

## Computerized corpus as a tool for educational technology and learning in the analysis of four-word recurrent expressions

Ali Şükrü Özbay<sup>a\*</sup> , Zehra Gürsoy<sup>a</sup> 

<sup>a</sup> Karadeniz Technical University, Turkey;

Suggested citation: Özbay, A. Ş., Gürsoy, Z. (2023). Computerized Corpus as a Tool for Educational Technology and Learning in the Analysis of Four-Word Recurrent Expressions. *Journal of Educational Technology & Online Learning*, 6(1), 249-272.

### Highlights

- We found that Turkish writers were selective in using structural expressions.
- There was a positive role of recurrent expressions in the learning of the disciplinary features.
- We found structural similarities between native and non-native writing in the two corpora.
- Corpus technology proved to be useful for learning as well as discovering new truths about language.
- Non-native writers mostly used less varied and limited number of expressions.

### Abstract

The study aimed to investigate the disciplinary variations and the structural/functional features of recurrent expressions in the introduction and literature sections of the Master of Arts (MA) theses in two corpora. It is significant since the bundles fulfil specific functions in the form of formulaic language. The sampling of the study was Turkish and British academic writers. Their essays mainly based on academic argumentative topics were selected as analysis data. Data collection included compilation and categorization of the computerized corpora for possible differences and similarities. Sketch Engine, an online text analysis tool, was used for the analysis. Exploring the usage patterns of recurrent expressions among the native and non-native corpora, we also analysed the four-word recurrent expressions and featured the existing variations of functional and structural aspects. Cross-linguistic analysis revealed that Turkish writers distinguish some features of recurrent expressions more than others. Structural expressions encompassed the verb-based, noun and propositional structures. The research, text, and participant oriented recurrent expressions built up the functional groups. The most frequently used functional categories were location and description in research-oriented recurrent expressions. The findings also indicated a positive role of recurrent expressions in the learning of the disciplinary features. Structural similarities were also found, indicating strong functional features for constructing discourse in research writing. The implication is that creating appropriate academic discursive practices with a focus on the properties of the recurrent expressions can be better analysed by employing several online corpus tools.

**Article Info:** Research Article

**Keywords:** *Recurrent expressions, Computerized corpora, Function, Structure, Native and non-native*

## 1. Introduction

The cross-linguistic influence, which is one of the most significant factors to consider in language acquisition, may have a greater role in understanding the various discursive practices, especially in academic settings. The realization of linguistic transfer, in return, creates a dual process where the intercourse between the Language 1 and Language 2 (L1 - L2 hereafter) is likely to affect each other (Brown & Gullberg, 2008). Jarvis (2000, p. 252) defined crosslinguistic influence as “one referring to any instance of learner data where a statistically significant correlation is shown to exist between some feature of

\* Corresponding author. Western Languages and Literature, Karadeniz Technical University, Türkiye.  
e-mail addresses: [ozbay@ktu.edu.tr](mailto:ozbay@ktu.edu.tr), [zehragursoy@ktu.edu.tr](mailto:zehragursoy@ktu.edu.tr)

learners' interlanguage performance and their L1 background". This indicates that several collocations and recurrent expressions used with similar meanings in different locations may also have several common characteristics that come out as a result of language transfer. Even in an academic genre such as research articles or reviews, multiword combinations and recurrent expressions are easily recognizable and distinguishable through computerized corpus tools such as AntConc, Sketch Engine, Wordsmith tools. While the use of L1 in written and spoken genre provides a great variety of multiword combinations, the exact opposite situation may result in less variation due to the nature, lexical and semantic connections of L2. However, these connections may be best processed and studied using computerized corpus tools given above to ease learning.

Being a dominant trend in the last two decades, computerized corpus linguistics has come into prominence and become widespread, enabling researchers to find on-spot results more effectively and effortlessly in their written or spoken data in an electronic format. This is because corpus tools and their functions may easily be used to observe how the words or expressions occur in context and why some others are misused by language learners (L2). They can also enable users examine authentic language data quantitatively and qualitatively as well as create concordance lines, clusters, collocates, wordlists, keyword lists and frequencies for further analysis in an electronic format. The analysis of keywords in contexts along with concordance lines provides a practical analysis of words and word combinations in terms of various aspects such as selecting new words with their appropriate semantic and pragmatic prosodies. This categorization of items may help language learners find the most-commonly used nativelike recurrent expressions and the commonest lexical patterns in various genres, and this is confirmed by Bennet (2010, p. 2), who emphasized the two broad questions which corpus linguists strive to answer: "What particular patterns are associated with lexical or grammatical features?" and "How do these patterns differ within varieties and registers?". Online computerized corpus use has given the dual advantages of exploring rich authentic language data with or without instruction. It has also focused on targeted language items such as grammatical structures and multiword combinations that can help users to develop themselves in academic/non-academic writing.

The study aimed to analyse the most frequently used recurrent expressions through an online corpus tool. The structurally and functionally similar and different bundle patterns along with the frequencies were analysed. The selection and the analysis of four-word recurrent expressions in this study is significant since

- (a) they were analysed previously by other researchers in different settings as a matter of concern (Granger, 2014).
- (b) they do not "only contribute to idiomaticity, but also contribute to demonstrating membership in a specific discourse community" (Adel and Erman, 2012, p. 81).
- (c) they fulfil specific functions in the form of formulaic language (Schmitt and Carter, 2004).
- (d) they correlate greatly with various formulaic language patterns used by native speakers (Lewis, 2009).
- (e) they are manageable in size (Chen and Baker, 2010; Hyland, 2008) as well as most of them are known to cover "three-word expressions in their structures" (Cortes, 2004, p. 401).
- (f) they were highly frequent and presented various structures and functions for more detailed investigations (Cortes, 2004; Hyland, 2008; Chen and Baker, 2010).
- (g) three-word bundles are largely seen very similar to four-word bundles while longer words are very few in the corpus (Csomay, 2022).
- (h) their frequencies are greater when compared to other types of word combinations and form a group of new structures and functions (Hyland, 2008).

While native corpora consisted of data from British (EngCorp) English MA theses, respectively, non-native corpus (TurkCorp) included data from the introduction and literature review parts of Turkish MA theses in Applied Linguistics. The aim was to investigate the four-word recurrent expressions and feature their similarities and differences both structurally and functionally. First of all, the overuse and underuse patterns of recurrent expressions were analysed in Turkish corpus by comparing them to those in the native corpora. Secondly, we observed the sequence of four-word recurrent expressions used in native corpora. While structural recurrent expressions included the verb-based, noun, and prepositional structures; functional recurrent expressions included the research-oriented, the text-oriented and the participant-oriented recurrent expressions. The quantitative and qualitative findings revealed structural and functional similarities and differences in four-word recurrent expressions.

## 2. Corpus, Educational Technology and Learning

As part of educational technology, the benefits and practicality of corpus applications have not been well explored especially in terms of English language teaching (Gong, 2019). The relation between computerized corpus as an “electronic library” (Gong, 2019), educational technology and learning is obvious in the sense that many corpus tools and their functions basically serve for the active involvement of modern educational technologies and learning outcomes (Fanilevna et al., 2020). Corpus technology is used as a valid research tool in many linguistic and quantitative studies in terms of methodology design and measuring linguistic changes. They can also be used as a means of testing prescriptive grammar rules as well as being an empirical component in the identification of meaning differences and similarities (Osipova, 2020). In English language teaching (ELT), corpus tools are largely used for several instructional purposes. For example, they can be used for facilitating vocabulary teaching and learning activities as well as improving learner language learning. After selecting an appropriate corpus, various activities can be performed. As a new corpus-based language pedagogy (CBLP), it aims to introduce a new pedagogy by integrating corpus into classroom to foster teachers’ and the students’ awareness, thus creating a new technology-oriented classroom pedagogy (Ma and Mei, 2021).

The instructional purposes can be categorized as the following:

- (a) The use of corpus for educational technology-based learning,
- (b) The use of corpus for computer-assisted learning and,
- (c) The use of corpus for data-driven learning (Friginal et al., 2020)

Especially, in terms of educational technology, the role of corpus tools and their active usage patterns in learning process play a significant role in fostering autonomy and learner-centred teaching. What is more, focusing upon software design and evaluation, corpus tools can also be used for reinforcing learning process (Friginal et al., 2020).

Learning process with a focus on corpus technology can be developed in various ways. In one example, teachers can have the responsibility to organize corpus based digital learning materials and tasks for students. This may be through selecting the corpus tasks and assign them to the students on a regular basis. Or students can use the corpus based digital tasks by using appropriate methods. They can carry out tasks by systematically analysing and experimenting language under the supervision of their teachers. The texts to be used in this process are called “text corpora” and their technology can be especially effective in learning new words and word combinations. The text corpora technology has the potential to support active and inductive learning by fostering learners to study and analyse the language from different dimensions and discover new truths about language as a result of guided discovery through corpus technology (Gong, 2019; Fanilevna et al., 2020;). In the learning process, the computerized corpora can also be very effective since it has the potential to develop learners’ critical attitude to learning language as well as fostering deeper understanding of language from various perspectives (Osipova, 2020).

### 3. Practical Implications

In Osipova's (2020) study, corpus is taken as a means for an "educational experiment" that aims to foster corpus-based idiom interpretation skills by creating and developing a set of corpus-based learning materials. The aim of preparing these materials was to contribute to the several practical implications such as developing linguistic and creative thinking, contextual guess, and to investigate its lexical and grammatical compatibility.

In her study of "Corpora and New Technologies in the Linguistics Classroom: A Pedagogical Use of a Clause Pattern Database", Martin et al. (2017) state that computerized corpus functions enable users to reach the data which can hardly be obtained without the use of corpus tools. Its practicality in collecting and analysing samples of authentic and naturally occurring data makes it all the more an effective educational tool to benefit from (Granger, 2001; Sinclair, 2004; Laso and Giménez, 2007; Aijmer, 2009; Reppen, 2009; Campoy et al., 2010; Urzua, 2015). Regarding its practical implications, corpus tools can also be used to ease learning by promoting the learning of targeted word combinations relatively faster and easier (Martin et al., 2017), enabling inductive and discovery-based learning (Ma et al., 2023), creating learner autonomy by giving them responsibility to examine corpus concordances (direct use), using computer software to analyse concordances prepared by teachers (indirect use), creating ground for hands-on and hands-off activities, obtaining quantitative data through text corpus to ease learning (Fanilevna et al., 2020), helping increase collocational competence (Fanilevna et al., 2020), compiling target-oriented computerized corpus in any discipline, to identifying the lexical and grammatical suitability of idioms (Osipova, 2020), using AntConc 4.0, Sketch Engine and LancsBox educational corpus tools, rapidly accessing the digital corpora from free web sources and helping to create meaning construction by providing huge context (Gong, 2019).

#### 3.1. Lexical Bundles

"Lexical bundle" is a linguistic term coined by Biber et al. (1999, p. 990), who defined the term as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" and "sequences of word forms that commonly go together in natural discourse" in the Longman Grammar of Spoken and Written English. For academic writers, the knowledge and use of various forms of formulaic language - used as a synonym with recurrent expressions- helps achieve naturalness in language use (Allen, 2009). Moreover, they carry significant roles in speech and academic writing because of their structural functions. Chen and Baker (2010, p. 31) argue that "most expressions in conversation are clausal, whereas most expressions in academic prose are phrasal". The workability of the data related to recurrent expressions has drawn attention, and many studies have been done in the last ten years. While several studies mentioned above investigated the distribution and lexical bundle usage patterns, others explored the relationship between expert and non-expert writing (Chen & Baker, 2010; Navarro & Martinez, 2019; Pan, 2019; Zhang et al., 2021).

In the last decade, the focus of computerized corpus-based academic writing studies has shifted into analysing multiword structured grammatical units and the lexical bundle patterns in various other fields by using several electronic corpus tools designed for educational purposes e.g., AntConc, LancsBox, Sketch Engine, Wordsmith, etc. For example, Adel and Erman (2012) in their analysis of recurrent expressions in L1 and L2 writing used a corpus and found that non-native speakers used fewer recurrent expressions in frequency and variation than native speakers. Functionally, the employment of participant-oriented expressions was more in native speakers, and they used less text-oriented expressions. Chen and Baker (2010) studied L1 and L2 writers and included novice and expert writing in native speakers' writing. The corpus data showed that native-novice students and non-native students used similar expressions in frequency, structure, and function. Hyland (2008) observed recurrent expressions across different disciplines' academic writings and indicated that types and functions of expressions varied between disciplines and played a significant part in the discourse of disciplinary-specific writing.

Recurrent expressions were split into two groups based on their grammatical types, structural characteristics and meanings, and functional characteristics. The studies by Biber et al. (1999) and Hyland's (2008) presented the most frequent structural patterns of 4-word expressions in English academic writing, and the results indicated a strong grammatical correlation in the formation of recurrent expressions. Over the years, recurrent expressions which are structurally classified by Biber et al.'s (1999) have been adapted for further studies with a focus on the aim and discipline of the data. In this study, depending on the Salazar's (2011, p. 50) five new categories adaptation such as "other noun phrases, other adjectival phrases, verb phrases with personal pronoun we, other passive fragments, and other verbal fragments", Güngör's (2016) exclusion of copula be + noun phrase/adjective phrase, and using (verb phrase or noun phrase) + that-clause fragment instead of (verb phrase +) that-clause fragment were adapted. To prevent any confusion, the sample of structural classifications are given in the following table.

**Table 1.**

## Structurally Classified Recurrent Expressions

Noun Structures	Noun phrase with of-phrase fragment, Noun phrase with other post-modifier fragments, other noun phrases
Prepositional Phrase Fragments	Prepositional phrase with embedded of-phrase fragment, Other prepositional phrases (fragment)
Verb Structures	Anticipatory <i>it</i> + verb phrase/adjective phrase, Passive verb + prepositional phrase fragment, Other passive fragments, (Verb phrase / noun phrase +) that-clause fragment, (Verb/adjective +) to-clause fragment, Pronoun/noun phrase + be (+...), other verbal fragments
Other Structures	Adverbial clause fragment, other adjectival phrases, other expressions

Adapted from Biber et al. (1999, pp. 1015-1024); Güngör, (2016, p. 3).

Structural categorization of recurrent expressions followed functional classification proposed by Biber, Conrad and Cortes (2003), later developed by Hyland (2008). Recurrent expressions are classified into three major functions according to their discourse objectives: research-oriented, text-oriented, and participant-oriented expressions. In Table 2 below, the functional classification presenting information about the functional features and explaining their most frequently used examples in discourse is given. Functional and structural characteristics have a significant role in academic writing and need recognition to contribute to the field because of their communicative purposes.

**Table 2.**

## Functionally Classified Recurrent Expressions

Research-oriented bundles: Help writers to Structure their activities and experience of the real world	Location-indicating time/place. Procedure, Quantification, Description, Topic-related to the field of research
Text-oriented bundles: Concerned with the organization of the text and its meaning as a message or argument	<i>Transition signals-establishing additive or contrastive links between elements</i> <i>Resultative signals-mark inferential or causative relations between elements</i> <i>Structuring signals-text-reflective markers which organize stretches of discourse and direct reader elsewhere in text</i> <i>Framing signals-situate arguments by specifying limiting conditions</i>
Participant-oriented bundles: These are Focused on the writer and reader of the text .	Stance features-convey the writer's attitude and evaluations, Engagement features-address readers directly

Adapted from Hyland, 2008, as cited in Güngör, 2016, p. 23

## 4. Methodology

To make a comprehensive analysis, this study explored the following questions below.

1. What are the most common four-word recurrent expressions found in the computerized corpora: TurkCorp, and EngCorp?
2. What do these recurrent expressions have in common based on the frequency data?
3. What are the similarities and differences between L1 and L2 corpora regarding structural and functional features?

### 4.1. Research Model/Design

The analysis was done by using corpus methodology by the use of online Sketch Engine tool and with two sub-corpora named as TurkCorp and EngCorp. Each corpus is made up of approximately ten academic articles. While EngCorp is native corpora, TurkCorp contained MA theses written by non-native writers. TurkCorp, a non-native corpus, was compared with the native-corpora EngCorp for possible similarities and differences of recurrent expressions.

First, the two corpora were classified according to various learner variables to measure the structural and functional differences. Second, the two sub-corpora were compiled, and the texts were converted to plain texts to be ready for use. The initial analysis yielded frequency of use and contextual data among the two corpora. The criteria for selecting the MA theses were several.

- (a) The first one was that the MA theses selected must be open access, which means that they are freely accessed and downloaded.
- (b) The second criterion was that the selected MA theses must be written in English only.
- (c) The third criterion was that the theses must be written by at least one native speaker of English, and the other must be written by a non-native speaker.
- (d) The final criterion was that the theses selected for analysis must be from the field of Applied Linguistics and/or English Language Teaching (ELT).

### 4.2. Data Collecting Tool and Sampling

An electronic online concordance tool was employed to investigate, classify and define recurrent expressions in various texts, and word lists were created to reveal the 50 most frequently used recurrent expressions for each corpus. In the last decade, employing software and online tools for the investigation of large samples of language gained attention in linguistic studies. Sketch Engine is one of the online tools that work with large samples of texts. As seen in Figure 1, the text analysis provides information on grammatical patterns of a word and phrase with several functions such as typical word combinations, keywords in contexts, synonyms, and translations. On a text-wide analysis, the tool offers word lists, bilingual terminology, and parts of speech tags to conduct further analyses. In addition to the grammatical functions, Sketch Engine calculates statistical information for the basic parameters of word frequencies in texts and statistics such as *T-score*, *MI-Score*, *log-likelihood*, *logDice*, etc. for the analysis and allows you to carry out analyses on personalized corpora.

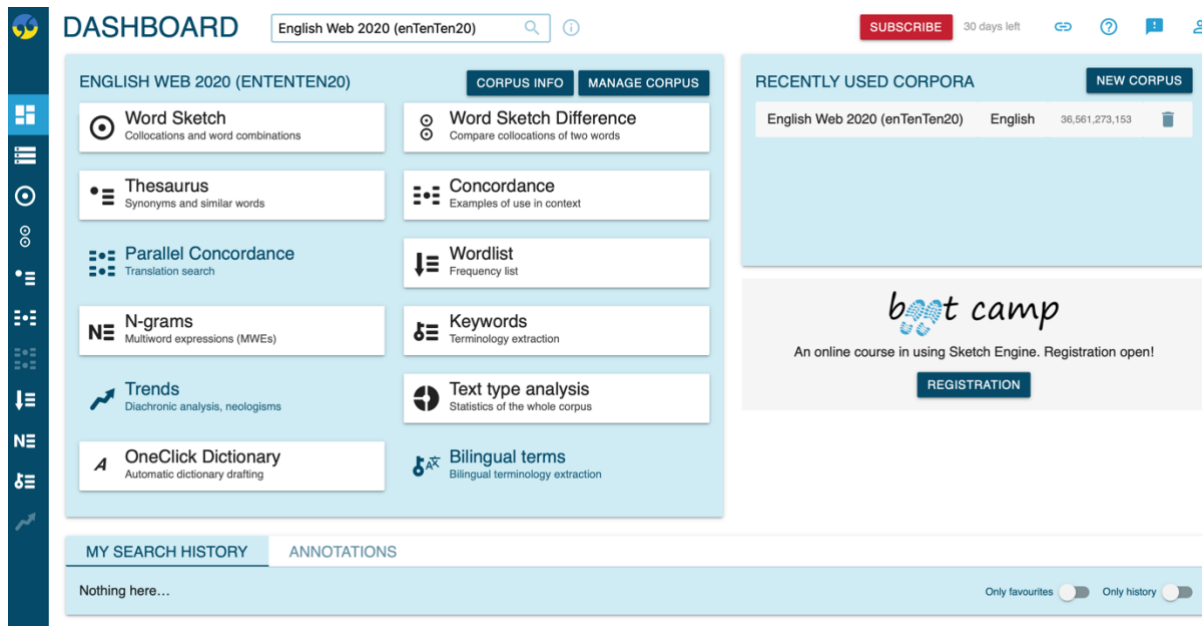


Fig. 1. Screenshot of the Sketch Engine dashboard, available for free at <https://www.sketchengine.eu>

Sketch Engine online tool is widely used by language students to observe word frequencies and constructions (Kilgarriff et al., 2014). As seen in Figure 2, the online tool allows learners to analyse frequency of the words in corpus to obtain basic ideas about it. The data reveal how language usage can vary depending on its context e.g., academic, spoken, written English.

The screenshot shows the Wordlist page for the English Web 2020 (enTenTen20) corpus. It displays a table of the most frequent words in the corpus. The table has 5 columns, each with 'Word' and 'Frequency' headers. The data is as follows:

Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency
1 the	2,180,184,828	11 with	282,999,141	21 at	178,746,595	31 will	111,834,957	41 if	82,481,967
2 and	1,099,770,995	12 on	280,411,737	22 from	173,725,845	32 his	109,230,647	42 n't	76,693,596
3 of	1,066,888,596	13 i	251,214,266	23 or	157,963,168	33 has	104,771,553	43 there	76,581,203
4 to	978,106,091	14 as	246,758,666	24 have	151,974,703	34 can	103,876,683	44 were	73,869,505
5 a	808,023,795	15 was	214,380,076	25 not	141,215,800	35 all	96,946,760	45 do	73,002,929
6 in	723,684,162	16 you	203,909,477	26 an	136,964,390	36 their	94,295,179	46 so	72,425,894
7 is	428,536,637	17 are	200,151,008	27 he	134,757,848	37 which	92,441,096	47 about	72,353,555
8 for	381,215,827	18 this	199,009,396	28 we	125,074,879	38 one	91,063,513	48 when	70,713,478
9 that	371,233,979	19 be	196,464,347	29 but	123,346,162	39 your	87,489,801	49 had	70,545,370
10 it	284,654,234	20 by	183,669,008	30 they	113,309,850	40 more	86,219,802	50 also	70,097,523

Fig. 2. Screenshot of the Wordlist page of the English Web 2020 (enTenTen20) corpus

Table 3 gives us quantitative data and a total of 20 academic articles were compiled. Then, the recurrent expressions were classified based on their structural and functional features. Structural analysis was based on Biber et al.'s (1999) structural taxonomy and the functional analysis was based on Hyland's (2008) functional taxonomy of academic registers. In this study, we focused on a specific academic register of MA theses and therefore this taxonomy was used in the classification of expressions. The general categories in this taxonomy are given in the analysis section.

**Table 3.**

Computerized Corpora Used for the Current Study and the Sampling

	<b>TurkCorp</b>	<b>EngCorp</b>
Tokens	364,197	486,068
L1	Turkish	British English
Genre	Academic	Academic
Number of texts	10	10
Tool	Sketch Engine	Sketch Engine

The sampling of the study was Turkish and British academic writers (native and non-native writers). Their essays mainly based on academic argumentative topics were selected as research data. For the study, 4-word recurrent word combinations were analysed and the rationale for focusing on four-word bundles were given in the previous section. The frequency cut-off used in the identification of expressions was set at 15 times per million words. For the purpose of avoiding idiosyncratic expressions, it was decided that an expression would have to appear in at least three different theses to be analysed. The expressions found during the analysis should not be accepted as complete structural units other than linking the two structural elements. Direct comparisons were made among all two sub-corpora to identify frequencies. The whole corpora compiled for the study are comprised of a total of twenty MA theses Applied Linguistics and which are published open-either in part or whole- in the freely accessible internet data sources or university online library sources. The TurkCorp and EngCorp corpora included data from similar topics such as “Academic and expository writing, the use of technology and language teachers, cooperative learning, perception of learner autonomy, content-based instruction, corpus linguistics”. The rationale for the selection of the MA theses was also that they contained several four-word recurrent expressions identified in the preliminary data collection process. Table 4 gives initial data related to the selected works.

**Table 4.**

Topical content of TurkCorp and EngCorp

<b>Corpus</b>	<b>Selected Topics</b>
T-1, T-3, E-1	Academic and argumentative writing and ESL/EFL learners
T-2, E-2, E-3	The use of technology and language teachers and learners
T-6, E-4	Cooperative Learning
T-4, E-5, E-6	Metacognitive Strategy Training
T-5, T-6, E-7,	Corpus Linguistics
T-7, T-10, E-9, E-10	Learner autonomy
T-8-E-8	Content-based instruction

## 5. Findings and Discussions

Osipova (2020) considered corpus as an “educational experiment” where corpus-based language learning materials are designed to meet the needs of the Turkish learners using corpus-based classroom materials which will be directed towards developing the learners’ linguistic and creative thinking abilities. In this study, we aimed at investigating the various recurrent expressions. Frequency analysis followed the structural and functional classification of the recurrent expressions with four-word strings. Significant differences were noted in the two corpora and the most frequent 50 recurrent expressions are given below.



**Table 5.**

The Most Frequent 50 Recurrent expressions

	<b>TurkCorp</b>	<b>Norm</b>	<b>EngCorp</b>	<b>Norm</b>
1	on the other hand	6.97	in the present study	1.60
2	the students in the	2.22	the end of the	1.52
3	the findings of the	1.83	of pre-service EFL teachers	1.48
4	on the basis of	1.75	computers in their teaching	1.25
5	the results of the	1.51	of the students in	1.21
6	the majority of the	1.45	levels of writing proficiency	1.21
7	that there is a	1.31	in the context of	1.19
8	at the end of	1.26	per cent of the	1.15
9	in their interaction with	1.23	on the other hand	1.15
10	to use the dyned	1.20	at the end of	1.15
11	is one of the	1.18	of learner autonomy in	1.11
12	the end of the	1.15	for the present study	1.02
13	at the same time	1.15	in terms of the	0.96
14	there is no significant	1.12	it is important to	0.94
15	as seen in the	1.12	in the field of	0.92
16	as a result of	1.12	in the English class	0.88
17	a chi-square was applied	1.09	in the current study	0.86
18	to find out the	1.07	in terms of their	0.84
19	to be able to	0.98	at the same time	0.84
20	the aim of the	0.98	the beginning of the	0.82
21	of the fact that	0.98	the results of the	0.80
22	in the process of	0.98	as well as the	0.80
23	the concept of learner	0.96	the extent to which	0.78
24	one of the most	0.96	in the proficient group	0.72
25	in the light of	0.96	as a result of	0.72
26	use of the dyned	0.93	an important role in	0.72
27	of the students in	0.90	a wide range of	0.72
28	the speech act of	0.87	the findings of the	0.69
29	the concept of autonomy	0.87	learner autonomy in the	0.69
30	seen in the table	0.87	a sense of efficacy	0.69
31	on the part of	0.87	the majority of the	0.65
32	it can be said	0.87	end of the course	0.65
33	can be said that	0.87	as shown in table	0.65
34	small number of the	0.85	to be able to	0.63
35	in a such way	0.85	in the case of	0.63
36	data obtained from the	0.85	there was no significant	0.61
37	the use of the	0.82	the ministry of education	0.61
38	in accordance with the	0.82	at the beginning of	0.61
39	is no significant diff.	0.79	in the next section	0.55
40	of learner autonomy in	0.76	to the present study	0.53
41	with respect to the	0.74	on the basis of	0.53
42	the data obtained from	0.74	in addition to the	0.51
43	it is seen that	0.74	to the fact that	0.49
44	it can be seen	0.74	the analysis of the	0.49
45	in the field of	0.74	most of the students	0.49
46	the part of the	0.71	in the process of	0.49
47	the main purpose of	0.71	as part of the	0.49
48	a small number of	0.71	the purpose of the	0.47
49	with the help of	0.63	for the purpose of	0.47
50	it is possible to	0.63	A small number of	0.39

The table above gives the most frequent four-word recurrent expressions based on their frequencies. When the normalized frequencies of these expressions were compared, it was found that *on the other hand* is the most frequent lexical expression in the Turkish corpus with an occurrence of 6.97 times per ten-thousand

words, but the same functional lexical expression was used 1.15 times in EngCorp. The second most frequent expression was *the students in the* in TurkCorp.

*On the basis of* is used 1.75 times by the TurkCorp and 0.53 times by the EngCorp, EFL learners largely used framing and causative expressions from the text-oriented category, which are *on the basis of* and *the results of the*. These two expressions were used more than a hundred times by TurkCorp. On the other hand, they were used with lower frequencies in EngCorp. Native speakers used them only 26 times. Another finding is that Turkish researchers used *is one of the* in the grouping category of the research-oriented expressions 43 times, but this expression was not used by native corpora at all. Unlike TurkCorp, native speakers seem to have used stance expression of *it is important to* from the participant-oriented expressions.

Indicating location from the research-oriented category, *the end of the* expression was used in the two corpora in varying frequencies. TurkCorp tended to use this expression with a frequency of 42 times, but it was more frequent in EngCorp with 74 times.

#### 4.1. Understanding Structural Characteristics through Online Sketch Engine Tool

Table 6 below shows the noun structures used in the two corpora. When the two sections are compared, it is seen that Turkish writers mostly used structural types such as *the findings of the*, *the results of the*, *use of the dynd*, *concept of learner autonomy* and *theme-based model of cbi* while English writers used more noun structured recurrent expressions such as *the beginning of the*, *the end of the*, *the results of the* and *the purpose of the* in comparison to other structures.

**Table 6.**

Noun Structures in TurkCorp and EngCorp

<b>Noun phrase + of-phrase fragment</b>	
TurkCorp	the findings of the, the results of the, the end of the, the aim of the, one of the most, the concept of learner, majority of the respondents, use of the dynd, concept of learner autonomy, the speech act of, the concept of autonomy, the majority of the, percent of the students, small number of the, the use of the, the part of the, the main purpose of, ministry of national education, a small number of, needs of the students, their choice of strategies, this part of the, the preparatory school of, interactions with members of
EngCorp	the beginning of the, the end of the, levels of writing proficiency, the results of the, a wide range of, the findings of the, a sense of efficacy, the majority of the, end of the course, the ministry of education, different levels of writing, the mean score of, the analysis of the, beginning of the course, the purpose of the, the promotion of learner, the writing proficiency of, promotion of learner autonomy, mean scores of the, most of the students
<b>Noun phrase with other post-modifier fragments</b>	
TurkCorp	a significant difference between, attitudes towards the dynd, the data obtained from, students in the experimental, their attitudes towards to, their interactions with members, the difference between the, the students in the, teachers' attitude towards the, positive attitudes towards the
EngCorp	the students in group, students in group b, students in group a, computers in their teaching, integrating computers in the, computers in the english, course integrating computers in, the extent to which, students in both groups, the students in the, an important role in, learner autonomy in the, the course integrating computers, the learner autonomy-focused instruction, raters in the proficient, technology in their teaching, significant difference between the, tertiary education in Vietnam, responsibility for their own
Other noun phrases	EFL and N.S.P. group, and N.S.P. groups in the learner training program, learning contract and learning, the pre-task and post-task, contract and learning diary

Another common structure among the two corpora was prepositional phrase fragments and this finding is concurrent with the findings of Biber et al. (1999). The prepositional phrase fragments in TurkCorp and EngCorp are shown below in Table 7 below.

**Table 7.**

## Prepositional Phrase Fragments in TurkCorp and EngCorp

<b>Prepositional Phrase Fragments in TurkCorp</b>	
Prepositional phrase + of-phrase fragment	on the basis of, as a result of, in the process of, in the light of, on the part of, in the field of, towards the promotion of, with the help of, at the end of
"of-phrase" fragment	of the fact that, of the students in, of learner autonomy in, of the respondents strongly
Other prepositional phrases	on the other hand, in their interactions with, in the experimental group, at the same time, as seen in the, in the preparatory school, in such a way, in accordance with the, from using the dynded, with respect to the, in the target language, in the control group, such a way as, in addition to these, as a foreign language, to the fact that, in the current study
<b>Prepositional Phrase Fragments in EngCorp</b>	
Prepositional phrase + of-phrase fragment	in the context of, in terms of the, in the field of, in terms of their, in the case of, on the basis of, in the form of, as part of the, for the purpose of, in the use of, to the development of, as a result of, at the beginning of, in the process of, at the end of,
"of-phrase" fragment	of the students in, of learner autonomy in, of the present study, of the learning process, of computer technology in, of computers in their, of pre-service efl teachers
Other prepositional phrases	for the present study, on in the present study, in the english class, in the current study, in the proficient group, on the politeness dimension, to the fact that, as shown in table, in the computer room, in the next section, for their own learning, to the present study, between the students in, in addition to the, on the other hand

The results of the native and non-native researchers' choice of prepositional phrase fragments reveal that the uses of prepositional phrases except for *prepositional phrase + of-phrase fragment* show a variety of patterned structure. While the native corpora illustrated the only "of-phrase" fragment such as *of the students in, of learner autonomy in, of the present study, of the learning process, of computer technology in, of computers in their, of pre-service efl teachers* the non-native TurkCorp used those such as *of the fact that, of the students in, of learner autonomy in, of the respondents strongly*.

The native and non-native researchers used nearly a number of prepositional phrases (prepositional phrase + *of*), other categories and noun phrases. This is indicated by Hyland (2008b), who stated that the *of*-phrase fragment carries various meaning and functions in academic discourse, identifying quantity (*the small number of*), place (*in the field of*), qualities (*the use of the*), etc. *Of*-phrase structure indicated logical relations between the propositions.

The last structural characteristics of recurrent expressions is based on verb-based structures which are also referred as clausal recurrent expressions. Verb-based structures can be composed of a verb or adjective with a to-clause fragment, or a verb phrase with a that-clause fragment (Bal, 2010). As shown in Table 8, the native and non-native researchers used a similar number of prepositional phrases.

**Table 8:**

## Verb-Based Structures in TurkCorp and EngCorp

<b>Verb-Based Structures in TurkCorp</b>	
Passive verb + prepositional phrase fragment	can be said that
Other passive fragments	a chi-square was applied, chi-square was applied to
that-clause fragment	that there is a, that make teachers abstain, the respondents state that it can be said, it is seen that, it can be seen, it is possible to
adjective phrase (verb + adjective +) to-clause fragment	to use the dynd, to find out the, to be able to
Pronoun/noun phrase + be (+...)	there is no significant
Other verbal fragments	seen in the table, teachers abstain from using, respondents state that they
<b>Verb-Based Structures in EngCorp</b>	
Passive verb + prepositional phrase fragment	will be discussed in, be pointed out that, should be pointed out
Other passive fragments (verb phrase/noun phrase +)	students were asked to, it should be pointed -
that-clause fragment	
adjective phrase (verb + adjective) to-clause	it is important to To be able to, to participate in the
Pronoun/noun phrase + be (+...)	there was no significant
Other verbal fragments	the present study is

However, the researchers in the native corpus used passive verb+ prepositional phrase fragments more than their non-native counterparts as well as a similar number of that-clause fragments. The passive verb + prepositional phrases show a locative or logical relation (Hyland, 2008b). Native speakers used fewer anticipatory it + verb phrase structures when compared to non-native researchers.

#### 4.2. Understanding Functional Characteristics

Hyland's (2008) taxonomy which is based on the taxonomy of Biber's (Biber et al., 2004, Biber, 2006) was used as a framework in the categorization of the functions of the expressions. It was Hyland, who classified the recurrent expressions in various functional categories. In Table 9 below, the expressions and their functions are presented. It is seen that the non-native and the native researchers used research-oriented expressions almost in similar number. These expressions are generally used for writers to structure their activities, including expressions referring to several categories. The first functional feature of recurrent expressions is research-oriented expressions. This expression consisted of five subdivisions as location, procedure, quantification, description and grouping.

**Table 9.**

## Research-Oriented Expressions

	<b>T.R. (f)</b>	<b>ENG. (f)</b>
Location	in the experimental group (43) in the preparatory group (40) in the control group (26) at karadeniz technical uni.(21) in this part of (17) the beginning of the (15) in the department of (15) at the beginning of (12) in the preparatory program (11)	in the English class (43) the beginning of the (40) in the proficient group (35) end of the course (32) in the computer room (31) at the beginning of (30) in the next section (27) in the efl classroom (15) in the next chapter (11) at the time of (11) by the end of (10)
Procedure	use of the dyned (34) data obtained from the (31) develop their skills of (27) was applied to the (22) applied to the data (19) in their learning process (18) in the data analysis (18) teachers were asked to (12) the study was conducted (11)	the analysis of the (24) participants were asked to (15) the development of the (13) used in this study (12)
Quantification	number of the respondents (59) the majority of the (53) percent of the students (32) small number of the (31) a small number of (26) average number of the (19) an average number of (18) a wide range of (13)	a wide range of (35) the majority of the (32)
Description	the aim of the (36) the concept of learner (35) the main purpose of (26) the role of writing (20) the use of concordance (20) is one of the (43) the part of the (26)	the role of the (21) the focus of the (20) the importance of the (18) a high level of (16) the quality of the (10)
Grouping	as one of the (15) the rest of the (10) one of the students (10)	as part of the (24) as one of the (22) the rest of the (10)

It is seen that Turkish and British researchers used different examples of research-oriented expressions in their articles. For example, Turkish researchers mostly used the expression *in the experimental group* (43 times) in the location section. On the other hand, British speakers mostly emphasized *in the English class* expression used 43 times in the location section. In the procedure section, however, the number and variety of recurrent expressions were limited in EngCorp but more in TurkCorp. In the quantification category, EngCorp contained almost no expressions. In both corpora, the description category included several recurrent expressions, but the grouping category included several expressions in TurkCorp and few others in EngCorp.

**Table 10.**

Similar Research-Oriented Expressions / Similarities with their frequencies

	<b>T.R. (f)</b>	<b>ENG (f)</b>
Location	at the end of (46) the end of the (42) in the field of (27)	the end of the (74) at the end of (56) in the field of (45)
Procedure	in the process of (36)	in the process of (24)
Quantification	the extent to which (18) the use of the (30)	the extent to which (38) the purpose of the (23)
Description	the nature of the (21) the purpose of the (11) the content of the (10)	the use of the (20) the content of the (19) the nature of the (17)
Grouping	one of the most (35)	one of the most (16)

In Table 10, similar research-oriented expressions are given. The most frequent expression from the location category is *at the end of* with a total frequency of 102. The most frequent use of this expression was in EngCorp with a frequency of 56. In the quantification category, the *the extent to which* expression is used almost two times more in EngCorp. In the grouping category, however, *one of the most* is used more in TurkCorp.

The variety in the number of lexical expression patterns and their frequencies observed in this similarity list may present indirect connections. For example, Hyland (2008b) argues that several functional lexical expression categories are closely related to the structural patterns. One of these categories is noun phrase + *of category* and they are very common in research-oriented functions. Prepositional phrase patterns on the other hand are used frequently as text-oriented functions. The fact that the non-native researchers used research-oriented and text-oriented expressions in similar number when compared to the native authors are concurrent with our findings.

**Table 11**

Text-Oriented Expressions / Differences with their frequencies

	<b>T.R. (f)</b>	<b>ENG (f)</b>
Additive	at the same time (42)	at the same time (41)
Comparative	no significant difference between (23)	—
Inferential	the difference between the (23) it is seen that (27) it was found that (17) the results showed that (10)	that there was no (21) there was a significant (14) that there was a (11)
Framing	on the basis of (64) in the light of (35) with respect to the (27) with the help of (23)	in the case of (31) on the basis of (26) in relation to the (22)
Structuring	as seen in the (41) can be seen that (23) in the present study (16) to find out the (39)	in the present study (78)
Objective	for the main purpose (21) in order to understand (19) in order to see (17) in order to make (16) in order to get (16)	

Text-oriented expressions were listed for each corpus in Table 11 above. The most frequent examples of text-oriented expressions were found in TurkCorp. In EngCorp, there were few expressions except for inferential category. The text-oriented expressions include inferential (*there was a significant, the results revealed that, the results showed that, it was found that, there were significant differences, and that there was no*), framing (*in the case of, on the basis of, in relation to the*) and the structuring (*as shown in table, as shown in the, in the present study*)

**Table 12**

Text-Oriented Expressions / Similarities with their frequencies

	T.R. (f)	ENG (f)
Additive	on the other hand (254) in addition to the (24) as well as the (16)	on the other hand (56) as well as the (39) in addition to the (25)
Comparative	a significant difference between (35)	-
Inferential	that there is a (48) there is no significant (41)	there was no significant (30) that there is a (14)
Causative	the findings of the (67) the results of the (55) as a result of (41)	the results of the (39) as a result of (35) the findings of the (34)
Structuring Framing	in the current study (23) in terms of the (17)	in the current study(42) in terms of the (47)
Objective	to be able to (36) for the purpose of (18) in order to be (17)	to be able to (31) for the purpose of (23) in order to see (10)

Similar text-oriented expressions are given in Table 12. *On the other hand* expression was mostly used in TurkCorp with a frequency of 254. The high tendency to use this additive category expression in the non-native corpora can be given to several factors. British researchers also seem to have used similar recurrent expressions in almost all categories with similar frequencies. In the framing and structuring categories, the uses of *in terms of the* and *in the current study* consecutively by EngCorp seems outstanding.

The third and final functional characteristic is the participant-oriented expressions. These expressions belonged to the least common functional category among the two corpora.

**Table 13.**

Participant-Oriented Expressions / Differences and Similarities

Differences	T.R. (f)	ENG. (f)
Stance	it is possible to (23) it is interesting to (15) an important role in (12)	an important role in (35) it is necessary to (15) it is likely that (13) it is useful to (11) were more likely to (10)
Engagement	it can be said (32)	it should be noted (14)
Similarities	----	----
Stance	to the fact that (23) the fact that the (20) it is important to (16)	it is important to (46) to the fact that (24) the fact that the (24)
Engagement	it can be seen (27)	it can be seen (20)

As seen in Table 13, EngCorp varied in stance expressions and used the most participant-oriented expressions among two corpora. The most frequent expression category was stance, and *it is important to* turned out to be the expression with a highest frequency of 46 by British researchers.

The overall results indicated that learners' use of recurrent expressions showed various similarities and differences in terms of genre-specific multiword combinations (recurrent expressions), requiring further attention and pedagogical focus. Determining the frequency data related to the recurrent expressions used by native and non-native writers in two electronic corpora are significant since

- (a) they illustrated how many times recurrent expressions were used in different contexts by different writer groups.
- (b) frequency information informed us about the native and non-native writers' usage patterns.
- (c) computerized corpus tools that have potential to be integrated into the language classrooms were used.
- (d) the possible changes as well as similarities and differences among the lexical bundles were clearly illustrated in two writer groups
- (e) the method of identifying lexical bundles in two different corpora through norming the frequencies to 15 cases per million words was applied using a fixed distribution as a cut-off criterion.

The results revealed that the non-native corpus (TurkCorp) had several shared features in terms of variations and frequencies in the type and number of expressions. The primary data obtained revealed that non-native writers mostly used less varied and limited number of expressions such as *on the other hand* and *the students in the, the findings of the, on the basis of, the results of the*. Comparisons between the two groups also revealed that TurkCorp and the EngCorp showed slightly different frequencies in the number of certain recurrent expressions. Native writers used mostly vernacular recurrent expressions when compared with the non-native writers, using several different recurrent expressions in their conceptual fields. It was seen that the British English researchers used more diverse expressions with a higher visibility. The overuse and underuse patterns in both groups should also be interpreted in terms of frequency data and the existing four-word recurrent expressions.

Recurrent expressions mostly employed were classified into three grammatical structures: noun structures, verb-based structures, and prepositional phrase structures. Being the most used structure of recurrent expressions, noun phrase structures were the primary structural characteristics of the recurrent expressions. In this regard, Biber et al. (2014, p. 9) focused on the role of noun phrases for "informational communicative purposes" in academic writing. They included noun phrase + of-phrase fragment, noun phrase with other post-modifier fragments and another noun phrase. These noun phrases have several characteristics in common. Such phrases as "*the course integrating computers, the learner autonomy-focused instruction, raters in the proficient, technology in their teaching*" are distinctive noun phrase recurrent expressions only intrinsic to native corpora. The phrases seen in non-native corpus can be interpreted as the variation of prepositional recurrent expressions between researchers who have different language backgrounds. To name a few, *as a result of, on the basis of, in the process of* and *on the other hand* turned out to be the most common prepositional phrase fragments of recurrent expressions structurally. The expressions *in terms of the, in the use of* and *as a result of* were mutually used in all corpora. Regarding the native and non-native researchers' choice of prepositional phrase fragments, the results revealed that the uses of prepositional phrases except for *prepositional phrase + of-phrase fragment* varied from native to non-native corpus.

The differences and similarities in terms of clausal expression usage patterns among the two corpora were also observed. In comparison to noun and prepositional structures, verb-based structural examples could not be found many in both TurkCorp and EngCorp. The usage of verb-based structures was different and only one expression was used in common in the group of (verb+adjective+) to-clause fragment. When compared, it was seen that TurkCorp's verb-based structures far outweighed. Therefore, similar to Güngör's



(2016) findings, it is seen that Turkish writers employed more verb phrase-based structures than the native English writers. Moreover, it is worth mentioning that EngCorp did not include any examples in the category of (verb phrase/noun phrase+) that-clause fragment, but the non-native corpus did.

Bychkovska and Lee's (2017) study on comparing lexical expression usage of Chinese and English students in argumentative essays showed similar results in verb phrase expressions. Chinese writers, unlike native writers, preferred to use more verb phrase expressions than prepositional and noun phrase expressions. Shin (2019) conducted a study between L1 and L2 writers, observing that both groups used similar recurrent expressions such as frequently used verb phrase expressions, stance expressions and prepositional phrase expressions. Muşlu (2018) conducted research on the usage of four-word stance recurrent expressions among Turkish and Japanese EFL learners in comparison with native writers and concluded that L2 writers showed overuse patterns with higher frequency. However, different from the present study, verb phrase expressions were the mostly used expression type. Pearson (2021), in his comparative study on recurrent expressions found out that novice L2 writers mostly used clausal expressions structurally. Text-oriented additive and stance expressions were the most functional category. Similarly, Hong (2019) observed in a longitudinal study that the more learning is progressed, the frequent use of clausal expression shifted to phrasal expression usage in academic writing. Cooper (2013), in a corpus-based study on four-word recurrent expressions, found that as the students' writing skills improved, their use of verb-based expressions decreased, and the noun-based expressions increased in usage. Akbulut (2020) found out that in non-native writing, verb-based and clause-based expressions were used more. Functionally, text and stance-oriented expressions were used at the highest level by the non-native writers, whereas native writers used mostly research-oriented expressions.

The next classification was based on functional characteristics with the three major categories. (Hyland, 2008a). Research-oriented and text-oriented functions were investigated and compared with structural characteristics. The research-oriented expressions were the most frequently used in the non-native corpus, followed by the text-oriented expressions. Furthermore, the two corpora were compared based on their four-word recurrent expressions' functional characteristics. Text-oriented expressions were displayed as the most common functional category in non-native corpus in comparison to other corpora. In the Turkish EFL context, Öztürk and Köse (2016) analysed recurrent expressions in terms of frequency, structure, and function among Turkish and English postgraduate students. The results indicated that Turkish writers used recurrent expressions more frequently but repeatedly. However, the structural and functional comparison showed that there was no significant difference in between the two groups of writers. The findings of this study showed that the causative oriented expressions were not used in any of the corpora. The limited number of stance expressions, on the other hand, may be due to the two reasons. The first may be that the writers may express stance and engagement in alternative ways rather than using four-word recurrent expressions (Hyland, 2005; Biber, 2006). Secondly, novice authors may abandon using stance expressions to argue their claims (Pérez-Llantada, 2014, as cited in Güngör, 2016, p.113).

## 6. Conclusion and Suggestions

The current study used online corpus software as part of educational technology and electronically analysed recurrent expressions used by native and non-native researchers and provided an insight into the different and similar usage patterns of the four-word recurrent expressions as evidenced in the two corpora. In the analysis of these expressions, an online corpus tool was used, and structural and functional aspects of expressions were analysed and presented in the form of tables. The analysis was done by using a very popular online corpus tool and the word list function was used to reveal the most frequent patterns. Corpus-based quantitative results showed the frequencies of the recurrent expressions both structurally and functionally and the resulting data revealed similar structural and functional characteristics of four-word recurrent expressions.

These findings are significant in that it is possible to claim that the research-oriented expressions can be considered as the most commonly used expressions, and the differences among the two corpora outweighed the similarities. In future studies, the usage patterns of various other recurrent expressions in L1 and L2 should be investigated in an interdisciplinary fashion.

Another significant finding was the demonstration of how corpus technology can be used for learning as well as discovering new truths about language. Being an active learning tool, corpus has a great potential for presenting extensive data and creating new tasks for language learners in the form of electronic data. Several educational experiments on the usefulness of corpus linguistic technology proved that this technology gives methodically reliable data and can be used effectively in the teaching and learning processes as well as promoting collaborative learning and interaction (Martin et al., 2017; Fanilevna et al., 2020).

The findings can be used for several purposes. First, from a pedagogical point of view, the preferences for the most frequent recurrent expressions by Turkish writers could be categorically used in the design of better classroom materials based on academic or expository content. Coursebook writers and material developers can design new materials with the target word combinations enhancing the learnability of these combinations. Second, since recurrent expressions constitute a significant part of academic and expository writing, writing teachers should be informed about the salience and frequency of the academic recurrent expressions so that they would prioritize the teaching of these words in their classrooms. Third, explicit teaching of the most frequent recurrent expressions could be fostering the Turkish learners' acquisition of recurrent word combinations

The findings obtained in this study should be treated with caution. First, recurrent expressions found in this analysis may not be the only ones as a discursive practice. Some other recurrent expressions may also be used for discursive practices and for creating meaning in various levels and performing function in different categories. Despite the analysis was limited to the 4-word recurrent expressions in MA theses in Applied Linguistics, it would be useful to extend the scope of recurrent expressions to be identified in different fields separately and compare them with each other in the future studies. What is more, yet other studies may be done to compare recurrent expressions functionally and structurally in different theses, dissertations, or articles in other fields as well.

## References

- Absalom, M. (2021). Digital corpora: Language teaching and learning in the age of big data. In T. Beaven & F. Rosell-Aguilar (Eds), *Innovative language pedagogy report* (pp. 97-101). Research-publishing.net. <https://doi.org/10.14705/rpnet.2021.50.1242>
- Adams, S. D. (1984). Medical discourse: Aspects of author's comment. *English for Specific Purposes*, 3(1), 25-36. [https://doi.org/10.1016/0272-2380\(84\)90004-0](https://doi.org/10.1016/0272-2380(84)90004-0)
- Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92. <https://doi.org/10.1016/j.esp.2011.08.004>
- Aijmer, K. (Ed.). (2009). *Corpora and language teaching*. John Benjamins Publishing Company.
- Akbulut, D., F. (2020). A bibliometric analysis of lexical expressions usage in native and non-native academic writing. *Journal of Language and Linguistic Studies*, 16(3), 1146-1166. <https://doi.org/10.17263/jlls.803583>
- Alcaraz Varó, E. (2000). El inglés profesional y académico. *El libro universitario. Serie filología y lingüística* (pp. 317-319). Alianza.

- Allison, D. (1995). Assertions and alternatives: Helping ESL undergraduates extend their choices in academic writing. *Journal of Second Language Writing*, 4(1), 1-15. [https://doi.org/10.1016/1060-3743\(95\)90020-9](https://doi.org/10.1016/1060-3743(95)90020-9)
- Bensaid, M. (2015). Arab ESL students and modals. *Arab World English Journal*, 6(4), 90- 97. <https://doi.org/10.2139/ssrn.2843929>
- Beute, O. (2020). *How Norwegian students of English and English native speakers use the modal auxiliary must and the quasi-modal have to: A contrastive study* [Master's thesis, University of Oslo]. University of Oslo Library.
- Bloor, M., & Bloor, T. (1991). Cultural expectations and socio-pragmatic failure in academic writing. In P. Adams, B. Heaton, & P. Howarth (Eds.), *Socio-cultural issues in English for academic purposes* (pp. 1-12). Modern English Publications/British Council.
- Brandt, S. (1999). *Modal verbs in Danish (Travaux du Cercle linguistique de Copenhague)*. C.A. Reitzel.
- Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1– L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in Second Language Acquisition*, 30(2). 225–251. <https://doi.org/10.1017/S0272263108080327>
- Btoosh, M.A. (2019). Modals in Arab EFL students' composition: A corpus-based approach. *College of Sciences and Humanities Linguistics and Literature Studies*, 7(3), 100-109. <https://doi.org/10.13189/lls.2019.070302>
- Bybee, J., & Fleischman, S. (1995). Modality in grammar and discourse: An introductory essay. *Modality in grammar and discourse*, 14, 503-517. <https://www.unm.edu/~jbybee/downloads/BybeeFleischman1995Introduction.pdf>
- Byloo, P. (2009). *Modality and negation: A corpus-based study* [Unpublished doctoral dissertation]. The University of Antwerp.
- Campoy, M. C., Bellés-Fortuno, B., & Gea-Valor, M. L. (Eds.). (2010). *Corpus-based approaches to English language teaching*. Continuum.
- Chen, H. I. (2010). Contrastive learner corpus analysis of epistemic modality and interlanguage pragmatic competence in L2 writing. *Arizona Working Papers in SLA & Teaching*, 17, 27–51. <https://journals.librarypublishing.arizona.edu/jslat/article/245/galley/239/view/>
- Coates, J. (1983). *The Semantics of modal auxiliaries*. Croom Helm. <https://doi.org/10.1017/S0022226700013980>
- Cook, W. A. (1978). Semantic structure of the English modals. *TESOL Quarterly*, 12(1), 5-16.
- Cooper, T. (2013). Can IELTS writing scores predict university performance? Comparing the use of lexical bundles in IELTS writing tests and first-year academic writing. *Stellenbosch Papers in Linguistics Plus*, 42, 63–79. <https://doi.org/10.5842/42-0-155>
- Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics*, 34(3), 369-388. <https://doi.org/10.1093/applin/ams045>
- DeCapua, A., & Dunham, J.F. (2007). The pragmatics of advice giving: Cross- cultural perspectives. *Intercultural Pragmatics*, 4(3), 319–342. <http://krpb.pbworks.com/f/advice+decapua.pdf>
- Depraetere, I., & Verhulst, A. (2008). Source of modality: A reassessment. *English Language & Linguistics*, 12(1), 1-25. <https://doi.org/10.1017/S1360674307002481>

- Deshors, S.C., & Gries, S.T. (2014). A case for the multifactorial assessment of learner language: The uses of may and can in French-English interlanguage. In D. Glynn & J.A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 179–204). John Benjamins. <https://doi.org/10.1075/hcp.43.07des>
- Diewald, G. (2001). A basic semantic template for lexical and grammaticalized uses of the German modals. In J. van der Auwera & P. Dendale (Eds.), *Modal verbs in Germanic and Romance languages* (pp. 23-41). John Benjamins.
- Dittmar, N., & Terborg, H. (1991). Modality and second language learning: A challenge for linguistic theory. In T. Huebner & C. A. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory* (pp. 347–384). John Benjamins.
- Dutra, D.P. (1998). *The acquisition of English root modality by non-native speakers* (Publication No. 9837383) [Doctoral dissertation, University of Florida]. ProQuest Dissertations Publishing.
- Efstathiadi, L. (2011). Epistemic modality markers in L1 and L2 discourse of modern Greek: A corpus-based study. *Proceedings of the 31st Annual Meeting* (pp. 138-156).
- Enriquez, M. A. T. (2012). *Philippine English and Taglish from a language contact perspective* [Doctoral dissertation, National University of Singapore]. ScholarBank@NUS Repository.
- Erton, I. (2018). Difficulties in teaching English modal auxiliaries to Turkish students: A cognitive pragmatic approach. *Arab World English Journal*, 9 (4), 56- 68. <https://dx.doi.org/10.24093/awej/vol9no4.4>
- Fanilevna Mukhamadiarova, A., Fidarisovna Caserta, L., Alexandrovna Kulkova, M., & Reuter, K. (2020). Use of corpus' technologies for the development of lexical skills. *Proceedings of the 7th International Conference on Applied Linguistics Issues*, 449-456. <https://doi.org/10.22055/rals.2020.16345>
- Flor, A. M. (2003). Non-native speakers' production of advice acts: The effects of proficiency. *Revista Espanola de Lingüística Aplicada*, 16(2), 139–153. <https://rael.aesla.org.es/index.php/RAEL/article/view/92>
- Friginal, E., Dye, P., & Nolen, M. (2020). Corpus-based approaches in language teaching: Outcomes, observations, and teacher perspectives. *Boğaziçi Üniversitesi Eğitim Dergisi*, 37(1), 43-68. <https://dergipark.org.tr/en/pub/buje/issue/55539/760480>
- Giacalone Ramat, A. (1992). Grammaticalization processes in the area of temporal and modal relations. *Studies in Second Language Acquisition*, 14(3). 297–322. <https://doi.org/10.1017/S027226310001113X>
- Gibbs, D. (1990). Second language acquisition of the English modal auxiliaries can, could, may and might. *Applied Linguistics*, 11(3), 297-313.
- Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction*, 1(1), 41–61. <https://doi.org/10.1075/etc.1.1.05gil>
- Gong, W. (2019). An innovative English teaching system based on computer aided technology and corpus management. *International Journal of Emerging Technologies in Learning (iJET)*, 14(14), 69–80. <https://doi.org/10.3991/ijet.v14i14.10817>
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp.3-189). Addison Wesley Longman.
- Granger, S. (2013). *Learner English on computer*. Routledge.

- Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 17-21. <https://doi.org/10.1111/j.1467-971X.1996.tb00089.x>
- Güngör, F. (2016). *Crosslinguistic analysis of lexical bundles in L1 English, L2 English, and L1 Turkish research articles* [Doctoral Dissertation, Gazi University].
- Heine, B. (1993). Agent-oriented vs. epistemic modality: Some observations. In J. Bybee & S. Fleischmann (Eds.), *Modality in grammar and discourse* (pp.17-53). John Benjamins.
- Hinkel, E. (1995). The use of modal verbs as a reflection of cultural values. *TESOL Quarterly*, 29 (2), 324-341. <https://doi.org/10.2307/3587627>
- Holmes, J. (1982). Expressing doubt and certainty in English. *RELC Journal*, 13(2), 9-28. <https://doi.org/10.1177/003368828201300202>
- Holmes, J. (1983). Speaking English with the appropriate degree of conviction. In C. Brumfit (Ed.), *Learning and teaching languages for communication: Applied linguistics perspectives*. CILT.
- Holmes, J. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics*, 9(1), 21-44. <https://doi.org/10.1093/applin/9.1.21>
- Hong, J. Y. (2019). Structural use of lexical bundles in the rhetorical moves of L1 and L2 academic writing. *English Teaching*, 74(3), 29-54. <https://doi.org/10.15858/engtea.74.3.201909.29>
- Hornby, A. S., Turnbull, J., Lea, D., Parkinson, D., & Phillips, P. (2010). *Oxford advanced learner's dictionary of current English*. Oxford University Press.
- Hu, Z., Brown, D., & Brown, L. (1982). Some linguistic differences in the written English of Chinese and Australian students. *Language Learning and Communication*, 1(1), 39-49.
- Huddleston, R. (2002). The verb. In R. Huddleston & G. K. Pullum (Eds.), *The Cambridge grammar of the English language* (pp.71-212). Cambridge University Press.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. <https://doi.org/10.1177/1461445605050365>
- Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students writing. *Journal of Second Language Writing*, 6(2), 83-205. [https://doi.org/10.1016/S1060-3743\(97\)90033-3](https://doi.org/10.1016/S1060-3743(97)90033-3)
- Imperial, R. A. (2014). Lexical variation in Philippine English: The case of deontic MUST and HAVE TO. *Philippine Journal of Linguistics*, 45, 1-18.
- Kiliç, V. (2013). A contrastive study of Turkish and English modality with reference to speech act theory. *Dialectologia*, 10, 1-22. <https://core.ac.uk/download/pdf/39122681.pdf>
- Kilgariff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1, 7-36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kilimci, A., & Can, C. (2008). TICLE: Uluslararası Türk Öğrenci İngilizcesi derlemi, TICLE: Turkish international corpus of learner English. *22<sup>nd</sup> Ulusal Dilbilim Kurultayı*, 1-11.

- Kızılay, Y. (2019). Semi-modal verb “Need to” and the modality of obligation “Must & Have to” in authentic corpus-based English. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, 240-257. <https://doi.org/10.29000/rumelide.648857>
- Klecha, P. (2014). *Bridging the divide: Scalarity and modality* (Publication No. 3639345) [Doctoral dissertation, University of Chicago]. ProQuest Dissertations Publishing.
- Larreya, P., & Rivière, C. (2005). *Grammaire explicative de l'Anglais*. Pearson Education Limited.
- Laso, N. J., & Giménez, E. (2007). Bridging the gap between corpus research and language teaching. In C. Perrián (Ed.), *Revisiting language learning resources* (pp. 49-64). Cambridge Scholars Publishing.
- Letica, S. (2009). Use of epistemic modality by non-native speakers of English. In R. Lugossy, J. Horváth & M. Nikolov (Eds.), *UPRT 2008: Empirical studies in English applied linguistics* (pp. 119-134). Lingua Franca Csoport.
- Lewis, M. (2009). *The idiom principle in L2 English: Assessing elusive formulaic sequences as indicators of idiomaticity, fluency, and proficiency*. VDM Verlag.
- Li, Q. (2017). A corpus-based study of modal verbs use in English: Writing by EFL students. *Canadian Social Science*, 13(11), 31-35. <http://dx.doi.org/10.3968/9975>
- Ma, Q., Chiu, M., Lin, S., & Mendoza, N. (2023). Teachers’ perceived corpus literacy and their intention to integrate corpora into classroom teaching: A survey study. *ReCALL*, 35(1), 19-39. <https://doi.org/10.1017/S0958344022000180>
- Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 177-190. <https://doi.org/10.1515/jccall-2021-2008>
- Martin, N. J. L., Pujadas, E. C., Villanueva, M. L. C., & Guinjoan, M. F. (2017). Corpora and new technologies in the linguistics classroom: a pedagogical use of a clause pattern database. *Revista de la Asociación Española de Lingüística Aplicada*, 15(1), 150-165.
- McEnery, T., & Kifle, A. (2002). Epistemic modality in argumentative essays of second-language writers. In J. Flowerdew (Ed.), *Academic discourse* (pp. 182-195). Pearson Education Limited.
- Milton, J., & Hyland, K. (1999). Assertions in students’ academic essays: A comparison of L1 and L2 writers. In R. Berry, B. Asker, K. Hyland, & M. Lam (Eds.), *Language analysis, description and pedagogy* (pp. 147-161). HKUST.
- Muşlu, M. (2018). Use of stance lexical bundles by Turkish and Japanese EFL learners and native English speakers in academic writing. *Gaziantep Üniversitesi Sosyal Bilimler Dergisi*, 17(4), 1319-1336.
- Myhill, J., & Laura, A. S. (1995). Discourse and interactive functions of obligation expressions. In J. Bybee & S. Fleischmann (Eds.), *Modality in grammar and discourse* (pp. 239-292). John Benjamins.
- Narrog, H. (1995). On defining modality again. *Language Sciences*, 2, 165-192. <https://doi.org/10.1016/j.langsci.2003.11.007>
- Ngula, R. S. (2017). Epistemic modal verbs in research articles written by Ghanaian and international scholars: A corpus-based study of three disciplines. *Brno studies in English*, 43(2), 5-27. <https://doi.org/10.5817/BSE2017-2-1>
- Nhat, T., & Minh, N. (2020). A study on modality in English-medium research articles. *VNU Journal of Foreign Studies*, 36(6), 74-92. <https://doi.org/10.25073/2525-2445/vnufs.4630>
- Osipova, E. (2020). Corpus linguistic technology as a digital tool in teaching idioms’ interpretation to EFL students. *IOP Conference Series: Materials Science and Engineering*, 940(1). <https://doi.org/10.1088/1757-899X/940/1/012135>

- Öztürk, Y., & Köse, G. D. (2016). Turkish and native English academic writers' use of lexical bundles. *Journal of Language and Linguistic Studies*, 12(1), 149-165. <https://dergipark.org.tr/pub/jlls/issue/36116/405567>
- Palmer, F. R. (1990). *Modality and the English modals* (2<sup>nd</sup> ed.). Longman.
- Palmer, F. R. (2013). *Modality and the English modals* (2<sup>nd</sup> ed.). Routledge.
- Panocová, R. (2008). Expression of modality in biomedical texts. *SKASE Journal of Translation and Interpretation*, 3(1), 82-90. [http://www.skase.sk/Volumes/JTI03/pdf\\_doc/Panocova.pdf](http://www.skase.sk/Volumes/JTI03/pdf_doc/Panocova.pdf)
- Pearson, W. (2021). A comparative study of lexical bundles in IELTS Writing Task 1 and 2 simulation essays and tertiary academic writing. *Journal of Academic Language and Learning*, 15(1), 27-52. Retrieved from <https://journal.aall.org.au/index.php/jall/article/view/717>
- Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84- 94. <https://doi.org/10.1016/j.jeap.2014.01.002>
- Perkins, M. (1983). *Modal expressions in English*. Frances Pinter.
- Reppen, R. (2009). English language teaching and corpus linguistics: Lessons from the American national corpus. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 204- 213). Continuum.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Saeed, A. T. (2009). Arab EFL students' acquisition of modals. *Research in Language*, 7, 75-98. <https://doi.org/10.2478/v10015-009-0006-5>
- Salazar, D., & Verdaguer, I. (2009). Polysemous verbs and modality in native and non-native argumentative writing: A corpus-based study. *International Journal of English Studies*, 9(3), 209–219. <https://doi.org/10.6018/ijes.1.1.99601>
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1-22). John Benjamins.
- Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching*. John Benjamins Publishing Company
- Smith, N. (2003). Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English. In R. Facchinetti, F. Palmer & M. Krug (Eds.), *Modality in contemporary English* (pp. 241-266). Mouton de Gruyter.
- Soon, D., Seod, Y., & Choi, I. (2018). Interlanguage development of young Korean EFL students' modal usage: A learner corpus study. *Linguistic Research*, 35, 83-103. <https://doi.org/10.17250/khisli.35..201809.004>
- Stamatović, M. V. (2016). Scaling deontic modality in parliamentary discourse. *Logos et Littera: Journal of Interdisciplinary Approaches to Text*, 2(3), 131-149.
- Torabiardakani, N., Khojasteh, L., & Shokrpour, N. (2015). Modal auxiliaries and their semantic functions used by advanced EFL students. *Acta Didactica Napocensia*, 8(2), 51–60. [http://padi.psiedu.ubbcluj.ro/adn/article\\_8\\_2\\_6.pdf](http://padi.psiedu.ubbcluj.ro/adn/article_8_2_6.pdf)
- Urzua, A. (2015). Corpora, context, and language teachers. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (pp. 99-122). John Benjamins Publishing Company.

- Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes*, 20(1), 83-102. [https://doi.org/10.1016/S0889-4906\(99\)00029-0](https://doi.org/10.1016/S0889-4906(99)00029-0)
- Verhulst, A., Depraetere, I., & Heyvaert, L. (2013). Source and strength of modality: An empirical study of root should, ought to and be supposed to in present-day British English. *Journal of Pragmatics*, 55, 210-225. <https://doi.org/10.1016/j.pragma.2013.05.010>
- Viana, V. (2006). Modals in Brazilian advanced EFL students' compositions: A corpus-based investigation. *Profile: Issues in Teachers' Professional Development*, 7(1), 77-86. <https://www.redalyc.org/pdf/1692/169213802006.pdf>
- Woodford, K. (2003). *Cambridge advanced learner's dictionary*. Cambridge University Press.
- Yang, X. (2018). A corpus-based study of modal verbs in Chinese students' academic writing. *English Language Teaching*, 11(2), 122-130. <https://doi.org/10.5539/elt.v11n2p122>