

The Use of Lexical Bundles and the Definite Article ‘the’: A Core Expression Analysis

Kutay Uzun¹

Type/Tür:

Research/Araştırma

Received/Geliş Tarihi: July 7 / 7
Temmuz 2018

Accepted/Kabul Tarihi:

September 14/14 Eylül 2018

Page numbers/Sayfa No: 269–286

Corresponding

Author/İletişimden Sorumlu

Yazar: kutayuzun@trakya.edu.tr



This paper was checked for plagiarism using iThenticate during the preview process and before publication. / Bu çalışma ön inceleme sürecinde ve yayımlanmadan önce iThenticate yazılımı ile taranmıştır.

Copyright © 2018 by

Cumhuriyet University, Faculty of Education. All rights reserved.

Abstract

Even though the learning and use of formulaic language is an important aspect of language learning, the study of lexical bundles, which are within the scope of formulaic language, appears to be an under-researched area in the domain of language learning and teaching. In this respect, the present study aimed to contribute to the literature by providing an account of the use of lexical bundles which included the definite article ‘the’ in a corpus of 160 literary analysis essays written by 2nd year English Language Teaching students throughout a compulsorily taken English Literature course at a Turkish university. The secondary aim of the study was to test the functionality of the Core Expression Analysis method proposed by Shin et al. (2018) for the detection of attempted but erroneous lexical bundles which were typically left undetected by concordance analysis software since they deviated from standard forms. Lastly, the study aimed to perform error analysis on the lexical bundles which involved inaccurate uses of the definite article ‘the’. The results showed that the use of the definite article ‘the’ within the corpus were generally accurate and all of the inaccurate uses were due to omission errors. The findings also confirmed the functionality of Shin et al.’s (2018) Core Expression Analysis method to analyse attempted lexical bundles through the analysis of core expressions.

Keywords: Lexical bundles, definite article, error analysis, core expression, literary analysis essay

Suggested APA Citation /Önerilen APA Atf Biçimi:

Uzun, K. (2018). The use of lexical bundles and the definite article ‘the’: A core expression analysis.

Cumhuriyet International Journal of Education, 7(3), 269–286.

<http://dx.doi.org/10.30703/cije.441596>

¹ Lecturer, Trakya University, Faculty of Education, English Language Teaching Department, Edirne/Turkey

e-mail: kutayuzun@trakya.edu.tr

ORCID ID: <https://orcid.org/0000-0002-8434-0832>

Sözcüksel Öbek ve Belirli Tanımlık 'the' Kullanımı: Çekirdek İfade Analizi

Öz

Kalıplaşmış ifadelerin dil öğrenimindeki yerinin önemi biliniyor olsa da dil öğrenimi ve öğretimine ilişkin araştırmalarda kalıplaşmış ifadelerden olan sözcüksel öbekler konusunu inceleyen çalışmaların sayıca az olduğu görülmektedir. Bu bağlamda çalışma 2. Sınıf İngilizce Öğretmenliği öğrencilerinin zorunlu olarak aldıkları İngiliz Edebiyatı dersinde yazmış oldukları 160 edebî analiz kompozisyonundaki 'the' belirli tanımlığını içeren sözcüksel öbeklerin kullanımı konusunu inceleyerek literatüre katkıda bulunmayı amaçlamaktadır. Çalışmanın ikinci amacı ise Shin ve diğerleri (2018) tarafından önerilen ve sözcüksel kullanımların hatalı kullanımlarının da tespitini yapabilen Çekirdek İfade Analizi yönteminin işlevselliğini incelemektir. Son olarak çalışma 'the' belirli tanımlığını içeren sözcüksel öbeklerin hatalı kullanımlarının ne tür hatalardan kaynaklandığını bulmayı amaçlamaktadır. Sonuçlar derlem içerisindeki 'the' belirli tanımlığının tüm hatalı kullanımlarının eksiklik türü hatalar olduğunu göstermektedir. Ayrıca, Shin ve diğerleri (2018) tarafından önerilen ve hatalı sözcüksel öbek kullanımlarını da tespit etmeyi amaçlayan Çekirdek İfade Analizi yönteminin işlevsel olduğu görülmüştür.

Anahtar Kelimeler: Sözcüksel öbek, belirli tanımlık, hata analizi, çekirdek ifade, edebî analiz kompozisyonu

Introduction

Processed more prevalently and ubiquitously by adult learners in comparison to individual items within language (McCauley & Christiansen, 2014), formulaic language can be defined as a language fragment composed of multiple lexical items or morphemes which are grasped and utilized together as a whole (Richards & Schmidt, 2013). An integral part of phraseological competence (Ellis, 2012), formulaic language appears to be of paramount importance in language learning. Nonetheless, the area is still an under-researched one especially in language learning contexts (Pan & Hu, 2017).

In relation to the aforementioned concept of phraseological competence, lexical bundles fall within the domain of formulaic language since they are multi-word units of a phraseological nature (Flowerdew, 2015). Biber, Johansson, Leech, Conrad and Finegan (1999) define lexical bundles as a repeated concatenation of three or more words irrespective of the structural or idiomatical features within the group of words (e.g. is one of the, as a result of the, extent to which the...). Setting the criteria for a word group to be counted as a lexical bundle, Cortes (2004) articulates that a word group of three or more words should occur 20 to 40 times per million words to be regarded as a lexical bundle. In addition to the frequency criterion, Biber et al. (1999) suggest a criterion of dispersion to avoid idiosyncratic hits in concordance searches. According to this criterion, a word group should occur at least 10 times per one million words in at least 5 different texts to be called a lexical bundle. Pan, Reppen and Biber (2016) suggest 40 occurrences per one million words in 5 different texts as a cut-off point while Hyland (2008) states that 20 occurrences per million words in 10% of the number of texts in the corpus should be taken as the cut-off point in the analysis of lexical bundles.

Biber, Conrad and Cortes (2004) identify three major functions of lexical bundles as expressing stance, organizing discourse and expressing reference. Among these, stance expressions are used to communicate attitude towards or for the

evaluation of a certain concept (e.g. I think it was, I want you to, can be used to...). The second function of lexical bundles, organizing discourse, is used to establish connection between the preceding and upcoming discourse (e.g. if you look at, on the other hand...). The last function of lexical bundles which is identified as referential expression refers to the feature of the lexical bundle referring to a concept or textual context for identification or bringing a concept to the forefront (Biber et al., 2004). In brief terms, the use of lexical bundles makes it possible to express and evaluate attitude, connect discourse and refer to concepts.

Having significant functions in terms of language use as mentioned above, the knowledge of and competence in using lexical bundles are also considered to be important since they are related to the production of rhetorical moves (Cortes, 2013), have pragmatic and discursual features that are customarily used by discourse communities (Biber et al., 2004; Hyland, 2008) and bear potential to reduce working memory load by being processed as a single item (Conklin & Schmitt, 2008; Wei & Ying, 2012). With those features, lexical bundles are integral constituents of written discourse (Li & Schmitt, 2009), making them significant elements of language to be learned and taught, too, as discourse competence both in spoken and written communication is a component of communicative competence as suggested by Canale and Swain (1980).

Indeed, lexical bundles may signal language development since their use show variety across different proficiency levels (Staples, Egbert, Biber, & McClair, 2013), level of expertise (Cortes, 2004) and the status of the language being used as L1 or L2 (Ädel & Erman, 2012). Moreover, it is possible to automate the identification of lexical bundles, clearly define their functions and observe the differences in their use across text types (Durrant, 2017), which makes them convenient to measure for developmental purposes. They also enable teachers or researchers to measure and assess the phraseological awareness and performance levels of learners through frequency counts and concordance analyses (Huang, 2015), contributing even further to their potential to be utilized in the tracking of language development. On this matter, however, Shin, Cortes and Yoo (2018) warn that the analysis of lexical bundles may be arduous and produce imperfect results since lexical bundles inaccurately attempted by learners might pose problems for software in their identification. For instance, an attempted lexical bundle which has a preposition error such as *in the other hand* would not be counted by concordance search algorithms even though it may be an erroneous use of *on the other hand*.

Despite the potential problems in their measurement, it is possible to come across with research studies from the Turkish context which scrutinize lexical bundles. For instance, Öztürk and Durmuşoğlu (2016) find that Turkish postgraduate students overuse lexical bundles in comparison to their native-speaker counterparts. Güngör and Uysal (2016) also identify differences in native and non-native uses of lexical bundles among scholars with non-native scholars using lexical bundles significantly more frequently than native scholars. In addition, Güngör and Uysal (2016) reveal that while the non-native use of lexical bundles primarily centres around clausal and verb phrase-based bundles, native scholars prefer noun phrase and prepositional phrase-based bundles more predominantly. In more general terms, Üstünbaş (2014) studies the spoken use of formulaic expressions among

undergraduate-level Turkish students and concludes that spoken fluency and overall proficiency level interact with the use of formulaic expressions. Contrarily, Mirici and Göksu (2016) put forth that the proficiency level of the learner makes no difference in the use of formulaic expressions, however, residence in an English-speaking country affects their use. In line with Mirici and Göksu (2016), Kılıç (2015) also indicates that writing coherence, proficiency in writing and overall proficiency in English are not related to the use of formulaic language among learners. In conclusion, there appears to be contradicting findings with respect to the role of proficiency level in the use of formulaic language, including lexical bundles, even though the number of studies dealing with the construct are rather scarce. Furthermore, none of the studies reviewed above seem to have taken attempted but erroneous lexical bundles into account.

Having considered the relevant literature on lexical bundles, it appears that they are under-researched within the Turkish context and worldwide especially in terms of accuracy in their use. To the researcher's knowledge, Shin et al.'s (2018) study, in which they reveal that the omission of the definite article '*the*' is the most common error which is followed by addition and misinformation, is the only one in the relevant literature which deals with the accuracy of lexical bundles in texts produced by language learners. This particular study is considered important especially because it points at a previously unexplored problem, which is the infrequent use of lexical bundles by language learners reported in previous studies. According to Shin et al. (2018), this may be the result of inaccurate lexical bundle attempts made by the learners that have been left undetected in the string searches performed by the search algorithms of different software since, in most cases, a hit can only be achieved in the case of an exact match and thus, erroneous attempts at producing lexical bundles have probably been skipped by those algorithms, potentially resulting in the loss of valuable data in terms of language development in previous studies. For the reasons of being under-researched and potentially ignoring valuable data, the scrutiny of lexical bundles with the consideration of erroneous attempts appears to be important in terms of both tracking language development and identifying the accuracy of those bundles.

Another noteworthy problem that Shin et al. (2018) identify within their study is related to the nature of the definite article '*the*' in English, whose relevance to lexical bundles arise from the fact that 160 of the 278 4-word bundles typically seen in academic prose naturally include the definite article, making it a prominent element in the lexical bundles used in academic register. According to Shin (2012) and Shin et al. (2018), even though the definite article is typically defined and operationalized through the noun following it, it also depends on the previous phrase as they are integral elements of lexical bundles. For this reason, Shin et al. (2018) argue that the previous studies may actually be limited in terms of the findings acquired regarding the definite article '*the*', which is among the most complex structures for the learners of English language (Chrabaszcz & Jiang, 2014), due to the absence of the investigation of inaccuracies in its use within lexical bundles.

The complexity of the definite article '*the*' for language learners arises from certain reasons pertaining to the nature of the article and the L1 background of the learner. For instance, the definite article '*the*' is typically unstressed in speech,

making it unnoticeable for learners (Master, 2002). Related to this, errors in the use of 'the' can be present even among advanced learners of English who have been learning the language for long years (Amuzie & Spinner, 2013). Furthermore, L1 speakers of languages without articles encounter difficulties in the acquisition of the definite article 'the' since its acquisition necessitates the internalization of numerous form-function mappings regarding its use (Chrabaszcz & Jiang, 2014; Zhao & Macwhinney, 2018). Elaborating on the same complexity, García Mayo (2009) suggests that L1 speakers of languages without the concept of definitiveness struggle in acquiring the use of the definite article 'the' as the concept cannot be transferred from their L1 to their L2 production. In short, the unnoticeability of the definite article in speech and being of an L1 background without articles or the concept of definitiveness may be at the core of the difficulties experienced by language learners in respect of the definite article 'the'.

Within the Turkish context, the complexities and complications arising out of the use of the definite article 'the' have also been the centre of attention, where research studies appear to be centred around the interaction of proficiency level with accuracy in the use of the definite article 'the' as well as the types of errors committed by learners of English language. For example, Dikilitaş and Altay (2011) analyse the potential interaction between accuracy in the use of the definite article and proficiency level and conclude that proficiency level is not associated with accuracy in the use of the definite article, however, it is related to the perception of difficulty with respect to its different uses. Conversely, between low and high-proficient learners of English, Dağdeviren (2010) observes that high-proficient learners are more accurate in their use of the definite article than their low-proficient counterparts with both groups committing addition errors. Contrary to Dağdeviren's (2010) findings, Koç (2015) identifies that intermediate and low-advanced learners of English do not differ in terms of addition errors, however, intermediate learners commit more omission errors than low-advanced learners. When the groups are combined, Koç (2015) observes that there are more omission errors than addition errors within the study group. Acquiring different findings than those of Koç's (2015) with the argumentative essays written by undergraduate students of English Language Teaching, Yalçın (2010) finds out that article errors are the most prominent ones among other types of errors such as preposition or verb errors and the types of article errors ordered by frequency are addition, omission and misinformation respectively. In a similar group but with a narrative essay writing task, Taşçı and Ataç (2018) reach different findings in that article errors are the third most common type of error after preposition errors and verb errors. Similar to Koç's (2015) findings, Taşçı and Ataç (2018) reveal that omission errors in the definite article 'the' are more frequent than addition errors while no misinformation error is reported in their study. Confirming the differences between Koç's (2015), Yalçın's (2010) and Taşçı and Ataç's (2018) findings, Çimen (2013) argues that task type interacts with the frequency of article errors since addition errors outnumber omission errors in a fill-in-the-article test and omission errors outnumber addition errors in a written production test in her study group. Considering the findings acquired within the Turkish context, it can be said that the proficiency level may or may not interact with

accuracy in the use of the definite article '*the*' and task type may also affect the findings with respect to the frequency of the type of errors committed by learners.

Taking into account the findings and suggestions regarding the definite article '*the*', it is understood that the structure is related to the concept of lexical bundles as it plays an integral role in their formation, however, both teaching and learning the structure may be an arduous task on behalf of teachers and learners due to its typically unstressed nature and the lack of positive transfer opportunities among learners from L1 backgrounds without articles or the concept of definitiveness. Finally, research results regarding the use of the definite article '*the*' seem to produce contradicting results, necessitating the further study of the structure in varying contexts.

Having considered the literature on both lexical bundles and the definite article '*the*', it can be seen that all studies within the Turkish context treat the definite article as a unit dependent on the following unit and there appears to be no study with respect to the role of the definite article in lexical bundles or the units preceding the article, on which it may also depend (Shin, 2012; Shin et al., 2018). In addition, the concept of lexical bundles can also be utilized for the teaching of the article system of English (Shin & Kim, 2017) which necessitates the treatment of both concepts together for language teaching purposes. In this respect, the present study aims to fill a gap within the context-specific literature by making use of Shin et al.'s (2018) method of analysing the uses of definite article within lexical bundles and to compare and contrast the findings obtained by Shin et al. (2018) in the Korean context with those obtained in the Turkish context, in both of which English is treated as a foreign language.

Purpose of the Study

In line with the relevant literature, the study aimed to find out the number of accurate and attempted lexical bundles which included the definite article '*the*' and are typically used in academic prose as identified by Biber et al. (1999) with the method proposed by Shin et al. (2018) in the Turkish EFL context. Moreover, the study aimed at identifying the accuracy of the definite articles embedded in these accurate and attempted lexical bundles. The last aim of the study was to investigate the types of errors in the use of the definite article '*the*' within the lexical bundles identified within the corpus of the study. The following research questions were formulated to meet the aims of the study:

1. How frequently are lexical bundles used in a corpus of literary analysis essays?
2. How frequently are the core expressions within lexical bundles used in the corpus of literary analysis essays?
3. What is the level of accuracy in terms of the definite article '*the*' within lexical bundles in the corpus of literary analysis essays?
4. How frequently can different types of definite article errors be observed in the corpus of literary analysis essays?

Methodology

Since the aim of the study was based on quantifying lexical bundles and the core expressions in those bundles as well as the accurate and inaccurate uses of the

definite article '*the*' within them, a quantitative and descriptive research design was preferred. In quantitative research, the researcher uses numbers or converts different types of data into numerical data for the standardization of variables in objective terms, allowing for the production of precise, reliable and replicable results (Dörnyei, 2007). Moreover, according to Creswell (2002), descriptive research designs are suitable for conditions in which the researcher is interested in collecting data about a given present situation and describe this particular situation without manipulating it. Since the study aimed at revealing the uses of lexical bundles and the definite article '*the*' in the literary analysis essays of the participants, a quantitative and descriptive design was considered suitable for the purposes of the study.

The Corpus

The corpus used in the study was composed of 160 literary analysis essays written in a compulsory English Literature course by 40 second year students of English Language Teaching (ELT) through the course of a compulsorily taken English Literature course at a public university in Turkey. Drawing upon Guthrie, Guthrie and Wilk's (2008) description that outliers in corpora are typically the texts which diverge from the focal topic or genre, all the essays written by the aforementioned students through the course of one academic year were included in the corpus since none of the essays were seen to have deviated from the rhetorical conventions of the literary analysis essay as a genre even though the amount of text that was produced and text quality appeared to have varied. Preliminary analyses showed that the corpus consisted of 48866 words producing an average of 305 words per essay, the shortest essay having 57 words and the longest one having 845 words.

The features of a literary analysis essay within the context of the study has been documented by Uzun (2016), who reveals that the genre is composed of three main parts as the introduction, main body and conclusion. In the introduction, the writer is expected to provide a brief background to the literary work being analysed and state a thesis which is a direct response to an essay prompt/question. In the main body, the points mentioned in the thesis are elaborated and supported or proven by providing specific examples from the text and/or the historical context and each of those points are discussed in a separate paragraph. The conclusion part consolidates the thesis and presents the subjective opinions of the writer with respect to the thesis of the essay (Uzun, 2016).

Data Analysis

As mentioned in the literature review section, Shin et al. (2018) pointed at a crucial problem in the identification of attempted lexical bundles which could be strenuous for software to detect as they potentially deviated from the standard use of those bundles. As a solution, Shin et al. (2018) proposed the identification of each 'core expression' within each lexical bundle, which they defined as "a phrase formed around a core that constitutes an important component, often lexical, of each LB [lexical bundle], along with a following word" (p. 32). In other words, Shin et al. (2018) argued that making a string search by eliminating the potentially problematic parts of lexical bundles by using only the core expressions in the search string (e.g. 'result of' instead of 'as a result of the'), and manually eliminating the hits that did not belong to a lexical bundle could provide accurate results in the detection of

erroneous lexical bundle attempts by learners. Taking this into account, 160 four-word lexical bundles which included the definite article 'the', and 97 core expressions belonging to these bundles, which were identified by Shin et al. (2018) in the list of lexical bundles provided by (Biber et al., 1999, Biber et al., 2004) were searched within the corpus of the present study using AntConc which is a concordance and text analysis software (Anthony, 2018).

To find out the level of accuracy in terms of the definite article 'the' in the corpus, the core expressions within the corpus, which belonged to the lexical bundles of interest, were investigated in terms of the use or misuse of the definite article 'the'. For example, while looking for the sum of accurate and inaccurate uses of the lexical bundles *at the end of*, *by the end of* and *the end of the*, the core expression *end of*, which is common in all three of them was sought for and thus, inaccurate uses of the definite article 'the' were also included in the hits. For the investigation and the detection of errors, Dulay, Burt and Krashen's (1982) framework of errors was used. In this framework, Dulay et al. (1982) classify errors as omission errors, addition errors and misinformation errors. Omission errors refer to the lack of an item required grammatically in a certain position within a sentence (e.g. 'at same time' instead of 'at the same time'). Addition errors, on the other hand, refer to the overuse of an item which cannot be in an accurate utterance (e.g. 'is the one of the' instead of 'is one of the'). Lastly, misinformation errors are those which misuse the form of a structure or morpheme (e.g. 'on this other hand' instead of 'on the other hand'). Taking this classification into account, each of the core expressions belonging to lexical bundles within the corpus was analysed manually to detect the errors and types of errors that were present. The reliability of the error analysis was sought for by having a second researcher repeat the analysis, which resulted in total agreement between two researchers who analysed the errors as understood by the computation of a Kappa Coefficient for inter-rater reliability ($K = 1.000, p < .001$).

Parallel to the aims of the study, the number of lexical bundles and core expressions, which belonged to lexical bundles, accurate and inaccurate uses of the definite article 'the' were presented in frequencies and percentages. In these analyses, the lexical bundles which exactly matched the uses in Biber et al.'s (1999) and Biber et al.'s (2004) lists were considered accurate. On the other hand, the lexical bundles in the corpus which deviated from the uses of the definite article 'the' in those lists were considered inaccurate. As suggested by Shin et al. (2018), prepositional deviations were not taken into consideration since the focus of analysis was the definite article use. As an example, both *at the end of* and *in the end of* were considered accurate while a use such as *at end of* was considered inaccurate as it lacked the definite article after the preposition. Moreover, the overlapping bundles were not counted twice. For example, in the cases where 'is one of the' and 'of the most important' overlapped in 'is one of the most important', the lexical bundle was counted as 'of the most important' only. Normalized frequencies were computed as the number of lexical bundles per million words.

Findings

The first research question aimed to identify the four-word lexical bundles which included the definite article 'the' as used within the corpus of 160 literary analysis essays. The results were quantified below in Table 1.

Table 1
Lexical Bundles in the Corpus

Lexical Bundle	<i>nf</i>	<i>f</i>	%	Lexical Bundle	<i>nf</i>	<i>f</i>	%
on the other hand	470.67	23	17.16	at the expense of	20.46	1	.75
of the most important	409.28	20	14.93	by the end of	20.46	1	.75
at the end of	327.43	16	11.94	in contrast to the	20.46	1	.75
at the beginning of	306.96	15	11.19	in the same way	20.46	1	.75
one of the most	286.50	14	10.45	on the one hand	20.46	1	.75
at the same time	225.11	11	8.21	the base of the	20.46	1	.75
to the fact that	143.25	7	5.22	the center of the	20.46	1	.75
is based on the	81.86	4	2.99	the end of the	20.46	1	.75
for the first time	40.93	2	1.49	the rest of the	20.46	1	.75
in the first place	40.93	2	1.49	the results of the	20.46	1	.75
is related to the	40.93	2	1.49	the role of the	20.46	1	.75
of the fact that	40.93	2	1.49	the time of the	20.46	1	.75
the beginning of the	40.93	2	1.49				
the origin of the	40.93	2	1.49				
				Types: 26	Token: 134		

As seen in the Table 1, the corpus included 26 types and 134 tokens of lexical bundles. Analyses showed that the most frequently used lexical bundles within the corpus were *on the other hand* (*nf* = 470.67, *f* = 23, % = 17.16), *of the most important* (*nf* = 409.28, *f* = 20, % = 14.93), *at the end of* (*nf* = 327.43, *f* = 16, % = 11.94), *at the beginning of* (*nf* = 306.96, *f* = 15, % = 11.19) and *one of the most* (*nf* = 286.50, *f* = 14, % = 10.45). On the other hand, some of the least frequently used lexical bundles were seen to have been *the end of the* (*nf* = 20.46, *f* = 1, % = .75), *the rest of the* (*nf* = 20.46, *f* = 1, % = .75), *the results of the* (*nf* = 20.46, *f* = 1, % = .75), *the role of the* (*nf* = 20.46, *f* = 1, % = .75) and *the time of the* (*nf* = 20.46, *f* = 1, % = .75). For illustrative purposes, some examples from the frequently used lexical bundles (underlined) were provided below.

Example 1. *On the other hand* although we know that Hamlet is faking his madness, he is that confused that he kills Polonius thinking that was his uncle Claudius.

Example 2. In general, there are some special themes affecting the whole novel like pride and prejudice but also the theme of love can be considered as one of the most important themes.

Example 3. However, at the end of the play, all of them are punished so every crime and criminal do not exist without punishment.

As seen in the concordance findings and examples, the hits received with the purpose of answering the first research question represented the uses of lexical bundles, which had no article, preposition or verb error, however, erroneous or attempted uses of the mentioned bundles were not present in the results.

The aim of the second research question was to identify the core expressions within the four-word lexical bundles which included the definite article 'the' to detect the potentially erroneous uses within the corpus. The results were tabulated below in Table 2.

Table 2
Core Expressions within Lexical Bundles

Core Expression	<i>f</i>	%	Core Expression	<i>f</i>	%
end of	25	15.72	first time	2	1.26
other hand	23	14.47	origin of	2	1.26
most important	20	12.58	results of	2	1.26
beginning of	17	10.69	base of	1	.63
one of	14	8.81	center of	1	.63
same time	11	6.92	expense of	1	.63
fact that	9	5.66	nature of	1	.63
based on	5	3.14	one hand	1	.63
role of	5	3.14	rest of	1	.63
contrast to	4	2.52	same way	1	.63
related to	4	2.52	terms of	1	.63
top of	3	1.89	time of	1	.63
course of	2	1.26			
first place	2	1.26			
			Types: 26	Tokens: 159	

As shown in Table 2, the most frequently used core expressions within lexical bundles were *end of* ($f = 25$, $\% = 15.72$), *other hand* ($f = 23$, $\% = 14.47$), *most important* ($f = 20$, $\% = 12.58$) and *beginning of* ($f = 17$, $\% = 10.69$), according to the concordance findings. Further, some of the least frequently used core expressions in lexical bundles were found to have been *base of* ($f = 1$, $\% = .63$), *nature of* ($f = 1$, $\% = .63$), *one hand* ($f = 1$, $\% = .63$) and *time of* ($f = 1$, $\% = .63$). When the exact match of lexical bundles (see Table 1), which was counted to have had 134 tokens, was compared with the number of the tokens belonging to the core expressions, which was 159, it was seen that two frequency counts had a difference of 25. Moreover, it was also seen that the core expressions of *nature of* ($f = 1$, $\% = .63$), *terms of* ($f = 1$, $\% = .63$) and *top of* ($f = 3$, $\% = 1.89$) were not present in Table 1. As a conclusion, these discrepancies were interpreted as having indicated erroneous uses of the lexical bundles investigated in the present study. Some attempted but erroneous uses were exemplified below:

Example 1. *In other words, the functions of women are restricted in some way but the author described some women characters like Lady Bracknell, Cecily and Gwendolen in terms of inversion of Gender Roles.*

Example 2. *Although so many years passed, it is top of the list of "best seller books" and it is continuing sell more than 20 million copies.*

Example 3. *Although, he has fears about what he'll face with next nobody could stop him about not to carry out his ideas until end of the story.*

As seen in all three examples, there were problems with the use of the definite article 'the' in some lexical bundles, which distorted the search results with respect to the four-word lexical bundles within the corpus. However, core expression analysis allowed for the identification of this discrepancy as well as the analysis of erroneous lexical bundles within the corpus.

The purpose of the third research question was to find out the level of accuracy in the four-word lexical bundles, which included the definite article 'the', by making use of the aforementioned discrepancy and attempted but erroneous lexical bundles within the corpus. Even though there were 25 core expression hits which deviated from the accurate uses of the lexical bundles of interest due to preposition, verb or article errors, 2 concordance lines which did not have definite article errors (e.g. 'in the end of' instead of 'at the end of') were considered as accurate since the analysis in terms of accuracy was limited to the use of the definite article 'the'. The results, grouped by core expressions, were given below in Table 3.

Table 3

Accuracy Levels of the Lexical Bundles in the Corpus Grouped by Core Expressions

Core Expression	Accuracy	Accurate	Inaccurate	Core Expression	Accuracy	Accurate	Inaccurate
other hand	100.00	23	0	same way	100.00	1	0
most important	100.00	20	0	time of	100.00	1	0
beginning of	100.00	17	0	based on	80.00	4	1
one of	100.00	14	0	end of	76.00	19	6
same time	100.00	11	0	related to	50.00	2	2
fact that	100.00	9	0	results of	50.00	1	1
first place	100.00	2	0	course of	50.00	1	1
first time	100.00	2	0	contrast to	25.00	1	3
origin of	100.00	2	0	role of	20.00	1	4
base of	100.00	1	0	nature of	.00	0	1
center of	100.00	1	0	terms of	.00	0	1
expense of	100.00	1	0	top of	.00	0	3
one hand	100.00	1	0				
rest of	100.00	1	0	TOTAL	85.53%	136	23

As demonstrated in Table 3, the average accuracy in the use of the definite article 'the' within four-word lexical bundles which included it was 85.53% within the entire corpus. Analysis also showed that the most frequently used core expressions which were *other hand* ($f = 23$), *most important* ($f = 20$), *beginning of* ($f = 17$), *one of* ($f = 14$) and *same time* ($f = 11$) achieved 100% accuracy along with some of the less frequent core expressions such as *fact that* ($f = 9$), *first place* ($f = 2$), *first time* ($f = 2$) and *origin of* ($f = 2$). On the other hand, a few errors were identified in the uses of the definite article 'the' in the lexical bundles which included *based on* ($f = 5$) and *end of* ($f = 25$), resulting in 80% and 76% accuracy respectively in these core expressions. In the core expressions *contrast to* ($f = 4$) and *role of* ($f = 5$), it was seen that the accuracy levels were 25% and 20% respectively, indicating that the lexical bundles which included these were predominantly erroneous. Lastly, the lexical bundles that included the core expressions of *nature of* ($f = 1$), *terms of* ($f = 1$) and *top of* ($f = 3$) were found to have been used only inaccurately.

The fourth research question attempted to identify the types of error in the inaccurate uses of the definite article 'the' within the corpus of literary analysis essays. The results were tabulated below in Table 4.

Table 4
Types of Error in the Uses of the Definite Article 'the'

Error Type	<i>f</i>	%
Omission	23	100.00
Addition	0	0.00
Misinformation	0	0.00

As seen in Table 4, error analysis revealed that all the errors in the use of the definite article 'the' within the corpus of literary analysis were omission errors ($f = 23$, $\% = 100.00$). The results indicated that none of the definite article errors within lexical bundles were addition ($f = 0$, $\% = .00$) or misinformation ($f = 0$, $\% = .00$) errors within the corpus.

Discussion, Conclusion and Suggestions

The present study aimed to identify the four-word lexical bundles which included the definite article 'the', the core expressions belonging to lexical bundles and definite article errors in the use of the definite article in a corpus of literary analysis essays written by 2nd year ELT students, using the method recently proposed by Shin et al. (2018). In general, the findings revealed that there was a scarcity of lexical bundles in the corpus since there were only 26 different four-word lexical bundles out of 160 and 26 different core expressions out of 97 as identified by Shin et al. (2018), representing only a small portion of the possible variety. Moreover, only 14 of 26 different bundles were above the cut-off frequency of 40 per million words (Pan et al., 2016) and 6 of those were only slightly above the cut-off limit. Furthermore, 12 out of 26 different lexical bundles were only slightly more frequent than the cut-off frequency of 20 per million words as suggested by Hyland (2008). According to the findings, the most frequently used lexical bundles within the corpus were *on the other hand*, *of the most important*, *at the end of*, *at the beginning of* and *one of the most*. Among the core expressions, *end of*, *other hand*, *most important*, *beginning of* and *one of* were among the most recurrent ones. In terms of definite article accuracy within lexical bundles, the corpus was found to be of high accuracy with 85.53%. Lastly, all the errors in the use of the definite article 'the' within the corpus were found to be omission errors. In brief, the findings indicated that there were not frequent instances of four-word lexical bundles including the definite article within the corpus, however, the ones that were present were generally accurate with some omission errors.

The findings with respect to the use of lexical bundles within the corpus were in line with those of Shin et al.'s (2018) in that the most frequently used bundles in both studies were the same even though the order of frequency ranks slightly differed. In the findings, it was seen that one of and beginning of, two core expressions among the most frequently used ones, overlapped, too. Furthermore, in both studies, it was seen that lexical bundles were only infrequently used by the

writers of the texts within the corpus. The infrequent use of lexical bundles found both in the present study and that of Shin et al.'s (2018) contradicted Öztürk and Durmuşoğlu's (2016) findings, which concluded that lexical bundles were overused by Turkish learners of English. This contradiction may have resulted from the difference in the text types used in the corpora of these three studies, the first two being essays and the last one being thesis (both MA and PhD), as the use of lexical bundles may differ across different types of text (Durrant, 2017) and the nature of academic register may have necessitated more frequent use of lexical bundles by the writers (Biber et al., 1999). Moreover, the difference may have stemmed from the gaps in the proficiency and expertise levels of the writers of the texts in the corpora (Staples et al., 2013; Cortes, 2004), as Öztürk and Durmuşoğlu (2016) used a corpus that included MA and PhD theses, the writers of which could typically be expected to be more proficient in English language. Having considered the similarities and differences in the findings in comparison to those of other studies, the scarcity in the use of lexical bundles within the corpus of the present study may have been consequent on the level of writing expertise and proficiency among the writers of the texts as well as the text type included in the corpus.

The findings in regards to the use of the definite article '*the*' were also in line with those of Shin et al.'s (2018), indicating that omission was the most frequent type of definite article error within the lexical bundles in the corpora of both studies. Moreover, in a similar context to the present study, Koç (2015) and Taşçı and Ataç (2018) also identified omission as the commonest type of definite article error, which was echoed in the present study. The parallelism among the findings of these studies and the present study appears to have resulted from the fact that neither Korean nor Turkish have an equivalent structure to the definite article '*the*' in English. For this reason, the writers of the texts in the corpus may have simply skipped the necessary uses of the definite article in their essays because they had no similar structure to transfer into their writing in English and they did not have the required form-function maps for the accurate use of the definite article '*the*' (Chrabasz & Jiang, 2014; Zhao & Macwhinney, 2018). The same reason can also account for the absence of addition or misinformation errors in the findings of the present study. Since the writers of the texts were of a native language background without the concept of definitiveness, they may not have been able to transfer the concept into their L2 texts (García Mayo, 2009) and the lack of a transferable structure may have resulted in the total absence of addition and misinformation errors. Nevertheless, Yalçın (2010) found in a previous study in a very similar context that addition errors regarding the definite article '*the*' were more common than omission errors, which contradicted the findings of the present study. The difference between the findings of this study and that of Yalçın's (2010) could have resulted from the fact that the present study dealt only with the definite articles within lexical bundles, while the latter investigated all uses of the definite article. Moreover, the size of the corpus in Yalçın's (2010) study, which was about one fifth of the present study, may have reduced the generalizability of the findings into broader contexts. In conclusion, the L1 background of the participants which did not have a definite article structure as in English may have been the reason behind omission error's being the most prevalent within the corpus of the study (Chrabasz & Jiang, 2014).

Regarding the accuracy level in terms of the definite article '*the*' within the corpus, the proficiency levels of the essay writers within the corpus may have resulted in the relatively high percentage in terms of the accurate uses of the definite article within lexical bundles. As all the writers of those texts were 2nd year ELT students who had taken language-intensive education in high school and either taken a year of English courses in the Prep School or been exempted from the Prep Year due to having been sufficiently proficient in English, they may have demonstrated a relatively high level of performance in the use of the definite article '*the*' in their essays. Regardless, the erroneous uses of the definite article were also expected since this particular structure is among the most difficult ones for learners (Chrabascz & Jiang, 2014) and errors may persist even after long years of learning the language (Amuzie & Spinner, 2013). In sum, the relatively long language education background of the text writers in the corpus may have accounted for the high level of accuracy in terms of the use of the definite article '*the*'.

Taking into account all the findings of the present study, it can be concluded that lexical bundles within literary analysis essays written by Turkish learners of English are rather infrequent but generally accurate in terms of the definite article '*the*' embedded within those bundles. In the cases of inaccuracy, the writers of the essays tend to omit the article instead of demonstrating addition or misinformation errors due to the potential unavailability of a similar structure that can be transferred from L1 into L2. Another practically useful conclusion of the study is that the method proposed by Shin et al. (2018) for the analysis of embedded items within lexical bundles, which includes the analysis of core expressions, appears to be a functional one as it allows for the analysis of erroneous uses. This method can be considered different from the previously performed concordance searches, which resulted only in the exact matches regarding lexical bundles and ignored the attempted ones. Thanks to the analysis of core expressions, it appears possible to analyse a given item within lexical bundles and examine the errors that may be present in those bundles.

Regarding the practical implications of the findings, it can be stated that phraseological awareness, of which lexical bundles pose an integral part (Huang, 2015), can be fostered through the teaching of lexical bundles and this may result in an increased level of writing performance among language learners by allowing the learner to use formulaic language more frequently and enabling the allocation of a larger amount of working memory resources for the content of a given text. Moreover, since lexical bundles have been shown to be associated with rhetorical moves, the teaching of the rhetorical moves with an English for Specific Purposes (ESP) approach, which attempts at teaching rhetorical moves to produce a text that pertains to a particular genre (Hyland, 2007), may also increase performance in terms of the production of lexical bundles. By analysing provided models as suggested in the ESP approach to genre, learners may familiarize themselves with the lexical bundles frequently used by a particular discourse community, resulting in their more frequent use and supporting the more efficient use of working memory resources as described above.

As a final point to consider, it should be noted that the study is limited to its own corpus of literary analysis essay written by 2nd year ELT students in the natural context of a compulsorily taken English Literature course. As Çimen (2013) and

Durrant (2017) also noted, analyses of different genres other than the literary analysis essay may produce different results, therefore, replicating the study with the same method on corpora of different text types can be suggested. Moreover, since each writer contributed to the corpus with multiple literary analysis essays, idiosyncratic uses of lexical bundles could not be controlled for in the study. It should also be kept in mind that the study is limited to the use of the definite article 'the'. Therefore, applying the same methodology with the purpose of analysing different items within lexical bundles may be useful. Lastly, the hypothesized method of manipulating performance in the production of lexical bundles through Genre-Based Instruction can be empirically tested to find out if an increase in the rhetorical performance of language learners results in an improvement in the production of lexical bundles.

References

- Ädel, 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, 81-92. <https://doi.org/10.1016/j.esp.2011.08.004>
- Anthony, L. (2018). *AntConc* (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software> on 25.06.2018.
- Amuzie, G., & Spinner, P. (2013). Korean EFL learners' indefinite article use with four types of abstract nouns. *Applied Linguistics*, 34(4), 415-434. <https://doi.org/10.1093/applin/ams065>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Canale M., Swain M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47
- Chrabascz, A., & Jiang, N. (2014). The role of the native language in the use of the English nongeneric definite article by L2 learners: A cross-linguistic comparison. *Second Language Research*, 30(3), 351-379. doi: 10.1177/0267658313493432
- Çimen, Ş. S. (2013). Acquisition of English article system by turkish learners in an EFL setting. *Muğla Sıtkı Koçman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 31, 90-105.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397-423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12, 33-43. <https://doi.org/10.1016/j.jeap.2012.11.002>

- Cresswell, J. W. (2002). *Educational research planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle Creek, NJ: Pearson Education.
- Dağdeviren, G. (2010). Use of English articles by speakers of Turkish in the EFL setting. *Novitas-ROYAL*, 4(2), 242-250.
- Dikilitaş, K., & Altay, M. (2011). Acquisition sequence of four categories of non-generic use of the English definite article the by Turkish speakers. *Novitas-ROYAL*, 5(2), 183-198.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. New York: Oxford University Press.
- Dulay, H., Burt, M., & Krashen, S. D. (1982). *Language two*. Oxford: Oxford University Press.
- Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2), 165-193. <https://doi.org/10.1093/applin/amv011>
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17-44. <https://doi.org/10.1017/S0267190512000025>
- Flowerdew, L. (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, 48(1), 99-116. doi:10.1017/S0261444813000037
- García Mayo, M. (2009). Article choice in L2 English by Spanish speakers: Evidence for full transfer. In M. García Mayo & R. Hawkins (Eds.), *Second language acquisition of articles* (pp. 13-35). Amsterdam: John Benjamins. <https://doi.org/10.1075/lald.49.05pil>
- Güngör, F., & Uysal, H. H. (2016). A comparative analysis of lexical bundles used by native and non-native scholars. *English Language Teaching*, 9(6), 176-188. doi: 10.5539/elt.v9n6p176
- Guthrie, D., Guthrie, L., & Wilks, Y. (2008). An unsupervised approach for the detection of outliers in corpora. Paper presented at *The International Conference on Language Resources and Evaluation, LREC 2008*, 3409-3413. Retrieved from www.lrec-conf.org/proceedings/lrec2008/pdf/866_paper.pdf on 10.09.2018.
- Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System*, 53, 13-23. <https://doi.org/10.1016/j.system.2015.06.011>
- Hyland K. (2007). Genre pedagogy: language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16, 148-164. <https://doi.org/10.1016/j.jslw.2007.07.005>
- Hyland, K. (2008). 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>
- Kılıç, S. Z. (2015). *The use of formulaic language by English as a foreign language (EFL) learners in writing proficiency exams*. (Unpublished Master's Thesis). Bilkent University, Ankara, Turkey.
- Koç, D. K. (2015). The non-generic use of the definite article the in writing by Turkish learners of English. *H. U. Journal of Education*, 30(2), 56-68.

- Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18(2), 85-102. <https://doi.org/10.1016/j.jslw.2009.02.001>
- Master, P. (2002). Information structure and English article pedagogy. *System*, 30, 331-48. [https://doi.org/10.1016/S0346-251X\(02\)00018-0](https://doi.org/10.1016/S0346-251X(02)00018-0)
- McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *The Mental Lexicon*, 9(3), 419-436.
- Mirici, İ. H. & Göksu, A. (2016). Influence of residence in English speaking countries on the effective use of the English routine formulas. In İ. H. Mirici, İ. H. Erten, H. Öz, & I. Vodopija-Krstanović (eds.), *Research papers on teaching English as an additional language* (pp. 17-30). Rijeka, Croatia: Faculty of Humanities and Social Sciences, University of Rijeka.
- 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.
- Pan, F., & Hu, R. (2017). David Wood, *Fundamentals of formulaic language: An introduction*. London: Bloomsbury, 2015. Pp. vii 198. *Journal of Linguistics*, 53(3), 693-697. doi:10.1017/S0022226717000081
- Richards, J. C., & Schmidt, R. (2013). *Longman dictionary of language teaching & applied linguistics (4th ed.)*. New York, NY: Routledge. <https://doi.org/10.4324/9781315833835>
- Shin, Y. (2012). A new look at determiners in early grammar: Phrasal quantifiers. *Language Research*, 48(3), 573-608.
- Shin, Y. K., & Kim, Y. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System*, 69, 79-91. <https://doi.org/10.1016/j.system.2017.08.002>
- Shin, Y. K., Cortes, V., & Yoo, I. W. (2018). Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing*, 39, 29-41. <https://doi.org/10.1016/j.jslw.2017.09.004>
- Staples, S., Egber, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12, 214-225. <https://doi.org/10.1016/j.jeap.2013.05.002>
- Taşçı, S., & Ataç, B. A. (2018). Written grammatical errors of Turkish adult learners of English: An analysis. *Journal of International Social Sciences Education*, 4(1), 1-13.
- Uzun, K. (2016). Developing EAP writing skills through genre-based instruction: An action research. *International Journal of Educational Researchers*, 7(2), 25-38.
- Üstünbaş, Ü. (2014). *The use of formulaic language by English as a foreign language (EFL) learners in oral proficiency exams*. (Unpublished Master's Thesis). Bilkent University, Ankara, Turkey.
- Wei, L., & Ying, H. (2011). On the role of formulaic sequences in second language acquisition. *US-China Foreign Language*, 9(11), 708-713.
- Yalçın, I. (2010). *Turkish speaking first year and third year ELT students' syntactic errors in their argumentative essays*. (Unpublished doctoral dissertation). Anadolu University, Eskişehir, Turkey.

Zhao, H., & Macwhinney, B. (2018). The instructed learning of form-function mappings in the English article system. *The Modern Language Journal*, 102(1), 99-119. doi: 10.1111/modl.124490026-7902/18/99-119

Authors' Biodata / Yazar Bilgileri

Kutay Uzun is a lecturer at the departments of Foreign Languages and English Language Teaching at Trakya University, Turkey. At the same time, the researcher is a PhD candidate in English Language Teaching at Çanakkale Onsekiz Mart University, Turkey. His research interests fall within the domains of academic writing, writing psychology, assessing writing and educational data mining.

Kutay Uzun Trakya Üniversitesi Yabancı Diller ve İngilizce Öğretmenliği bölümlerinde Öğretim Görevlisi olarak görev yapmaktadır. Araştırmacı aynı zamanda Çanakkale Onsekiz Mart Üniversitesi İngilizce Öğretmenliği bölümünde doktora çalışmalarına devam etmektedir. İlgili alanları arasında akademik yazma becerisi, yazma psikolojisi, yazmanın değerlendirilmesi ve eğitimsel veri madenciliği bulunmaktadır.