

CORPUS UTILITY IN TEACHING LEGAL ENGLISH VOCABULARY TO LAW STUDENTS

Hương Thanh Nhac*

Faculty of legal foreign languages, Hanoi Law University, Viet Nam

*correspondence: nhacthanhhuong@gmail.com

<https://doi.org/10.24071/llt.v28i1.11295>

received 10 January 2025; accepted 16 April 2025

Abstract

Corpora usage has been increasingly employed in facilitating specialized English for students in English-for-specific-purposes classes. This paper focuses on the application of a corpus-based approach as a supplementary method to traditional ones in instructing legal English vocabulary in terms of its meaning and collocations to law students in Vietnam. To achieve this aim, a group of 29 students of the experimental group was assigned to corpus-based work along with a traditional teaching method, while 27 students in the control group experienced the sole later one in legal English classes during 15 weeks. The pedagogical experiment, as well as the assessment and scoring criteria, were discussed and presented at the beginning of the course. Participants' tests, survey questionnaires, and focus-group discussions were utilized as the data collection instruments to ensure triangulation and a multidimensional analysis of the research issue. Results demonstrated that the experimental group performed much better than the control group in terms of vocabulary understanding and interpretation, as well as collocations. Students of the treatment group also showed positive attitudes, acknowledging the potential benefits of the corpus-based approach for language-related enhancement in their future profession.

Keywords: corpus utility, experimental and control group, law students, legal English vocabulary

Introduction

The technology innovation has led to the enhancement of new educational approaches in academic settings. In the field of linguistics, corpus, closely connected with the digital industry, has been acknowledged for its provision of rich and reliable language data (Baker, 2013; Lai, 2015; Wu, 2016). As part of this trend, the integration of corpora, i.e, corpus-based learning, has been recently introduced to English language teaching, which is expected to benefit various stakeholders. According to Kadirbekova (2023), the initial function of the application of corpus aims to assist specialists in comprehending the language for communicative purposes i.e., to help linguists and English-for-specific-purposes (ESP) teachers compare the results taken from this corpus with textbooks and other materials, or English-as-a-foreign language (EFL) teachers to get used to common collocations, as well as enable translators to choose the appropriate words and collocations when

translating a certain phrase from one language into English. In other words, corpora are noteworthy to improve input language resources (Huang, 2014; Liu & Jiang, 2009), thereby leading to a significant pedagogical achievement (Butarbutar, 2021; Oktavianti et al., 2022; Vyatkina, 2018). For learners, corpus utility has earned a reward on vocabulary learning and teaching (Cresswell, 2007; Stevens, 1991). Similarly, ESP corpora enable learners to acquire potential specialized collocations, serving to facilitate their discipline learning (Abedi, 2014; Özdemir, 2014) as noted by Belcher (2006, p.142), the analysis of corpora “pushes students towards empirically based understanding of language used for specific purposes”. Simultaneously, the corpus provides essential information that may not be found in a dictionary or other printed materials. Furthermore, accessibility of collocation can be helpful to solve several issues related to the production of correct lexical units, and also to characterize the usage of vocabulary. As Dario (2014) reported that memorizing vocabulary, separating it from the context, is not recommended. Likewise, a meaningful and real-life situation is likely to push up the vocabulary acquisition process. Therefore, choosing appropriate teaching materials creating learners’ opportunity to be exposed to real-life use of English within a specific subject area is deemed as a vital factor contributing to the success of an ESP course (Lubina, 2023). Together with needs analysis and course objectives, teaching materials are the elements that draw attention from the learners most as the aforementioned frame the course content, the teaching methodology, as well as the evaluation and assessment criteria. Tenieshvili (2023) also stated that the combination of different language teaching approaches proves to be pivotal to ensure the achievement of expected results in an ESP course. In such cases, the application of a corpus-based approach proves to be of great necessity.

In the literature, several studies have been carried out to address what educational purposes a corpus is likely to offer (i.e, Boyko et al., 2022; Friginal, 2013; Lai, 2015; McCarthy, 2004; O’Keeffe et al., 2007; Oktavianti et al., 2022); nonetheless, a few language teachers seem to be attracted by the corpora resources (Römer, 2010). Yet, in Vietnam, the application of corpora is relatively new among teachers in pedagogical language teaching contexts, except for a few studies either on the use of corpora on English text-book evaluation (Huynh & Nguyen, 2020), in EFL context in general (Le & Do, 2021), or the possibility of adaptation of corpus linguistic for primary EFL education in particular (Le & Pham, 2023). Nearly few research has been conducted in ESP classrooms, especially in English for legal purposes settings. Therefore, this study aims to address students’ perception of the corpus as well as its efficiency in enhancing learners’ legal English vocabulary.

Legal English vocabulary, a type of specialized vocabulary in ESP

Among learning issues in ESP, specialized vocabulary known as “recognizably specific to a particular topic, field, or discipline” (Nation, 2001, p.64) is a special group of low frequency words limited to a certain area in which they appear with a fairly high frequency (Kadirbekova, 2023; Nation, 2001). It is noted that specialized vocabulary acquisition is pivotal in assisting students to gain academic literacy. Therefore, delving into the meaning of such terms is a requirement for deep understanding and usage. Waring and Nation (2004) emphasized that knowing the neighboring words contributes to reckoning the target term's meaning in a context. In other words, a knowledge of collocation, which is

“a sequence of two or more consecutive words, that has characteristics of a syntactic and semantic unit, and whose exact and unambiguous meaning cannot be derived directly from the meaning of connotation of its components” (Bennet, 2010, p. 610)

Legal English vocabulary is a subtype of specialized vocabulary in legal fields, which bears distinctive lexical and syntactic features, according to Veretina-Chiriac (2012). Lexical aspects of legal English vocabulary include the use of either old-fashioned, technical, borrowed terms or words used daily with unfamiliar meanings. As concerns syntactic features, nominalization and third person are preferable to verb usage in legal writing, which makes the text longer and becomes non-dynamic. Such distinctions resulted in vagueness and miscomprehension among learners, even legal majors.

Corpus as a methodological approach in language teaching

Bowker and Pearson (2002, p.9) stated that “a corpus can be described as a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria”. Sharing a similar viewpoint, Reppen (2010) specified the corpus as “a large and principled collection of naturally occurring texts (written and spoken) stored electronically” (p. 2).

From such definitions, four aspects, including authentic, large, electronic, and specific criteria, need to be considered. In the first sense, corpus collection includes the language used in real situations, either in written or spoken form (for example: letters, contracts, legal cases, or speech), allowing for the frequency distribution rather than limited uses. Secondly, concerning the size of a corpus, there is no specific rule, yet it is addressed as “large enough” for the purpose of systematic analysis of target linguistic patterns. Furthermore, due to technological advancement, the corpus has been stored electronically, thereby easily shared and explored among stakeholders. Lastly, the corpus is stored subject to a specific principle, which serves various purposes. Accordingly, corpus is understood as “the collections of written or spoken language-in-use which are stored, analyzed and utilized for a variety of purposes by researchers, teachers and learners” (Friginal et al., 2020, p.45)

As a methodological approach, corpora can be utilized in different contexts that make use of real language databases and computer-assisted learning (Al-Mahbashi et al., 2015). Corpora as data-driven learning (DDL) can be directly applicable to vocabulary learning (Oktavianti, 2015), writing instruction (Huang, 2014), as well as in ESP classrooms (Boulton, 2016). Friginal (2018) also “identifies corpus approach as emerging tools, online resources, classroom activities, and emphasizes the important contributions of the English teacher as a corpus-based materials developer and researcher to teachers and learners globally” (Friginal, 2018) as cited in Friginal et al. (2020, p.47).

In terms of its potential effects, corpus utility is said to be beneficial to learners’ input language, especially vocabulary, collocations, and phraseology. As Abduramanova (2020) noted that vocabulary is like the bricks of a high building; thus, understanding its thorough meaning is of great importance in developing other skills. Usually, professional terms are not fully provided with sufficient examples in a traditional bilingual dictionary in a variety of cases. In such a situation, relying on a pertinent corpus is necessary for word choice or collocations. Furthermore, corpus tools earned the reputation of promoting learners’ capacity to comprehend

and envisage that “a word is not limited to the word itself but also the words around it” (Firth, 1957). In other words, students are more likely to grasp the “meaning” or the “usefulness” of a word authentically, thereby enhancing the language acquisition process and improving language proficiency (Varley, 2008). According to Madalena (2001), real-life language usage is much more valuable than examples in the book or created by the language instructor, which is deemed not to cultivate authentic use. As a result, students find it challenging when dealing with real and complex language in real life. Additionally, Johns (1994) reported that teachers can design and develop teaching materials and activities to nurture learners’ hands-on learning experience and autonomy to discover authentic language patterns. Corpora are found to be an encouraging reference instrument for flexible use due to their easy access notwithstanding anytime or anywhere (Lee & Swales, 2006). With the assistance of computer storage, language databases can be researched through the tools of frequency indicators and concordances (Enayati & Pourhosein Gilakjani, 2020). Also, corpora were said to be practical for students to scrutinize the usage of language components such as verbs, articles, adverbials, collocates, and even sentence structures used in the specific disciplines through comparing their writing to other writers’ (Friginal, 2013; Lee & Swales, 2006). Similarly, Cortes (2011) created an opportunity for students to compile, analyze the linguistic patterns of selected texts of published research articles, allowing them to “inductively discover language structures”. Specifically, through comparing and contrasting their pieces of writing to professional ones, students find it much simpler to opt for appropriate word choices, enabling them to enhance their overall writing abilities (Friginal, 2013; Gilquin et al., 2007).

Similarly, in the case of ESP settings, the usefulness of corpora is attainable. Belcher (2006, p. 142) affirmed that corpora analysis “pushes students toward an empirically based understanding of language used for specific purposes”, simultaneously furnishing them with complementary information which may not be included in a dictionary or textbook. In particular, available ESP learning materials in the coursebooks do not cater for the actual learners’ needs due to their general and broad knowledge (Harwood, 2005). Accordingly, making use of the authentic texts could lead to the custom-made educational resources from the worthwhile approach, as Biber et al. (1998, p.3) specify:

“The corpus-based approach [...] provides a means of handling large amounts of language and keeping track of many contextual factors at the same time. It therefore has opened the way to a multitude of new investigations of language.”

However, as a methodological approach in language teaching (Biber et al., 2010; Lee & Swales, 2006), the corpus-based approach is considered a challenging task for both teachers and students. Although the incorporation of corpus-based learning resources in the classrooms has increasingly grown, several researchers (Flowerdew, 2015; Geluso & Yamaguchi, 2014; Meunier & Reppen, 2015) emphasized that instructors, regardless of being trained in corpus linguistics, are not ready to apply corpus-based activities due to their lack of confidence, assessment criteria, time constraints, as well as the obstacles in re-designing the suitable courses (Friginal, 2018).

Method

Participants

To figure out corpus application efficiency, the current research employed the qualitative and quantitative methods with the participation of fifty-six students assigned into two groups, namely the control group and experimental group. All participants provided consent and voluntarily participated in the study. To ensure confidentiality, all data collected was handled anonymously and securely stored. Participants were also fully informed of their right to withdraw from the study at any point without any consequences.

The control group, consisting of twenty-seven students, was provided with the traditional teaching method, while twenty-nine students in the experimental group experienced a corpus-based approach as an additional teaching technique besides the traditional one. All of the participants were sophomores, ranging from 19 to 21 years of age. Concerning gender, 69.6% (39) of the participants were female, while the rest were male, reflecting the nature of the social science of law areas. One thing noted is that all of the Vietnamese participants were law majors in a high-quality training program required to attend a legal English course in the 2022-2023 academic year as a core subject. Furthermore, for the high-quality program in law, learners were required to hold a B1 level of English according to CEFR or equivalent as a minimum compulsory requirement for admission. Specifically, when seeking their demographic information relating to their English proficiency level, it reveals that 92.8 % (52) of the surveyed students achieved IELTS band scores in the range of 5.5 to 6.5 or equivalent. Only four students got 7.0 IELTS or higher. Such statistics mean that the majority of the participants were homogenous regarding their general English proficiency.

Instruments and data collection procedure

To obtain the research objectives, three investigative tools were employed.

Participants' tests

Participants were required to do pre-tests and post-tests to measure their legal English vocabulary in terms of meaning, collocations. The tasks of the tests, including multiple choice gap-filling, matching collocation and meaning interpretation forms, were adapted from those in TOLES - a test of legal English skills - higher and advanced levels. The test contents concentrate on three areas of law delivered during the course, namely tort law, contract law, and company law. The mark distribution among task types is 3-3-4. The validity of the tests was scrutinized by two experienced legal English instructors.

Survey questionnaire

Survey questionnaire consisting of three sections: demographic information, fourteen statements concerning students' attitudes towards corpora benefits and difficulties, as well as three open-ended questions seeking learners' clarifications for their experience. Statements were adapted from the study of Oktavianti et al., (2022) with necessary changes. The validity of the questionnaire was reviewed by two independent lecturers of legal English, while Cronbach's alpha coefficient, reaching 0.804, was considered to be a reasonable level of reliability.

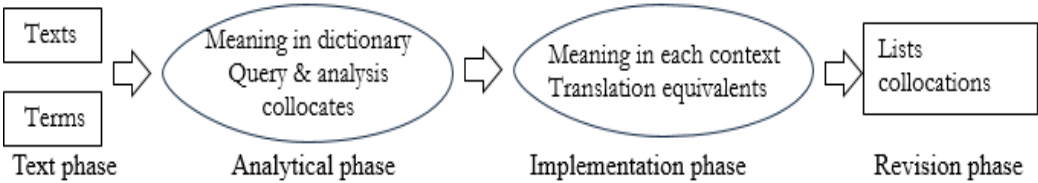
Focus-group discussion

Three questions concerning the familiarity, benefits, and difficulties of corpus linguistics in acquiring legal English vocabulary were formed to explore further details for students’ choices.

Procedure

The research was conducted in the second semester of the academic year 2022-2023. At the beginning of the semester, the experimental group was informed the utilization of corpus besides traditional textbooks as well as the steps needed to be gone through. A pre-test was given to the experimental and control group at the beginning of the semester to evaluate students’ knowledge of legal vocabulary. Then, both classes were instructed by traditional teaching methods with the explanation of new terms and practicing exercises in the legal English course book by the publisher Cambridge, covering tort law, contract law, and company law. The experimental class, besides the conventional teaching method, was provided with corpus-based activities. It is noted that the corpus used in this course is Law and corpus linguistics established by Brigham Young University linguistics professor Mark Davies. It is an academic sub-discipline containing large language databases for better grasp of words or phrases in legal texts in the form of statutes, constitutions, or court decisions. Until now, eight corpora have been developed and hosted on a website called BYU Law & Corpus Linguistics, the URL for which is <https://lawcorpus.byu.edu/>. Such corpora serving as “entry-point” into corpus linguistics have played a vital role in legal interpretation (Goldfarb, 2018).

Specifically, the experimental teaching approach was adapted from Boyko et al. (2022) with minor modifications represented as follows.



To be more specific, at the first stage, students were required to skim and scan the reading texts to identify the newly must-learnt terms/ vocabulary related to the certain law areas (See Table 1).

Table 1. A list of “must-learnt” terms

Reading text	Must-learnt terms
Reading 1: Contract law	Party
	Consideration
	Formation
	Counter-offer
	Essential terms
	Oral contract
	Breaching party
	Non-breaching party
	Injured party
	Breach of contract
	Damages

Reading text	Must-learned terms
	Specific performance
	Assignment

Then the primary meanings of the terms were figured out with the use of a dictionary. After that, students working in groups of three students to use corpora a query term is added in the search bar of a suggested corpus to search for the frequency of the term, collocates and concordances. (See Figure 1 as an example)

--	--	--	--	--
1 entire de novo, and that as the	damages	were entire, the judgment must	43.U.S.263	1844 43 U.S. 263
2 tort actions for the recovery of	damages	for personal injuries, id., at 27	526.U.S.687	-- 526 U.S. 687119 S.Ct. 16241...
3 chment of property, as well as	damages	. Although it alleged that the p	458.U.S.886.81-202	1982 458 U.S. 886
4 harmful conduct, while larger	damages	would "over-deter" by leading	517.U.S.559.94-896	-- 517 U.S. 559116 S.Ct. 15891...
5 e to do so, to pay all costs and	damages	caused to the owner or other	148.U.S.1.1.117	1893 148 U.S. 1
6 s Court's precedents. Punitive	damages	are "private fines levied by civi	492.U.S.257.88-556	1989 109 S.Ct. 2909
7 J.F. \$75,000 in compensatory	damages	and \$25,000 in punitive dama	491.U.S.524.87-329	1989 109 S.Ct. 2603
8 um up anew on the subject of	damages	, referring to the evidence, and	54.U.S.363	1851 14 L.Ed. 181
9 ges, 9th ed. § 332; Sutherland,	Damages	, 3d ed. § 337; Black, Tax Title	232.U.S.261.66	-- 58 L.Ed. 596
0 of the parties, or to assess the	damages	complained of in the bill. It wi	56.U.S.233	1853 56 U.S. 233
1), Siegart sought \$4 million in	damages	against Gilley, contending tha	500.U.S.226.90-96	1991 111 S.Ct. 1789
2 eral officer in a suit for money	damages	is without recourse in the eve	444.U.S.527.77-1546.78-303	1980 444 U.S. 527
3 breached free from liability for	damages	. We need not address the for	463.U.S.491.81-1966	1983 463 U.S. 491
4 o shall proceed to assess the	damages	or benefits, as the case may b	217.U.S.547.157	1910 217 U.S. 547
5 ry: "Should you award punitive	damages	, in fixing the amount, you mu	499.U.S.1.89-1279	1991 111 S.Ct. 1032
6 ministratrix, as aforesaid, prayed	damages	, setting forth that she was th	227.U.S.434.854	1913 227 U.S. 434
7 ed to absolute immunity from	damages	liability predicated on his offic	520.U.S.681.95-1853	1997 520 U.S. 681117 S.Ct. 16361...
8 tory provision for recovery of '	damages	' only, but would be at war wit	377.U.S.476.75	1964 12 L.Ed.2d 457
--	--	--	--	--
Guaranteed rights: "Historically,	damages	have been regarded as the ori	457.U.S.731.79-1738	1982 102 S.Ct. 2690
to any rule for the measure of	damages	in equity receiverships. In the	305.U.S.493.63	1939 305 U.S. 493
give the value of the horse in	damages	, being 8, and so they did. And	132.U.S.406	1889 33 L.Ed. 393
Clause as applying to punitive	damages	when those documents them	492.U.S.257.88-556	1989 109 S.Ct. 2909
y, defendant below, to recover	damages	for injuries done to two of the	125.U.S.260	1888 31 L.Ed. 731
ght within the state to recover	damages	for injury to the person. The st	177.U.S.172.174	1900 20 S.Ct. 617
el,' and that the claim for such	damages	shall constitute a lien upon su	189.U.S.158.98	1903 189 U.S. 158
a law of the place provides for	damages	'only punitive in nature.' 28 U.S	352.U.S.128.31	1956 352 U.S. 128
a shipowner the entirety of his	damages	, after adjusting for the plainti	511.U.S.202.92-1479	1994 511 U.S. 202114 S.Ct. 14611...
esioneers, as and for liquidated	damages	, the sum of \$25 for every day	133.U.S.473	1890 33 L.Ed. 674
Prosecutions and suits for treble	damages	authorized by Sec. 205(e), wit	321.U.S.414.374.375	1944 321 U.S. 414
argo could recover their whole	damages	from the New York. The New	204.U.S.220.134	1907 27 S.Ct. 246
his rulings upon the subject of	damages	and interest. Judgment affir	91.U.S.618	1875 91 U.S. 618
es the United States liable for	damages	on account of personal injurie	330.U.S.446.69.514	1947 67 S.Ct. 847
All the plaintiffs claimed treble	damages	for the monetary loss they ha	392.U.S.134.733	1968 20 L.Ed.2d 982
, delivering judgment, said: 'All	damages	directly arising from the imper	120.U.S.630	1887 120 U.S. 630
illy that " 'punitive or vindictive	damages	, or smart money, [are] not to b	456.U.S.556.80-1765	1982 456 U.S. 556
on a libel in rem, to be liable in	damages	to the owners of that vessel. T	157.U.S.386.46	1895 15 S.Ct. 657

Figure 1. Concordances of the term "damages"

At this stage, students chose from five to seven (even more) collocates to discuss and ascertain their meaning by analyzing the authentic context in the concordance (see Table 2 as an example). Also, students were encouraged to find Vietnamese equivalents, if possible, to clarify their natural usage.

Table 2. A list of “damages” collocations

Terms	Collocations	Interpretation
Damages -Meaning in dictionary: (monetary compensation)	1. larger damages	- bigger amount of money for compensation.
	2. punitive/ vindictive damages	- Exemplary damages are awarded separately from actual damages to punish the wrongdoer.
	3.compensatory damages	- designed to compensate the injured party for the actual losses.
	4. liquidated damages	- Estimated damages to be paid out in the case of breach of contract.
	5. prayer damages	- specific amount claimed as damages at the end of a complaint or petition
	6. assess the damages	- the process for determining the nature and extent of the loss, suffering, or harm.
	7. recover damages	- get an amount of money that a court requires someone to pay to make up for loss or harm.
	8. claim for damages	- demand the other party for compensation

In the final phase, students presented the target terms identified with their collocations, meaning interpretation and usage in an authentic context.

After the treatment period, both groups did a final post-test as a part of their final course grade, and simultaneously, to evaluate potential effects of the corpus-based approach on legal English vocabulary acquisition of students in the experiment group. Notably, there was no change in the content, tasks designed, total points (10 points), time allowance in the pre- and post- tests. Furthermore, a survey questionnaire was delivered to the experimental group to delve into their perspectives towards the corpora's utility in general. The survey questionnaire designed in Google forms was sent to the participants via emails, allowing them to respond in seven days.

Subsequently, based on students' notable answers from the survey, five of them were selected randomly for a follow-up group focus discussion seeking their voices for further explanations. Yet, it should be noted that the researcher chose the five interviewers purposefully in terms of their English certificate to ensure the involvement of three with an IELTS band of 5.5 to 6.5 and two with the higher band. A combination of random and purposive sampling methods was employed to examine whether differences in students' further explanations could be identified based on their English proficiency levels. The group-focused discussion was led by the researcher, who raised questions regarding specific aspects of the benefits and obstacles of the corpus utility.

Data analysis

Data was collected and treated with the assistance of SPSS software. Concerning the marks of the pre-test and post-test, paired sample t-tests and independent sample tests were utilized to address any notable discrepancies between the retention of legal English vocabulary and its collocations of the two groups. As regards learners' viewpoints, descriptive statistics were employed to address their responses relating to the benefits and difficulties when applying corpus-based activities using 5-point Likert scales. The result was interpreted as particularly *strongly disagree* (1.0–1.80), *disagree* (1.81–2.60), *neutral* (2.61–3.40), *agree* (3.41–4.20), and *strongly agree* (4.21–5.00).

Findings and Discussion

Findings

The efficacy of corpus-based activities

As stated, prior to the implementation period, a pre-test was assigned to all participants to address their knowledge of legal English vocabulary related to three areas of law: contract law, tort law, and company law, whose knowledge had not been delivered to them before.

Table 3. Paired samples t-test analysis for pre-test scores of the two groups

Group Statistics		N	Mean	Std. Deviation	Std. Error Mean
group	group				
Pre-test	Experimental group	29	3.470	.5432	.0690
	Control group	27	3.624	.5232	.1046

Table 4. Pre-test scores - Experimental group and control group

Paired Differences		Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
Mean	Std. Deviation		Lower	Upper			
Control group vs. Experimental group	-2.7760	.8095	.16	-3.11 -2.44	-17.1	24	.000

Table 4 illustrates little disparity in the mean scores of the experimental and control groups, which accounted for 3.470 and 3.624, respectively. The mark range within each group fluctuates from 2.310 to 6.458, with the majority of grades below average. This statistic is easy to explain due to a high number of unfamiliar legal English terms and vocabulary of the law branches. Furthermore, that the sig. Statistics in Levene's Test value varied slightly between the two groups specifying a small disparity. Based on such value, it can be drawn that students of both groups showed similar limited knowledge of legal English vocabulary concerning the three aforementioned law fields.

Table 5. Post-test scores - experimental and control groups

		Mean	N	Std. Dev.	Std. Error Mean
Post -test	Experimental group	8.593	29	.7430	.1059
	Control group	7.370	27	1.0746	.1037

After the treatment duration, all participants were required to take the post-test, whose results are shown in Table 5. As noticed, students in both experimental group and control group demonstrated the enhancement in their knowledge of legal English vocabulary with much higher grades of $M=8.593$; $SD=0.7430$ and $M=7.370$; $SD=1.0746$, respectively. In comparison with the control group, the experimental one performed much better, as shown by the high mean score. Specifically, the members of the experimental group showed their outstanding capacity in finishing the test with an overall mark of around band 8.5. Moreover, the Standard deviation reached 0.743, indicating that there was no big difference between members' marks. Undeniably, the sole application of the traditional teaching method by reading, lecturing, and finishing tasks in the course-book was effective, yet did not bring the positive results as the integration of new strategies, i.e., corpus-based activities.

In particular, concerning collocation and meaning interpretation in context task types, students provided with corpus-based activities proved their greater ability in comprehending terms meaning and their collocates.

Table 6. Results of each task type in post-test and pre-test

	Gap-filling (3pts)		Matching collocation (3pts)		Term Interpretation (4pts)	
	M	Std.dev	M	Std.dev	M	Std.dev
Control group	2.890	0.972	2.574	0.992	1.917	1.2580
Experimental group	2.911	0.945	2.895	0.691	2.799	0.6380

Table 6 displayed the significant difference in detailed results of each task in the post-test of the experimental group and the control group. Noticeable distinction can be witnessed in the mark of task 3 "Term Interpretation" with the value $M=2.799$, $SD=.6380$; $M=1.917$; $SD=1.2580$, accordingly. The statistic standard deviation value was 0.6380 in the experimental group, revealing that students' marks were quite close to each other. In contrast, within the control group, SD reached 1.258, which witnessed the bigger gap among members' performance. Understandably, the integration of a corpus-based approach as an additional teaching method has had a tremendous positive effect on learners' understanding and interpretation of legal terms.

Students' views on the corpus utility

Besides measuring the impacts of the corpus-based approach on students' acquisition of legal English vocabulary as well as its meaning and collocations, the experimental group's perceptions were also investigated.

Understandably, when being asked for their prior experience of using corpora, all of the students in the treatment group expressed unfamiliarity with corpora application in classes. In other words, corpus is a completely novel notion among students, therefore, the attitudes towards corpus utilization primarily focused on corpus-based activities generally applied in this legal English course.

Table 7. Benefits of using a corpus

Statement	N	Mean	Std.dev	Interpretation
1. I believe corpus is beneficial for legal English language learning.	29	4.32	1.072	Very high
2. Corpus-based instruction is exciting and enjoyable.	29	3.89	.725	High
3. Corpus-based activity makes me more responsible in learning legal English.	29	3.43	1.282	High
4. I find it easier to understand the meaning of legal English terms.	29	4.15	.769	High
5. I am able to acquire several collocations.	29	3.53	.748	High
6. Corpus provides me a huge number of legal English terms, thus helping me gain competence in writing	29	3.31	1.025	High
7. I find it easier to remember the new terms.	29	3.59	.800	High
8. Corpus can be used as a useful companion to the existing legal English textbooks	29	4.05	1.127	High
9. Corpus-based instruction is informative	29	3.67	.963	High

Table 7 showed that students assumed the benefits of corpus utility. According to the statistics, the majority of learners believed that corpus is beneficial for legal English language learning ($M=4.32$; $SD=1.072$). This finding is similar to Okatavianti et al.,’s research (2022), in which 96.8% of the participants agreed that the corpus benefits students, leaving few blank options. Moreover, the corpus lesson also made students feel excited and enjoyable ($M=3.89$; $SD=.725$), as well as become more responsible in legal English learning ($M=3.43$; $SD=1.282$). Such a positive opinion on corpus-based activity portrays that the corpus is considered to be a substantial resource in language learning and teaching. Accordingly, learners expected the corpus to result in good effects on learners’ acquisition of the meaning and collocation of legal English vocabulary, with the mean score of 3.89 and 3.43, respectively. Specifically, for further explanation, follow-up interviews confirmed:

In basic legal English courses, we were treated with a traditional teaching method, which makes use of knowledge in the academic course book. Coming to this semester, we had a chance to analyze and research reliable, authentic materials that are in the form of case law or decisions of the Supreme Court. I can acquire a number of collocations. (P.A)

Through group discussion, I can figure out the meaning of the word in each context, helping me understand and remember it more deeply. (L.N)

Following this trend, learners highly supported the informative feature of corpora, reaching a high rate of agreement ($M=3.67$, $SD=.963$). Realizing the strong points of corpus application, the majority of learners viewed the corpus as a useful companion to the existing textbooks ($M=4.05$; $SD=1.127$). Such results highlighted

the necessity of utilizing corpus as learning materials in addition to current coursebooks.

Legal English is so complicated that I find it difficult to understand the meaning of legal terms, as well as its uses in legal contexts. However, the corpus helps me to consult its meaning and collocations. (M.L)

Concerning challenges encountered during the application of corpora as additional activity, Table 8 highlights several aspects relating to the familiarity and complexity of corpora usage.

Table 8. Difficulties in using the corpus

Statement	N	Mean	Std.dev	Interpretation
10. I am not used to corpus application.	29	4.16	.745	High
11. Corpus use is too complex.	29	3.87	.954	High
12. Using corpus is time-consuming.	29	3.21	.875	Medium
13. I find it difficult to access the corpus due to the internet connection.	29	3.52	.774	High
14. I find it difficult to comprehend the meaning of the target terms due to unfamiliar vocabulary.	29	3.92	.784	High

Notwithstanding the positive perceptions of the corpus-based approach, students in the treatment group showed their strong agreement with not being used to it (M=4.16, SD=.745) as they admitted that,

It is new and novel. I have no prior background knowledge of corpus linguistics. Sometimes it is so complex that I cannot know how to use it. Therefore, it is better if we are trained to utilize corpora. (Q.P)

I need concentration and time to find out the meaning of the words. As you know, the corpus used in this course is legal matters included in legal cases or court decisions, thus, analyzing and understanding a term is not an easy task at all. (M.L)

Sharing a similar viewpoint, P.A added that, “unfamiliar, and specialized vocabulary surrounding the target term makes me find it even more difficult to comprehend its meaning”

Those above ideas are partly explanations for the high rate of agreement on the statement relating to corpus complication and stable internet connection requirement (M=3.87; SD=.954; M=3.52; SD=.774, respectively).

The corpus is online website, so sometimes corpus search is interrupted due to poor internet connection, even disconnection. This fact causes several disadvantages to us, whose mobile phones are not always connected to internet. (L.N)

Learners were neutral when mentioning the time consumption when using the corpus. As two students confessed,

Actually, it takes a huge amount of time when searching and finding the meaning of the term in each context. However, when fully understand the term, you realize that it is valuable and it deserves. It is not academic knowledge we learn from textbooks, yet it comes from real-life documents, which we need to get accessed for future job. Corpus should be encouraged in legal English teaching and learning. (H.M)

Difficulties when using corpus are undeniable. However, corpus benefits seem to outweigh disadvantages. If possible, a corpus course should be designed in a language program to get students acquainted its use. From this, I think our learning autonomy can be enhanced. (M.L)

Obviously, despite the fact that difficulties may be met in the process of using corpus linguistics, interviewed students exhibited their consensus and support for its potential application.

It can be argued that, although the mean rate is high, reflecting the agreement among the majority of the participants, a large standard deviation in students' responses is observed. Such statistics indicate that the participants held divergent views on the value and effectiveness of corpus tools in their legal English learning. This variability may illustrate the difference in individual learning preferences, familiarity with corpus use, English proficiency levels, or exposure to corpus-based instruction. The high dispersion implies that while several students find such tools highly beneficial for language learning and legal terminology acquisition, others may perceive them as less accessible or less relevant.

Discussion

Generally, the majority of the survey students addressed the corpus as a beneficial tool in learning legal English, especially in enhancing legal vocabulary. Also, the efficiency of the corpora approach has been shown in the language improvement of the treatment group when compared with the traditional teaching method. Such findings are in line with previous studies.

Concerning the effectiveness of the corpus-based approach on learners' specialized vocabulary, that is, legal English terms, the experimental group manifested more outstanding performance in the post-test than the control group. This result asserted positive points of the corpus, which is similar to that of Boyko et al. (2022) in an ESP technical context. Boyko et al. (2022) stated that ESP technical majors not only improved their vocabulary acquisition but also substantially enhanced their technical translation. Moreover, learners were likely to expand lexical skills, determine different meanings of words in various contexts, as well as identify word combinations, contributing to professional practice. Similarly, despite the deviation in teaching context, this paper affirmed the efficiency of corpus utility in collocations acquisition and retention, as in Kadirbekova's research (2023).

In terms of learners' views, the result parallels with the study of Oktavianti et al. (2022) in which almost all students were quite favorable to corpus use despite having no prior experience. Additionally, the students agreed that corpus-based activities are beneficial in language learning, which is similar to previous research (Kilimci, 2017; Lai, 2015; Yoon & Hirvela, 2004). The strength of the corpus included the enhancement of vocabulary and its collocations, as well as helping

learners gain competence in other skills such as writing or speaking. Such positive responses can be attributed to the practicality of authentic materials, ameliorating learning and teaching quality (Oktavianti et al., 2022; Römer, 2011). Also, students showed their satisfaction when being allowed to analyze language data, which cannot be carried out if using only a traditional coursebook. This result is congruent with the studies of Römer (2011) and Timmis (2015). Undoubtedly, the context of this current research is different from others, yet all the studies reported that the benefits outweighed the obstacles including unfamiliarity, complexity and other related issues (Hirata et al., 2013; Kilimci, 2017; Leńko-Szymańska, 2015; Oktavianti et al., 2022; Yoon & Hirvela, 2004). For such findings, students expressed their optimistic attitude towards the encouragement of corpora utility as teaching materials and resources in language education. Even students recommended the integration of a corpus and corpus linguistic course in a language education program so that both teachers and learners could maximize the efficacy of corpora utility. This view is strongly confirmed in previous studies in the literature (Leńko-Szymańska, 2015; Oktavianti et al., 2022). In other words, it is realized that linguistics should not be overlooked in English language teaching. Instead, there need for a combination of coursebook and corpus to enrich the linguistic knowledge of both teachers and learners (Conrad, 2000; McEnery & Xiao, 2013; Timmis, 2015), leading to the enhancement of the teaching and learning process.

Conclusion

The current research focuses on the efficacy of the corpus-based approach as an additional method and students' perspectives on its application in legal English vocabulary teaching. Overall, the findings elaborate that corpus utility opens up new avenues in language teaching, having positive impacts on learners' vocabulary and collocation acquisition as well as a deeper understanding of its meaning in different contexts. Such findings propose that the integration of corpus-based approaches into the teaching of legal English vocabulary and collocations carries several important pedagogical implications. Firstly, the use of corpora provides learners with authentic, context-rich language input, enabling them to observe how legal terms and collocations are used in real-life texts. This exposure enhances lexical awareness and supports the acquisition of discipline-specific vocabulary more effectively than isolated word lists or decontextualized examples. Secondly, corpus tools can empower learners to become more autonomous in their language learning. Through learning how to use concordancers and corpus search interfaces, students are more likely to investigate patterns of use, frequency, and collocational behavior of legal terms. This promotes data-driven learning (DDL), where students take an active role in discovering language rules and usage patterns, thereby deepening their understanding and retention. Furthermore, from a pedagogical standpoint, incorporating corpus-based tasks into legal English instruction encourages learner autonomy and critical thinking. Tasks that require students to analyze concordance lines, compare usage across different legal genres, or identify common collocations can facilitate both linguistic and cognitive development, especially in legal English contexts where students need to simultaneously process complex legal content and academic language. Finally, the application of corpus tools offers a valuable resource for curriculum design and materials development.

Instructors may design or select corpus data that aligns with the specific needs, levels, and professional interests of their students. Such targeted instruction enhances the relevance of learning materials and increases student motivation.

Nevertheless, corpus-based activities are troublesome to students due to their unfamiliarity with corpus activities and a huge number of new terms. Those aforementioned problematic issues need to be addressed to maximize the corpus efficiency in English language teaching in general and legal English teaching in particular.

References

- Abduramanova, D.V. (2020). Innovative techniques of teaching vocabulary in the second language classroom. *Science and Education*, 1(2), 34-38.
- Abedi, Z., & Mobaraki, M. (2014). The effect of grammatical collocation instruction on understanding ESP texts for undergraduate computer engineering students. *Language Teaching and Research*, 5(3), 631-641. <http://dx.doi.org/10.4304/jltr.5.3.631-641>
- Al-Mahbashi, A., Noor, N. M., & Amir, Z. (2015). The effect of data driven learning on receptive vocabulary knowledge of Yemeni University learners. *3L, Language, Linguistics, Literature: The Southeast Asian Journal of English language studies*, 21(3), 13-24.
- Baker, P. (2013). From gay language to normative discourse: A diachronic corpus analysis of Lavender Linguistics conference abstracts 1994–2012. *Journal of Language and Sexuality*, 2(2), 179–205. <https://doi.org/10.1075/jls.2.2.01bak>.
- Belcher, D. (2006). English for specific purposes: Teaching to perceived needs and imagined futures in worlds of work, study, and everyday life. *TESOL Quarterly*, 40(1), 133-156. <http://dx.doi.org/10.2307/40264514>.
- Bennet, G.R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan: University of Michigan Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. New York: Cambridge University Press.
- Biber, D., Reppen, R., & Friginal, E. (2010). Research in corpus linguistics. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd Ed., pp. 548-570). Oxford: Oxford University Press.
- Boyko, S. A., Koltsova, E. A., & Spiridonova, V. A. (2022). Application of a corpus-based approach in teaching English for specific purposes to master's degree students of engineering and technical majors. *Global Journal of Engineering Education*, 24(1), 40-45.
- Boulton, A. (2016). Integrating corpus tools and techniques in ESP courses. *ASp*, 69, 113–137. <https://doi.org/10.4000/asp.4826>.
- Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. New York: Routledge.
- Butarbutar, R. (2021). Learner's perception of task difficulties in technology-mediated task-based language teaching. *Englisia: Journal of Language, Education, and Humanities*, 9(1), 129-144. <http://dx.doi.org/10.22373/ej.v9i1.10079>.

- Cresswell, A. (2007). Getting to 'know' connectors? evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 267-288). Amsterdam: Rodopi.
- Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34(3), 548–560. <https://doi.org/10.2307/3587743>.
- Cortes, V. (2011). Genre analysis in the academic writing class: With or without corpora? *Quaderns de Filologia-Estudis Lingüístics*, 16, 65-80. <https://doi.org/10.7203/qfilologia.16.3932>.
- Dario, A. R. R. (2014). *TV series and movies in the teaching of legal English vocabulary* (Maestria en Didáctica del Inglés con Énfasis en Ambientes de Aprendizaje Autónomo). Retrieved from <https://intellectum.unisabana.edu.co/handle/10818/10811>.
- Enayati, F., & Pourhosein Gilakjani, A. (2020) The impact of computer assisted language learning (CALL) on improving intermediate EFL learners' vocabulary learning. *International Journal of Language Education*, 4(2), 96-112. <http://dx.doi.org/10.26858/ijole.v4i2.10560>.
- Firth, J. (1957). *Papers in linguistics*. Oxford, UK: Oxford University Press.
- Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15-36). Philadelphia, PA: John Benjamins North America.
- Friginal, E. (2013). Developing research report writing skills using corpora. *English for Specific Purposes*, 32(4), 208-220.
- Friginal, E. (2018). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities*. New York: Routledge.
- Friginal, E., Dye, P., & Nolen, M. (2020). Corpus-based approach in language teaching: Outcomes, observations, and teacher perspectives. *Bogazici University Journal of Education*, 37(1), 43-68.
- Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Learner attitudes and efficacy. *ReCALL*, 26(2), 225-242. <https://doi.org/10.1017/S0958344014000044>.
- Gilguin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319-335. <http://dx.doi.org/10.1016/j.jeap.2007.09.007>.
- Goldfarb, N. (2018). The BYU law corpora. *Historical linguistics, Language and the law*. Retrieved from <https://languageolog.ldc.upenn.edu/nll/?p=38155>.
- Harwood, N. (2005). "Nowhere has anyone attempted In this article I aim to do just that": A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatic*, 37(8), 1207-1231. <https://doi.org/10.1016/j.pragma.2005.01.012>.
- Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexicogrammatical patterns in L2 writing. *ReCALL*, 26(2), 163–183. <https://doi.org/10.1017/S0958344014000020>.

- Hirata, Y., Hirata, Y., & Thompson, P. (2013). Two different types of corpora: Japanese students' perceptions. In J. Lam, K. C. Li, S. K. S. Cheung, & F. L. Wang (Eds.), *Knowledge sharing through technology* (Vol. 407, pp. 1–15). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-45272-7_1
- Huynh, T. T. N., & Nguyen, V.L. (2020). Applying corpus linguistic to English textbook evaluation: A case in Vietnam. *Journal of foreign studies*, 26(6), 109-121. <https://doi.org/10.25073/2525-2445/vnufs.4632>.
- Le, T. T., & Pham, T. T. (2023). Adapting corpus linguistics for primary EFL education in Vietnam: Perspectives from linguistics scholars. *Journal of Innovation and Research in Primary Education*, 2(2), 55-66. <https://doi.org/10.56916/jirpe.v2i2.563>.
- Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, 93(1), 61-78. <https://doi.org/10.1111/j.1540-4781.2009.00828.x>.
- Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odline (Ed.), *Perspectives on pedagogical grammar* (pp. 27-45). Cambridge: Cambridge University Press.
- Kadirkbekova, D. (2023). Corpus-based approach in learning collocations in information-communication technologies (ICT). *Oriental Renaissance: Innovative, educational, natural and social sciences*, 3(2), 433-466.
- Kilimci, A. (2017). Learners' perspectives towards corpus use in vocabulary learning. *International Journal of Language Academy*, 5(6), 343–359. <http://dx.doi.org/10.18033/ijla.3765>.
- Lai, S. -L. (2015). EFL students' perceptions of corpus-tools as writing references. *Proceedings of the 2015 EUROCALL Conference*, 336–341. <https://doi.org/10.14705/rpnet.2015.000355>.
- Le, L.T., & Do, T. X. D. (2021). Application of corpus linguistics to foreign language teaching: Teaching Vietnamese vocabulary to foreigners through English-Vietnamese corpus. *Tap chi Khoa hoc Dai hoc Hue: Khoa hoc Xa hoi va Nhan van*, 130(6E), 67-81. <http://doi.org/10.26459/hueunijs.130i6E.6386>.
- Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56-75. <https://doi.org/10.1016/j.esp.2005.02.010>.
- Leńko-Szymańska, A. (2015). A teacher-training course on the use of corpora in language education: Perspectives of the students. In A. Turula, B. Mikołajewska, & D. Stanulewicz (Eds.), *Insights into technology enhanced language pedagogy* (pp. 135-150). Frankfurt: Peter Lang.
- Lubina, B. L. (2023). Applying content-based material in ESP courses: The example of Deborah Potter's handbook of independent journalism. *The Journal of teaching English for specific and academic purposes*, 11(2), 319-327. <https://doi.org/10.22190/JTESAP230322024L>.
- Madalena, S. R. (2001). *An investigation into how corpus analysis may be used in the second language classroom to solve some of the problems surrounding non-native speakers' understanding of seemingly synonymous words*. Japan: Reports- Research. <https://eric.ed.gov/?id=ED458795>.

- McCarthy, M. (2004). *Touchstone: From corpus to course book*. Cambridge: Cambridge University Press.
- McEnery, T., & Xiao, R. (2013). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 364–380). New York: Routledge. <https://doi.org/10.4324/9780203836507.ch22>.
- Meunier, F., & Reppen, R. (2015). Corpus vs. non-corpus informed pedagogical materials: Grammar as the focus. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 498–514). Cambridge: Cambridge University Press.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Oktavianti, I. N. (2015). *Data-driven learning in the classroom: The use of British national corpus in teaching vocabulary*. Paper presented at The 62nd TEFLIN International Conference. The 62nd TEFLIN International Conference, Udayana University.
- Oktavianti, I. N., Triyoga, A., & Prayogi, I., (2022). Corpus for language teaching: Students’ perceptions and difficulties. *Professional Journal of English Education*, 5(2), 441-455. <https://doi.org/10.22460/project.v5i2.p441-455>.
- Özdemir, N. Ö. (2014) Using corpus data to teach collocations in medical English. *Journal of Second Language Teaching and Research*, 3(1), 37-52.
- Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In M. C. Campoy, M. L. Gea-valor, & B. Belles-fortuno (Eds.), *Corpus-based approaches to English language teaching*. London: Continuum.
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205–225. <https://doi.org/10.1017/S0267190511000055>.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.
- Stevens, V. (1991). Concordance-based vocabulary exercises: A viable alternative to gap-fillers. *ELR Journal*, 4, 47-63.
- Tenieshvili, A. (2023). Application and combination of different foreign language teaching methods in ESP classroom. *The Journal of teaching English for specific and academic purposes*, 11(1), 203-213. <https://doi.org/10.22190/JTESAP221221013T>.
- Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. London: Routledge. <https://doi.org/10.4324/9781315715537>.
- Varley, S. (2008). I’ll just look that up in the concordance: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2), 133-152. <http://dx.doi.org/10.1080/09588220902778294>.
- Waring, R., & Nation, I. S. P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English Speaking World*, 4, 97-110.

- Wu, Y. J. (2016). The effects of utilizing corpus resources to correct collocation errors in L2 writing – Students’ performance, corpus use and perceptions. In S. Papadima-Sophocleous, L. Bradley, & S. Thouësny (Eds.), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 479-484). <https://doi.org/10.14705/rpnet.2016.eurocall2016.610>.
- Veretina-Chiriac, I. (2012). *Characteristics and features of legal English vocabulary*. *Revistă Științifică a Universității de Stat din Moldova*, 4(54), 103-107. <https://humanities.studiamsu.md/wp-content/uploads/2022/01/20.-p.103-107.pdf>.
- Vyatkina, N. (2018). Language corpora for L2 vocabulary learning: Data-driven learning across the curriculum. In P. Ecke & S. Rott (Eds.), *Understanding vocabulary learning and teaching: Implications for language program development* (pp. 121-145). Boston, MA: Cengage Learning.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13, 257–283. <https://doi.org/10.1016/j.jslw.2004.06.002>.