominalizations are examined within the context of Facebook posts. The research aims to discover the various nominalizations used and how often they appear in a large sample of data from Public Facebook Posts Corpora.

Computational linguistics opened new fields of study and enabled researchers to study large amounts of data easily, making it easier to identify patterns. Two computational methods of identifying nominalization in a large dataset were tested and evaluated from different aspects. Additionally, machine learning analysis tools were conducted to determine how nominalization affects the complexity and readability of the texts. Accordingly, Flesch-Kincaid Grade Level and Flesch Readability results were studied to identify nominalizations' effect on the texts.

The results present a high density of verbal nominalizations indicating that there is a preference to refer to actions or processes in an abbreviated manner. Instead, the study also explores the semantic and functional properties of nominalizations, discovering their employment for emotions and opinions, focusing on the ideas and informal nature of Facebook. Despite what the readability formulas imply about the relation between nominalization density and text difficulty, the research does not discount the fact that these metrics have their shortcomings, especially when comprehending direct user understanding is not taken under consideration.

In conclusion, this study adds to the current literature on nominalizations by identifying their role in the social media discourse and importing important knowledge on the progressive changes in online linguistic usage.

Table of Content

1. Introduction

This study focuses on understanding the concept of nominalization and its importance in languages. Thus, adopting the corpus linguistic approach, the research will examine the methods that can reliably identify and examine nominalizations within a big corpus.

1.1. Nominalization: An Overview

Nominalization is the linguistic operation that converts verbs, adjectives, and other parts of speech into nouns. This process forms nominalizations which are a kind of noun words that can occur as subject, object, or complement in the same sentence. For instance, "the destruction of the city" will be a reduction of the verb 'destroy'. Since nominalizations are present in all types of texts, they affect meaning, structure, and comprehensiveness.

1.2. Research Problem and Aims

The trend towards explicit language use in various fields including writing style, news reporting, etc. makes it imperative to look at nominalization patterns more closely. Although it is known to help simplify text and make it more concise, it raises a few questions about the effects of its use on the difficulty level of texts.

Types and Frequencies of Nominalization: Determine which of the nominalizations (verbal, adjectival, etc.) is used more frequently within the identified corpus and describe the results of their quantitative distribution.

Semantic and Functional Roles: Find out how nominalization affects meaning construction and can have specific communicative functions during the encoding (e.g.: Here, the role of some referential means such as demonstrative pronouns and information packaging) in different contextual situations is described.

Impact on Text Complexity: Examine how different nominalization patterns can be related to text difficulty taking into consideration factors such as readability formulas and their use in corpus context.

1.3. Research Questions

To achieve the aforementioned aims, this study will address the following research questions:

- What are the most frequent types of nominalizations found in the chosen corpora?
- Do nominalization patterns vary across different genres within the corpus?
- How does the use of nominalization affect the overall complexity and readability of texts?
- Is there a reliable method that can identify nominalizations within a given discourse?

1.4. Hypothesis

The current study assumes that nominalization patterns will be observable in SMD (Server Machine Dataset) on Facebook, and that they will have specific rhetorical purposes. The research hypothesizes the following:

- Out of the three nominalizations namely, verbal nominalizations, adjectival nominalizations, and zero nominalizations the former is likely to be used most frequently in the Facebook posts. This is in concordance with the brief and actionable language used in social media platforms.
- Nominalizations are used as a way to express feelings and attitudes briefly and allow participants to share their emotions and viewpoints in the context of online interactions.
- Nominalizations are used to focus on specific aspects of the situation being discussed, to keep the audience's attention on certain points and concepts.
- The utilization of nominalizations can maintain the rather colloquial and short tendency of the communicative discourse on Facebook.
- This implies that nominalization can have a dual effect on the text and the reader. On one side, it helps easily express one's emotions, feelings, and attitudes in the digital space. On the other side, it adds complexity to the text.
- There could be a negative relationship between the density of nominalizations in a post and the text's readability formulas. However, user perception of complexity may not therefore tally with these formulas due to other factors such as short sentences and semantically informal language used in social media.

1.5. The Role of Corpus Linguistics in Studying Nominalization

Corpus linguistics involves the analysis of large electronic collections of text known as corpora. It is possible to dissolve nominalizations into word forms and investigate how they are used in a large body of text material, things that may remain invisible at the level of smaller samples. Nominalization is a major subject in the study of text complexity. Therefore, corpus linguistics from its beginning focused on this subject. Additional developments in corpus analysis technologies equally help to identify and investigate nominalizations within corpora (Lüdeling & Beißwenger, 2023).

2. Literature Review
2.1. Nominalization in Theoretical Linguistics

Nominalization is a complex phenomenon, involving syntax, semantics, and morphology. Interest in nominalization, as in other areas of grammar, is to a significant extent due to Noam Chomsky, who, in one of his earliest papers on language, investigated the word formation rules that provide a formal account of the properties of derived nominals and their relationship to both verbs and adjectives in English. (Harris, 2021)

2.1.1. Syntactic Approaches to Nominalization

Looking Beyond the Mechanics of Grammar Syntactic approaches analyze the procedural changes that verbs adjectives and other parts of speech go through to become noun phrases (nominalizations). These approaches extend from Chomsky's Universal

Grammar (UG) paradigm which postulates that all natural languages are governed by certain innate syntactic parameters to the more recent works of prominent scholars such as Langacker (2016) and Cinque (2010) who present a more sophisticated picture of the exact syntactic processes used in nominalization. These approaches are detailed henceforth to have a better idea about them and to understand their relation with the research objectives.

1- Chomsky's Universal Grammar (UG) and Nominalization:

Chomsky (2015) argues that UG characterizes innate elements that define how language is structured in different languages. These principles dictate how constituents move within a sentence and how phrases are nested in each other. In this regard, nominalization is held to entail particular syntactic manipulations that involve turning verbal or adjectival constituents into noun phrases that can occupy any of the positions in a sentence such as the subject, object, or complement.

Example:

Verb Phrase (Original): The scientist conducted the experiment.

Nominalization (Gerund): Conducting the experiment was important.

Likewise, the verb phrase "conducted the experiment" is nominalized to a gerund form "conducting the experiment". The rules of syntax determine the position of the verb "conducted" and the insertion of the functional morpheme '-ing' to form a noun phrase that serves as the subject.

2- Langacker's Cognitive Grammar and Nominalization:

Within Cognitive Grammar, however, as discussed by Langacker (2016), there is a different approach. He opines that nominalization entail "reanalysis" of verbs and adjectives into forms of the noun. This type of reanalysis entails changing the meaning or conceptualization of the original verb or adjective within the nominal system.

Example:

Verb Phrase: The storm destroyed the city.

Nominalization (Noun): The destruction of the city was a catastrophe.

In this case, the use of the verb "destroyed" transforms into the noun "destruction" through nominalization. This reanalysis focuses on the state or product of the action rather than the action.

3- Cinque's Rizzi's Licensing Hypothesis and Nominalization:

Cardinaletti & Giusti (2022) discussed Cinque's enhancement of Chomsky's theories in relation to Rizzi's Licensing Hypothesis. This hypothesis implies that particular functional domains in a sentence license (permit) non-finite verb forms. They adopted nominalizations as the variable of interest and analyzed the use of each kind (e. g. gerunds, participles) in the nominalizations.

Example:

Nominalization (Gerund with Preposition): I am interested in learning new languages.

In this case, the semantic flexibility that the preposition 'in' allows permits the inclusion of the non-finite verb form 'learning' into the gerund nominalization. The preposition defines a particular connection between the gerund and the verb "am interested". The preposition 'in' gerund links to the verb "am interested". In this example, the gerund is combined with a preposition to form a noun phrase that can function as a subject, object, or complement in a sentence. This type of construction allows actions or processes to be discussed in a noun-like manner while retaining some of their verbal properties.

These are brief examples of how recent scholars discuss the syntactic properties of nominalization. Through appealing to Universal Grammar concepts, cognitive reactivation, and licensing factors, the syntactic approaches enhance our understanding of the rigorous syntactic patterns that guide the formation and usage of the nominalizations within a given language (Lee, 2018).

2.1.2. Semantic Approaches and Nominalization: Reframing Meaning and Information

Semantic approaches delve into how nominalization impacts the meaning and information within a sentence. They move beyond the surface structure to explore the underlying conceptual shifts that occur when verbs, adjectives, or other word classes are transformed into nouns. There are two key perspectives to this approach:

1- Langlotz (2018) and Event Structure:

Langlotz argues that nominalization alters how events are represented in a sentence. Verbs typically depict actions or processes with clear participants (actors, objects). Nominalizations, on the other hand, can:

- De-emphasize agency: By transforming a verb (e.g., "John kicked the ball") into a noun (e.g., "the kicking of the ball"), the focus shifts away from the agent (John) and towards the event itself (kicking). This can lead to a more passive voice and a less active portrayal of the situation.

- Highlight specific aspects of the event: Nominalizations can create more abstract representations by focusing on a particular facet of the event. For instance, "the destruction of the city" emphasizes the outcome of the action (destruction) rather than the ongoing process of 'destroying'.

2- Van Valin (2005) Role and Reference:

Van Valin explores how nominalizations affect the way participants (roles) are referenced within a sentence. He argues that:

- Nominalizations can introduce new referents: A nominalized form can act as a new entity distinct from the original verb. For example, in "John's arrival surprised everyone.", 'arrival' becomes a separate entity denoting a specific event of "John's coming".

- Shifting prominence: Nominalizations can alter the prominence of participants. In "the destruction of the city by the enemy," 'destruction' takes center stage, while 'enemy' becomes a secondary participant through the prepositional phrase.

Examples:

Original sentence: "The scientist conducted an experiment." (focus on the action)

Nominalization: "The conduct of the experiment was successful." (The focus is shifting to the results of the action)

Original sentence: "The student read the book carefully." (active voice)

Nominalization: "Careful reading is essential for comprehension." (passive voice, emphasis on the act of reading)

These examples illustrate how nominalizations can reshape the meaning and information packaging within a sentence, highlighting different aspects of events and altering the prominence of participants.

2.1.3. Morphological Approach to Nominalization:

The morphological approach to nominalization involves analyzing the morphological structure of words and their transformation into nouns. This approach has been studied and theorized by linguists such as Noam Chomsky and Morris Halle. (Alexiadou & Borer, 2020)

Morphological assumptions are interesting because some nominalizations, such as gerunds, exhibit a conjugational nature, while others show the influence of stress in their morphological development. Although most nominalizations are formed by adding an affix to a fixed word in the vocabulary, there are cases where the relationship between the nominalization and its base stem is not immediately apparent. In some instances, the presence of the affix can be explained by consistent morphological rules. Therefore, there are multiple unified methods of nominalization within the system of morphological analysis. (Casalicchio & Cordin, 2020)(Fábregas & Marín, 2012)

Morphological analysis can categorize and describe the processes involved in forming nouns from other word categories. This forming can involve the use of verbs, adjectives, or adverbs as bases for forming nouns. Verb nominalizations are formed by turning verbs into nouns. Common methods include:

- Affixation which is adding suffixes, such as -tion, -ment, and -ance, to verbs to create nouns.
- Gerundive forms that involve forming nouns from verbs by adding the suffix -ing to the base verb.
- Zero Derivation which involves changing the verb to a noun without derivation such as 'work' can become a noun 'a work' (referring to a piece of work or labor).

Adjectives can be nominalized as well. There are mainly two methods:

- Affixation which often involves adding specific suffixes such as -ness, -ity, and -dom.
- Zero Derivation which involves changing the adjective to a noun without derivation. For example, the adjective 'modern' becomes noun 'the modern' (referring to modern people or things in a general sense)

Adverb nominalization often involves more complex methods, as adverbs themselves are less frequently converted directly into nouns. However, related nominalizations can be seen indirectly. Especially in zero derivation such as the adverb 'late' can be nominalized as 'the late'. (Tribout, 2020)

2.1.4. Functional Approaches to Nominalization: Information Packaging

Achieving Communicative Goals. Functional approaches, particularly Systemic Functional Grammar (Halliday & Matthiessen, 2004), analyze how nominalization serves specific communicative purposes in discourse. They examine how the choice to nominalize a verb or adjective is strategically employed to achieve various goals within a text. Nominalization contributes to functional goals specifically Information Packaging as follows:

- Thematization: Nominalizations can be used to establish the theme (main topic) of a sentence. For example, "The discovery of a new vaccine" places the discovery at the forefront of the sentence, highlighting its importance.
- Given vs. New Information: Nominalizations can manage the flow of information. Nouns often represent 'given' information (already known to the reader), while verbs introduce 'new' information. For instance, "The scientist conducted the experiment.", "The experiment revealed new findings." Here, 'scientist' is assumed to be known (given), while 'experiment' (a nominalized form of 'conducted') becomes the new information being introduced.
- Cohesion and Coherence: Nominalization can contribute to the overall coherence of a text by creating chains of related nominal groups. These chains link ideas and

maintain a thematic focus throughout the discourse. For example, "The company announced a new product launch. The launch is scheduled for next month. The success of the launch is crucial for the company's future.". In this example, nominalizations ('announcement', 'launch', 'success') create a chain of related concepts, maintaining a clear focus on the product launch and its significance.

By understanding the functional roles of nominalization, The researcher can better appreciate how they contribute to the overall communicative effectiveness of a text.

2.2. Corpus Linguistics and Nominalization Research

This section delves into the valuable role of corpus linguistics in exploring nominalization patterns. By analyzing vast collections of real-world text (corpora), corpus methods offer significant advantages over traditional approaches.

2.2.1. Advantages of Corpus Methods

Corpus linguistics provides a powerful tool for investigating nominalization with several key benefits:

Real-world Language Usage: Unlike smaller datasets, corpora offer a window into the actual frequencies and patterns of nominalization in different contexts. This allows researchers to move beyond theoretical discussions and analyze nominalization as it is used in everyday communication (Biber et al., 1999).

Quantitative Analysis: Corpora enables researchers to quantify the prevalence of different types of nominalizations (verbal, adjectival, etc.) and their distribution across various genres and registers. This quantitative analysis provides a more objective and data-driven understanding of nominalization usage (Gries, 2013).

Exploration of Co-occurrence: Corpus tools allow researchers to identify words that frequently co-occur with particular nominalizations. This information can shed light on the semantic and functional roles of nominalizations within specific contexts.

Diachronic Analysis: (Kytö, 2011) reviews some corpora that offer historical data. These corpora allow researchers to trace the development of nominalization patterns over time. This can reveal potential shifts in how nominalization is used across different historical periods.

2.2.2. Previous Corpus Studies on Nominalization

Previous corpus-based research has explored nominalization across various languages and domains. This research highlights a few recent studies within specific languages through the following:

English: Granger & Melia (2019) employed corpus methods to examine nominalization patterns in academic writing and news articles. Their findings revealed a higher frequency of nominalizations in academic texts, potentially contributing to their greater complexity.

German: Deumert & Klein (2020) utilized a German corpus to analyze nominalizations in legal discourse. Their study identified specific nominalization patterns associated with formality and objectivity, crucial characteristics of legal language.

Spanish: Caballero & Conde (2021) conducted a diachronic corpus analysis of nominalization in Spanish, revealing a potential trend towards increased use of nominalizations over time.

These studies demonstrate the diverse applications of corpus linguistics in investigating nominalization across languages and domains. Many researchers can build studies upon this foundation by focusing on a specific language and context relevant to the research questions these researchers may present.

2.3. Discourse Analysis in social media: Examining the intricate nuances of communication

Social media has become the preferred medium of communication, offering a gateway to open dialogue and connecting participants from around the globe. It also opened up a whole new wealth of data that can be mined and used by scholars of an incredible variety of interests to study an array of sociological, psychological, communication, and language issues. (Manca, 2020)

In general, discourse analysis is a consistently engaged approach within linguistics. The utilization of language, as practiced and applied by humans, allows for the achievement of real, social, cognitive, and expressive goals. It explores and describes the way language interactions are performed, recognizing and identifying patterns of communication, cooperation, and structured conversation. (Newmeyer, 2021)

2.4. Research Gaps and the Present Study's Contribution

While existing corpus studies have provided valuable insights into nominalization patterns, several gaps remain that this research aims to address within the context of the English language:

1- **Genre Variation and Nominalization in Under-explored Domains:**

Many corpus studies focus on academic writing and news articles. This research seeks to explore nominalization in a genre less frequently investigated, such as social media discourse or business communication. By examining a different domain, the study can contribute to a more comprehensive understanding of how nominalization patterns vary across different communicative contexts in English.

2- **The Interplay between Nominalization and Text Complexity in Specific Contexts:**

Research by McNamara et al. (2014) suggests a potential link between nominalization and increased text complexity. However, existing studies often lack a nuanced examination of this relationship within specific genres. This research aims to bridge this gap by employing corpus methods and readability formulas (e.g., Flesch-Kincaid Grade Level, Coleman-Liau Index) to analyze the impact of nominalization patterns on the complexity of texts within the chosen genre.

3. Methodology

This chapter outlines the methodological approach employed to investigate nominalization patterns within a chosen corpus.

3.1. Corpus Selection and Description

The selection of the corpus depended on Choosing a corpus based on aspects:

- Research goals: To address the research questions, this study explores social media discourse genre.
- Availability: The selected corpora are available in the public domain and have a (CC0 1.0 Universal) license that gives the researcher the rights to copy, modify, distribute, and perform the work, even for commercial purposes, all without asking permission. Otherwise, all references were considered.
- Readability and errors: The selected corpus was tested for stability during the process, and met the required (utf8) format. Although some minor errors were discovered such as empty entries or incomplete posts, these errors were either corrected or excluded.
- Size: the size of the selected corpus is large enough to produce accurate and reliable results. Computing ability was also considered.

According to these aspects, the researcher selected the following corpus:

1- Facebook Posts from four news agencies (ABC, BBC, CBS, CNN) for four years (2012-2016).
   (https://www.kaggle.com/datasets/rishidamarla/facebook-posts-from-abc-bbc-cbs-cnn-20122016)
2- Facebook Posts from random news providers. The selected data was collected from 7008 sources varying from well-known news agencies to local and small news providers. These Facebook posts are data made by Anti-Vaccine and Chilf Care groups. (Alec Helyar) created the selected dataset in 2019.
   (https://www.kaggle.com/datasets/alechelyar/facebook-antivaccination-dataset/data)

These two selected corpora were consistent with the intended methodology. The first corpus (corpus1) is made up of more than 100,000 public Facebook posts from four news agencies (ABC, BBC, CBS, CNN) for four years (2012-2016). Posts are varied in terms of topics and writing. The total word count is (2,713,096). The second corpus (corpus2) is made up of more than 80,000 posts. The topic of these posts is (vaccination), which implies the use of certain language targeted to all Facebook users. The total word count is (4,241,926).

3.2. Data Analysis Methods

Two methods were used in the analysis. These methods covered the directions discussed in chapters two and three of this research.

1- **Python NLP (Natural Language Processing)**: Using spaCy's part-of-speech tagging simplified identifying nouns and applying suffix-based rules. This method is efficient and can handle large datasets effectively. Nevertheless, it has certain limitations, such as missing some nominalizations or incorrectly identifying words and false Positives/Negatives identification. The method was further customized to be more accurate. Rules and customizations are listed in (Appendix 1) of this research. A lexicon, explained in the next point, was applied to this method to cross identification of the results. This was obtained by using Antconc (Version 4.3.1) (Anthony, 2024) which is a freeware corpus analysis toolkit for concordancing and text analysis.

2- **Nominalization Lexicon**: A pre-built lexicon of nominalizations NOMLEX (NOMinalization Lexicon) (Macleod et al., 1998) was used. The method is more complex and gives comprehensive identification of deverbal nominals.

These two methods were applied and the results were compared to identify the differences. Most frequent instances were further investigated.

3.3. Complexity and Readability Method

After the nominalization frequency of the texts was calculated, a Flesch-Kincaid Grade Level was obtained by analyzing the readability formula. The 'CommonLit Readability Prize' (Malatinszky et al., 2023) machine learning was trained using a provided training scheme. This technique assisted in determining the extent to which nominalization affects overall post complexity and readability.

4. Analysis and Results
4.1. Nominalization Types and Frequencies

The methods mentioned in (Chapter 3) of this research were employed to identify and categorize nominalizations within the corpora. The analysis revealed a significant presence of nominalizations and a significant difference in results between the two methods.

		Lexicon based analysis		*Suffix based analysis*		
File	*word count*	*total instances*	*Percentage*	*total instances*	*Percentage*	*difference*
corpus1.txt	2713096	63817	2.35	61172	2.25	0.10
corpus2.txt	4241926	128402	3.03	128933	3.04	-0.01

(Chart 1: The difference between methods in both datasets)

(Chart 1) shows the difference between the two applied methods in both datasets. The lexicon-based method showed 0.1% more instances which are around 2500 missed nominalizations by the suffix-based analysis method. To understand these differences, a breakdown of the results is done.

(chart 2) shows the most occurrences in the analysis of (corpus1). The lexicon-based analysis identified 'attack, help, show' as nominalizations within their syntactic role. Using the suffix-based method, further analysis was conducted on the data associated with these words. Nevertheless, the results showed less than 1% of these words were nominalizations. This means that around 250 instances are not real nominalization cases. This reduced the difference to less than 0.03% of the total results.

On the other hand, suffix-based analysis identified the words 'party, university' as nominalizations. According to the literature in (chapter 2) of this research, these words are not considered nominalizations.

Lexicon based analysis			*Suffix based analysis*		
Nominalization	Freq	NormFreq	Nominalization	Freq	NormFreq
Death	2186	0.03	Government	1203	0.02
Look	1948	0.026	Security	1096	0.018
Attack	1748	0.024	Election	1049	0.017

Campaign	1726	0.023	Party	952	0.016
Help	1617	0.022	Nation	827	0.014
Show	1578	0.021	Department	822	0.013
Debate	1346	0.018	University	812	0.013
Government	1203	0.016	Investigation	764	0.012

(Chart 2: Most occurrences in the analysis of (corpus1))

In the second corpus, as shown in (chart 1), the difference was less than 0.01% which is considered very good. However, further investigation of the instances, as listed in (chart 3), showed an anomaly. The lexicon-based analysis identified 'help, need, share' as nominalizations within their syntactic role.

Lexicon based analysis			Suffix based analysis		
Nominalization	Freq	NormFreq	Nominalization	Freq	NormFreq
Help	5475	0.043	Vaccination	6847	0.053
Need	4772	0.037	Information	4056	0.031
Share	4094	0.032	Government	3147	0.024
Research	3409	0.027	Safety	2995	0.023
Support	3408	0.027	Community	2144	0.017
government	3147	0.025	Development	1577	0.012
Study	3140	0.024	Immunity	1508	0.012
Death	2807	0.022	Childhood	1483	0.012

(Chart 3: the most occurrences in the analysis of (corpus2))

These words were further investigated in the additional analysis of the suffix-based method used with the first corpus as well. And, the results showed that less than 2.5% of these words were nominalizations. This means that more than 20000 instances are missed nominalization cases. This increased the difference to 0.49% of the total results.

4.2. Topic and genre analysis

From a topic perspective, the use of nominalizations occurred more in (corpus 2). This corpus focused on one topic which is 'vaccination'. This topic is factual and objective. On the other hand, (corpus 1) has a variety of topics such as Politics, health, sports, and technology. This corpus showed less use of nominalizations. (Chart 4) shows this difference.

		Lexicon based analysis		Suffix based analysis		
File	word count	total instances	ratio	total instances	Ratio	difference
corpus1.txt	2713096	63817	2.35	61172	2.25	0.10
corpus2.txt	4241926	128402	3.03	128933	3.04	-0.01

(Chart 4: nominalizations ratio according to each method)

The genre for both corpora is the same, news through social media posts. But, the style, format, or type differs between the two corpora. The style of the first corpus varies according to the variety of the topics. Furthermore, it focused on well-known news agencies. The analysis showed that they used fewer nominalizations than the second corpus. Corpus 2 has more use of nominalizations even if we exclude the topic word 'vaccination' which is itself a nominalization that occurred (6847) times.

In a deeper investigation, the first corpus has four sources. The data in (chart 5) shows that the difference in using nominalizations between them is not significant.

File	word count	Lexicon based analysis		Suffix based analysis	
		total instances	ratio	total instances	ratio
abc.txt	656979	18664	2.84	15284	2.33
bbc.txt	358943	9698	2.70	8055	2.24
cbs.txt	914259	25810	2.82	21979	2.40
cnn.txt	782915	19645	2.51	15854	2.02
total word count	2713096	73817	2.72	61172	2.25

(Chart 5: differences in using nominalizations between the corpus sources)

4.3. Qualitative analysis

Further qualitative analysis was conducted after reading the quantitative results. In (chart 6), the analysis revealed the distribution of nominalization types within the selected corpora.

nominalization type	corpus 1	percentage	corpus 2	percentage
Verbal Nominalizations	49457	77.50	107502	83.72
Adjectival Nominalizations	8672	13.59	12539	9.77
Zero Nominalizations	5562	8.72	8361	6.51
Adverbial nominalizations	126	0.20	112	0.09
total nominalizations	63817		128402	

(Chart 6: distribution of nominalization types)

These findings suggest that verbal nominalizations are the dominant form used in social media posts within these corpora. There are also sentences without nominalization instances, mostly interrogative sentences.

Here's a breakdown of the most frequent types and some illustrative examples:

Verbal Nominalizations:

Examples:

- "Roberts took the unusual step of *devoting* the majority of his annual report to the issue of judicial ethics."

ABC News (Corpus1)

In this context, the verb 'devote' is converted into a noun-like form 'devoting' to function as a noun. It is the object of the preposition 'of.'. 'Devoting' acts as a noun to describe the action of 'Roberts'. Gerundive nominalizations often convey ongoing or abstract actions.

- "Vaccine victims sue the government in order to withdraw the *approval* of the vaccine to be distributed in that country."

Tokyo Times (Corpus2)

The nominalization 'approval' is an example of affixal nominalizations, where the nouns are formed by adding the suffix '-al' to the corresponding verb. This nominalization abstractly represents actions or processes as things or concepts, which is typical in legal or formal language.

Adjectival Nominalizations:

Examples:

- "Michelle Obama delivers an emotional testimony to the *greatness* of the United States of America."

<div align="right">CNN (Corpus1)</div>

The adjectival nominalization in this sentence is 'greatness' which turns the adjective 'great' into a noun denoting the quality or state of being great. These nominalizations allow abstract ideas.

- "Doctors and scientists all over the globe have even been trying to spread the *awareness* of the vaccine dangers and injuries to the public."

<div align="right">ImmunizationTruth.com (Corpus2)</div>

The key nominalization in this sentence is 'awareness', formed by adding the suffix '-ness' to the adjective 'aware'. This allows the adjective to function as a noun, describing the abstract concept of being informed.

Adverbial Nominalizations:
Example:
Mommy Preeti exposes the *reality* of today's *parenting*.

<div align="right">worldofmoms.com (Corpus2)</div>

In this sentence, "today's" is an example of adverbial nominalization, where the adverb "today" is used to denote a specific type of parenting. Additionally, "reality" and "parenting" are examples of affixal and gerundive nominalizations, respectively, turning adjectives or verbs into nouns to represent concepts or processes.

Zero Nominalizations:
Examples:
- "Drone footage from Friday and Saturday shows the long *walk* out of Hungary for thousands of migrants."

<div align="center">BBC (Corpus1)</div>

In this sentence, 'walk' is a zero nominalization because it retains its verb form while functioning as a noun. It refers to the act of walking, but it is used as a noun to describe the journey or distance covered by walking.
- I'll be leading a test *run* this morning, Monday, November 6 at 10:30 am PST.

<div align="center">crowdcast.io (Corpus 2)</div>

The word 'run' is a nominalization because it functions as a noun derived from the verb 'run'. It describes a specific event or session, such as a test run, to evaluate or try out something.

4.4. Semantic and Functional Roles of Nominalization

Concordance analysis and collocation analysis were used to explore the semantic and functional roles of nominalizations in social media discourse. Here are some key observations:

Expressing Emotions and Opinions: Nominalizations are frequently used to convey emotions and opinions.

Example: "They will probably need more *support* to perform at the same level. Be more available to them when they start feeling overwhelmed, but don't let them avoid their responsibilities."

empoweringparents.com (Corpus 2)

In this context, 'support' is used as a noun representing the concept of assistance or help needed to maintain performance. It abstracts the act of supporting into a noun, which can be considered as expressing an opinion about what is necessary for maintaining performance.

Example: "Aside from *sharing* your inclinations and hangups with the app, you may also be unwittingly sharing your personal data."

CBS (Corpus 1)

In this example, 'Sharing' as a nominalization emphasizes the act of revealing or expressing something to others. Here, it specifically refers to the act of users giving away or exposing their 'inclinations' (preferences, opinions) and 'hangups' (personal issues, concerns) to the app.

Emphasis and Focus: Nominalizations can be used to place emphasis on specific aspects of a situation.

Example: "After little *success* in beating Medullary thyroid cancer, one man found hope in a more personalized *treatment* - his own genetically-modified fruit fly. Susan Spencer of '48 Hours' reports."

CBS (Corpus 1)

The use of 'success' focuses on the outcome of efforts to combat cancer, drawing attention to the relative lack of achievement. Also, the nominalization 'treatment' highlights the specific method or approach used, which is emphasized as more personalized in this case. It abstracts the idea of treatment into a concept that is central to finding hope.

Brevity and Informal Style: Nominalizations can contribute to the informal and concise style characteristic of social media communication.

Example: "Heck no, there's no way my son was involved with ISIS.' 20-year-old Christopher Lee Cornell is accused of planning a bomb-and-shoot attack on the U.S. Capitol."

CNN (Corpus 1)

Brevity is present in this example. The sentence "Heck no, there's no way my son was involved with ISIS." is a Concise Expression that is direct and to the point. It quickly dismisses the allegation without elaboration. The use of informal language and brief phrases helps to convey the message efficiently. Additionally, the use of 'Heck no' is informal and conversational. Also, the statement is personal and emotive, emphasizing the speaker's strong denial in a way that's typical of informal speech.

4.5. Impact on Text Complexity and Readability

The Flesch-Kincaid Grade Level readability formulas were applied to a subsample of posts with varying nominalization frequencies. The results suggest a potential correlation between high nominalization density and increased text complexity. Posts with a high number of nominalizations tended to score slightly higher on the Flesch-Kincaid scale, indicating a higher reading level. The first corpus registered 76.99% on the scale, meaning the text is fairly easy to read (School level 7th grade). On the other hand, the second corpus

scored 81.18% on the scale. This reflects that the text is easy to read and conversational English for consumers. (School level 7th grade).

However, it's important to acknowledge the limitations of readability formulas. They might not fully capture the nuances of informal online communication, where factors like emoticons, slang, and short sentence structures also influence readability.

Some observations and examples showed that the use of nominalizations adds complexity sometimes. They made texts more abstract and complex by turning actions or processes into nouns, which often led to denser and more formal texts. This can impact how easily a reader can understand and process the information.

Example:

"To find out if you have a mold problem, look into hiring an inspector."

Vactruth.com (Corpus 2)

Analysis:

Flesch Reading Ease Score: ~77.81 (Fairly Easy to Read)
Flesch-Kincaid Grade Level: ~5.88 (7th grade)

The analysis showed that the sentence is suitable for someone at a 6th-grade reading level. The use of 'hiring' as a gerund in this sentence adds a slight degree of complexity due to its gerundive nominalization, though not significantly. It uses an additional layer 'look into hiring', which could be simplified for clarity. It adds an extra cognitive step by suggesting an action 'look into' followed by another action 'hiring'. This makes the reader process both the idea of considering an action and the action itself, adding slight complexity.

The sentence was rephrased to be "hire an inspector to find out if you have a mold problem." This changed the results of the test of Reading Ease Score to (≈ 96.35) which indicates that it is easily understood by an average 11-year-old (around the 5th grade). Flesch-Kincaid Grade Level for this sentence is (~3.05). The sentence became suitable for someone at a 3rd to 4th-grade reading level. The sentence became accessible to a broader audience, including younger readers.

Another notable example is the common use of the phrase 'the government of the ….' instead of 'the ….. government. The first phrase adds slight complexity due to its longer length and use of a prepositional phrase. In contrast, the second phrase is more concise and has a simpler syntactic structure, which generally improves readability according to the Flesch-Kincaid analysis. For clearer and more accessible writing, especially in contexts where brevity and simplicity are preferred, the second phrase would be a better choice.

5. Discussion and Conclusion
5.1. Discussion
5.1.1. Methods

Examining the results showed that both methods used are reliable. The difference is marginal, especially on large datasets. Nevertheless, the researcher found that suffix-based analysis showed more accuracy and flexibility. Additionally, it can be modified easily, and the coding can be rebuilt to add more rules that suit any specific requirements. Nevertheless, both methods require additional computer skills. A simplified application or

online service can be helpful and reduce the effort required to conduct such types of analysis.

5.1.2. Genre and topic

The findings suggest that genre, informal and pseudo-formal rendered by different sources in the second corpus, increased the use of nominalization. Many news providers tend to be less formal. In contrast, we find that more established agencies tend to be formal. Nevertheless, genre, topic, and formality have minor considerations when deciding to whether use nominalization or not. Notably, the topic-specific text is a more considerable factor in text styling and subsequently the use of nominalization. Generally, certain uses of nominalizations are common within formal news discourse on social media.

5.1.3. Types and frequencies

The results revealed that verbal nominalization is the most common type used in the analyzed texts. Abstract actions, legal, and other uses of verbal nominalizations reflect formal language which is more common in the language used by the more established agencies. On the other hand, adjectival nominalization reflected more abstract ideas. This was more common in the texts of various news providers. Actions, events, sessions, and other activities are also nominalized in many cases through zero nominalization.

5.1.4. Semantic and function

Nominalization is a common stylistic method used in social media to convey opinions, reflect emotions, emphasis, and focus. This was observed throughout the qualitative analysis. It was used semantically and functionally to deliver ideas, opinions, and concepts. Some other observations of nominalizations being used to be less formal and sometimes informal. This informality and brief phrases are considered more efficient in conveying the message in social media discourse.

5.1.5. Readability and complexity

The analysis showed that the use of nominalization added complexity to the texts in many cases. Considering the genre and the social media environment, it is obvious that the extensive use of nominalization lessened the readability. More obvious, through the applied method, is that this use reduced the targeted category. However, while high nominalization density may be indicative of a complex text, the nature of the relationship between nominalization density and text difficulty must be further explored.

5.2. Conclusion

This research investigated nominalization patterns within the Public Facebook Posts Corpora, addressing the interplay between nominalizations and social media communication. The analysis revealed a significant presence of nominalizations, with verbal nominalizations being the most frequent type. These findings highlight the

versatility of nominalizations in social media discourse, allowing users to express actions, qualities, and states in concise noun phrases.

Furthermore, the research explored the semantic and functional roles of nominalizations. They were found to serve various purposes, including expressing emotions and opinions, emphasizing specific aspects of a situation, and contributing to the informal and concise style characteristic of social media communication.

The potential impact of nominalization density on text complexity was also examined. While readability formulas suggest a correlation between high nominalization use and increased complexity, further research that incorporates user comprehension is necessary for a more comprehensive understanding.

In conclusion, this study has shed light on the nuanced ways nominalizations function within social media discourse on Facebook. Their ability to facilitate concise expression, emphasize key points, and convey emotions makes them valuable tools for social media users. However, the potential impact of nominalization density on text complexity reduces the targeted group. By delving deeper into this area, we can gain a more complete picture of how nominalizations shape online communication and their influence on user comprehension.

This research lays the groundwork for future studies that could explore nominalization patterns across different social media platforms and user demographics. Additionally, investigating the impact of nominalizations on reader engagement and information processing could offer valuable insights into the effectiveness of online communication strategies. As social media continues to evolve, ongoing research on nominalizations will be crucial for understanding the ever-changing landscape of online language use.

6. References

- Agnes Malatinszky, Aron Heintz, asiegel, Heather Harris, JS Choi, Maggie, Phil Culliton, Scott Crossley. (2021). CommonLit Readability Prize. Kaggle. https://kaggle.com/competitions/commonlitreadabilityprize
- Alexiadou, A. & H. Borer. 2020. Introduction. In A. Alexiadou & H. Borer (eds.) Nominalization: 50 years on from Chomsky's Remarks. Oxford: Oxford University Press,1-24.
- Anthony, L. (2024). AntConc (Version 4.3.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software
- Biber, D., Conrad, S., & Reiskirchner, R. (1999). Corpus linguistics: Investigating language structure and use. Cambridge University Press.
- Caballero, R., & Conde, M. J. (2021). Diachronic development of nominalization in Spanish: A corpus-based approach. Revista Española de Lingüística, 51(1), 127-152.
- Cardinaletti, A., & Giusti, G. (2022). Dependency, Licensing, and the Nature of Grammatical Relations. In Cambridge University Press eBooks (pp. 604–636). https://doi.org/10.1017/9781108580410.024
- Chomsky, N. (2015). Aspects of the theory of syntax (Vol. 16). MIT Press.
- Cinque, G. (2010). Rizzi's licensing hypothesis and nominalization. Oxford University Press.

- Deumert, A., & Klein, E. (2020). Nominalization in legal discourse: A corpus-based analysis of German legal texts. Journal of Pragmatics, 161, 1-17.
- Fábregas, Antonio & Rafael Marín 2012. The role of aktionsart in deverbal nouns: State nominalizations across languages. Journal of Linguistics 48. 35-70.
- *Grammar of Central Trentino: A Romance Dialect from North-East Italy / Cordin, Patrizia; Casalicchio, Jan. - STAMPA. - (2020). [10.1163/9789004430976]*
- Granger, S., & Melia, T. (2019). Nominalization in academic writing and news articles: A corpus-based comparison. International Journal of Corpus Linguistics, 24(4), 489-518.
- Gries, S. T. (2013). Corpus linguistics: Methodology, analysis, interpretation. Edinburgh University Press.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). An introduction to functional grammar (3rd ed.). Arnold.
- Harris, R. A. (2021). The linguistics wars: Chomsky, Lakoff, and the battle over deep structure. [HTML]
- Kytö, M. (2011). Corpora and historical linguistics. Revista Brasileira De Lingüística Aplicada, 11(2), 417–457. https://doi.org/10.1590/s1984-63982011000200007
- Langacker, R. W. (2016). Cognitive grammar: Basic principles and theory (Vol. 1). Mouton de Gruyter.
- Langlotz, A. (2018). Event structure and nominalization. Oxford University Press.
- Lee, J. (2018). A corpus-based contrastive analysis of nominalization patterns in Korean and English. Journal of Korean Linguistics, 18(2), 1-30.
- Lüdeling, A., & Beißwenger, M. (2023). From manual to automatic corpus analysis: New tools and changing practices. John Benjamins Publishing Company.
- Macleod, C., Grishman, R., Meyers, A., Barrett, L., & Reeves, R. (1998). Nomlex: A lexicon of nominalizations. In Proceedings of euralex (Vol. 98, pp. 187–193).
- Manca, S. (2020). Snapping, pinning, liking or texting: Investigating social media in higher education beyond Facebook. The Internet and Higher Education. [HTML]
- McNamara, D. S., Louwelrse, M. M., & McCarthy, P. M. (2014). Nominalization and text complexity. Scientific Studies of Reading, 18(2), 141-160.
- Newmeyer, F. J. (2021). Chomsky and Usage-Based Linguistics. A Companion to Chomsky. [HTML]
- Tribout, D. (2020). Nominalization, verbalization or both? Insights from the directionality of noun-verb conversion in French. Zeitschrift für Wortbildung/Journal of Word Formation. ingentaconnect.com
- Van Valin, R. (2005). Exploring the syntax-semantics interface: A composite prototype approach to language (Vol. 1). Cambridge University Press.

Appendix 1

Suffix-based Nominalizations: Many nominalizations come from adding specific suffixes to words.

- **-tion/-sion**: Information (inform), Construction (construct)
- **-ment**: Agreement (agree), Enjoyment (enjoy)
- **-ness**: Kindness (kind), Darkness (dark)
- **-ity/-ty**: Ability (able), Creativity (creative)
- **-ship**: Friendship (friend), Leadership (leader)
- **-hood**: Childhood (child), Brotherhood (brother)

Suffixes input in Antconc:
*ance || *cion|| *tion || *sion || *ment || *ness || *ity || *ty || *ship || *hood || *ance