IDEAS

Journal of Language Teaching and Learning, Linguistics and Literature

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Volume 14, Number 2, December 2026
pp. 7900 - 7931

Issued by English study program of UIN Palopo

Recurrent Multi-word Sequences: Syntactic Structure and Function in EFL Student Writing Texts

Indah Nuraini Gusasi
English Language Education Study Programme,,Faculty of Letters and Culture,
Universitas Negeri Gorontalo

Article Info

Received: 2026-04-07
Revised: 2026-05-19
Accepted: 2026-09-18

Keywords:
RecurrentMulti-word Sequences; EFL Writing; Syntactic Structure; Discourse Function; Learner Corpus

DOI:
10.24256/ideas. v14i2. 12429

Corresponding Author:
Indah Nuraini Gusasi
Indh.gusasiiii@gmail.com
English Language Education Study Programme,,Faculty of Letters and Culture, Universitas Negeri Gorontalo

Abstract

This study addresses the limited empirical attention to how Recurrent Multi-word Sequences (RMS) reflect syntactic structure and function in EFL students' writing, particularly in the English Language Education Study Programme (ELESP), Universitas Negeri Gorontalo. It examines how students use RMS in terms of syntactic structure and function and compares the writing patterns of third-semester and fifth-semester students. A qualitative comparative design was employed using a learner corpus consisting of 143 student writing texts, with 98 texts from third-semester students and 45 texts from fifth-semester students. Data were analyzed using AntConc to identify four-word RMS based on frequency and range, followed by concordance analysis to examine their syntactic structures and discourse functions. The findings reveal a clear difference in syntactic patterns between the two cohorts. Third-semester students predominantly used noun phrase (NP)-structured RMS, accounting for 66.7% of their total RMS frequency, whereas fifth-semester students predominantly used verb phrase (VP)-structured RMS, accounting for 66.3%. Functionally, discourse organizers were dominant in both cohorts, representing 79.5% of RMS in the third-semester corpus and 64.2% in the fifth-semester corpus. However, fifth-semester students showed a higher proportion of stance expressions, increasing from 20.5% in the third-semester corpus to 35.8% in the fifth-semester corpus. The findings suggest a shift from topic-centered and referential language patterns toward more verb-centered, general academic, and

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

evaluative patterns as students progress to a higher semester. Pedagogically, the findings imply that EFL writing instruction should provide greater attention to recurrent academic language patterns, particularly varied noun and verb phrase constructions, discourse organizers, and stance expressions, to support students' syntactic development, cohesion, and ability to express academic arguments more effectively.

1. Introduction

Writing is one of the most important skills for EFL students because it allows them to communicate ideas, opinions, and knowledge in a clear and organized way (Marlina, 2024). In academic writing, students are expected not only to use correct grammar and vocabulary but also to produce texts that are coherent, natural, and well-structured (De Smedt et al., 2022). However, many EFL students still experience difficulties in organizing ideas cohesively and using appropriate language patterns in writing, resulting in texts that sound less natural and less connected (Nora Fitria et al., 2024; Ismi et al., 2025). For this reason, examining the language patterns that students frequently use in their writing has become an important area of research in EFL writing studies.

English Department students at Universitas Negeri Gorontalo still face difficulties in writing texts, especially in organizing ideas clearly and using appropriate grammatical structures. Their writing often lacks cohesion and natural flow, even when they understand basic grammar and vocabulary (Syarif, 2022; Aljoza & Almustafa, 2024). This problem may be caused by limited use of effective language patterns, such as recurrent multi-word sequences (RMS), which are important for building fluent and well-structured texts.

Recurrent multi-word sequences (RMS) are strings of words that repeatedly occur together across multiple texts and registers in language corpora (Cortes, 2013a; Douglas Biber et al., 2004). Recent studies in EFL writing research have highlighted that learners' language development can be understood more clearly through patterns that appear repeatedly in their texts. According to (Li & Fang, 2022), RMS provide a useful window into how students naturally combine words to form meaning in real writing situations.

RMS are also considered reliable indicators of cohesion, fluency, and syntactic control (Douglas Biber et al., 2004). (Casal & Yoon, 2023) also found that students who frequently use RMS tend to produce more coherent and structured texts. These RMS often reflect the way learners build cohesion, showing how ideas are linked across sentences. Therefore, analyzing RMS helps researchers understand how students develop their writing skills from both a cohesive and structural perspective.

In line with this, RMS are closely related to syntactic structure and function because they show how words are commonly organized and used in real

communication. In terms of structure, RMS can appear as patterns like noun phrases (the importance of the study) or verb phrases (is based on the), while in terms of function, they help express meaning such as transitions (on the other hand) or organizing ideas (the purpose of this study is). Recent studies show that mastering these patterns helps learners write more fluent and cohesive texts. However, many UNG students still struggle to use these patterns effectively, which causes their writing to lack cohesion and sound less natural (Aljoza & Almustafa, 2024; Hamdani & Abid, 2025). Therefore, analyzing RMS in terms of syntactic structure and function can help explain and improve students' writing problems.

Previous research on RMS has also been conducted in the English Language Education Study Programme (ELESP) of Universitas Negeri Gorontalo, such as the study by Mahading (2024). However, the study did not examine in depth how RMS are used in terms of syntactic structure and function in students' writing. It is still unclear whether students' ability to use RMS appropriately in these aspects can make their writing more complex and varied. Therefore, this study aims to explore this issue more deeply by focusing on how RMS reflect both structure and function in writing. Another important gap addressed in this research is that it specifically investigates the writing of students from the ELESP at Universitas Negeri Gorontalo.

Even so, understanding the relationship between RMS and syntactic structures & functions is crucial for portraying students' writing ability. Recent findings (Appel et al., 2024) show that analyzing RMS allows researchers to identify consistent patterns that reflect both

linguistic development and functional choices. These patterns help explain how students' writing evolves, not only in terms of vocabulary, but also in how sentences are organized and linked. As emphasized by (Yang et al., 2022), examining RMS within syntactic functions provides a more balanced view of behavior in writing, particularly in relation to recurrent syntactic patterns and discourse functions. because it integrates form and meaning. Therefore, the present research uses RMS analysis to understand how students' writing reflects their syntactic structure and the function created through their writing. This research aims to examine how students use RMS in their writing in terms of syntactic structures and functions, as well as to provide a clearer description of students' writing based on the patterns of syntactic structures and functions reflected through their use of RMS.

Limitation of the study

This study has several limitations that should be considered when interpreting the findings. First, the two learner corpora differ considerably in size, with the Class of 2020 corpus containing substantially more tokens than the Class of 2021 corpus. Although normalized frequency was considered to support comparison, differences in corpus size may still influence the number and distribution of identified RMS. Second, the writing texts were produced in different courses and writing tasks, meaning that topic, genre, and task requirements may

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

have influenced the recurrent language patterns found in each corpus.

Third, the study uses a cross-sectional comparison of two different cohorts rather than following the same students over time. Therefore, differences between the cohorts cannot be interpreted as direct evidence of students' linguistic development across semesters. Fourth, the analysis focuses specifically on four-word RMS and selected syntactic and functional categories. Other sequence lengths, grammatical structures, and discourse functions may reveal additional patterns that are not captured in this study.

Finally, the study focuses on phraseological patterns found in students' writing and does not use RMS as a direct measure of overall writing quality or proficiency. The findings therefore describe recurrent language behavior within the selected learner corpora and should not be generalized to all EFL students or all academic writing contexts.

2. Method

This study employs a comparative corpus-based research design to examine the use of Recurrent Multi-word Sequences (RMS) in EFL students' writing. A corpus-based approach is used because the study identifies recurrent sequences through frequency and range and examines their distribution in two learner corpora. The study compares the RMS identified in the writing of third-semester and fifth-semester students in terms of syntactic structure and function. The comparison is intended to describe similarities and differences in the language patterns found in the two corpora rather than to measure students' overall writing ability or language development.

The analysis consists of two complementary stages. First, AntConc is used to identify RMS based on frequency and range criteria. Second, the identified RMS are examined through concordance and contextual analysis to determine their syntactic structures and discourse functions. The first stage provides corpus-based evidence concerning the recurrence and distribution of the sequences, while the second stage provides a qualitative interpretation of how the sequences are used in context. Therefore, the study combines corpus-based quantitative information with qualitative contextual interpretation within a comparative design.

The data in this study come from a specialized corpus that contains samples of students' academic writing. This corpus represents real writing produced by EFL learners in an authentic classroom context. By using corpus-based analysis, this study looks closely at how students naturally use RMS in their writing and how these sequences make their texts more cohesive and grammatically stronger. The main focus is on identifying the most frequent and varied RMS used by both groups of students.

This study applies a two-layered analysis to examine students' use of RMS in their writing. The first layer focuses on identifying RMS. Because the two learner corpora differ considerably in size, the frequency and range criteria were

considered in relation to corpus size. The initial extraction used a minimum frequency of five occurrences and a minimum range of three texts to identify recurrent four-word sequences. However, raw frequency alone may favor sequences from the larger corpus.

Therefore, normalized frequencies were also calculated to allow a more balanced comparison between the two corpora. Range was retained as an additional criterion because it indicates how widely an RMS is distributed across the students' texts rather than whether it occurs repeatedly in only one or a few texts. The selected criteria were therefore used as identification criteria, while normalized frequency and distribution across texts were considered when interpreting differences between the two corpora.The researcher analyzes the frequency (how often the sequences appear) and range (how widely they are distributed across students' texts) in two groups: third-semester students (2020 academic year) and fifth-semester students (2021 academic year). This analysis helps to identify the most common RMS used by students and to compare their usage across different academic levels.

The second layer focuses on examining how these identified RMS are used in terms of syntactic structure and function. In this study, syntactic structure is limited to noun phrase patterns, while function is focused on stance expressions. This study focuses on noun phrases and verb phrases in terms of syntactic structure, as well as stance expressions and discourse organizers in terms of function. These categories were selected because they appeared most frequently in the students' writing texts that will be analyzed in this study. Noun phrases and verb phrases help show how students construct grammatical patterns and develop sentence complexity, while stance expressions and discourse organizers show how students' express opinions, attitudes, and organize ideas in their writing. These functions are important because they help reveal students' writing styles and how they build meaning in their texts.

Although other structures and functions were also found in the analysis, this study only focuses on these four categories to keep the discussion more focused and systematic. Analyzing all RMS categories at once would make the study too broad and difficult to discuss in depth. By limiting the analysis to the most frequently used structures and functions, the researcher can provide a clearer description of how students use RMS in their writing. Through this analysis, the study aims to compare whether third-semester or fifth-semester students use RMS in a more complex, organized, and formal way.

Source of Data

The data of this study consist of a corpus of students' writing texts of third- and fifth-semester students of ELESP, which were collected from two separated corpora:

1. Corpora 1: EFL students in the third semester (2020 academic year). Data collected from these corpora consisted of written assignments.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

2. Corpora 2: EFL students in their fifth semester (2021 academic year). Data collected from these consisted of written assignments.

The two learner corpora were constructed from students' writing produced in their respective academic courses and writing tasks. Therefore, differences in course, task, topic, and genre may influence the recurrent language patterns found in each corpus. For example, topic-specific assignments may encourage students to repeatedly use terminology related to a particular subject, whereas more general academic tasks may encourage broader academic expressions. Consequently, differences between the two corpora are interpreted as differences in observed RMS patterns within the respective corpus contexts rather than as evidence that semester level alone caused the differences. This consideration is particularly important because the study uses a cross-sectional comparison of two different student cohorts.

3. Result

The findings are presented in a systematic way to answer the two research questions. First, this chapter provides an overview of the RMS identified in the corpus. It then describes how the identified RMS are used in terms of syntactic structure, particularly noun phrase and verb phrase patterns. Finally, it explains how the RMS function as stance expressions and discourse organizers in students' writing. The discussion of these findings is presented in the following sections.

Coding Validation

To improve the reliability of the qualitative classification, the identified RMS were examined through concordance lines and their surrounding contexts. The classification of syntactic structure and function was based on the definitions and categories provided by the selected theoretical framework. The researcher first classified the RMS independently and then rechecked the classifications against their concordance contexts. Ambiguous cases were reviewed by comparing the sequence with the definitions and structural examples in the framework. This process was used to reduce inconsistent classification and ensure that the final categories reflected the actual use of the RMS in the students' texts.

Overview of RMS Identified in the Corpus

The Recurrent Multi-word Sequences (RMS) were extracted using the N-Gram feature in AntConc (version 4.4.0). The analysis focused on four-word sequences. Because the two corpora were different in size, the extraction settings were adjusted for each group. Using the same extraction criteria for corpora of different sizes could produce unreliable results, such as identifying too many unimportant sequences in the smaller corpus or missing important patterns in the larger corpus. Therefore, the extraction parameters were adapted to match the size of each corpus, following the framework of Biber et al. (2004). Unlike Wray (2002),

who focuses more broadly on formulaic language as prefabricated sequences, Biber provides specific corpus-based criteria such as frequency and range for identifying RMS. Therefore, Biber's framework was more suitable for this study because it supports the systematic identification and comparison of RMS across students' texts.

For the Class of 2020, the corpus contained 98 texts with a total of 86.15 occurrences per 10,000 words. A minimum frequency of 5 occurrences and a minimum range of 3 texts were used to identify the RMS. Meanwhile, the Class of 2021 corpus consisted of 45 texts with 23.49 occurrences per 10,000 words. Since this corpus was smaller, the same minimum frequency of 5 occurrences and a minimum range of 3 texts were applied to ensure that enough recurrent sequences could be identified for further analysis.

After the extraction process, the identified RMS were classified into two categories. First, they were analyzed based on their syntactic structure, namely Noun Phrase and Verb Phrase. Second, they were classified according to their discourse function as Discourse Organizers or Stance Expressions. Before discussing these categories in detail, it is important to provide an overview of the total number of RMS identified in each corpus. This overview helps explain the amount of data analyzed and provides a foundation for interpreting the structural and functional patterns discussed in the following sections. Table 3.1 presents the corpus information and the total number of RMS extracted from each group before the classification process.

Table 3.1. Overview of Corpus Specifications and Total RMS Identified per Cohort

Group	**Number of Texts**	**Total Tokens**	**N-Gram Size**	**Min. Co-occurrence**	**Min. Range**	**Total RMS**
Class of 2020 (3rd Semester)	98	88,576	4 words	5	3	78
Class of 2021 (5th Semester)	45	25,969	4 words	5	3	18

As shown in Table 3.1, Class of 2020 yielded a considerably larger number of RMS (78 sequences) compared to Class of 2021 (18 sequences), a difference that is largely attributable to the substantial difference in corpus size between the two groups 88,576 tokens versus 25,969 tokens, a ratio of more than three to one. The higher token count and greater number of texts in the 2020 corpus naturally produce a wider range of recurrent sequences at the specified thresholds, since a larger corpus offers more opportunities for the same four-word combination to recur across different texts and writers. This difference in raw output does not by itself indicate a difference in writing quality or sophistication between the two cohorts; rather, it sets the scale against which the structural and functional distributions discussed in the following sections must be read, since proportions

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

and percentages rather than raw counts provide the more meaningful basis for cross-cohort comparison.

In addition to frequency, each of the RMS tables presented in the following sections also reports the range of every sequence, that is, the number of distinct texts in which the bundle occurs at least once, as opposed to the total number of times it occurs across the corpus as a whole. Range is a standard companion measure to frequency in corpus-based lexical bundle research (Biber et al., 2004), because frequency alone cannot distinguish between a sequence that is genuinely widespread across many student writers and one that is highly frequent only because it happens to be repeated many times within a small number of individual texts.

A bundle with a high frequency but a low range, for instance, may reflect one or two writers' idiosyncratic phrasing rather than a pattern shared across the cohort, whereas a bundle with a comparatively high range relative to its frequency indicates that the sequence is used consistently by a broad cross-section of students. Reporting range alongside frequency therefore allows the following tables to be read not only for how often a sequence recurs, but also for how representative that recurrence is of the cohort as a whole. With this overview established, the following sections turn to a detailed analysis of each group's RMS, first in terms of syntactic structure and then in terms of discourse function.

Syntactic Structure of RMS in Students' Writing

After identifying the Recurrent Multi-word Sequences (RMS) in the corpus, the next step was to analyze their syntactic structure. This analysis answers the first research question by examining how students use RMS as grammatical patterns in their writing. Following the framework of Biber et al. (2004), the identified RMS were classified based on their syntactic structure. In this study, the analysis focuses on two structural categories: noun phrase patterns and verb phrase patterns. These categories were selected because they were the most frequently found in the students' writing corpus. After classifying the identified RMSs into noun phrase and verb phrase patterns, the distribution of these syntactic structures for each cohort is presented in the following figures. The charts provide a visual comparison of the frequency of each syntactic pattern in the students' writing.

Figure 3.1. Distribution of Syntactic Structures in RMS occurrences/frequency of the Class of 2020

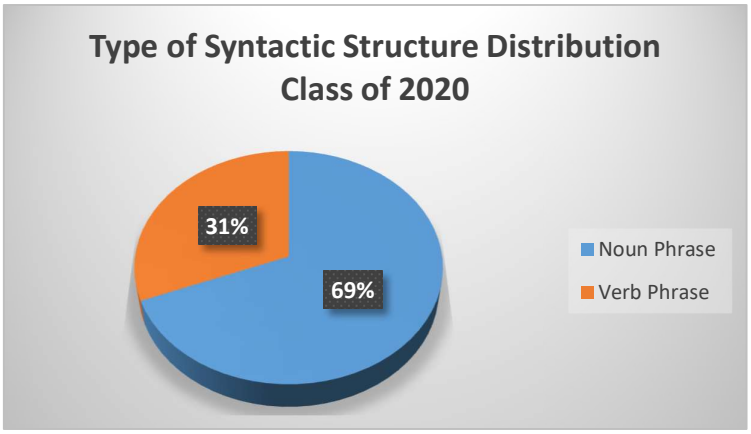
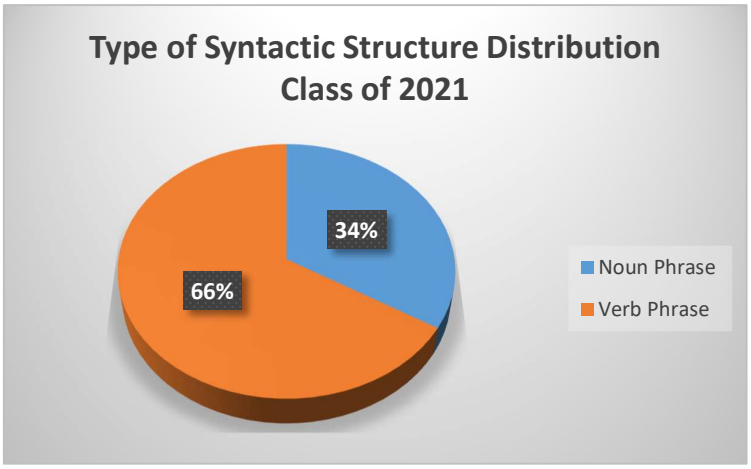


Figure 3.2. Distribution of Syntactic Structures in RMS occurrences/frequency of the Class of 2021



As shown in Figure 3.1, the Class of 2020 demonstrates a greater use of noun phrase structures, accounting for 62% of the occurrences/frequency identified RMSs, while verb phrase structures represent 38%. This indicates that noun phrase constructions were more frequently employed by students in this cohort. The dominance of noun phrases suggests that the students tended to construct recurring multiword sequences centered around nominal expressions, which are commonly used to present concepts, objects, or abstract ideas in academic writing.

In contrast, Figure 3.2 shows that the Class of 2021 exhibits a different pattern. Verb phrase structures account for 66% of the occurrences/frequency identified RMS, whereas noun phrase structures make up only 34%. This suggests that students in this cohort relied more heavily on verb phrase constructions than noun phrase constructions. The higher proportion of verb phrases indicates a greater tendency to use recurring expressions centered on actions or processes, reflecting a different pattern of syntactic preference compared to the Class of 2020.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

a) Noun Phrase Patterns

Noun Phrase (NP) - structured RMS represents sequences in which the main grammatical element is a noun. These sequences usually consist of a head noun together with other grammatical elements, such as determiners, adjectives, prepositional phrases, or other modifiers that provide additional information. In academic writing, NP-structured RMS help writers introduce, explain, and develop important ideas in a clear and organized way. The frequent use of these sequences in students' writing is often considered a sign that students are developing their ability to produce more complex noun phrases, which is an important feature of academic writing. It also shows that they are becoming more frequent at organizing information into clear and meaningful units (Biber et al., 2004; Yang et al., 2022).

In Class of 2020, a total of 52 NP-structured RMS were identified, with a combined frequency of 671 occurrences across the corpus, a figure that already signals how central nominal phrasing is to this cohort's writing. These sequences reflect the topical and conceptual vocabulary of the students' writing assignments, which centred heavily on themes of children's literature, creative writing, and educational contexts for students with special needs.

Representative NP bundles include in the form of (f=69), children s literature is (f=32), of children s literature (f=28), literature and creative writing (f=26), and the results of the (f=19). Read together, these five bundles alone trace the shape of the assignment: institutional or procedural description (in the form of), disciplinary definition (children s literature is, of children s literature), thematic linkage (literature and creative writing), and empirical reporting (the results of the).

The high frequency of these sequences suggests that students in this cohort frequently used topic-related noun phrases in their writing. This indicates that they were familiar with the key concepts and vocabulary commonly used in their field of study. Table 3.2 presents the complete list of NP-structured RMS identified in Class of 2020, ranked by frequency, so that the relative weight of each construction within the cohort's writing can be easily identified.

Table 3.2. Frequency Distribution of Noun Phrase-Structured RMS in the Class of 2020 Corpus

Recurrent Multi-word Sequence	**Frequency**	**Range**
in the form of	69	26
children s literature is	32	8
of children s literature	28	7
literature and creative writing	26	10
president of the country	25	9
blind junior high school	22	3
creative writing is a	20	11

Recurrent Multi-word Sequence	Frequency	Range
the results of the	19	10
in the learning process	18	9
students with special needs	18	5
the blind junior high	18	3
between literature and creative	15	9
in junior high school	15	4
relationship between literature and	15	9
children s literature in	14	7
creative writing and literature	14	10
from the results of	14	10
the form of a	13	9
english teacher at the	12	5
in the process of	11	8
of the blind junior	11	3
students in junior high	11	3
literature in teaching english	10	4
of writing this paper	10	5
the use of the	10	6
on the right side	9	6
one of the most	9	9
the use of language	9	4
this type of literature	9	3
at the same time	8	5
children s literature has	8	4
in the world of	8	8
the end of the	8	8
the relationship between literature	8	5
at the age of	7	6
at the end of	7	6
children literature in teaching	7	5
children s literature and	7	5
children s literature can	7	5
during the learning process	7	4
function of children literature	7	6
in terms of language	7	5
in the context of	7	4
information related to the	7	7
of children literature in	7	5
on the results of	7	7
results of the research	7	3

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Recurrent Multi-word Sequence	Frequency	Range
s literature in the	7	5
the color of the	7	4
the english teacher at	7	3
with special needs in	7	5
an analysis of the	6	3

Beyond the raw counts in Table 3.2, it is instructive to see how the two most frequent bundles actually surface in running text, since frequency alone cannot show whether a sequence is used mechanically or meaningfully. The following concordance lines, drawn directly from the student texts, illustrate how in the form of and the results of the the two highest-ranking NP bundles in this cohort appear in authentic writing contexts:

"...Institution also has teaching staff in the form of teachers who are competent in educating students with strategies tailored to students same as schools in general..."

"... The general description of children's literature is that children's literature is specially made for children's reading..."

These two examples show that the NP bundles used by the Class of 2020 mainly help students introduce and explain concepts related to their field of study. For example, in the form of is used to describe the specific form of an institutional provision, while the results of the is used to present the findings of the students' own research. Most of these sequences have a referential function because they help writers identify and describe important ideas or information. This pattern reflects the descriptive and explanatory nature of the writing tasks completed by this group. After examining the NP bundles used by the Class of 2020, the next section explores whether students in the Class of 2021 show a similar pattern or use different types of NP bundles in their writing.

In Class of 2021, a total of 7 NP-structured RMS were identified, with a combined frequency of 32 occurrences. While the absolute count is lower than that of Class of 2020 a difference attributable to the smaller corpus size and higher extraction threshold the NP patterns identified in this group demonstrate a greater degree of structural variety and thematic generalizability. Sequences such *as one of the most* (f=7), *for a long time* (f=5), *an example of a* (f=4), and *the purpose of this* (f=4) are not topic-bound to a single subject area but rather represent general academic NP constructions applicable across a range of argumentative and analytical writing contexts. Table 4.3 presents the NP-structured RMS identified in Class of 2021.

Table 3.3. Frequency Distribution of Noun Phrase-Structured RMS in the Class of 2021 Corpus

Recurrent Multi-word Sequence	Frequency	Range
one of the most	7	6

Recurrent Multi-word Sequence	Frequency	Range
for a long time	5	4
an example of a	4	3
country in the world	4	3
country is the most	4	3
in the form of	4	4
the purpose of this	4	4

As with Class of 2020, it is worth grounding these frequencies in actual usage. Concordance lines from Class of 2021 illustrate the deployment of the two most frequent NP sequences, one of the most and the purpose of this, in context:
"...And there are many festivals and traditions in Chinese culture. One of the most important is the Lunar New Year..."
"...The purpose of this essay is to discuss both perspectives and will continue to the conclusion..."

Unlike the examples from the Class of 2020, which refer to specific concepts or topics, these two concordance lines show that one of the most and the purpose of this are used as more general academic expressions. The phrase one of the most is used to introduce or compare items within a broad category, while the purpose of this is commonly used to introduce the main objective or argument of a piece of writing. These expressions are not limited to one specific topic and can be used in many different academic contexts. This suggests that fifth-semester students are more familiar with general academic writing patterns and are able to use them to organize and present their ideas effectively. In contrast, third-semester students tend to use noun phrase patterns that are more closely related to the specific topic of their writing assignment. After discussing the noun phrase patterns used by both groups, the next section examines the second syntactic category, namely Verb Phrase patterns.

b) Verb Phrase Patterns

Verb Phrase (VP)-structured RMS represents sequences in which the main grammatical element is a verb. These sequences may include main verbs, modal verbs, linking verbs, and other elements that complete the meaning of the verb. In academic writing, VP-structured RMS help writers express opinions, explain ideas, describe processes, and present logical reasoning. Therefore, examining these sequences helps show how students use verb patterns to communicate their ideas clearly and appropriately in an academic context. It also provides an understanding of their ability to use different types of verb constructions in their writing. (Biber et al., 2004).

Class of 2020, 26 VP-structured RMS were identified, with a combined frequency of 302 occurrences. The most frequent sequences include *is one of the* (f=25*), it can be concluded* (f=21), *can be concluded that* (f=20), *to be able to* (f=16),

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

and *there is also a* (f=16). The high frequency of it can be concluded and can be concluded that, which appeared 41 times in total, shows that students frequently used common expressions to write conclusions in their academic texts. This suggests that they were already familiar with one of the common conventions of academic writing, even at this stage of their study. In addition, the frequent use of being able to see (f=15) and not being able to (f=15) reflects the topics discussed in their writing. These expressions were commonly used because many of the writing assignments focused on issues related to visual impairment, particularly the experiences and abilities of students with visual disabilities. Table 3.4 presents the complete list of VP-structured RMS for Class of 2020.

Table 3.4. Frequency Distribution of Verb Phrase-Structured RMS in the Class of 2020 Corpus

Recurrent Multi-word Sequence	**Frequency**	**Range**
is one of the	25	19
it can be concluded	21	13
can be concluded that	20	13
is a type of	16	7
there is also a	16	8
to be able to	16	13
being able to see	15	3
is a form of	15	8
not being able to	15	3
on the other hand	14	10
be in the form	10	6
above it can be	9	6
as well as a	9	5
that can be used	9	6
writing is a form	9	5
and creative writing are	8	5
in accordance with the	8	6
is expected to be	8	8
one of which is	8	8
teach students with special	8	5
there are still many	8	5
based on the data	7	5
be seen from the	7	6
it can be seen	7	6
teaching english for young	7	5
teaching english to young	7	3

To illustrate how these bundles related to conclusions and ability

constructions actually function in student writing, the following concordance lines show it can be concluded and not being able to, two of the cohort's most frequent VP bundles, in their original textual context:

"...Based on the discussion presented in the previous sections, it can be concluded that literature for children with special needs must be adapted to accommodate their sensory and cognitive needs..."

"...Thus, being blind means not being able to see in the conventional sense, though this does not equate to a complete absence of perceptual experience..."

These examples show that the VP bundles used by the Class of 2020 mainly serve two purposes. First, they help students connect ideas and draw conclusions in their writing. Second, they are used to discuss the specific topics given in the writing assignments. The frequent use of it can be concluded and can be concluded that in many students' texts shows that they commonly relied on these expressions to write conclusions. Although this indicates that students were familiar with common academic writing conventions, it also suggests that they depended on the same expressions rather than using a wider variety of verb phrase patterns. The next section examines whether the Class of 2021 students show a similar pattern or use a more varied range of VP bundles in their writing.

In Class of 2021, 11 VP-structured RMS were identified, with a combined frequency of 63 occurrences significantly higher in proportion relative to the group's NP count than observed in Class of 2020. The most frequent VP bundles are *is one of the* (f=14), *on the other hand* (f=8), *as well as the* (f=6), and *to be able to* (f=5). A notable feature of the Class of 2021 VP set is the presence of evaluative and epistemic sequences such as is the most important (f=4), *it is also possible* (f=5), and *make it difficult for* (f=4), which encode the writer's assessment and judgment regarding the propositions being presented. This suggests a greater deployment of modalizing and evaluative VP structures in fifth-semester writing, indicative of more complex argumentative positioning. Table 3.5 presents the VP-structured RMS identified in Class of 2021.

Table 3.5. Frequency Distribution of Verb Phrase-Structured RMS in the Class of 2021 Corpus

Recurrent Multi-word Sequence	**Frequency**	**Range**
is one of the	14	10
on the other hand	8	8
as well as the	6	4
is an example of	5	3
it is also possible	5	3
to be able to	5	4
can be concluded that	4	4
have the right to	4	3
is the most important	4	4
it can be concluded	4	4

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Recurrent Multi-word Sequence	Frequency	Range
make it difficult for	4	3

Concordance lines from Class of 2021 illustrate the richer evaluative and contrastive functions of two of this cohort's characteristic VP bundles, on the other hand and it is also possible:

"...On the other hand, country has a unique tradition, namely drinking soju, or drinking liquor; this tradition has long been known..."

"...However, it is also possible for values to contradict and conflict with each other, especially in complex modern industrial societies..."

Where the Class of 2020 examples above show it can be concluded closing off an argument and not being able to describing a fixed physical condition, these Class of 2021 examples show on the other hand actively staging a contrast between two positions and it is also possible hedging a claim about competing values. These concordance lines demonstrate that fifth-semester students employ VP bundles not only to frame conclusions but also to introduce contrasting perspectives, evaluate claims, and signal epistemic possibility functions that reflect a more widely distributed command of argumentative discourse conventions than those observed in the Class of 2020 VP patterns.

Having examined the NP and VP patterns separately for each cohort, it is now possible to bring the two structural categories together and compare their relative weight within each group. A comparison of the NP and VP patterns across both cohorts reveals a clear structural divergence between the two groups. In Class of 2020, NP-structured sequences were substantially dominant, constituting 52 out of 78 total RMS (66.7%) and occurring with a combined frequency of 671 more than twice the total frequency of VP sequences (302, 33.3%). This strong dominance of NP patterns suggests that third-semester students mainly organize their writing by using noun-based structures. They rely on repeated noun phrases to connect ideas and maintain the main topic throughout their writing.

In contrast, Class of 2021 exhibited the opposite structural tendency: VP-structured sequences were dominant, accounting for 11 out of 18 total RMS (61.1%) with a combined frequency of 63 (66.3%), compared to only 7 NP sequences (38.9%, frequency 32, 33.7%). The dominance of VP patterns in the fifth-semester group suggests that these students tend to use more verb-based structures in their writing. This shows that they are more frequent able to express logical relationships, present their opinions or arguments, and explain their ideas through a wider range of verb constructions. Table 3.6 provides an overall structural comparison between the two cohorts.

Table 3.6. Comparative Distribution of Noun Phrase and Verb Phrase Structures Across Class of 2020 and Class of 2021

Structure	Class of 2020 — No. of RMS	Class of 2020 — Total Freq.	Class of 2020 — %	Class of 2021 — No. of RMS	Class of 2021 — Total Freq.	Class of 2021 — %
Noun Phrase (NP)	52	671	66.7%	7	32	33.7%
Verb Phrase (VP)	26	302	33.3%	11	63	66.3%
Total	**78**	**973**	**100%**	**18**	**95**	**100%**

As shown in Table 4.6, the structural profiles of the two cohorts are markedly different. Class of 2020 demonstrates an NP-dominated pattern consistent at an early stage of academic writing development, where students mainly use noun-based structures to present their ideas rather than more complex sentence structures. Class of 2021, by contrast, shows a VP-dominated profile that aligns with a more advanced stage of syntactic development, suggesting that fifth-semester students are more capable of constructing verb-centred arguments and employing modal and inferential constructions to advance their positions. These findings carry implications for the evidence of students' phraseological behavior in writing, particularly in relation to recurrent syntactic patterns and discourse functions and will be discussed further in Section 4.2.1. Having addressed the first research question in full, the discussion now proceeds to the second dimension of analysis: the discourse functions that RMS perform in students' writing.

Functional Categories of RMS in Students' Writing

The second research question examines how students use Recurrent Multi-word Sequences in terms of their discourse functions in writing. Following the functional taxonomy adopted from Biber et al. (2004) as adapted in this study, all identified RMS were classified into two functional categories: Stance Expressions (SE) and Discourse Organizers (DO). Stance Expressions are sequences that encode the writer's epistemic, attitudinal, or evaluative position toward the information being presented. Discourse Organizers, on the other hand, are sequences that primarily serve to structure the flow of information, connect textual units, or guide the reader through the logical organization of the text. Where the structural analysis in Section 3.2 asked what grammatical form students' RMS take, this functional analysis asks what communicative work those sequences are doing a distinction that is essential, since a bundle's syntactic category does not by itself determine its rhetorical role. After identifying the functional categories of the identified RMSs, the distribution of discourse organizers and stance expressions

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

for each cohort is presented in the following figures. These charts provide a visual comparison of the functional patterns found in the students' writing.

Figure 3.3. Distribution of Functional Categories of RMS occurrences/frequency in the Class of 2020

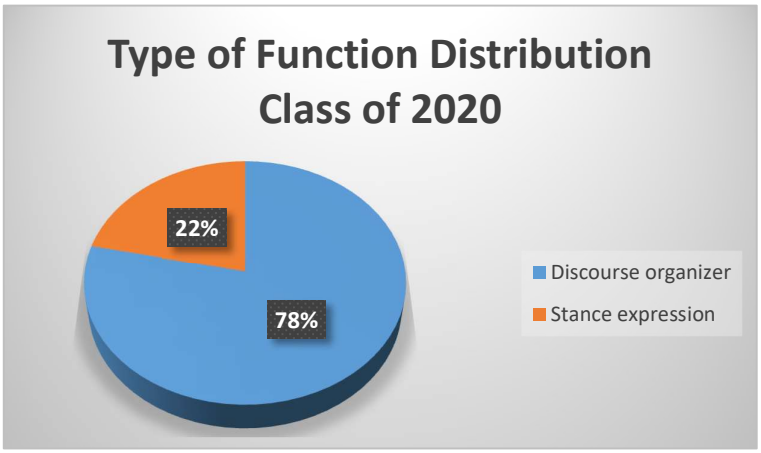
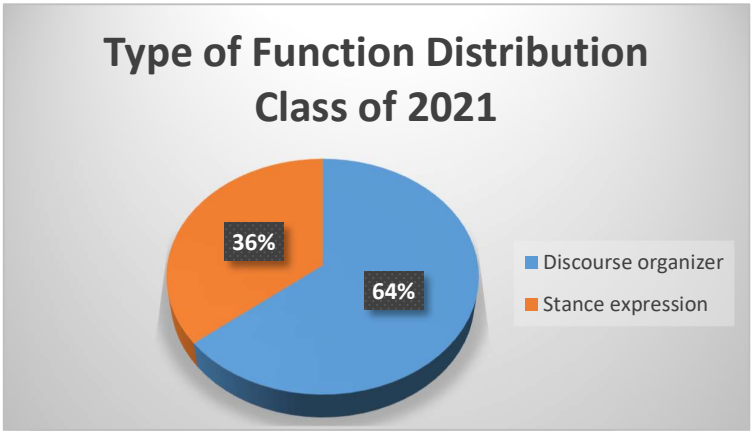


Figure 3.4. Distribution of Functional Categories of RMS occurrences/frequency in the Class of 2021



As illustrated in Figure 3.3, discourse organizers account for 76% of the identified RMSs in the Class of 2020, whereas stance expressions represent only 24%. This indicates that discourse organizers were used considerably more frequently than stance expressions in this cohort. The predominance of discourse organizers suggests that the students primarily employed recurring multiword sequences to organize ideas, connect information, and structure the flow of their writing, while stance expressions were used less frequently to convey personal attitudes or evaluations.

As shown in Figure 3.4, the Class of 2021 also demonstrates a higher proportion of discourse organizers, which account for 64% of the occurrences/frequency identified RMS. In comparison, stance expressions

comprise 36% of the total. Although discourse organizers remain the dominant functional category, the proportion of stance expressions is relatively higher than that of the Class of 2020. Overall, the results indicate that students in both cohorts relied more frequently on discourse organizers to structure and connect ideas in their academic writing than on stance expressions to express evaluation or personal perspective.

a) Stance Expressions

Stance Expressions are multi-word sequences that signal the writer's subjective positioning toward the content of their writing. They may encode certainty, possibility, necessity, evaluation, or attitudinal commitment, and are therefore central to the academic practice of argumentation. Research in academic discourse has consistently demonstrated that the ability to deploy stance expressions appropriately is a key marker of writing development in EFL contexts, as it indicates the writer's growing capacity to move beyond descriptive prose toward evaluative and argumentative writing (Biber et al., 2004; Cortes, 2013).

In Class of 2020, 16 stance expression sequences were identified, with a combined frequency of 210 occurrences. The most frequent SE bundles include *children s literature is* (f=32), *it can be concluded* (f=21), *can be concluded that* (f=20), *to be able to* (f=16), *and being able to see* (f=15). The high frequency of *children s literature is* shows that students in this cohort often focused on the main topic of their writing assignments. They frequently used this noun phrase to express their ideas and opinions about the nature and importance of children's literature. In addition, the frequent use of *it can be concluded* and *can be concluded that* shows that students were familiar with common expressions used to write conclusions in academic writing. However, they tended to rely on the same expressions repeatedly instead of using a wider variety of phrases to express their conclusions or viewpoints. Table 3.7 presents the complete list of SE-functioning RMS for Class of 2020.

Table 3.7. Frequency Distribution of Stance Expression RMS in the Class of 2020 Corpus

Recurrent Multi-word Sequence	**Frequency**	**Range**
children s literature is	32	8
it can be concluded	21	13
can be concluded that	20	13
there is also a	16	8
to be able to	16	13
being able to see	15	3
not being able to	15	3
of writing this paper	10	5
above it can be	9	6
that can be used	9	6

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Recurrent Multi-word Sequence	Frequency	Range
this type of literature	9	3
children s literature has	8	4
is expected to be	8	8
there are still many	8	5
children s literature can	7	5
it can be seen	7	6

The following concordance lines illustrate how children s literature is and can be concluded that, the two dominant SE bundles identified above actually function in the writing of Class of 2020 students:

"...Children s literature is often considered a simple genre, yet it plays a crucial role in shaping the moral and emotional development of young readers..."

"...From the discussion of the data presented above, it can be concluded that the adaptation of literary texts for students with special needs requires careful attention to both content and format..."

These examples show that SE bundles in Class of 2020 are primarily deployed to assert evaluative propositions about their essay topics and to signal inferential conclusions: *children s literature is* opens a generalized value-claim about the subject matter, while *can be concluded that* closes an argument built on the evidence discussed earlier in the text. The frequent recurrence of such formulaic frames suggests that while students demonstrate an awareness of academic stance conventions, their range of stance-marking strategies remains somewhat limited and formulaic. This raises the question of whether the same limitation holds for the more advanced cohort, which the following paragraphs address.

In Class of 2021, 8 stance expression sequences were identified, with a combined frequency of 34 occurrences. The SE bundles identified in this group — including *it is also possible* (f=5), *to be able to* (f=5), *can be concluded that* (f=4), *have the right to* (f=4), *is the most important* (f=4), and *make it difficult for* (f=4) demonstrates that fifth-semester students use a wider and more varied range of stance expression strategies than the students in the Class of 2020. Instead of relying on only a few common expressions, they use different stance expressions to communicate their ideas more effectively. These expressions are used to show possibility (it is also possible), make evaluations (is the most important), and express rights or obligations (have the right to, make it difficult for). This suggests that fifth-semester students are able to use a greater variety of stance expressions in their academic writing. Table 3.8 presents the SE-functioning RMS identified in Class of 2021.

Table 3.8. Frequency Distribution of Stance Expression RMS in the Class of 2021 Corpus

Recurrent Multi-word Sequence	Frequency	Range
it is also possible	5	3
to be able to	5	4
can be concluded that	4	4
country is the most	4	3
have the right to	4	3
is the most important	4	4
it can be concluded	4	4
make it difficult for	4	3

Concordance lines from Class of 2021 illustrate the more varied evaluative functions of two representative SE bundles, to be able to and make it difficult for, in this cohort's writing:

"...Having good and correct language skills can encourage a child to be able to continue reading books independently as they grow older..."

"...However, unnecessary changes to the register of a text will make it difficult for readers to follow the flow of the dialogue..."

Unlike the conclusion-related and topic-specific expressions SE bundles typical of Class of 2020, *to be able to* here frames a developmental capability and *make it difficult for* expresses a negative evaluative judgment about a specific writing choice both markedly more analytical uses of stance. These concordance lines reflect a more argumentatively engaged and evaluatively widely distributed deployment of stance expressions among fifth-semester students, suggesting greater competence in using multi-word sequences to convey intellectual positioning and critical judgment. Having examined Stance Expressions in both cohorts, the discussion now turns to the second functional category, Discourse Organizers.

b) Discourse Organizers

Discourse Organizers are multi-word sequences that help organize ideas and connect different parts of a text. They guide readers by showing how information is related and how the text is structured. These sequences can be used to introduce topics, show transitions, indicate order or contrast, refer to previous information, and connect sentences or paragraphs. In students' writing, the use of Discourse Organizers reflects their ability to organize ideas clearly and logically. A wider and more appropriate use of these sequences generally indicates that students are able to produce well-structured and coherent academic writing that is easy for readers to follow (Halliday & Hasan, 1976; Biber et al., 2004).

In Class of 2020, 62 discourse organizer sequences were identified, with a combined frequency of 763 occurrences, making this the dominant functional category for this cohort by a wide margin. The most frequent DO bundles include *in the form of* (f=69), *of children s literature* (f=28), *literature and creative writing*

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

(f=26), *is one of the* (f=25), and *the results of the* (f=19). The sheer number and frequency of DO sequences in Class of 2020 suggests that students rely heavily on multi-word sequences to anchor their writing to specific topics and to structure information within and across paragraphs. However, the topically bound nature of many of these sequences such as *blind junior high school* (f=22*), the blind junior high* (f=18), and *students with special needs* (f=18) indicates that a significant portion of the discourse organizing function is driven by subject-specific vocabulary repetition rather than generalizable academic organizational patterns. Table 3.9 presents a selection of the most representative DO-functioning RMS identified in Class of 2020.

Table 3.9. Frequency Distribution of Discourse Organizer RMS in the Class of 2020 Corpus

Recurrent Multi-word Sequence	**Frequency**	**Range**
in the form of	69	26
of children s literature	28	7
literature and creative writing	26	10
is one of the	25	19
president of the country	25	9
blind junior high school	22	3
creative writing is a	20	11
the results of the	19	10
in the learning process	18	9
students with special needs	18	5
the blind junior high	18	3
is a type of	16	7
between literature and creative	15	9
in junior high school	15	4
is a form of	15	8
relationship between literature and	15	9
children s literature in	14	7
creative writing and literature	14	10
from the results of	14	10
on the other hand	14	10
the form of a	13	9
english teacher at the	12	5
in the process of	11	8
of the blind junior	11	3
students in junior high	11	3
be in the form	10	6
literature in teaching english	10	4
the use of the	10	6

Recurrent Multi-word Sequence	Frequency	Range
as well as a	9	5
on the right side	9	6

To ground this pattern in authentic text, the following concordance lines show in the process of and the results of the — a topic-bound and a more transitional DO bundle, respectively — in their original context:

"...In the process of teaching and learning, especially in the development of children's vocabulary, repetition and visual aids play a central role..."

"...The writers also realize that there are still many mistakes in the process of writing this paper, particularly with respect to presenting the results of the analysis in a fully systematic manner..."

These examples show that the DO bundles used by the Class of 2020 have two main functions. First, they help students keep the discussion focused on the main topic, as shown by in the process of. Second, they help introduce or connect information, as seen in the results of the, which is used to present the students' research findings. This pattern is consistent with the frequent use of on the other hand (f=14), which helps students show contrast between ideas. However, the high frequency of topic-related sequences suggests that students mainly organize their writing by referring to the specific topic of the assignment, rather than using a wider range of general academic expressions to connect and develop their ideas. The next section examines whether the same pattern is found in the more advanced cohort.

In the Class of 2021, 10 discourse organizer sequences were identified, with a combined frequency of 61 occurrences. The DO bundles used by this group, including is one of the (f=14), on the other hand (f=8), one of the most (f=7), as well as the (f=6), and the purpose of this (f=4), are more general and more typical of academic writing than those found in the Class of 2020. Instead of being closely related to one specific topic, these sequences can be used in many different writing tasks and academic contexts. For example, the purpose of this is commonly used to introduce the aim of a piece of writing, on the other hand is used to show contrast between ideas, and as well as the is used to add supporting information within an argument. Table 3.10 presents the DO-functioning RMS identified in Class of 2021.

Table 3.10. Frequency Distribution of Discourse Organizer RMS in the Class of 2021 Corpus

Recurrent Multi-word Sequence	Frequency	Range
is one of the	14	10
on the other hand	8	8
one of the most	7	6
as well as the	6	4
for a long time	5	4
is an example of	5	3
an example of a	4	3

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Recurrent Multi-word Sequence	Frequency	Range
country in the world	4	3
in the form of	4	4
the purpose of this	4	4

Concordance lines from Class of 2021 illustrate the more conventionally academic discourse organizing functions performed by for a long time and is an example of, two representative DO bundles from this cohort:

"...This colonization lasted for a long time until it passed through World War II..."

"...The definition of status is a person's standing, position, or state in society. Middle class is an example of a person's financial status..."

Unlike the examples from the Class of 2020, which focus on specific topics, for a long time is used to describe a period of time that can be applied to many different topics, while is an example of is used to introduce a general example rather than information related to one specific subject. These concordance lines show that fifth-semester students use DO bundles not only to organize their writing but also to introduce essays, show contrast, add information, and describe time relationships. This suggests that they have developed a general understanding of how to organize ideas effectively in academic writing.

After discussing Stance Expressions and Discourse Organizers separately, the overall functional distribution of RMS can now be compared between the two cohorts. In the Class of 2020, Discourse Organizers were the most common functional category, with 62 out of 78 RMS (79.5%) and a total frequency of 763 occurrences. This number was much higher than the frequency of Stance Expressions (210, 20.5%).

The dominance of Discourse Organizers suggests that third-semester students mainly used these sequences to organize information and keep their writing focused on the main topic. In the Class of 2021, Discourse Organizers were also the largest category (10 bundles, 61 occurrences, 64.2%). However, the proportion of Stance Expressions was higher (8 bundles, 34 occurrences, 35.8%) than in the Class of 2020. This finding suggests that fifth-semester students were more able to express opinions, evaluations, and arguments while still using Discourse Organizers to organize their writing effectively. Table 3.11 provides an overall functional comparison between the two groups.

Table 3.11. Comparative Distribution of Stance Expression and Discourse Organizer Functions Across Class of 2020 and Class of 2021

Function	Class of 2020 — No. of RMS	Class of 2020 — Total Freq.	Class of 2020 — %	Class of 2021 — No. of RMS	Class of 2021 — Total Freq.	Class of 2021 — %
Stance Expression (SE)	16	210	20.5%	8	34	35.8%

Function	Class of 2020 — No. of RMS	Class of 2020 — Total Freq.	Class of 2020 — %	Class of 2021 — No. of RMS	Class of 2021 — Total Freq.	Class of 2021 — %
Discourse Organizer (DO)	62	763	79.5%	10	61	64.2%
Total	**78**	**973**	**100%**	**18**	**95**	**100%**

The data in Table 3.11 confirm that both cohorts rely more on DO-functioning sequences than on SE-functioning sequences, consistent with findings in previous corpus studies of EFL student writing that have noted the prevalence of referential and organizational bundles over stance-marking bundles in undergraduate writing (Cortes, 2013; Yang et al., 2022). However, the proportionally higher SE usage in Class of 2021 suggests a developmental progression toward greater argumentative sophistication. Having now presented the full set of findings for both research questions syntactic structure in Section 3.2 and discourse function in Section 3.3 the chapter turns next to a detailed discussion of what these patterns mean in light of existing theory and prior research.

4. Discussion

The findings indicate a difference in the distribution of syntactic and functional RMS patterns between the two cohorts. The discussion draws on foundational work in corpus linguistics (Biber et al., 2004), research on lexical bundles in academic discourse (Cortes, 2013; Yang et al., 2022), and the theoretical framework of cohesion in written language (Halliday & Hasan, 1976) to illuminate what the observed patterns reveal about the writing evidence of students' phraseological behavior in writing, particularly in relation to recurrent syntactic patterns and discourse functions. of students across the two cohorts.

Students' Use of RMS in Terms of Syntactic Structure

The findings show a clear difference between the two cohorts, class of 2020 mostly used Noun Phrase (NP) sequences, which made up 66.7% of their total RMS, while Class of 2021 mostly used Verb Phrase (VP) sequences, which made up 66.3% of their total RMS frequency. This difference is meaningful because syntax is the system that shows how learners combine words into complete and organized ideas (Ha, 2022; Yang et al., 2022). The following paragraphs explain this pattern in more depth, drawing on the theories of multi-word sequences and syntax.

The strong NP use in Class of 2020 fits Biber et al.'s (2004) description of referential bundles, which are sequences that mainly name and describe a topic rather than build an argument. In other words, Class of 2020 students tend to build their writing by "mentioning and explaining something." Their frequent use of noun phrases shows that they often construct their texts around a particular topic, object, concept, or piece of information. Many of these RMS are also closely

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

connected to the topics of their assignments, such as children's literature is and students with special needs. Therefore, these RMS do not simply appear as grammatical patterns; they help students maintain and explain the topic throughout their writing. This can also be explained through Nattinger and DeCarrico's (1994) idea that learners store and reuse ready-made phrases because doing so lowers the mental effort needed to build a sentence from scratch. Since Class of 2020 students are still in an earlier stage of academic writing, they likely lean on these ready-made noun phrases as a practical way to talk about their topic. This supports the view that formulaic phrases help reduce the cognitive load of writing, especially for less experienced writers (Wray, 2013).

This pattern becomes clearer when looking at the actual bundles students used. Sequences such as in the form of and children s literature is, the two most frequent NP bundles in this cohort, work mainly to name or point to something rather than to argue a point. This topic-centered pattern helps explain why NP has a higher percentage in Class of 2020. Rather than mainly using RMS to express what they think about the information, students tend to organize their writing by repeatedly referring to the concepts or information they are discussing. According to Yang et al. (2022), this kind of NP-heavy use is typical of descriptive and content-focused writing rather than analytical writing. This matches the nature of the assignments given to Class of 2020, which mostly asked students to describe and explain their topic rather than to argue a position.

In contrast, Class of 2021 showed VP dominance, with bundles such as it can be concluded, it is also possible, and to be able to appearing frequently. This suggests that Class of 2021 uses RMSs in a somewhat different way. Their verb phrases are not only used to mention or describe a topic, but also to "do something with the information." For example, these sequences allow students to express that something may happen, that something is important, or that something causes difficulty. They can also use expressions such as on the other hand to connect and compare ideas. Thus, their writing moves beyond simply presenting a topic and its information toward expressing a view about that information, showing its relationship with other ideas, and making an evaluation. These sequences do more than name a topic; they express conclusions, possibility, and ability, which are closer to argumentative and analytical language. This fits Cortes's (2013, 2015) observation that the more native-like and flexible use of recurring verbal sequences often marks writing that is more advanced. It also aligns with the broader view of syntax in EFL writing, where complex sentence structures that use subordination are considered more widely distributed than simple ones (Ha, 2022; Wang et al., 2024).

Looking closer at the VP bundles also shows a wider range of communicative purposes in Class of 2021. Bundles like is the most important and have the right to add an evaluative or judgmental tone, while it is also possible expresses possibility rather than certainty. Therefore, the higher percentage of VP in Class of 2021 can

be understood not simply as a difference in grammatical structure, but as a reflection of how students use RMSs to perform different communicative actions in their writing. Instead of mainly using recurrent expressions to maintain the topic, they increasingly use them to express relationships, evaluation, possibility, and personal positioning. This variety suggests that fifth-semester students are starting to use verb phrases not just to state facts but to reason and evaluate. Such use reflects a more developed command of the verbal constructions needed for academic argumentation (Biber et al., 2004; Cortes, 2013).

It is important to note that the number of VP sequences in Class of 2021 (11) is smaller than in Class of 2020 (26) in raw terms. However, this must be read alongside the very different corpus sizes of the two groups, as shown in Table 4.1. Because Class of 2021's corpus is much smaller, its 66.3% VP proportion still represents a genuine structural tendency rather than a result of having more data to work with. This is in line with the idea that frequency findings in corpus studies should always be interpreted in proportion to corpus size, not by raw counts alone (Biber et al., 2004).

Taken together, these findings suggest a general pattern of growth from NP-centered writing in the third semester to VP-centered writing in the fifth semester. This supports Wray's (2002, 2013) idea that as learners gain more experience, they do not abandon formulaic phrases altogether, but instead build a wider and more flexible set of them, including ones that carry argument and evaluation. For teaching, this suggests that introducing useful VP frames, such as it can be argued or this suggests that, earlier in students' academic writing training may help them develop these skills sooner rather than later. Having discussed the syntactic dimension of RQ1, the discussion now turns to the second research question, which looks at the discourse functions of RMS in students' writing.

Students' Use of RMS in Terms of Functions

The findings show that both cohorts relied more on Discourse Organizer (DO) sequences than on Stance Expression (SE) sequences, but Class of 2021 used a noticeably higher proportion of SE sequences (35.8%) than Class of 2020 (20.5%). This connects to the concept of cohesion, which refers to the grammatical and lexical ties that hold a text together as a unified whole (Halliday & Hasan, 1976). The following paragraphs explain what this pattern reveals about students' functional evidence, drawing on the cohesion and RMS theories.

The strong presence of DO sequences in both cohorts is consistent with earlier corpus-based studies on student academic writing (Biber et al., 2004; Cortes, 2013). Discourse organizers work to connect ideas and guide readers through a text, which is a central part of what makes writing cohesive (Halliday & Hasan, 1976; Hyland, 2008). Because both cohorts rely heavily on DO sequences, this shows that students at both levels already understand the basic need to organize and connect their ideas. However, the way this organizing function is carried out differs between the two groups, as the next paragraphs explain.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

In Class of 2020, many DO sequences are tied closely to the assignment topic, such as in the form of, of children's literature, and students with special needs. When we look more closely at these RMS, many of them remain strongly connected to particular topics. This means that Class of 2020 students do use RMS to organize information, but they often do so by repeating or maintaining information that is directly related to the topic of their writing. In other words, these sequences work more like topic vocabulary than general organizing tools. This matches Cortes's (2013) distinction between bundles that are specific to one subject area and those that can be used more broadly. The findings also suggest that part of this organizing function is driven more by the repetition of topic-specific vocabulary than by general academic organizational patterns. This also reflects a common issue in EFL cohesion, where learners may repeat the same lexical items too often instead of varying their cohesive devices (Ahmed, 2010). This suggests that third-semester students' ability to organize their writing still depends heavily on how familiar they are with the topic.

Class of 2021, on the other hand, used DO sequences that are more general and could apply to almost any topic, such as on the other hand, the purpose of this, and for a long time. This suggests that, although DO is still the most frequently used function in Class of 2021, the types of RMS they use begin to show a more general academic pattern. Rather than mainly repeating words from the topic being discussed, these students appear to use RMS as tools for organizing their arguments. These sequences work as genuine organizational tools rather than topic-specific vocabulary, since they can introduce contrast, state purpose, or mark time in any kind of academic text. For example, on the other hand can be used to compare ideas or introduce another point of view, the purpose of this can be used to explain a purpose, and is an example of can be used to introduce an example. Therefore, these RMS can be applied across different topics rather than being closely tied to one particular assignment. This fits Halliday and Hasan's (1976) view that using varied and appropriate cohesive devices is a sign of a more mature writing style. It also matches findings that texts with more varied cohesive devices tend to be judged as more coherent and of higher quality (Alawerdy & Alalwi, 2022; Quoc Tuan et al., 2023).

Turning to Stance Expressions, the higher proportion found in Class of 2021 is meaningful because stance markers show a writer's opinions, judgments, and level of certainty (Hyland, 2008). Bundles such as it is also possible, is the most important, and make it difficult for reflect the writer's evaluation of the content, not just a description of it. The difference becomes clearer when the actual stance expressions used by the two cohorts are examined. Class of 2020 frequently used expressions such as it can be concluded and can be concluded that. This shows that these students already know how to express a conclusion in an academic way, but they tend to rely on the same patterns repeatedly. In contrast, Class of 2021 used more varied expressions, such as it is also possible, is the most important, have the

right to, and make it difficult for. These expressions allow students to show different meanings, including possibility, evaluation, importance, rights, and judgment. Therefore, the difference is not simply that Class of 2021 uses more SE, but that their SE use shows a more varied way of positioning themselves toward the information they discuss. This kind of evaluative language is associated with a more developed academic voice and stronger argumentative skills (Biber et al., 2004). This suggests that fifth-semester students are beginning to move beyond describing ideas and are starting to comment on and judge them as well.

This does not mean Class of 2020 lacks stance awareness completely. The frequent use of it can be concluded and can be concluded that (f=21 and f=20) shows that these students already know how to signal a conclusion. However, the repeated use of these expressions also shows that their way of expressing stance is still relatively limited and formulaic. They have already learned a recognizable way of "speaking academically," but they have fewer recurring expressions for showing different types of evaluation or positioning. In contrast, the wider range of stance expressions in Class of 2021 suggests that these students are beginning to use recurrent language not only to state conclusions, but also to express possibility, evaluation, importance, and judgment. However, because these two phrases are repeated so often on their own, they function more like a fixed formula than a flexible stance tool. This matches the idea that overreliance on the same formulaic sequence, without variation, can limit how naturally and effectively a writer expresses different shades of meaning (Wray, 2002; Hyland, 2008).

Overall, these functional findings point to a similar pattern of growth as the structural findings discussed Fifth-semester students use a wider variety of stance expressions and more broadly applicable discourse organizers, while third-semester students rely more on repeated, topic-bound sequences. Both patterns reflect the same underlying development: as students gain more experience with academic writing, their use of RMS becomes less tied to memorized, topic-specific phrases and more flexible, evaluative, and generalizable. This has a practical implication for teaching, since explicitly noticing and practicing different RMS functions in class, rather than leaving them to be picked up naturally, can help students build this flexibility earlier (Cortes, 2013; Uçar, 2017).

5. Conclusion

This study concludes that students' use of Recurrent Multi-word Sequences (RMS) in academic writing develops from a more topic-centered and formulaic pattern toward a more flexible and analytical pattern across the two cohorts. Class of 2020 students relied more heavily on Noun Phrase (NP) sequences, which function mainly to introduce, describe, and maintain information related to the writing topic. Their RMS use reflects an earlier stage of academic writing development, where students depend on familiar and ready-made expressions to reduce cognitive effort and organize ideas. In contrast, Class of 2021 students showed greater use of Verb Phrase (VP) sequences, indicating a shift toward using

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

RMS to express relationships between ideas, conclusions, evaluations, and academic positioning. Although both cohorts used Discourse Organizer (DO) sequences as the dominant functional category, Class of 2021 demonstrated more varied and general academic organizing patterns, while Class of 2020 relied more on topic-specific expressions. The higher use and wider range of Stance Expression (SE) sequences among Class of 2021 further suggest that these students have developed a stronger ability to express judgment, possibility, importance, and evaluation in academic writing.

The study is limited to two cohorts of students from one educational context and focuses specifically on the use of RMS in academic writing. Therefore, the findings may not fully represent students from different institutions, disciplines, educational levels, or cultural backgrounds. In addition, the study analyzes the frequency, syntactic structure, and functional categories of RMS, but it does not investigate other possible factors influencing students' evidence of students' phraseological behavior in writing, particularly in relation to recurrent syntactic patterns and discourse functions., such as individual learning strategies, writing instruction, exposure to academic texts, or students' proficiency levels. Since the analysis is based on written products, it may not capture students' actual knowledge of RMS use in other academic communication contexts, such as speaking or classroom interaction.

Based on the findings, students are encouraged to expand their academic language resources by learning RMS not only as fixed expressions but also as flexible tools for organizing arguments, expressing stance, and developing academic voice. Writing instruction should provide explicit teaching of different RMS functions, including discourse organization, evaluation, and argument construction, rather than focusing only on grammatical accuracy. Lecturers are recommended to introduce commonly used academic phrase frames, provide opportunities for students to analyze authentic academic texts, and encourage variation in the use of formulaic expressions. Future researchers may conduct longitudinal, comparative, or mixed-method studies involving broader participants to examine how evidence of students' phraseological behavior in writing, particularly in relation to recurrent syntactic patterns and discourse functions.

The findings have implications for EFL academic writing instruction, particularly for corpus-informed vocabulary and phraseological teaching. Teachers can use learner-corpus examples to introduce students to recurrent academic expressions rather than teaching vocabulary only as individual words. Classroom activities can include identifying frequent four-word sequences, examining their concordance lines, and comparing how the same or similar sequences are used across different contexts. Such activities can help students recognize how recurring expressions contribute to syntactic construction, discourse organization, and the expression of stance.

The findings also suggest the importance of developing phraseological diversity in academic writing. Students can be exposed to different types of noun phrase and verb phrase sequences as well as discourse-organizing and stance expressions. Instead of requiring students to memorize fixed lists of expressions, teachers can encourage them to observe recurrent patterns in authentic academic texts and use them appropriately according to context. This approach can help students develop greater awareness of phraseological choices while avoiding the assumption that one RMS pattern is inherently better than another.

6. References

Ahmed, A. H. (2010). *Students' Problems with Cohesion and Coherence in EFL Essay Writing in Egypt: Different Perspectives*.

Alawerdy, A. S., & Alalwi, F. S. (2022). Enhancing English as a foreign language university students' writing through explicit instruction of conjunctions as cohesive devices: An experimental study. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1053310

Albelihi, H. (2022). *A corpus-based investigation of phrasal complexity in a corpus-based investigation of phrasal complexity in l1, l2 phd students, and expert writers in education l1, l2 phd students, and expert writers in education field field*. https://digitalrepository.unm.edu/educ_llss_etds/

Appel, R., Geluso, J., & Feng, H. H. (2024). An examination of phrase-frames in L2 english academic writing: Exploring relationships with writing quality. *System*, *123*. https://doi.org/10.1016/j.system.2024.103349

Casal, J. E., & Yoon, J. (2023). Frame-based formulaic features in L2 writing pedagogy: Variants, functions, and student writer perceptions in academic writing. *English for Specific Purposes*, *71*, 102–114. https://doi.org/10.1016/J.ESP.2023.03.004

Chen, Y.-H., & Baker, P. (2010). *LEXICAL BUNDLES IN L1 AND L2 ACADEMIC WRITING*. *14*(2), 30–49. http://llt.msu.edu/vol14num2/chenbaker.pdf

Cortes, V. (2013a). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, *12*(1), 33–43. https://doi.org/10.1016/j.jeap.2012.11.002

Cortes, V. (2013b). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, *12*(1), 33–43. https://doi.org/10.1016/j.jeap.2012.11.002

Douglas Biber, Susan Conrad, & Viviana Corted. (2004). *Lexical Bundles in University Teaching and Textbooks*.

Ellis, R. (2012). Language Teaching Research and Language Pedagogy. *Language Teaching Research and Language Pedagogy*. https://doi.org/10.1002/9781118271643

Ha, M. J. (2022). Syntactic Complexity in EFL Writing: Within-Genre Topic and Writing Quality 1. In *Computer Assisted Language Learning Electronic Journal (CALL-EJ)* (Vol. 23, Number 1).

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Halliday, M. A.K. & Hasan, K. (1976). *1976_Cohesion in English .pdf*. 374.

Husnaini, H. (2022). Development of Self Esteem-Oriented Micro Teaching Materials for IAIN Palopo English Education Students. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 10(1), 538-560.

Ismayanti, D., & Syam, A. T. (2022). The Community Manipulation through Big Brotherâ€™ s Tyranny in George Orwellâ€™ s Nineteen Eighty-Four. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 10(2), 1556â-1569.

James R. Nattinger and Jeanette S. DeCarrico. (1994). LEXICAL PHRASES AND LANGUAGE TEACHING Oxford: Oxford University Press, 1992. Pp. xvi + 218. *Studies in Second Language Acquisition*, *16*(2), 254–254. https://doi.org/10.1017/S0272263100013000

Ken Hyland. (2008). *As can be seen: Lexical bundles and disciplinary variation*. https://studylib.net/doc/8902747/as-can-be-seen--lexical-bundles-and-disciplinary-variation?utm_source=chatgpt.com

Li, Y. K., & Fang, A. B. (2022). Synergistic effects of multiword sequences structure, function, frequency and association on raters' evaluations of essay quality. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1026658

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149. https://doi.org/10.1017/S0267190512000098

Quoc Tuan, N., Thi Bao Trang, N., & Vu Quynh Nhu, N. (2023). USE OF COHESIVE DEVICES IN PARAGRAPH WRITING BY EFL STUDENTS AT ENGLISH LANGUAGE CENTERS IN VIETNAM. In *VNU JOURNAL OF FOREIGN STUDIES* (Vol. 39, Number 3).

Wang, W., Chen, W., Shi, S., Tong, L., & Duan, M. (2024). A Corpus-Based Study on Cohesion Development in Writing by EFL Novices. *Journal of Language Teaching and Research*, *15*(3), 772–786. https://doi.org/10.17507/jltr.1503.10

Wray, A. (2013). Formulaic language. *Language Teaching*, *46*(3), 316–334. https://doi.org/10.1017/S0261444813000013

Yang, Y., Yap, N. T., & Ali, A. M. (2022). A Review of Syntactic Complexity Studies in Context of EFL/ESL Writing. *International Journal of Academic Research in Business and Social Sciences*, *12*(10). https://doi.org/10.6007/ijarbss/v12-i10/14843