Issue 45, Enero 2026 59-77

A corpus-based analysis of patterned structures and semantic types used by learners of English

Buse Uzuner (corresponding author)
Bahcesehir University

Ali Şükrü Özbay
Karadeniz Technical University

Ceyhun Yükselir
Osmaniye Korkut Ata University

Received: 2024-04-13 / Accepted: 2025-06-20
https://doi.org/10.30827/portalin.vi45.30573
Porta Linguarum ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: The use of patterned structures by language learners of different native language backgrounds needs to be investigated to better understand the degree of formulaicity. This mixed-method study investigates how EFL learners with Turkish, Japanese, and Norwegian L1s use verb-noun patterns in argumentation compared to native speakers of English. Data was obtained from the International Corpus of Learner English (version 3), which include essays written by university students with diverse L1 backgrounds and the LOCNESS, a native corpus. The research was conducted between Fall 2022 and Spring 2023 as part of a Master's thesis. Using Contrastive Interlanguage Analysis (CIA) and Corpus Pattern Analysis (CPA), the study reveals that accept and allow are statistically significant for further qualitative examination. The qualitative findings indicate inconsistent usage patterns for Turkish and Norwegian learners, while Turkish and Japanese learners showed more consistency. These results highlight the importance of raising learners' awareness of verb-noun patterns. One pedagogical implication is that teaching materials could integrate verb-noun collocations to enhance formulaic competence in academic writing.
Keywords: Corpus Analysis, learner corpora, verb-noun patterns, CIA, CPA.

Análisis basado en corpus de estructuras patrón y tipos semánticos utilizados por estudiantes de inglés

RESUMEN: El uso de estructuras patrón por parte de estudiantes de idiomas con diferentes lenguas maternas debe investigarse para comprender mejor el grado de formulariedad. Este estudio de métodos mixtos analiza cómo los estudiantes de inglés como lengua extranjera (EFL) con lenguas maternas turca, japonesa y noruega utilizan patrones verbo-sustantivo en la argumentación, en comparación con hablantes nativos de inglés. Los datos se obtuvieron de la versión 3 del International Corpus of Learner English (ICLE), que incluye ensayos redactados por estudiantes universitarios de diversos orígenes lingüísticos, y del LOCNESS, un corpus de hablantes nativos. La investigación se llevó a cabo entre el otoño de 2022 y la primavera de 2023 como parte de una tesis de maestría. Utilizando Análisis Contrastivo de Interlengua (CIA) y Análisis de Patrones de Corpus (CPA), el estudio revela que accept y allow son estadísticamente significativos para un análisis cualitativo más detallado.

Los hallazgos cualitativos indican patrones de uso inconsistentes en los estudiantes turcos y noruegos, mientras que los estudiantes turcos y japoneses mostraron mayor coherencia. Estos resultados subrayan la importancia de sensibilizar a los aprendices sobre los patrones verbo-sustantivo. Una implicación pedagógica es que los materiales didácticos podrían integrar colocaciones verbo-sustantivo para mejorar la competencia formularia en la escritura académica.
Palabras clave: Análisis de corpus, corpus de aprendices, patrones verbo-sustantivo, CIA, CPA

1. Introduction

We aimed to contrastively examine the verb-noun pattern use of Turkish, Norwegian, and Japanese learners along with native speakers of English using the Pattern Dictionary of English Verbs (PDEV) and corpus patterns analysis (CPA) from a collocational perspective. Linguistic features in learner language offer insights into linguistic differences (Flowerdew, 2015), and corpus-based studies have proved to be potential alternatives for providing effective teaching methods. However, the structural diversity between native and target language may cause learners to experience difficulties in comprehension and lack awareness of authentic and formulaic language (McEnery & Brezina, 2022). The correct use of multi-word combinations is one of the basics for more natural language production (Ellis et al., 2015; Erman et al., 2016; Gablasova et al., 2017), while incorrect usage might lead to grammatical errors or semantic shifts (Laufer & Waldman, 2011). Built on the premise that "meanings can be associated with phraseological patterns as well as or instead of words in isolation" (Hanks, 2012, p. 68), the CPA is used to make pattern descriptions (Hanks & Može, 2019). The study of verb patterns in learner language is not only an academic endeavour but a doorway to unravelling intricate layers of language acquisition.

CPA is a useful method for analysing English word patterns systematically by "mapping meaning onto words in text" (Hanks, 2004, p. 87). The method focuses on identifying patterns of actual use and associating implicatures (Cinková & Hanks, 2010), with a second order set of rules controlling the exploitation of the norms (Hanks, 2013). The large corpora and statistical evidence provide a better understanding of the line between norm and exploitations (Hanks & Može, 2019). We used CPA to group concordance lines based on their lexical, semantic, and syntactic features. PDEV was developed as a new kind of online pattern dictionary to define verb patterns and the meanings emerging from the patterns of each verb (Hanks, 2013; Hanks & Ma, 2020). Most learner language research includes different language backgrounds for comparing the effects of native language on English production (Ädel, 2015). One of the most effective methods of comparative corpus studies is Contrastive Interlanguage Analysis (CIA), developed as "the core of International Corpus of Learner English (ICLE) project whose main aim is to uncover 'foreign soundingness' in learner writing" (Granger, 1996, p. 43). The qualitative and quantitative nature of CIA enables the comparison of the diversity of the same language to evaluate differences between L2 English from L1 English (Hasselgård & Johansson, 2011).

2. Previous Studies

Altenberg and Granger (2001) investigated high frequency verbs used by Swedish and French learners and found that learners both underuse and produce incorrect usages of high frequency verbs. Swedish learners overused *make + adjective* structures more than French learners because English and Swedish share a similarity in this structure. Nesselhauf (2003) found that German EFL learners had difficulty in the production of verb-noun collocations due to the influence of their L1. Laufer and Waldman (2011) compared the use of verb-noun collocations by three L2 proficiency groups and the native corpus LOCNESS. All proficiency level learners used fewer collocations compared to the native corpus, and advanced learners used collocations more accurately than others although they still tended to make errors. Tsai (2020) examined the effectiveness of form-focused instruction (FFI) and concept-based instruction (CBI) in verb-noun collocation teaching to Chinese learners of English and found that the CBI group demonstrated superior performance over the FFI group. Kartal and Yangın-Ekşi (2018) investigated the effectiveness of corpus tools in teaching verb-noun collocations and found a statistically significant difference between experimental and control groups in terms of collocation production. Yuvayapan & Yükselir (2021) investigated Turkish EFL learners' perceptions towards collocation and collocational errors and found that learners were familiar with collocations and that errors stem from both the lack of collocation automation and L1 interference. Du et al. (2022) found that proficient learners tend to use semantically complex and long collocations. Previous literature has also shown positive attitudes towards corpus support in collocation learning (Ding et al., 2024; Saeedakhtar et al., 2020). Matte and Sarmento (2024) investigated Brazilian learners' collocation use and found that learners used fewer collocations. Recently, research by Gablasova and Brezina (2025) focused on the use of adjective-noun collocations and L2 proficiency in spoken English. Their study confirmed the systematic development of collocation use with increasing L2 proficiency.

Rees (2018) used CPA to investigate the differences in meaning and use of the Academic Word List between various academic disciplines and found that the AWL has discipline specific meanings based on their context. Alqarni (2019) used the PDEV to investigate its potential to describe learner corpora through the use of patterns and meanings. To apply the CPA method, the Swedish and Chinese sub-corpus of ICLE, and LOCNESS were used and 16 verb patterns were analysed. The verb *allow* was found the most noticeable for further investigation. The findings also showed that the PDEV can be successfully used for describing learners' pattern choices. Franklin (2020) applied the CPA method to analyse the use of "killing" terms. The People, Products, Pests and Pets (PPPP) corpus, consisting of animal-related texts, was used to examine animal discourse using PDEV.

All in all, there is a large number of corpus-based studies on collocations exploring the relationship between collocation use and L2 proficiency. However, there is still a need for investigating L2 learners' collocation use. Building upon this gap, this study aims to investigate how L2 learners from different L1 backgrounds use verb-noun collocations and how they differ from native English speakers and each other. Moreover, the study employs a novel approach by utilising the PDEV, which provides a more systematic and data-driven method for analysis.

The following research questions guided the study:

RQ1. What are the similarities and differences between verb patterns used by learners?

RQ2. What are the probable divergent patterns in the use of verb-noun patterns by learners?

RQ3. How can CPA be used to describe verb-noun patterns in learner corpora?

RQ4. To what extent is the use of PDEV for the description and identification of target divergent patterns possible in learner corpus analysis?

3. METHOD

The research was conducted between the 2022 fall term and 2023 spring term as part of the first author's Master's thesis and employs a sequential embedded mixed-method design as shown in Figure 1. In sequential mixed research, the phases follow each other chronologically and one analysis follows the previous one (Teddlie & Tashakkori, 2011). After conducting statistical analyses, the CPA provided deeper understanding of the use of verb-noun patterns. Moreover, CIA was used for qualitative analyses. The aim was to identify the over/under/divergent use of verb-noun patterns.

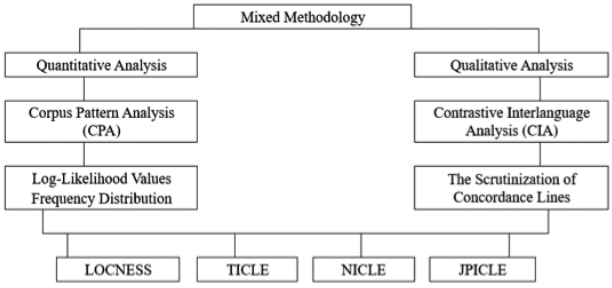

Figure 1. *Research Design*

3.1. Comparability measures

Since the heterogeneous nature of learner language may cause misinterpretation, the the data for this study were extracted from the International Corpus of Learner English (ICLE), version 3 (Granger et al., 2020). ICLEv3 includes data collected over several years by national teams and covers essays written approximately between the 1990s and 2018, with the final version released in 2020. Three sub-corpora, TICLE, NICLE, JPICLE as learner corpora, and LOCNESS, a native learner corpus, were chosen, and the PTE Academic Collocation List was used to decide the verb-noun collocations for further analysis. The selection of the target corpora and the comparability levels were based on a range of variables, as shown in Table 1.

Table 1. *Comparison between the target corpora ICLE and LOCNESS*

Variables	*ICLE*			*LOCNESS*	*Comparability**
	TICLE	*NICLE*	*JPICLE*		
Essay type	Argumentative	Argumentative	Argumentative	Argumentative Literary-mixed	++++
Number of Essays	280	317	366	322	++
Size	200,601 words	213,701 words	200,958 words	324,304 words	+++
Length of each essay (tokens)	+500	+300	+390	=/+ 500	++
The average age of students	22	23	20	17-22	++++
Level of study	University level (third to above years)	University level (first to the third year)	University level (first to fifth year)	University level (A-level students)	++
Topics	Abortion Sex Equality Nuclear Energy Suicide Capital Punishment Money Euthanasia Human Rights	Natural equality Feminism Science and technology Crimes Equality of men Prison system Abortion	Language acquisition English as a second language American people	Abortion Euthanasia Capital punishment Nuclear power Sex equality	+++
Age	22.08 (19-38)	23.94 (18-55)	20.06 (18-36)	17- 48	++++
Gender	F=81.43% / M=18.57%	F=74.13 / M=% 25.87%	F=72.95 / M=% 26.50%	NA	+++
English Level	B2 & Lower	C1+	B2 & Lower	Native	+++

* +: low; ++: between low and average; +++: between average and high; ++++: high

One learner corpus was selected from Asia, Europe, and Eurasia to represent different linguistic backgrounds. The ICLE aims to provide non-native varieties of English (Granger et al., 2020) and comprises written texts of argumentative essays by EFL learners. TICLE was selected due to the researchers' empirical observations indicating a lack of awareness on the part of Turkish English learners of word combinations. NICLE was selected because English is widely used in Norway, with high exposure through education, media and daily communication, and because Norwegian learners show more proficiency in English than Turkish learners (Granger et al., 2020). English is used as a second language in Norway and is a part of daily life through frequent exposure to television, music and films, not just being taught in schools. JPICLE was selected because Japanese learners show similar proficiency in TICLE and NICLE and have a similar total word count (Granger et al., 2020; Schanding & Pae, 2018).

3.2. Selection of target verbs

The PTE Academic Collocation List, covering 2,469 most frequent written academic collocations extracted from the Pearson International Corpus of Academic English (PICAE), was used. Sixty-three verb-noun collocations were found from the list as a first step (see Table 2).

Table 2. *Verbs and their raw frequencies in four target corpora (verb-noun collocations)*

LOCNESS	*ICLE*							
	TICLE		*NICLE*		*JPICLE*			
V	RF	V	RF	V	RF	V	RF	
have	1281	have	1478	have	1227	have	1341	
make	631	give	430	make	404	make	474	
take	424	make	407	do	390	use	464	
use	382	do	360	take	300	take	318	
give	323	use	313	use	266	do	266	

Further verb selection criteria were employed as the raw frequencies were inadequate for investigating 21 verbs at the same time. To do this, the verbs were searched for in PDEV for common shared usages. The selection criteria demanded that every selected word from the PDEV project had at least two patterns. At the end of the search, 10 words and frequencies were found to be common in the PTE Academic Collocation List and PDEV project (see table 3).

Table 3. *Frequency of verbs in each source*

VERBS	*PDEV*	*PTE*	*LOCNESS*	*TICLE*	*NICLE*	*JPICLE*
Allow	8	1	147	24	25	17
Accept	9	1	100	57	18	20
Follow	20	3	49	16	25	22
Affect	2	6	47	61	18	21
Adopt	10	2	30	12	4	4
Maintain	5	1	27	10	17	6
Add	8	1	23	9	7	5
Conduct	6	5	20	6	1	5
Attend	3	1	19	18	17	34
Apply	8	3	17	39	2	9

4. RESULTS

We used Sketch Engine for analysis and PDEV for determining the verb patterns created by tagging verbs with their semantic types based on a custom ontology, where each

entry includes corpus-based CPA patterns and implicatures. This provided more detailed descriptions of verb patterns and matching meanings with each verb's unique patterns. The mixed-method approach and sequential embedded design were employed in two phases. In sequential mixed research, the phases follow each other chronologically and one analysis follows the previous one (Teddlie & Tashakkori, 2011). Firstly, frequency and statistical analysis of the verbs was conducted in a contrastive fashion. Another quantitative analysis was carried out to explain the diversity in concordance lines through log-likelihood and overuse or underuse of the verbs in specific patterns. Lastly, divergent usages of the verb-noun patterns were examined, and their patterns were matched through referring to sources such as Grammar Patterns1: Verbs (GR1: Verbs) (Francis et al., 1996) and British National Corpus (BNC). In the quantitative section, CPA was carried out and the log-likelihood and frequency distribution of the verb-noun patterns across three learner corpora was calculated.

4.1. Identifying Patterns

Three verbs, *adopt*, *apply*, and *conduct*, were eliminated as they had very few frequencies and seven verbs were chosen for further investigation as their frequencies were higher than their pattern numbers in the PDEV. Figure 2 shows the distribution of target verbs and pattern numbers.

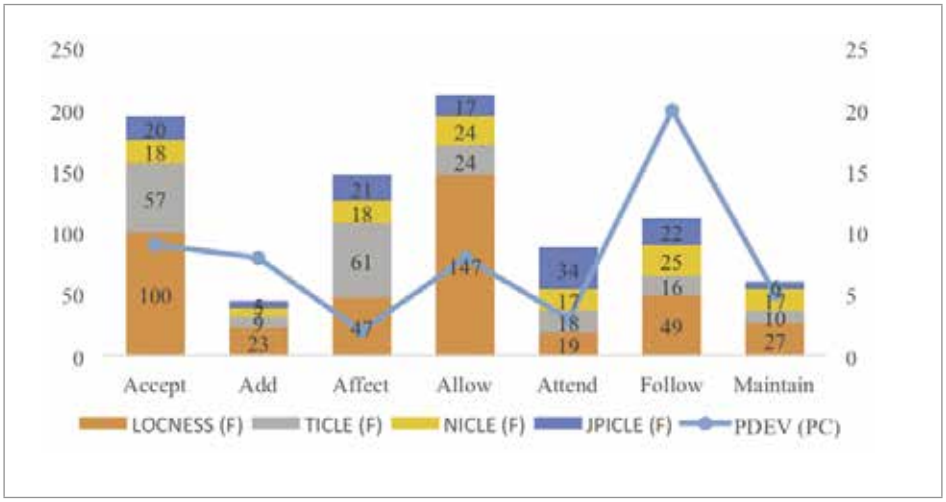

Figure 2. *Target verb-n distribution*

Log-likelihood and overuse and underuse values of these target verbs were calculated to determine their statistical significance in each corpus.

Table 4. *Log-likelihood values and over and under representations*

Verbs	TICLE		NICLE		JPICLE	
	Overuse Underuse	Log-Likelihood	Overuse Underuse	Log-Likelihood	Overuse Underuse	Log-Likelihood
Accept	-	0.40 / p<0.05	-	33.28 / p < 0.0001	-	26.67 / p < 0.0001
Add	-	1.59 / p<0.05	-	3.55 / p < 0.01	-	5.50 / p < 0.01
Affect	+	13.83 / < 0.001	-	4.01 / p < 0.0 1	-	1.60 / p<0.05
Allow	-	50.93 / p < 0.0001	-	53.82 / p < 0.0001	-	65.08 / p < 0.0001
Attend	+	1.49 / p<0.05	+	0.87 / p<0.05	+	14.52 / < 0.001
Follow	-	5.83 / p < 0.01	-	1.05 / p<0.05	-	1.62 / p<0.05
Maintain	-	2.25 / p<0.05	-	0.02 / p<0.05	-	6.26 / p < 0.01

The statistical investigation of target verbs shows that the verb *attend* is over-used in target corpora, and the verb *affect* is only over-used in TICLE. The verb *allow* is under-used in the target corpora relative to LOCNESS, with LL values of 50.93 in TICLE, 53.82 in NICLE, and 65.08 in JPICLE. Similarly, the verb *accept* is under-used in the target corpora, with LL values of 33.28 in NICLE, 26.67 in JPICLE.

4.2. Identifying *accept-n* patterns

The raw frequency of *accept-n* usages in target corpora is 195 after manual examination and pattern match in the PDEV. The log-likelihood (LL) scores and overuse and underuse values were calculated to decide whether the frequency differences were statistically significant.

Table 5. *Statistical analysis of* accept-n *patterns*

Corpora	Frequency	Log-likelihood Scores	Overuse/Underuse
TICLE	57	0.40 p<0.05	-
NICLE	18	**33.28 p<0.0001**	-
JPICLE	20	**26.67 p<0.0001**	-

There were 57 occurrences of *accept-n* patterns with underused patterns. Although being underused in each corpus, *accept-n* use in TICLE shows the highest significance with an LL score of 0.40. There are nine patterns of the verb *accept* in PDEV; however, only six patterns match the instances of the target corpora.

Table 6. *Log-likelihood and over-under representations of patterns for* accept-n

Pattern Number	1	2	5	8	9
TICLE	2.13 / -	2.25 / +	0.16 / +	**10.98** / +	-
NICLE	0.47 / -	0.26 / +	1.32 / +	2.35 / +	-
JPICLE	1.10 / -	1.53 / +	**9.69** / +	-	-

Table 6 shows that the LL value of pattern 9 was not calculated. This is because there were no instances in LOCNESS. Pattern 8 is overused in TICLE, with an LL value of 10.98. Moreover, pattern 5 is also overused in JPICLE, with an LL value of 9.69. An investigation of concordance lines for pattern 8 in TICLE shows 14 occurrences.

4.2.1. Identifying allow-n Patterns in Target Corpora

Allow-n pattern investigation shows that allow is the most noticeable verb with eight patterns in PDEV and is the highest frequency in the target corpora, making it a good choice for further investigation.

Table 7. *Statistical analysis of* allow-n *patterns*

Allow-n	Frequency	Log-likelihood Score	Overuse/Underuse
TICLE	24	50.93 / $p < 0.0001$	-
NICLE	24	53.82 / $p < 0.0001$	-
JPICLE	17	65.08 / $p < 0.0001$	-

The LL values in Table 7 indicate that *allow-n* patterns are underused in the target corpora relative to LOCNESS. As mentioned previously, *allow* has eight patterns in the PDEV; however, four patterns match the instances in the target corpora, seen in Table 8.

Table 8. *PDEV entries of* allow-n *patterns*

Pattern Number	Pattern and Implicature	LOCNESS (*f*)
1	**Pattern: [[**Human 1 | Institution 1 | Eventuality]] **allow** [[Human 2 | Institution 2 | Self]] [to+INF] **Implicature:** [[Human 1 | Institution 1 | Eventuality]] gives [[Human 2 | Institution 2 | Self]] the opportunity, time, or permission {to/INF [V]} **Example:** You should not **allow the money to** rule you, you should rule it according to your desires. (TICLE) **Example:** These games **allow people to** use their imagination at a high level. (NICLE) **Example:** The US should **allow local people to** establish their own government**.** (JPICLE)	97
2	**Pattern: [[**Human 1 | Institution 1 | Eventuality]] **allow** [[Institution 2 | Self | Privilege | Human 2]] **Implicature:** [[Human 1 | Institution 1 | Eventuality]] gives [[Human 2 | Institution 2 | Self]] permission or the opportunity to have [[Privilege]] **Example:** the risk of euthanasia is … **allowing individuals the freedom** to say "here is as far as I go". (TICLE) **Example:** I think that they should be **allowed more play time**, where they can use their imagination more freely. (NICLE) **Example:** Such education is not education that does not **allow students opportunities** to learn real life. (JPICLE)	8

Pattern Number	Pattern and Implicature	LOCNESS (f)
3	**Pattern: [[**Eventuality | Human | Institution]] **allow** [[Physical_Object]] {to be V-ED} [to+INF] **Implicature:** [[Human | Institution | Eventuality]] causes or provides the opportunity for [[Physical_Object]] {to be V-ED} **Example:** On the other hand, huge public support that animal testing is necessary to **allow drugs to be safely tested** on and used by humans. (TICLE) **Example:** Why can't children be **allowed to be** bored sometimes… (NICLE)	7
4	**Pattern: [[**Human | Eventuality 1 | Institution]] **allow** [[Eventuality 2]] **Implicature:** [[Human | Institution | Eventuality 1]] gives the opportunity, time, or permission for [[Eventuality 2]] to occur **Example:** Remember that just standing idle when you could prevent death is a decision to **allow suicide**. (TICLE) **Example:** Even so, we **allow a certain amount of censorship** in our media. (NICLE) **Example:** Japan administration obey the OECD's logic and doesn't **allow the necessity of "Indication**". (JPICLE)	31

There are only four patterns for the verb and no instances for the pattern numbers 5, 6, 7 and 8 and pattern 3 in JPICLE. A detailed investigation was conducted to reveal the frequency distribution among patterns. Table 9 summarises the statistical values of the patterns.

Table 9. *Log-likelihood and over-under representations of patterns for* allow-n

Pattern Number	1	2	3	4
TICLE	**4.19 / -**	0.07 / -	1.68 / +	**5.58 / +**
NICLE	0.00 / +	0.27 / +	0.02 / -	0.00 / -
JPICLE	0.12 / -	0.01 / +	-	1.19 / +

The patterned analysis of the target verb reveals that the frequencies in the corpora are less than those in LOCNESS. Pattern 4 is overused in TICLE, with an LL value of 5.58. Another finding is that pattern 1 is under-used in TICLE, with an LL value of 4.19. There is a total of 24 instances for *allow-n* and an investigation for pattern 4 in TICLE reveals 12 matches. On the other hand, of the 147 instances for the target verb in LOCNESS, 31 match patterns 4 in the PDEV.

In pattern 4, the implicature is that Human or Institution or Eventuality 1 gives the opportunity, time, or permission for Eventuality 2 to occur. The description of example 43. is Eventuality 1 allowing Eventuality 2.

> 4T. Thus, not all the societies want to **allow the euthanasia** because of their cultural and moral norms I'm agree with the application of the euthanasia. (TICLE:43)
> 4.N. This is also a religious issue for many people, because of the fact that most religions don't **allow abortion**. (NICLE:67)
> 4.J. In big cities most apartments do not **allow pets**. (JPICLE:78)

4.2. Divergency of *Accept* and *Allow*

There are nine patterns in PDEV for *accept* based on the 250 instances in BNC50 data. Table 10 shows the raw frequencies. Examining the 371 concordance lines of *accept* shows that only 25 do not match any of the nine patterns in PDEV. TICLE has 117 occurrences for *accept* and nine of them are possible divergent instances.

Table 10. *Raw frequency and LL values of* accept

Accept-v	Raw Frequency		
LOCNESS	172	Log-likelihood Score	Overuse/Underuse
TICLE	117	0.36/p < 0.05	+
NICLE	49	**30.22/ p < 0.0001**	-
JPICLE	33	**48.25/ p < 0.0001**	-

Further investigation was carried out to decide the divergency length. The results show that seven of nine instances match the patterns found in GR1: Verbs but not in PDEV. The classification of their patterns in GR1: Verbs (Francis et al., 1996) is given in Table 11.

Table 11. *Pattern classification of* accept *in TICLE instances not matched in PDEV*

Classification	Pattern	Example	BNC (*f*)
Complex patterns with prepositions and adverbs	V n *to* n	*accept* them to internship programs	39
	V n *as* n	*accept* euthanasia as a remedy	57
	V n *for* n	*accept* a real value for the real world	200

Since the verb patterns in Table 11 were not found in GR1: Verbs, they were searched for in BNC and occurred 39 times indicating a low frequency for such a large database and was left for further examination. NICLE has 49 instances of *accept* and only three do not match any pattern in PDEV.

In the first classification (V n as adj), a noun group, a prepositional phrase with *as*, and an adjective follow the verb, indicate a specific quality or description of the object of the sentence. Then, in the use of the "V n for n" pattern, the verb of the example (*accept*) is followed by a noun and a prepositional phrase *with* for describing the noun. Lastly, the pattern "V n from n" refers to acquiring something from someone or somewhere, meaning being willing to receive support from someone. Five instances of *accept* in JPICLE do not match any pattern in the PDEV. According to the investigation of these instances, the patterns of four instances occurred in GR1: Verbs, and one instance could not be defined. The first four instances matched patterns in Grammar Patterns as complex patterns and complex patterns with prepositions and adverbs. However, the fifth occurrence could not be defined because the term *sudi* was not found in BNC and marked as an error. Table 12 shows the classification of the patterns for the remaining instances in NICLE with their BNC frequencies.

Table 12. *Pattern classification of* accept *in NICLE instances not matched in PDEV*

Classification	Pattern	Example	BNC (*f*)
Complex Patterns	V n *to-inf*	*accept* using English to teach	44
Complex patterns with prepositions and adverbs	V n *to* n	*accept* nuclear energy to society	30
	V n *at* n	*accept* having pets at home	8
	V n *from* n	*accept* benefit from other countries	88

The patterns of *accept* are classified as complex patterns and complex patterns with prepositions and adverbs. In its first pattern, *accept* is followed by noun group and infinitive (V n to-inf) as a complex pattern. The example (accept using English to teach) shows the tendency of approving a noun group for a specific purpose, and "using English" is used as a noun group. Three patterns for *accept* in JPICLE are complex patterns with prepositions and adverbs. The first pattern of this classification (V n to n) shows that *to* is used as a preposition rather than infinitive and contributes to the passive meaning of the sentence. *To* in this instance shares a similar meaning with *by* in the passive form to indicate who will accept the eventuality. Other prepositions, *at* and *from*, are used to refer to the place of the eventuality.

4.2.2. Focus on allow

Eight patterns were recorded for *allow*. Table 13 shows the raw frequency and log-likelihood values in each corpus.

Table 13. *Raw frequency and LL values of* allow

Allow-v	Raw Frequency		
LOCNESS	270	Log-likelihood Score	Overuse/Underuse
TICLE	60	65.58 / $p < 0.0001$	-
NICLE	72	53.39 / $p < 0.0001$	-
JPICLE	42	94.38 / $p < 0.0001$	-

The log-likelihood values of *allow* indicate an association between the observed and the LOCNESS data with significant association. In total, 444 concordance lines of *allow* were examined, with only eight of them not matching any of the eight patterns present in the PDEV. The raw frequency of *allow* in TICLE is 60 and three instances are potential divergent patterns. *Allow* is underused in TICLE, with an LL value of 65.58. Three of the 60 instances in TICLE are possible divergent usages. Thus, they were further examined by checking their patterns in GPB1: Verbs and BNC. Each instance matches patterns in Grammar Patterns within three classification types, which are simple patterns (v to-inf), simple patterns with prepositions and adverbs (v with n) and complex patterns with prepositions and adverbs (v n in n). Simple verb patterns have meaning groups such as *begin*, *appear*, *try*, *manage*, *fail*, *regret to say*, *hasten*, *chance*, and *tend.* The TICLE example with the simple pattern primarily focuses on the initiation and maintenance of an action. Moreover,

the second pattern group consists of *allow*, followed by a prepositional phrase comprising *with* and a noun group (allow with humans). Finally, the verb *allow* is followed by a noun group and a prepositional phrase with *in* followed by a noun group. Further investigation of *allow* was conducted on NICLE, involving 72 instances. Two of the 72 instances do not match any pattern in the PDEV. The verb *allow* is underused in NICLE, with an LL value of 53.39, significant at 99.99%. As two instances do not match any pattern in the PDEV, further investigation was undertaken by checking their patterns. Both instances match patterns in Grammar Patterns as a simple pattern. *Allow* is followed by *to*-infinitive. In the second instance and the verb (be) is followed by an object (airplanes). Finally, the same procedure was carried out with instances in JPICLE. Three of the 42 instances do not match any pattern in the PDEV. *Allow* is underused in JPICLE, with the highest LL value of 94.38, and the three instances match patterns in Grammar Patterns. The instances are classified as simple pattern (v to-infinitive) and complex pattern with preposition (v n as n).

5. Discussion

We found that the verbs *accept* and *allow* were underused, revealing that not all corpora have instances for each pattern in the PDEV. This finding broadly corroborates the ideas of Sinclair (1991) about the relationship between sense and meaning, suggesting the meaning of a specific word is related to a specific pattern. Proficiency level may be another reason. In ICLEv3 Manuel, Granger et al. (2020) indicate that English is widely used in Norway. Thus, NICLE and LOCNESS were expected to show similar pattern use in their analysis as Norwegian learners of English exhibited a more proficient learner profile than Turkish and Japanese learners. The lack of patterns found for 4, 5, 8, and 9 indicate that Turkish learners show wider use of *accept-n* patterns than native speakers. This may be a counterpart to Alqarni's (2019) findings that Swedish learners of English show narrower use of patterns of *allow*. Unlike the findings of Altenberg and Granger (2001) that revealed learners of English from different language backgrounds abstain from using high-frequency verbs, Turkish learners showed a wider pattern use and the frequencies show that learners with different proficiencies and language backgrounds use fewer verb-noun patterns than native speakers (Laufer & Waldman, 2011). Moreover, learners did not use patterns 5 and 6 of *allow* while native speakers did use these patterns. This may be because Turkish and Japanese are agglutinative languages, differing from English as a preposition language. The verb-noun pattern use of Turkish learners indicates the superiority of lexical collocational knowledge over grammatical.

For the second question, which asked what the probable divergent patterns are in learners' use of verb-noun patterns, learners produced some divergent patterns when the PDEV was taken as a reference and when they were checked in GP1:Verbs and searched in BNC, showing that Turkish learners produce more possible divergent patterns, indicating a L1 transfer effect on collocation use (Laufer & Waldman, 2011; Peters, 2016). Barfield (2006) found that Japanese learners have better knowledge of single verbs than a combination of verbs and nouns. Altenberg and Granger (2001) also highlight that learner errors mostly occur due to their L1. Lu et al. (2022) found similar results with L3 learners and concluded that verb-noun combination errors stem from the L1 effect. The divergent pat-

tern use of Turkish and Japanese learners can be explained with the Yamashita and Jiang (2010) study, demonstrating that learners with lower proficiency in English tend to make more errors. The L2 learners' divergent patterns can be seen as unnatural, and they can be judged as non-native; however, the correct use of collocations is directly proportional to proficiency level (Granger et al., 2006). Nevertheless, even advanced learners may not use verb-noun patterns correctly, according to several studies supporting the idea that advanced learners misuse or use fewer collocations (Granger, 1998; Nesselhauf, 2003). L2 learners either overuse or underuse these patterns or misuse them, resulting in collocational errors or divergent usages (Altenberg & Granger, 2001; Bağcı, 2014; Nesselhauf, 2005). These errors or divergent usages of patterns may also stem from either the lack of pattern information in dictionaries or the learners' lack of awareness of the significance. Franklin (2020) states that the patterns and meaning of a verb are more important than the verb itself, as the patterns of a specific verb may be used interchangeably. Our findings also show that learners do not use some patterns of a verb as frequently as other patterns of the same verb, making some divergent pattern use. Finally, the divergent instances in learner corpora may be beneficial for the future intensification of PDEV entries and the development of EFL classrooms for enhancing collocation awareness.

For the third question, which focused on how CPA describes verb-noun patterns, we investigated verb+noun patterns although some instances could not be matched with the patterns existing in the PDEV. Alqarni (2019, p. 244) explains the absence of patterns in learner data as "the task the learners were asked to complete rather than a lack of knowledge". The overuse and underuse of the patterns indicate a tendency of the target learners towards the specific patterns in the target verbs. The possible reason for the representation of the target verb patterns can be given to the verb pattern itself or to the learner's knowledge about the target verb. Alqarni (2019) indicates that the frequency of a verb may be low in the native corpus, making it an uncommon pattern in learner data. Based on previous research, it can be concluded that learners with higher proficiency levels tend to use more native-like collocations, as B and C level learners differ from other proficiency groups in terms of frequency, and they are more flexible in combining lexical features (Brezina & Fox, 2021; Gablasova & Brezina, 2025).

The final question relates to the use of PDEV for the description and identification of target divergent patterns. These verb patterns consist of the node word and the additional elements. For example, the 20 pattern entries of the verb *follow* includes different elements such as semantic information (eventuality) and structural information (prepositional phrase or that clause). On the other hand, implications helped in identifying the patterns according to their meanings given in these entries. The PDEV also helps to identify learner errors in verb pattern use by applying CPA method. Hanks and Ma (2020) argue that the effectiveness of the PDEV is because it is a dictionary of English verbs using the authentic data to demonstrate learners how verbs are used in native language. CPA/PDEV includes both syntactic and semantic analysis based on corpus data. The former investigates the idiomatic expressions, passive voice, and phrasal uses, and the latter investigates the semantic features shared by collocations (El Maarouf & Baisa, 2013). Previous literature found that CPA is large enough to sample general language that is not based on a specific discourse (Franklin, 2020; Rees, 2018).

It is possible to extend these findings on the basis of analyzing diverse verb-noun patterns from different learner corpora, covering diverse linguistic backgrounds and proficiency levels. Employing sizable balanced samples of native and non-native English speakers may increase more detailed and nuanced understanding of collocations and the mutual comparison between verbal speech and academic written text is likely to spot possible differences in how the participants use verb-noun patterns.

6. Conclusion and Implications

We investigated EFL learners' use of verb-noun patterns through CPA. It builds on the Sinclairian approach to lexical analysis by examining English verb use based on evidence from the BNC (Hanks, 2013). The findings show that Turkish learners use a wider range of verb-noun patterns than Norwegian and Japanese learners and that CPA and PDEV provide a significant description of learner corpora. The unknown patterns may be indicative of L2 language proficiency and learner difficulties in language production (Ortega, 2009). Some verb-noun usages not found in BNC may also show that they might be L1-related if learners translate multi-word expressions directly from their native language into English (Granger et al., 2006).

The study advanced our knowledge of non-native English speakers' use of verb-noun patterns in their academic papers. Conducted with minimal limitations, the study ensured a comprehensive exploration of the research topic. However, one limitation is that PDEV is still not a complete tool and does not include all English verb patterns. Thus, the study was limited to certain verbs which had patterns in PDEV. On the other hand, the study was also based on a specific dataset, three learner corpora and one native corpus, which presented challenges for determining the verbs in each corpus. Future studies could examine more patterns not found in PDEV to increase the generalisability. It is also recommended that further research may include more language backgrounds. While this study is limited to those verb patterns included in the PDEV, the findings suggest several pedagogical implications for English classrooms. Both CPA and PDEV may be related to Data-Driven Learning (DDL) activities, a popular teaching and learning method in ELT helping learners explore patterns. Boulton (2012) states that there are online corpus-based tools providing instant benefits to language learners through DDL activities. The application of DDL requires teacher-created concordances to enable learners to comprehend language structures such as vocabulary, grammar, and collocations (Crosthwaite et al., 2021, p. 1). Language teachers could use DDL to teach verb-noun patterns to their learners through various classroom activities (Pinto et al., 2023), enabling learners to discover language patterns based on authentic language data with or without the help of their teachers (Boulton, 2007, p. 1). Through CPA, learners can observe the rules of the language and learn how to combine their knowledge with real-life applications. Moreover, the analysis of real language patterns enables students to understand how language can change in different contexts. Students can learn the correct use of verbs in communication by observing these patterns in the PDEV. Moreover, corpus-based English lessons are more practical and motivational for students, leading them to be more autonomous while benefiting from the real-life examples from credible sources (Şimşek & Can, 2023). The tool can also be used to detect learner errors in verb use through learner corpus analysis. These errors might help language teachers to understand whether or why

learners have difficulty in understanding and using verbs in their learning process and to apply new methods or strategies to overcome these problematic language uses.

Future studies may explore a more diverse range of verb-noun patterns not found in PDEV to enhance the generalizability of the findings. Additionally, it is recommended that further research incorporate a more diverse set of linguistic backgrounds, allowing for a more nuanced understanding of how non-native English speakers from different language groups utilize verb-noun patterns in academic writing. Expanding the scope to include various academic disciplines and genres could also provide deeper insights into the influence of contextual factors on phraseological choices.

Funding information

This work was supported by the Office of Scientific Research Projects of Karadeniz Technical University. Project number: SYL-2023-10667. This work is based on the first author's Master's thesis titled "A corpus analysis of learner English from different backgrounds: Investigating verb-noun patterns through the pattern dictionary of English verbs", which is accessible via National Thesis Center at https://tez.yok.gov.tr/UlusalTezMerkezi/

7. References

Ädel, A. (2015). Variability in learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 401-422). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414

Alqarni, A. S. (2019). Applying corpus pattern analysis to learner corpora: investigating the pedagogical potential of the pattern dictionary of English verbs [Doctoral dissertation, University of Birmingham].

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, *22*(2), 173-195. https://doi.org/10.1093/applin/22.2.173

Bağcı, N. D. (2014). Turkish university level EFL learners' collocational knowledge at receptive and productive levels [Master's thesis, Middle East Technical University].

Barfield, A. W. (2006). An exploration of second language collocation knowledge and development [Doctoral dissertation, University of Wales, Swansea].

Boulton, A. (2007). DDL is in the details...and in the Big Themes, In M. Davies, P. Rayson, S. Hunston, & P. Danielsson (Eds.), *Proceedings of Corpus Linguistics 2007*.

Boulton, A. (2012). What data for data-driven learning? *EuroCALL Review*, *20*(1), 23-27. http://doi.org/10.4995/eurocall.2012.16038

Brezina, V., & Fox, L. (2021). Adjective + noun collocations in L2 and L1 speech: Evidence from the Trinity Lancaster Corpus and the spoken BNC2014. In S. Granger (Ed.), Perspectives on the L2 phrasicon: The view from learner corpora (pp. 152-177). Multilingual Matters.

Cinková, S., & Hanks, P. (2010). Validation of Corpus Pattern Analysis: Assigning pattern numbers to random verb samples. http://nlp.fi.muni.cz/projects/cpa/CPA_valiman.pdf

Crosthwaite, P., Luciana, & Wijaya, D. (2021). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning*, 1-29. https://doi.org/10.1080/09588221.2021.1995001

Ding, C., Reynolds, B. L., & Ha, X. V. (2024). Learners' perceptions of a collocation instruction and practice component in a Chinese EFL context. *The Language Learning Journal*, *52*(1), 37-48. https://doi.org/10.1080/09571736.2022.2098366

Du, X., Afzaal, M., & Al Fadda, H. (2022). Collocation use in EFL learners' writing across multiple language proficiencies: A corpus-driven study. *Frontiers in Psychology*, *13*, 752134. https://doi.org/10.3389/fpsyg.2022.752134

Ellis, N. C., Simpson-Vlach, R., Römer, U., O'Donnell, M., & Wulff, S. (2015). Learner corpora and formulaic language in second language acquisition. In S. Granger, G., Gilquin & F. Meunier. (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 357-378). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414.016

El Maarouf, I., & Baisa, V. (2013). Automatic classification of semantic patterns from the Pattern Dictionary of English Verbs. In *Proceedings of the Joint Symposium on Semantic Processing: Textual Inference and Structures in Corpora*, November 2013. (pp. 95-99).

Erman, B., Forsberg Lundell, F., & Lewis, M. (2016). Formulaic language in advanced second language acquisition and use. In K. Hyltenstam (Ed.), *Advanced Proficiency and Exceptional Ability in Second Languages* (pp. 111-147), De Gruyter Mouton. https://doi.org/10.1515/9781614515173-007

Flowerdew, L. (2015). Learner corpora and language for academic and specific purposes. In S. Granger, G. Gilquin, & F. Meunier. (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 465-484), Cambridge University Press. http://doi.org/10.1017/CBO9781139649414.021

Francis, G., Hunston, S., & Manning, E. (Eds.). (1996). *Grammar patterns 1: Verbs*. HarperCollins.

Franklin, E. (2020). Acts of killing, acts of meaning: An application of corpus pattern analysis to language of animal-killing [Doctoral dissertation, Lancaster University].

Gablasova, D., & Brezina, V. (2025). Adjective+ noun collocations in L2 spoken English: How robust is the role of proficiency? *International Journal of Learner Corpus Research*, *11*(1), 79-113. https://doi.org/10.1075/ijlcr.23034.gab

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus□based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, *67*(S1), 155-179. https://doi.org/10.1111/lang.12225

Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds), *Languages in Contrast: Text-based Cross-linguistic Studies* (pp. 37-51). Lund University Press.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In Cowie, A. (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 145-160). Oxford University Press. https://doi.org/10.1093/oso/9780198294252.003.007

Granger, S., Dupont, M., Meunier, F., Naets, H., & Paquot, M. (2020). *The International Corpus of Learner English. Version 3.* Louvainla-Neuve: Presses universitaires de Louvain. https://dial.uclouvain.be/pr/boreal/object/boreal:229877

Granger, S., Paquot, M., & Rayson, P. (2006) Extraction of multi-word units from EFL and native English corpora. The phraseology of the verb 'make'. In A. Häcki Buhofer and H. Burger (Eds.), *Phraseology in Motion I: Methoden und Kritik. Akten der Internationalen Tagung zur Phraseologie* (pp. 57-68). Schneider Verlag Hohengehren.

Hanks, P. (2004). Corpus Pattern Analysis. In G. Williams & S. Vessier (Eds.), *11th Euralex International Congress Proceedings* (pp. 87-97). Université de Bretagne-Sud.

Hanks, P. (2012). Corpus Evidence and Electronic Lexicography. In S. Granger, & M. Paquot (Eds.), *Electronic Lexicography* (pp. 57-82), Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199654864.001.0001

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations,* MIT Press. https://doi.org/10.7551/mitpress/9780262018579.001.0001

Hanks, P., & Ma, W. (2020). Meaning and grammar in the light of corpus pattern analysis. *International Journal of Lexicography*, *34*(1), 135-149. https://doi.org/10.1093/ijl/ecaa027

Hanks, P., & Može, S. (2019). The way to analyse 'way': A case study in word-specific local grammar. *International Journal of Lexicography*, *32*(3), 247-269. https://doi.org/10.1093/ijl/ecz005

Hasselgård, H., & Johansson, S. (2011), Learner corpora and contrastive interlanguage analysis. In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (Eds.), *A Taste for Corpora in Honour of Sylviane Granger*, (pp. 33-61). John Benjamins. https://doi.org/10.1075/scl.45.06has

Kartal, G., & Yangineksi, G. (2018). The effects of using corpus tools on EFL student teachers' learning and production of verb-noun collocations. *PASAA: Journal of Language Teaching and Learning in Thailand*, *55*, 100-125.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647-672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Lu, H.-C., Chung Cheng, A., & Yeh, M-H (2022). Acquisition of L3 Spanish combinations: development in multilingual contexts. *Porta Linguarum An International Journal of Foreign Language Teaching and Learning, 38*, 159-172. https://doi.org/10.30827/portalin.vi38.22338

Matte, M. L., & Sarmento, S. (2024). Investigating Brazilian English learners' use of academic collocations: A corpus-based study. In S. Sarmento, R. Rebechi, M. L. Matte (Eds.), *English for Academic Purposes* (pp. 178-203). Editora Zouk.

McEnery, T., & Brezina, V. (2022). *Fundamental Principles of Corpus Linguistics*. Cambridge University Press. https://doi.org/10.1017/9781107110625

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223-242. https://doi.org/10.1093/applin/24.2.223

Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. John Benjamins. https://doi.org/10.1075/scl.14

Ortega, L. (2009). *Understanding Second Language Acquisition*. Routledge.

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113-138. https://doi.org/10.177/1362168814568131

Pinto, P. Crosthwaite, C. T. Carvalho, F. Spinelli, T. Serpa, W. Garcia, & A. O. Ottaiano (Eds.). (2023). *Using Language Data to Learn about Language: A Teachers' Guide to Classroom Corpus Use*. UQ Pressbooks. https://doi.org/10.14264/3bbe92d

Rees, G. P. (2018). A phraseological multi-discipline approach to vocabulary selection for English for academic purposes [Doctoral dissertation, Universitat Pompeu Fabra].

Saeedakhtar, A., Bagerin, M., & Abdi, R. (2020). The effect of hands-on and hands-off data-driven learning on low-intermediate learners' verb-preposition collocations. *System*, *91*, 102268. https://doi.org/10.1016/j.system.2020.102268

Schanding, B., & Pae, H. K. (2018). Shell noun use in English argumentative essays by native speakers of Japanese, Turkish, and English: Frequency and rate of noun-pattern attraction. *International Journal of Learner Corpus Research*, *4*(1), 54-81. https://doi.org/10.1075/ijlcr.16014.sch

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Şimşek, T., & Can, C. (2023). Integration of corpus literacy consultation into language teacher education: Course design, implementation, and evaluation. *Porta Linguarum An International Journal of Foreign Language Teaching and Learning*, *39*, 193-211. https://doi.org/10.30827/portalin.vi39.23885

Teddlie, C., & Tashakkori, A. (2011). Mixed method research: Contemporary issues in an emerging field. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (pp. 285-300). Sage.

Tsai, M. H. (2020). The effects of explicit instruction on L2 learners' acquisition of verb–noun collocations. *Language Teaching Research*, *24*(2), 138-162. https://doi.org/10.1177/1362168818795188

Yamashita, J., & Jiang, N. A. N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*(4), 647-668. https://doi.org/10.5054/tq.2010.235998

Yuvayapan, F., & Yükselir, C. (2021). Understanding Turkish EFL students' perceptions about collocations and investigating their collocational errors in descriptive and argumentative essays: Understanding Turkish EFL students' perceptions about collocations. *International Journal of Curriculum and Instruction*, *13*(3), 2179-2194.