The Distribution of Verb Tenses and Modal Verbs in *Life Today*:
A Corpus-Based Analysis of Phase F English Textbooks

Fallianda
Universitas Sunan Gresik
Corresponding Author. E-mail: fallianda@lecturer.usg.ac.id

Abstract. This study investigates the distribution of verb tenses and modal verbs in Phase F English textbooks in *Life Today* textbook using a corpus-based approach. A self-compiled mini-corpus was constructed from four textbook units representing narrative, argumentative, hortatory exposition, and discussion texts. The corpus consists of reading passages, selected listening and video transcripts provided in the teacher's handbook, with the total of 10,438 unigram tokens and 2,736 unigram types. Corpus analysis was conducted using AntConc 4.3.1, focusing on frequency and dispersion patterns of the 100 highest- occurrence of verb forms, modal auxiliaries, and particular lexical verbs. The findings indicate that the language input reflects genre-specific grammatical patterns, for example the dynamic tense variation characteristic of narrative texts or the grammar complexity and lexical variation in discussion texts. However, discrepancies are identified between the grammatical patterns evidenced in the corpus and the explicit language features presented in the textbook, particularly in terms of grammatical scaffolding and the progression of text complexity across units. These findings highlight the need for more explicit and corpus-informed grammatical scaffolding to support genre-based language learning and to enhance the development of students' receptive and productive skills at the senior secondary level.

Keywords: Corpus-based analysis; verb tense; modality; textbook evaluation; genre-based pedagogy

Abstrak. Penelitian ini mengkaji distribusi tense kata kerja dan kata kerja modal dalam buku teks Bahasa Inggris Fase F pada buku teks Life Today menggunakan pendekatan berbasis korpus. Sebuah mini-korpus yang disusun sendiri dibangun dari empat unit buku teks yang mewakili teks naratif, argumentatif, hortatif, dan diskusi. Korpus tersebut terdiri dari teks bacaan, transkrip mendengarkan, dan video yang disediakan dalam buku panduan guru, dengan total 10.438 token unigram dan 2.736 jenis unigram. Analisis korpus dilakukan menggunakan AntConc 4.3.1, dengan fokus pada pola frekuensi dan dispersi dari 100 bentuk kata kerja, kata kerja modal, dan kata kerja leksikal tertentu yang paling sering muncul. Temuan menunjukkan bahwa masukan bahasa mencerminkan pola gramatikal spesifik genre, misalnya variasi tense dinamis yang khas pada teks naratif atau kompleksitas gramatikal dan variasi leksikal pada teks diskusi. Namun, ketidaksesuaian teridentifikasi antara pola gramatikal yang teramati dalam korpus dan fitur bahasa eksplisit yang disajikan dalam buku teks, terutama terkait dengan penyangga gramatikal dan perkembangan kompleksitas teks antar unit. Temuan ini menyoroti kebutuhan akan penyangga gramatikal yang lebih eksplisit dan didukung korpus untuk mendukung pembelajaran bahasa berbasis genre serta meningkatkan pengembangan keterampilan reseptif dan produktif siswa pada tingkat sekolah menengah atas.

Kata Kunci: Analisis berbasis korpus; tense kata kerja; modalitas; evaluasi buku teks; pedagogi berbasis genre

INTRODUCTION

Language use in the context of language learning refers to the actions performed by language learners, both as individuals and social agents. They develop a variety of skills both general and particular communicative language competences. They draw on the competences in various contexts under various conditions and constraints, which are later processed to comprehend and/or produce texts in oral, written form, or both. Communicative language competence comprises of several elements: linguistic, pragmatic, sociolinguistic, strategic, and intercultural. Each of these elements is understood to include knowledge, skills, and practical expertise. Linguistic competence involves knowledge and skills related to vocabulary, pronunciation, syntax, and other aspects of language as a system.

In the Indonesian context, English language learning in Kurikulum Merdeka allows the students to elaborate their socio-cultural and intercultural insights, which further improve their critical reasoning skills. Regarding the language proficiency levels, the government has adopted The Common Reference Levels (CEFR) to measure students' proficiency. Introduced in 2022 during the COVID-19 pandemic, *Kurikulum Merdeka* aims to optimize the national education system during the crisis. The adoption of CEFR can be seen in its learning phases of presenting and viewing, in addition to listening, reading, speaking, and writing, emphasizing English communicative competence through various text genres and six language skills (Pujianto et al., 2024). Such curriculum objectives differ from the previous curriculum, which focused on basic and core competencies.

At the Secondary Education level (*SMA/MA/SMK/MAK/Package C Program*), students are expected to achieve Level B1 of CEFR after they graduate from high school, and Level B2 for those who take Elective English (*Bahasa Inggris Peminatan*) (Hardini et al., 2022). Level B1 in the curriculum requires students to be able to: (1) maintain interaction and explain something in many different situations with clear articulation; (2) convey main ideas to be comprehensively expressed; and (3) maintain communication despite gaps in between. English proficiency level B2 is a competency level English level four (Upper Intermediate) in CEFR. In everyday conversation, this level is commonly referred to as the level when students are able to use English with confidence. Phases F, which are for 11-12 grade students, require students to reach the expected level of B1 and/or B2 (Emilia et al., 2025). The present study focuses on Phase F, which are for 11-12 grade students, in which Phase F primarily highlight students' written and oral competences at certain levels.

Table 1. Lists of topic units and skill focus of Phase F for 12th grade in *Life Today* textbook (Hardini et al., 2022)

Unit	Topic	Skill Focus
I.	Narrative Text: *The Story of a Friendly Future*	**Skill Focus:** **Listening and Speaking** •　　Listening to a conversation about energy saving •　　Performing a role play about conserving energy **Reading and Viewing** •　　Reading a narrative text about *The Breakers of Limitations* from Gurung Mali •　　Watching videos about renewable energy and fossil fuels

		Writing and Presenting • Writing a simple narrative text about renewable energy • Retelling the narrative text **Vocabulary** • Renewable Energy **Schematic Structures and Language Features** • Past Tense • Action Verbs
II.	Argumentative Text: *E-Money*	**Skill Focus:** **Listening and Speaking** • Listening to a podcast about e-money • Giving opinions on cash money and e-money **Reading and Viewing** • Reading an argumentative text about a cashless society: *What do we gain and lose when cash is no longer king?* • Watching videos about homemade banking and digital payment **Writing and Presenting** • Writing an argumentative text in the form of a podcast about e-money • Performing an argumentative text in the form of a preferred media **Vocabulary** • Digital Money **Schematic Structures and Language Features** • Abstract Nouns • Action Verbs • Temporal Connectives • Technical Terms • Modal Auxiliaries
III.	Hortatory Exposition: *Netiquette*	**Skill Focus:** **Listening and Speaking** • Listening to a dialogue about social media • Discussing posting mean things on social media **Reading and Viewing** • Reading a hortatory exposition text about *Are Social Media Platforms Safe?* • Observing an infographic about netiquette **Writing and Presenting** • Writing a hortatory exposition text about netiquette • Doing a group presentation **Vocabulary** • Netiquette **Schematic Structures and Language Features** • Present Tense • Emotive Phrases

IV.	Discussion Text: *Carbon Footprints*	**Skill Focus:** **Listening and Speaking** • Listening to a dialogue about reducing carbon footprints • Interviewing friends on how to limit carbon footprints **Reading and Viewing** • Reading a discussion text about *Do we need to stop eating meat?* • Watching videos about carbon footprints **Writing and Presenting** • Writing a discussion text about carbon footprints • Doing a group presentation **Vocabulary** • Carbon Footprints **Schematic Structures and Language Features** • Generic Participants • Simple Present • Thinking Verbs • Modality • Contrastive Conjunctions • Passive Voice

Given these clearly defined learning targets, instructional materials—particularly textbooks—become a crucial medium through which these competencies are interpreted, operationalized, and realized in classroom practices. English textbooks have always played a crucial role in supporting students' language acquisition and their ability to apply new knowledge (Brown, 2001; López-Barrios & de Debat, 2014; Maral & Ergül, 2025). Their value lies in providing a structured environment in the learning process. Littlejohn (2011)point out that textbook units are typically built around one or two major texts that function as scaffolds for language presentation, practice, and production. In relation, Richards(2001) explains what matters is whether the text as a material in the textbook effectively reflects the intended grammatical, lexical, and discourse elements. Furthermore, Nunan (2004)emphasizes, language input must be meaningful and systematically aligned with learning objectives.

Suitable grammar presentation and instruction are essential for each text type, for instance essays require complex sentences and formal grammar, while informal letters allow more relaxed structures. Moreover, the features of writing those of argument text could be different from those of narrative type (Melissourgou & Frantzi, 2017). Understanding these distinctions helps learners choose the appropriate style and tone, supporting their text comprehension and production. To examine how such grammatical features are realized in instructional materials, the present study adopts a corpus-based methodology, compiling and analyzing representative texts from Phase F English textbooks. This allows for a data-driven examination of the frequency, distribution, and authenticity of grammatical patterns—particularly verb tenses and modal verbs—as they appear in pedagogical materials (Biber et al., 1998; McEnery & Hardie, 2012).

Genre and grammar approach proposed here, along with the corpus-based methodology, offers an objective approach to assess and evaluate learners understanding and productive skills on particular generic, structural and grammatical features of genres (Table 2). Despite the interd ependence between the two skills, this relationship has not been sufficiently examined in textbook-based studies. The existing studies tend to treat language input (reading texts and grammar presentation) and language output (writing activities) as separate components, focusing either on textual analysis or on learners' written performance. As a result, there is limited empirical evidence on whether the grammatical patterns embedded in textbook reading texts are systematically aligned with the grammatical demands of writing tasks, particularly in senior secondary textbooks designed to support CEFR-oriented outcomes. In the context of Phase F English textbooks under *Kurikulum Merdeka*, it remains unclear whether the grammatical input provided through reading texts adequately prepares students for genre-based writing production.

Table 2.. The genres and grammar of school writing in infants and primary classrooms (Knapp & Watkins, 1994)

Reading and Writing Skills
I. Generic Structures
-   Genre or Text Types
-   Theme
-   Rhetorical and Language Features
-   Vocabulary
II. Textual language
-   Connectives
-   Reference
-   Tense
-   Sentence Structure
III. Syntactical Language
-   Clause Pattern
-   Agreement
-   Verb Form
-   Prepositions
-   Articles
-   Plurals
-   Punctuation
IV. Spelling

Among the various components in grammar, the distribution of verb tenses and modals constitutes the remark of particular of text types. Research shows that tenses and modals are essential for conveying time, stance, and argument structure in academic discourse (Biber et al., 1999; Carrio-Pastor, 2014; Swales & Feak, 2012), and that their usage patterns vary across genres and disciplines, requiring context-sensitive instruction (O'Keefe et al., 2007; Winkler & Metherell, 2012). For example, factual descriptions in descriptive texts are generally written consistently in the present tense, whereas narratives and arguments can move between present and past tense. Therefore, it is important for students to be aware of such changes in text types where reading or writing may move between narrating, arguing, describing, and so on.

On this basis, the present study investigates how grammatical features in senior high-school English textbooks—specifically those used at Phase F in *Kurikulum*

Merdeka. Employing corpus-based approach, this study also determine the grammatical input, sequencing, and level of complexity. To address these aims, the study is guided by the following research questions:

1. How are grammatical patterns represented across text types in Phase F English Textbook?
2. To what extent are the grammatical patterns aligned with the language features in the textbook units?

METHODS

This study employed a quantitative corpus-based research design complemented by qualitative interpretation to examine the distribution of verb tenses and modal verbs across four text genres (Bowen, 2009; Gries, 2008). This study aims to evaluate their alignment with the language features presented in an English textbook for senior high school students. The population of the study consisted of all reading texts and instructional materials contained in the four units of the selected textbook, which represent the genres mandated in the curriculum.

A corpus was compiled from selected texts representing different genres within the textbook units of *Life Today* (Hardini et al., 2022). Corpus-based methods were selected because they allow for systematic and empirical analysis of linguistic patterns based on authentic textual data rather than intuition (Biber et al., 1998; McEnery & Hardie, 2012). The corpus consisted of various types of pedagogical input, particularly language exposure drawn from reading texts as well as video and listening transcripts available in the teacher's handbook. The data were collected from four textbook units, yielding 2,736 unigram types and 10,438 unigram tokens. The analysis was conducted using a word-based, case-sensitive unigram search (N-gram size = 1), with results sorted by frequency. To identify dominant grammatical patterns, the distribution of the 100 highest-frequency word types was examined, and concordance (word-in-context) analysis was subsequently carried out to determine the verb tenses involved.

The analysis focused on identifying frequently occurring verb forms and modals, and selected lexical verbs found during the data collection process. Frequency counts were calculated to determine dominant grammatical patterns across units. In addition to frequency, dispersion values were observed to examine how evenly each linguistic feature was distributed across texts, allowing the analysis to distinguish between forms that represents the characteristic of a genre and those that are only episodic appear in a genre. The findings were then interpreted qualitatively by relating frequency–dispersion patterns to observe patterns of schematic structures and language features of each unit, allowing for an evaluation of corpus–textbook alignment.

The primary instrument used in this study was a self-compiled mini-corpus constructed from the textbook texts. The corpus was digitized and processed using corpus analysis software, AntConc 4.3.1 (Anthony, 2024), to identify verb forms and modal auxiliaries. The list of target word types of verb tenses and modals was determined based on their relevance to grammatical features commonly associated with the four genres (Biber et al., 1999). To ensure accuracy, the automatically generated frequency lists were manually checked to confirm grammatical function and to exclude irrelevant homographs.

Data collection procedures involved several stages. First, the texts from each unit were categorized according to genre. Second, word frequency counts were generated for each target word type. Third, dispersion values were calculated to

determine how evenly each word type was distributed across texts. Dispersion analysis was used to complement frequency data, as frequency alone does not reveal whether linguistic forms are spread across a text or concentrated in limited sections (Gries, 2008).

The analysis plan focused on descriptive statistical methods, specifically frequency counts and dispersion values. Frequency was used to identify dominant grammatical forms, while dispersion values were used to assess distribution patterns across texts. These descriptive statistics are appropriate for corpus-based discourse analysis, as they allow researchers to identify systematic patterns of language use without inferring causality (Biber et al., 1998; McEnery & Hardie, 2012). No inferential statistical tests were applied because the study aimed to describe and interpret usage patterns rather than to test hypotheses.

To enhance validity and reliability, several measures were taken. Construct validity was supported by grounding the selection of word types and genres in established genre theory and corpus linguistics literature. Reliability was strengthened through repeated checks of frequency outputs and cross-verification of dispersion results. In addition, interpretations were consistently triangulated with textbook language features to avoid overgeneralization.

FINDING AND DISCUSSION

This section presents and discusses the findings derived from the corpus-based analysis of verb tenses and modal verbs used across the textbook units. By examining frequency and dispersion patterns, the analysis highlights the dominant grammatical forms to which learners are exposed and how these forms function across text genres. The results provide insight into the relationship between actual language use in the textbook corpus and the schematic structures and language features outlined in each unit, therefore revealing areas of alignment as well as pedagogical gaps that may affect learners' grammatical awareness and genre competence.

The verb tenses and modals used across units

Table 3. The number of frequencies of verb tenses and modals in *Life Today* textbook

No.	Word Type	Freq	No.	Word Type	Freq
1.	is	128	11.	know	26
2.	are	73	12.	like	24
3.	s	70	13.	said	24
4.	was	66	14.	use	23
5.	can	61	15.	had	21
6.	have	59	16.	make	20
7.	be	49	17.	may	18
8.	will	29	18.	been	17
9.	has	28	19.	were	17
10.	do	27			

The corpus data indicate a strong dominance of present-tense and copular constructions across units, with *is* (128) and *are* (73) emerging as the most frequent word types. This pattern reflects the central role of present tense in expressing definitions, general statements, and evaluations, which are characteristic of argumentative, hortatory, and discussion texts. The high frequency of *be* (49), *has* (28),

and *have* (59) further highlights the prominence of copular and auxiliary verbs, supporting grammatical structures such as passive voice, evaluative statements, and aspectual meanings.

Past-tense forms, including *was* (66), *were* (17), *had* (21), and *said* (24), are also well represented, indicating the importance of narrative sequencing and reporting events, particularly in narrative texts. However, these forms occur less frequently than present-tense verbs, suggesting that past-tense usage is more genre-specific rather than evenly distributed across units.

Modal auxiliaries such as *can* (61), *will* (29), and *may* (18) appear consistently, signaling the frequent use of modality to express ability, prediction, obligation, and possibility, especially in argumentative, hortatory, and discussion texts. The presence of do (27) and lexical verbs like *make* (20), *use* (23), *know* (26), and *like* (24) indicates a preference for action-oriented and cognitive processes that support reasoning, explanation, and stance-taking.

To examine how grammatical features are realized across different genres in the Life Today textbook, Table 4 presents a corpus-based distribution of verb tenses, auxiliary verbs, modals, and selected lexical verbs across four instructional units. The distribution table will be used to indicate specific grammatical role in text constructions.

Table 4. The distribution of verb tenses and modals in *Life Today* textbook

No.	Word Type	Unit 1. Narrative			Unit 2. Argumentative			Unit 3. Hortatory			Unit 4. Discussion		
		Freq	Normal Freq	Disp	Freq	Normal Freq	Disp	Freq	Normal Freq	Disp	Freq	Normal Freq	Disp
1.	is	26	7276.798	0.718	30	13286.094	0.752	22	11071.968	0.811	50	19083.969	0.837
2.	are	8	2239.015	0.592	28	12400.354	0.74615	15	7549.069	0.715	22	8396.947	0.837
3.	s	29	8116.429	0.851	14	6200.177	0.677	13	6542.526	0.675	14	5343.511	0.695
4.	was	43	12034.705	0.718	11	4871.568	0.684	7	3522.899	0.435	5	1908.397	0.463
5.	can	24	6717.045	0.707	6	2657.219	0.556	13	6542.526	0.655	18	6870.229	0.704
6.	have	17	4757.907	0.722	9	3985.828	0.651	16	8052.340	0.649	17	6488.550	0.767
7.	be	14	3918.276	0.714	12	5314.438	0.701	13	6542.526	0.696	10	3816.794	0.789
8.	will	10	2798.769	0.606	7	3100.089	0.628	3	1509.814	0.491	9	3435.115	0.548
9.	has	8	2239.015	0.514	8	3542.958	0.637	5	2516.356	0.667	7	2671.756	0.695
10	do	11	3078.645	0.467	8	3542.899	0.628	7	3522.899	0.628	1	381.679	0.000
11.	know	4	1119.507	0.236	5	2214.349	0.385	11	5535.984	0.630	6	2290.076	0.631
12.	like	12	3358.522	0.792	4	1771.479	0.592	5	2516.356	0.553	3	1145.038	0.289
13.	said	19	5317.660	0.385	-	-	-	2	1006.543	0.333	3	1145.038	0.491
14.	use	4	3573.507	0.236	9	3985.828	0.651	3	1509.814	0.289	7	2671.756	0.628
15.	had	18	5037.783	0.671	-	-	-	-	-	-	3	1145.038	0.491
16.	make	8	2239.015	0.688	7	3100.089	0.571	2	1006.543	0.333	3	1145.038	0.491
17.	may	-	-	-	8	3542.958	0.417	7	3522.899	0.571	3	1145.038	0.289
18.	been	6	1679.261	0.728	3	1328.609	0.491	3	1509.814	0.491	5	1908.397	0.553
19.	were	7	1959.138	0.435	4	1771.479	0.447	5	2516.356	0.667	1	381.679	0.000

The corpus-based approach of the narrative text reveals word types which exhibit frequency and dispersion values. The copular verbs *was* (43/66; dispersion = 0.718) and *is* (26/128; dispersion = 0.718) are consistently distributed, with *was* dominating past-time narration to highlight the genre's focus on past events and *is* is used in direct speech and descriptive segments. The time marker of past narration can also be seen in the use of *had,* which appears 18 times from the total of 21 occurrences. Modality of *can, have to,* and *will* were also typically found in this genre, which are used to shape characters' actions and unfolding of events, providing a nuanced layer of meaning that contributes to the dynamic nature of the narrative.

The reporting verb *said* shows relatively high frequency (19 out of 24 in total occurrences) but the lowest dispersion value (0.385), reflecting its role as a dialogue marker in sections involving direct or indirect speech. The table representing the distribution of verb tenses, modal verbs, and specific lexical verbs in each genre will focus on the use of the active voice. A particular mention of its counterpart, the noticeably high frequency of passive voice, will be provided separately. The expression *like* shows the highest frequency and its dispersion value (12/24; 0.792), reflecting its frequent use in descriptive or comparative expressions instead of expressions of likes or dislikes. Table 5 and 6 below are then presented such information in narrative genre.

Table 5. Distribution of verb tenses in narrative text

Present simple		Present progres-sive	Present Perfect	Present Perfect progres-sive	Past Simple		Past progres-sive	Past Perfect	Past Perfect progres-sive
is	*are*	*is/am/are + verb-ing*	*has/have + verb 3*	*has/have + been + verb-ing*	*was*	*were*	*was/were + verb-ing*	*had + verb 3*	*had + been + verb-ing*
24	6	3	7	0	7	7	8	6	0

Table 6. Distribution of modals and lexical verbs in narrative text

Modal Verbs				Lexical Verbs						
can	*have /has/had to*	*will*	*may*	*do*	*know*	*said*	*use*	*has/ have*	*had*	*make*
22	6	10	0	10	4	19	4	10	18	8

While past progressive and past perfect tenses can be found in the corpus data, there is no explicit presentations or explanations of their use in the textbook (Hardini et al., 2022, p. 22.. The textbook refers to either "past tenses" or "past tense" without explicit instruction on what types of past tenses will be covered. Additionally, although the present simple tense is widely used, the language features do not specify its application or the contexts in which students should use the present tense forms. As a result, teachers are left to interpret for themselves which tenses or grammar structures will be addressed during the learning process in the classroom. Given to the fact that specific tenses are repeatedly covered in particular genre, this repetition is still needed, particularly in light of the current *Kurikulum Merdeka*, which is designed to address students' evolving needs and align with the latest skill requirements.

The findings in table 7 below confirm that argumentative texts rely heavily on present-tense copular, auxiliary verb, and action-oriented verbs such as *use, make,* and *has/have*. The copular verbs *is* (30/128; dispersion = 0.752) and *are* (28/73; dispersion = 0.746) are the most frequent and widely distributed forms, reflecting their function in expressing definitions, claims, and general statements throughout the text. Moreover, the auxiliary form of *to be* is also observed in its present progressive form, which is used to indicate ongoing actions or states, further emphasizing the dynamic nature of the argumentation process in the text. Other lexical forms, including *use* (9/23; dispersion = 0.651) and *make* (7/20; dispersion = 0.571) suggests more localized but still recurrent use across the argumentative text.

Table 7. Distribution of verb tenses in argumentative text

Present simple		Present progres-sive	Present Perfect	Present Perfect progres-sive	Past Simple		Past progres-sive	Past Perfect	Past Perfect progres-sive
is	*are*	*is/am/are + verb-ing*	*has/have + verb 3*	*has/have + been + verb-ing*	*was*	*were*	*was/were + verb-ing*	*had + verb 3*	*had + been + verb-ing*
29	19	11	5	0	8	2	1	0	0

Table 8. Distribution of modals and lexical verbs in argumentative text

Modal Verbs				Lexical Verbs						
can	*have /has/had to*	*will*	*may*	*do*	*know*	*said*	*use*	*has/ have*	*had*	*make*
7	2	7	8	0	5	0	9	8	0	7

A unique pattern that can be found in this genre is the use of "do" as an auxiliary verb, which is highly frequent (8 out of 13 in total), and also has a high dispersion value of 0.637. We can see that *do* as lexical verbs is nowhere to be found in the text (see Table 8). This suggests that students may practice forming questions or negative sentences to express their opinions.

- La Ode: "So, what *do* you think about their habits?"
- What *do* you think of a cashless world?
- La Ode: "What *do* you mean?"

Sri: "It's very convenient. And this is the reason why I started my online shop.

- Going cashless, on the other hand, may be more difficult than useful. First of all, electronic payments *do* not have the same level of privacy as cash payments.
- Customers can easily withdraw their cash from the bank if they *do* not like the fees, therefore banks are constrained in their capacity to pass on those expenses.

Another noticeable pattern is the use of passive voice in argumentative text. Fourteen passive voices in the form of present simple tense are found. This fact highlights the need for passive voice exposure in the learning process. However, no explicit mention is provided in the presentations of language features in argumentative

text (Hardini et al., 2022, pp. 77-80). The textbook only mentions features, such as: abstract noun, action verbs, technical terms, modal auxiliaries, temporal connectives, and present tense. Although modality also plays an important role in this genre, we cannot deny the fact that it is inevitable to avoid using passive voice in this genre. One reason for this is that passive constructions allow the focus to shift from the subject performing the action to the action itself or the object receiving the action, which is often necessary in argumentative writing to emphasize ideas or concepts rather than individual agents

The corpus analysis shows that hortatory texts are characterized by the frequent and evenly distributed use of present-tense and modal constructions, reflecting their persuasive and advisory function. Modal verbs play a central role in shaping hortatory meaning. The other noticeable patterns that can be found are higher use of present perfect tense and lexical verbs of *know* (see Table 9 and 10 below), as well as passive voice in its present simple forms. The verb *know* is frequently used to highlight the importance of awareness or understanding, which is crucial in persuasive texts that aim to inform or convince the audience.

Table 9. Distribution of verb tenses in hortatory text

Present simple		Present progressive	Present Perfect	Present Perfect progressive	Past Simple		Past progressive	Past Perfect	Past Perfect progressive
is	*are*	*is/am/are + verb-ing*	*has/have + verb 3*	*has/have + been + verb-ing*	*was*	*were*	*was/were + verb-ing*	*had + verb 3*	*had + been + verb-ing*
18	11	6	18	0	4	5	0	0	1

Table 10. Distribution of modals and lexical verbs in hortatory text

Modal Verbs				Lexical Verbs						
can	*have /has/had to*	*will*	*may*	*do*	*know*	*said*	*use*	*has/ have*	*had*	*make*
13	0	3	7	3	11	2	3	3	0	2

The present perfect tense is used to connect past actions or experiences to the present, indicating relevance or importance to the current discussion or argument in hortatory texts.

- Dean: Hi Nana, *have* you *heard* the news about a netizen who is sued for posting rude words online to an actress? I *have posted* the same thing to someone's status but luckily, no one sued me. What about you?
- Social media platforms, as one of the new media, *have become* a phenomenon in the world including in Indonesia with an amazing growing number of users.
- This initiative to turn Sumba into an iconic island of clean energy <u>*has*</u> since *attracted* additional support from other development partners, the private sector and most recently the ADB.
- For village headman, Lede Dauga, having electricity at home <u>*has changed*</u> his life.

Nevertheless, the language focus in the textbook features present tense and emotive phrases. It cannot be determined whether there are specific features of the use of present tense forms that distinguish them from the use of present tense in narrative

texts, for instance. Separately, the textbook distinguishes features of facts and opinions which plays a significant role in helping students understand how to differentiate between objective information and subjective viewpoints, a key aspect in constructing arguments (Hardini et al., 2022, pp. 123-127.

The corpus analysis of the discussion text shows a strong dominance of present-tense copular indicating their central role in constructing balanced and objective discourse. Modal verbs such as *can* (18/61; 0.704) suggesting stable use in expressing ability and/or possibility. The variety of lexical words like *has/have, know*, *said*, *use*, and *make* in discussion texts means that the language is flexible and nuanced, allowing for the expression of a wide range of concepts, ideas, and arguments. In other words, the lexical variety might also indicate the complexities of the discussion. The variety of tenses in discussion texts appears because they help convey different perspectives and relationships between events or ideas. All the information of the distribution can be seen in Table 11 and 12 below.

Table 11. Distribution of verb tenses in discussion text

Present simple		Present progres-sive	Present Perfect	Present Perfect progres-sive	Past Simple		Past progres-sive	Past Perfect	Past Perfect progres-sive
is	*are*	*is/am/ are + verb-ing*	*has/ha ve + verb 3*	*has/have + been + verb-ing*	*wa s*	*wer e*	*was/were + verb-ing*	*had + verb 3*	*had + been + verb-ing*
38	2	7	7	2	2	1	1	0	1

Table 12. Distribution of modals and lexical verbs in discussion text

Modal Verbs				Lexical Verbs						
can	*have /has/had to*	*will*	*may*	*do*	*know*	*said*	*use*	*has/ have*	*had*	*make*
18	1	9	3	1	6	3	7	14	3	3

More importantly, the complexity of ideas and structures in the discussion text has been highlighted in the textbook. However, learning such complex grammar in just one unit session may not be sufficient. This is not only because students have not yet been introduced to the active-passive voice in previous units (Narrative, Argumentative, and Hortatory texts), but also because in this unit, they are expected to learn active-passive structures across various tenses (Present simple, Present progressive, Past simple, Past progressive, Present perfect, Past perfect, Future simple, Infinitive, and Modals) (Hardini et al., 2022, pp. 174-185). Given this, it is crucial to acknowledge that mastering such a wide range of structures in a single units may overwhelm students, making it difficult for them to fully grasp and apply these concepts effectively. A more gradual and focused approach could help ensure deeper understanding and retention of these grammatical structures.

- At one stage, beef production *is* the leading cause of deforestation in tropical rainforests which adds to the environmental impact of beef from that part of the world.

- In the report issued by ourworldindata.org, it *is said* that half of all habitable land *is used* for agriculture, and three-quarters of that land *is used* to feed and raise livestock.
- Now, customers *are* increasingly *buying* online, where they *can* order whatever they want directly to their door with the click of a mouse.
- In addition, eating meat is the number one contributor to the heat we have *been* experiencing lately: global warming.
- It *will take* huge changes but following this plan *will lower* our risk of cancer, strokes, and diabetes.

Corpus–Textbook Alignment

The corpus-based findings indicate that the alignment between the linguistic patterns found in the texts and the explicit language features presented in the textbook is partial rather than comprehensive. While the textbook successfully introduces genre-specific topics and skills, it does not consistently reflect the grammatical complexity revealed by the corpus data.

In narrative texts, the corpus shows frequent shifts between past tense, present tense, and passive constructions, which function to manage narration, description, and reported speech. However, the textbook limits its language features to Past Tense and Action Verbs, without explicitly explaining how present-tense forms (e.g., *is, are*) and simple passive voice are used alongside past tense to support orientation, description, and evaluation in narratives. This omission may reduce learners' awareness of how tense variation operates as a discourse strategy rather than a grammatical inconsistency.

In hortatory exposition texts, the corpus data reveal a high frequency and wide dispersion of modal auxiliaries such as *can*, *may*, and *will*, which are central to expressing recommendation, obligation, and possibility. Despite this, the textbook lists Present Tense and Emotive Phrases as the primary language features, without explicitly including modality. This creates a noticeable gap between the linguistic reality of hortatory discourse and its pedagogical representation, potentially limiting students' ability to construct persuasive and appropriately nuanced recommendations.

By contrast, discussion texts show a stronger alignment between corpus findings and textbook features Nevertheless, despite showcasing the complexity of tenses such as the present perfect and present perfect continuous, as well as passive forms of complex tenses, students are likely to face difficulties, and teachers may find it challenging to provide adequate guidance. As a result, students may encounter these structures in reading texts without sufficient metalinguistic support to understand their evaluative and temporal functions.

Overall, the findings suggest that while the textbook broadly aligns with genre conventions at a schematic level, it underrepresents key grammatical features that are empirically frequent and functionally important in the corpus. This misalignment highlights the need for a more corpus-informed approach to textbook design, where language features are selected not only based on traditional genre descriptions but also on actual usage patterns. Explicit attention to tense shifts, modality, and complex verbal constructions would enhance learners' grammatical awareness and better support accurate and meaningful genre-based language production.

CONCLUSION

This study demonstrates that frequency and dispersion analyses provide complementary insights into how grammatical forms function across text genres in

English textbooks, revealing distinct grammatical patterns that align closely with each genre's communicative purpose. The interaction between frequency and dispersion confirms that grammatical choices are not random but genre-sensitive, functioning as key resources for realizing meaning, stance, and rhetorical purpose. However, the study is limited by its reliance on a single textbook and phase, which may not fully represent the diversity of genre-specific grammatical patterns across different textbooks or educational contexts. It is recommended that textbook developers and teachers incorporate corpus-informed grammatical explanations into genre-based instruction, with explicit attention to tense and lexical variation, to enhance learners' grammatical awareness and enable more accurate language use across genres. Future research should explore multiple textbooks and phases to provide a broader understanding of these grammatical patterns and their teaching implications.

REFERENCES

Anthony, L. (2024). *AntConc: Vol. 4.3.1*. Waseda University.
    https://www.laurenceanthony.net/software/AntConc

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Brown, H. D. (2001). *Teaching by Principles: An Interactive Approach to Language Pedagogy* (2nd ed.). Longman.

Carrio-Pastor, M. L. (2014). Lexical and grammatical patterns in academic discourse. *Journal of English for Academic Purposes*, *15*, 45–58.

Emilia, E., Hamied, F. A., & Musthafa, B. (2025). Genre-based pedagogy and CEFR alignment in Indonesian senior high school English curriculum. *Indonesian Journal of Applied Linguistics*, *15*(1), 1–18.

Gries, S. Th. (2008). *Quantitative corpus linguistics with R: A practical introduction*. Routledge.

Hardini, S. R., Susanti, Retno, Hardini, A., Merdianto, Marjenny, Rani, Nurhayati, Isry, Laila, & Syathroh. (2022). *Bahasa Inggris: Life Today*. Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia.

Knapp, P., & Watkins, M. (1994). *Genre, text, grammar: Technologies for teaching and assessing writing*. UNSW Press.

Littlejohn, A. (2011). *The analysis of language teaching materials: Inside the Trojan Horse*. Cambridge University Press.

López-Barrios, M., & de Debat, E. V. (2014). Global vs. Local: Does It Matter? In S. Garton & K. Graves (Eds.), *International Perspectives on Materials in ELT* (pp. 37–52). Palgrave Macmillan. https://doi.org/10.1057/9781137023315_3

Maral, B. N., & Ergül, H. (2025). CEFR Proficiency Levels of Secondary School English Textbooks' Word Lists in Türkiye. *International Journal of Curriculum and Instruction*, *17*(3), 892–909.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Melissourgou, M., & Frantzi, K. (2017). Genre identification based on syntactic and lexical features. *Educational Technology & Society*, *20*(2), 154–165.

Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.

O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Pujianto, Nurkamto, J., & Suparno. (2024). Implementing CEFR in the Merdeka Curriculum: Implications for English language teaching in Indonesia. *Indonesian Journal of Applied Linguistics*, *14*(1), 1–15.

Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge University Press.

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students*. University of Michigan Press.

Winkler, I., & Metherell, B. (2012). *Academic writing and genre: A systematic approach*. Routledge.