

## Doublets in Legal Discourse

### Data-Driven Insights for Enhancing the Phraseological Competence of EFL Law Students

<https://doi.org/10.3991/ijet.v15i20.13985>

Waheed M. A. Altohami

Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia  
Mansoura University, Mansoura, Egypt  
w.altohami@psau.edu.sa

**Abstract**—This paper explores legal doublets in the Corpus of US Supreme Court Opinions in order to identify the key linguistic features which govern their use. Based on the framework of data-driven learning approach (DDL), it assumes that getting EFL law students to use online available, user-friendly online corpus tools would help to enhance their phraseological competence. The data set (total frequency = 25.388 doublets) was restricted to doublets coordinated by 'and' as they proved to be highly frequent and linguistically variant. The paper has reached three findings. First, these doublets tend to follow one of six grammatical patterns: verb/verb, noun/noun, adjective/adjective, preposition/preposition, determiner/determiners, and conjunction/conjunction. Second, legal writings tend to put certain morphosyntactic constraints on the use of doublets. Finally, a full-fledged analysis of the phraseological profile of these doublets includes the identification of their grammatical patterns, their various lemmas, the nuances of meaning between their component lexemes, and the remarkable collocates used with them.

**Keywords**—EFL law students; data-driven learning; legal discourse; phraseological competence; corpus tools.

## 1 Introduction

Legal texts are important documents that involve articles, conditions, terms and judgments which organize the relationship among individuals, companies, corporations and states. They state rights, duties, procedures and commitments. Hence, any sort of ambiguity, confusion or contradiction is not tolerated. That is why legal practitioners are requested to be keen on using clear structures and expressions that are rendered empirically reliable and meaningful. One of the distinctive linguistic features of legalese or legal discourse is the use of strings of usually synonymous and near synonymous lexical items known as 'doublets' which act as independent phraseological units within written discourse. The grammatical patterns of these doublets are assumed to be problematic for EFL law students. Instead of having students as merely

users of these doublets, this paper uses a data-driven learning (DDL) approach, being a pattern-based approach to grammar and vocabulary, to have law students undertaking discovery tasks of available corpora to learn about the accurate use of doublets in their writings. It aims at exploring the use of doublets in authentic texts represented by the Corpus of US Supreme Court Opinions (COSCO-US) in order to identify these doublets, their key grammatical patterns, and the context-based constraints on their forms (lemmas), collocates, and syntactic variations which are assumed to form the phraseological profile of the target doublets.

Therefore, the present study, by means of quantitative and qualitative methods, seeks to answer three questions:

- a) What are doublets in COSCO-US?
- b) What are the key grammatical patterns in which these doublets are used?
- c) How students could analyze the phraseological profile of these doublets?

Question (a) would offer a list of the frequency of doublets coordinated by the coordinator 'and' which proved to be highly frequent and more linguistically variant in the corpus. Question (b) is limited to the exploration of the various grammatical patterns of the identified doublets. Based on a close reading of the concordance lines in which some representative examples of the identified doublets are used. Question (c) would provide the key linguistic features remarkable of the use of doublets. By answering these three questions, EFL law students would be able to enhance their phraseological competence with regard to the use of legal doublets.

The significance of this study is that it would provide empirical evidence on the pivotal role played by corpora in enhancing students' learning practices in general and in improving their phraseological competence in particular. It is geared towards providing the tenets of the use of available data to offer a holistic view of the method these doublets could be implemented in the design of writing courses for EFL law students based on authentic language use. Applying DDL approach to the investigation of doublets in legal written discourse would empower both students and teachers to move from the superficial textual features of the legal register to understanding its exclusive structural features and stylistic conventions.

The rest of this paper is organized as follows. Section 2 reviews literature related to the use of DDL approach for enhancing learners' phraseological competence in legal discourse. Section 3 discusses the theoretical background of the present study, including phraseology and phraseological units, DDL approach, and legal discourse. Section 4 presents the study methodology. Section 5 summarizes key findings.

## **2 Literature Review**

Many studies emphasized the efficacy of integrating corpus-based and corpus-driven methods into common learning and teaching practices to mediate insights to the development of learners' autonomy [1-8]. These studies stressed the role played by corpora in improving learning and teaching practices. Corpora are claimed to help teachers to improve teaching materials so as to be remarkably authentic and pedagog-

ically relevant [9]. Also, teachers could develop so many activities and exercises that they monitor, or that are entirely learner centered. Other studies were much concerned with what corpus-based and corpus-driven methods could offer to enhance various dimensions of language proficiency as well as a variety of competences in ESL and EFL contexts [10-13].

Though there is a surge of publications in the area of applying corpus-based methods to enhancing learners' phraseological competence by means of DDL approach [14-17], most of the publications which explored the use of doublets in legal discourse were much concerned with translating such doublets from and into different languages [18-20]. To my best knowledge, the only study that addressed the use of doublets by EFL law students is conducted by Ref. [21], and it was quantitatively limited to the phraseological units based on the colligational patterns of prepositions among law undergraduates in Malaysia. The basic research design included pre- and post-tests administered to a control group and an experimental group. Findings showed that students in the experimental group generally showed notable advance in the use of collocational patterns of prepositions.

It could be then subsumed then that no study targeted doublets in legal discourse in its entirety. Related literature did not provide a clear, empirical method of analyzing the linguistic aspects characteristic of these doublets. Also, it did not offer any insights of how the linguistic aspects of doublets could be manipulated to enhance the phraseological competence of EFL students of law. Hopefully, the present study would fill in this gap.

### **3 Theoretical Framework**

This section discusses the key theoretical background underlying the present study. It starts with an overview of legal English. Then, it explains the scope of phraseology and the linguistic profile of doublets. Finally, it shows how corpus-based and corpus-driven methods are used to enhance teaching and learning practices through DDL approach.

#### **3.1 Legal English**

All legal practices depend on the tools of linguistic analysis. In the last fifty years, new linguistic disciplines such as legal linguistics and forensic linguistics emerged as challenging research areas in applied linguistics. These research areas are much concerned with the study of spoken and written legal discourse with special reference to social justice issues, legal translation, interpreting, and legal English teaching and learning. The language used in legal contexts (also known 'legalese') is claimed to form a language variety of its own. A close investigation of such variety shows the thumbprints of different versions of English including Latin, Old French and Old English [18].

As stated in [22], legalese is remarkably obscure, circumlocutory, archaic and torturous. That is, a word in a legal document may have different senses and legal con-

sequences [23]. As stated in [24], legalese could be "innovative, casual and purposefully vague" (p. 24). It has its own lexicon, some of which is mainly borrowed from Latin and French and is manipulated to perform specific textual functions. Relatedly, legal lexicon is claimed to be neither completely archaic nor innovative. Unlike traditional language, legalese is stylistically verbose and inherently involves a liberal use of jargon which cannot be easily grasped by common readers. Even, there might be a tendency to use expressions with flexible meanings.

With regard to the salient linguistic features of legal English, Ref. [25] and Ref. [24] mention five syntactic features of legalese: sentence length, nominal character, complex prepositional phrases, and binomial and multinomial expressions. In general, sentences tend to be markedly lengthy and complex, with more embeddings, when compared to sentences in other registers. Sometimes, legal discourse tends to involve wordy phraseological units and ponderous phrases which could be substituted with units involving fewer words. The syntax of legal discourse is also claimed to be non-standard and is generally impersonalized. Nouns are sometimes over-repeated to avoid ambiguous reference and sexist deictic expressions. Since legal practitioners tend to offer regulations by prohibition, negative forms also abound.

The next subsection discusses the nature of 'doublets' being one of the characteristic lexicon-grammatical features of legal discourse.

### **3.2 Phraseological units, doublets and phraseological competence**

A phraseological unit (also known as 'phraseme' and 'phraseologism') is a multi-word lexical unit such as idioms, phrasal verbs and collocations which tend to occur in particular formal patterns. As stated in [26], a phraseological unit is a lexicalized, reproducible billexemic or polylexemic word group in common use, which is syntactically relative, semantically stable, and stylistically neutral. The area which is concerned with exploring the different categories of phraseological units is known as 'phraseology'. The introduction of corpus tools in linguistics caused major developments in phraseological studies, adding new categories of phraseological units such as lexical bundles, collocations, and multi-word lexical units [27]. Based on the motivation of the unit, Ref. [28] classifies phraseological units into:

- a) Phraseological fusions whose meaning cannot be deduced from the meaning of their component parts (e.g. 'red tape')
- b) Phraseological units whose meaning can be deduced from the meaning of their component parts (e.g. 'to show one's teeth')
- c) Phraseological collocations which contain two parts, one is metaphorical and the other is direct (e.g. 'to meet the demand')

Though phraseological units abound in all types of discourse, legal discourse is claimed to have specific sets of phraseological units. A particular type of these phraseological units is known as 'doublets' (also known as 'binomials', 'couplets', and 'twin formulae'). A doublet simply refers to a string of two words which are lexically paralleled, syntactically coordinated, semantically related and structurally fixed. Doublets are perceived as a 'style marker in law language' [30].

As stated in [23], during the Anglo-Saxon era, legal doublets were tended to be alliterative, e.g. 'part and parcel'. Furthermore, the introduction of French words into English after the Roman Conquest is a direct reason for having legal pairs such as 'null and void' and 'cease and desist' which were translated from French in the late medieval era. Then, these doublets were codified and used in the mainstream language as fixed expressions. Hence, they are not simply lexical collocates which are "usage-determined or preferred syntagmatic relations between two lexemes in a specific syntactic pattern. Semantically autonomous, the base of a collocation is selected first by language user for its independent meaning. The second element is selected by and semantically dependent on the base" [27].

As stated in [31], English doublets, based on their order, could be classified into formulaic (fixed) and non-formulaic (reversible). Also, based on the lexical semantic link of the components of the doublet, Ref. [32] classifies doublets into five categories:

- a) Near synonymous (e.g. 'null and void')
- b) Complementary (e.g. 'assault and battery')
- c) Opposites (e.g. 'assets and liabilities')
- d) Subdivisions (e.g. 'months and years')
- e) Consequence (e.g. 'shoot and kill')

Similarly, Ref. [31] classifies doublets into three categories:

- a) Synonymous, e.g. 'last will and testament'
- b) Antonymous, e.g. 'in person or by proxy'
- c) Complementary, e.g. 'shoot and kill'.

Phraseological competence is an integrative part of the general linguistic competence. It refers to the ability of a language user to understand and use different phraseological units in communication. Being exposed to authentic language use, such competence is largely claimed to be effectively improved. One reliable approach which augments this claim is the use of large corpora to learn about phraseology. A practical realization of using corpus-based and corpus-driven methods in language learning and teaching is the approach of data-driven learning (DDL).

### **3.3 Data-driven learning approach**

Data-driven learning approach (DDL) emerged as a practical step to bridge the gap between the process and the product of learning. As stated in [33], in DDL "language learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data" (p. 2). It involves the direct or indirect application of corpus tools by means of which learners would be able to explore massive quantities of data where they learn inductively about language by calculating the frequencies of the target phenomenon, its distribution in concordance lines, and its patterns in different registers and across multilingual corpora. As stated in [34], having the opportunity to interact with corpora, students would be able to learn about the language-oriented

inquires that could be raised and addressed based on corpus investigation. DDL approach was claimed to help teachers to identify "patterns of specialized phraseology, which are barely mentioned in the general bilingual and monolingual dictionaries used by their students" [35].

DDL approach was reported to be practically reliable in learning about vocabulary depth and usage, collocations [4], and phraseology [36] as well as in improving skills of error correction [37], retention and call [38], and writing skills [37, 39-43]. The outcome of corpora investigation is then used for designing exercises around the target items based on the concordance lines. As stated in [33], DDL methods get students to "explore regularities of patterning in the target language", and it is further concerned with "the development of activities and exercises based on concordance output".

One of the critiques leveled at the use of DDL approach is the issue of focusing on accuracy rather than fluency which might affect learner's communicative competence [44]. Also, it is claimed that DDL techniques do not suit low-level students, and that students become less motivated. One final caveat of DDL is that classrooms with poor internet connection would face many technical problems. To overcome some of these difficulties, teachers are recommended to blend direct and indirect DDL methods as follows. Teachers are encouraged to base their discussions on free available corpora which do not involve complex technical issues. Also, teachers could prepare handouts to help their students to pursue their exploration and reading into corpora [1]. In some cases, teachers can edit output concordance lines for readability purposes. Furthermore, teachers can open discussion over the concordance lines involving the target item, and intervene in case students need help [45]. If students do not have access to computers in the classroom, teachers can print some of the concordancing output and bring them to the classroom [1]. It can be concluded then that DDL approach could be reconfigured in a way that help to maximize the use of corpus tools.

## **4 Methodology**

The study is a descriptive research based on a corpus-based/corpus-driven inductive approach and it is mainly intended to offer an empirical method for implementing a set of doublets in writing course for EFL students of law to enhance their phraseological competence. In what follows, I offer a description of the data. Then, I proceed to show how data analysis is conducted.

### **4.1 Data**

The data I make use of in this this paper is based on the [Corpus of US Supreme Court Opinions](https://www.english-corpora.org/scotus/) (COSCO-US), available at <https://www.english-corpora.org/scotus/>. In general, this corpus is a specialized corpus that involves 32.000 Supreme Court decisions covering the period from 1790 to the present. It contains around 130 million words. Though limited to the American legislative language, the target corpus in general is an asset that practically suits the purposes of this paper as it allows to concord-

ance any item, sorts concordance lines, and gives access to the full context in which the searched item is used. All the features of the corpus are offered with clear instructions and accordingly would be suitable for EFL students. An initial reading into the corpus showed that doublets coordinated with 'and' proved to be highly frequent and lexically variant. Hence, the study would be limited to doublets coordinated with 'and'. It could be subsumed that most of the claims and expected findings this paper would reach are to be seen as confined to the limits of this corpus.

## **4.2 Procedure of analysis**

The procedure followed in the analysis of doublets involves three major steps. Firstly, COSCO-US is accessed and analyzed to identify all doublets and their frequency based on the collocates display tool available at the corpus website as it helps to display the concordance lines in which doublets are used. Secondly, all extracted doublets are examined with reference to their total frequency and various forms. Then, the target doublets are classified based on their grammatical patterns. Thirdly, I selected the most frequent doublet in each grammatical pattern to be meticulously investigated to highlight the salient linguistic features characteristic of the use of doublets in legal writings. These linguistic features would form the phraseological profile of each doublet. Based on the findings of data analysis, a set of pedagogical implications are offered to maximize the value of the adopted DDL approach in enhancing the phraseological competence of EFL law students.

# **5 Analysis**

## **5.1 Quantitative analysis**

Quantitative analysis in this section is concerned with calculating the total frequency of the doublets in COSCO-US, the raw frequency of each doublet, the key grammatical patterns of the extracted doublets, and the total frequency of the doublets in each grammatical pattern. This step would help to identify the most frequent and hence the most significant doublets in the corpus to decide on the target items to be implemented in the recommended activities designed for writing courses for EFL law students.

Using the collocates display tool, table (1) shows the frequency of the extracted doublets in a descending order. It only displays the stem of each doublet with no reference to its various morphological forms (lemmas). By default, frequency of nouns covers singular and plural forms; for verbs, it covers the base form, the second and third conjugations as well as gerund; for adjectives, it covers the base form as well as the comparative and superlative forms.

**Table 1.** Frequency of doublets in COSCO-US

No.	Doublet	Freq.	No.	Doublet	Freq.
1	Term and condition	2412	43	True and correct	97
2	Full faith and credit	1985	44	Do and perform	96
3	Null and void	1727	45	Able and willing	76
4	Any and all	1610	46	Give and grant	72
5	Each and every	1392	47	Save and except	65
6	Fact and circumstance	1276	48	Control and regulation	62
7	Heirs and assigns	909	49	Release and discharge	61
8	Over and above	766	50	Keep and perform	58
9	Will and testament	741	51	Keep and maintain	56
10	Cease and desist	729	52	Lien and encumbrance	53
11	Due and payable	685	53	Break and enter	48
12	Power and authority	654	54	Kind and nature	46
13	Final and conclusive	628	55	Care and attention	42
14	Make and provide	625	56	Deem and consider	33
15	Cost and expense	543	57	Depose and say	29
16	Full and complete	535	58	Taxes and charges	26
17	Goods and chattels	533	59	Few and far	26
18	Have and receive	495	60	Mind and memory	22
19	Ways and means	453	61	Hue and cry	21
20	Full force and effect	442	62	Cease and terminate	18
21	Sole and exclusive	428	63	Known and described as	17
22	Authorize and direct	407	64	Lewd and lascivious	16
23	By and with	385	65	Alter and change	15
24	If and when	367	66	Bind and obligate	14
25	Illegal and void	357	67	Furnish and supply	13
26	Part and parcel	352	68	Indemnify and hold harmless	12
27	By and between	325	69	Finish and complete	10
28	Have and hold	305	70	Construe and interpret	9
29	Successor and assign	303	71	Perform and discharge	9
30	Due and owing	272	72	Govern and construe	8
31	Object and purpose	250	73	Cases and circumstances	7
32	Free and clear	178	74	Demise and lease	7
33	Covenant and agree	153	75	Touch and concern	7
34	Make and enter into	149	76	Sound mind and memory	6
35	Law and order	146	77	Cancel and set aside	4
36	Order and direct	127	78	Agree and covenant	3
37	Legal and valid	122	79	Appropriate and proper	3
38	Custom and usage	113	80	Acknowledge and confess	2
39	Fit and proper	104	81	Uphold and support	2
40	Aid and abet	103	82	All and sundry	1
41	Sale and transfer	103	83	Course and method	1
42	Pay and discharge	101			

As demonstrated in table (1), the identified doublets totaled 83. A close investigation of these doublets shows that they could be syntactically structured into six basic

grammatical patterns: 'verb (and) verb', 'noun (and) noun', 'adjective (and) adjective', 'preposition (and) preposition', 'determiner (and) determiners', and 'conjunction (and) conjunction'. See table (2).

**Table 2.** Number of doublets in all six grammatical patterns in COSCO-US

No.	Grammatical Pattern	Types	Tokens
1	Verb + and + verb	33	3780
2	Noun + and + noun	26	11426
3	Adjective + and + adjective	17	5272
4	Preposition + and + preposition	4	1541
5	Determiner + and + determiner	2	3002
6	Conjunction + and + conjunction	1	367
	Total	83	25,388

As demonstrated in table (2), the most frequent grammatical patterns are 'verb (and) verb', 'noun (and) noun' and 'adjective (and) adjective' respectively. It could be claimed then that legal discourse tends to involve more repetition and less variations. That is, it is not semantically rich. Verbs, nouns and adjectives are content words which carry the meaning of a legal text. The high frequency of verbs could be attributed to the fact that legal texts tend to act as action plans to keep the rights of different parties. They cover a wide variety of speech acts including assertives which represent a state of affairs (e.g. Under these grants, they have and hold the rightful and exclusive possession of the premises), directives which get different parties to do something (e.g. The Commission ordered the defendants to cease and desist from charging or collecting any rate), accreditives which are manipulated to transfer permission or authorization from one agent to another (e.g. the Board shall forthwith authorize and direct the Secretaries or any of them to eliminate such excessive profits), commissives which commit to a future course of action on the part of different parties (e.g. The Contractor shall indemnify and hold harmless the Owner) and expressives which express the attitudes of different parties about a state of affairs (e.g. the shares of the capital stock of the company should be deemed and considered personal estate).

Additionally, the high frequency of nouns could be attributed to the fact that legal texts are highly informative. In total, the extracted doublets based on nouns signify the key topics and issues addressed in legal texts, such as terms (e.g. the legislature, in proposing the terms and conditions of the act, use the word "banks" with reference to the consent), circumstances (e.g. We also decline to conjecture broadly on the significance of possession in cases and circumstances not before this Court), cost (e.g. the commissioners would construct the same at the cost and expense of the company), beneficiaries (e.g. they have allotted to the said plaintiffs their heirs and assigns forever), regulations (e.g. It imposes a structure for the control and regulation of abortions in Missouri during all stages of pregnancy), etc.

All of the adjectives used in the identified doublets are meant to state specific organizational situations including validity (e.g. That there was no legal and valid assessment for taxes of the land), payability (e.g. the whole debt became due and paya-

ble as a personal obligation of Nelson), conclusiveness (e.g. their decision in the premises is final and conclusive as to the dutiable value of the importation), exclusiveness (e.g. the statutes constitute him the sole and exclusive judge of the existence of these facts) and completeness of terms and conditions (e.g. No objection is made that this seizure was not full and complete). Furthermore, doublets that follow the 'preposition-preposition' pattern are used to name specific limits and numerical values (e.g. The expenses of the defendant over and above taxed costs are usually as great as those of plaintiff), agents (e.g. And it is mutually agreed by and between the parties hereto that during the remainder of the term), instruments (e.g. The judges of these are appointed by the President, by and with the advice and consent of the Senate) and limitations (e.g. The complainant and appellant is excused from printing the record in this suit, save and except such portions).

In what follows, I would show how EFL law students could master the use of these doublets through a systemic analysis based on DDL methods.

## 5.2 Qualitative analysis

Since this section is qualitatively oriented, I would only analyze some representative examples of the doublets. Such examples include the top frequent doublet in each grammatical pattern. These doublets are 'cease and desist', 'term and condition', 'null and void', 'over and above', 'any and all', and 'if and when'. For each of these doublets, I would offer a holistic view including its grammatical pattern, different forms (lemmas), the semantic relation between the component lexical items, and finally the most frequent collocates that are used with such doublets.

**Cease and desist:** The phraseological unit 'cease and desist' is the most frequent doublet in the corpus which follows the grammatical pattern 'verb (and) verb' (total frequency= 729). The corpus shows two lemmas for this unit represented by the base form 'cease and desist' (frequency= 725) and the past form 'ceased and desisted' (frequency= 4). The simple present form 'ceases and desists' and the progressive form 'ceasing and desisting' are never used. See the following example concordance lines.

: make an order requiring the carrier to **cease and desist** from construction and operation  
Western Meat Company, shall forthwith **cease and desist** from violating the provisions of section 5  
a notice to said common carrier to **cease and desist** from such violation, and to make  
at the union headquarters that the union has **ceased and desisted** in the manner aforesaid  
. even though they are 'unfair,' **from any cease and desist** order by the Commission

As shown in the previous concordance lines, the doublet 'cease and desist' is only used with the preposition 'from' to refer to enforcing somebody or an institution to stop engaging in a particular activity (cease), and not to do it again (desist). Hence, the two lexical items are near synonyms. Whenever such doublet is used with the preposition 'from', it must be followed by a gerund. Also, it used as a modifier to the lexical item 'order' in reference to an order issued by a court stating that a body or someone stops doing something entirely. Finally, with reference to the collocates used

with the node 'cease and desist', the corpus displays 59 significant collocates with frequency more than 3 occurrences (MI>3). The rationale behind limiting the ratio of mutual information (MI) to 3 is to avoid weak collocates. The top-ten collocates used with this doublet are shown in table (3).

**Table 3.** Top-ten collocates used with 'cease and desist'

No.	Collocate	Freq.	MI	No.	Collocate	Freq.	MI
1	From	404	4.20	6	Orders	38	5.35
2	Order	155	4.86	7	Practices	30	6.05
3	Labor	49	4.94	8	Violation	30	4.21
4	Unfair	48	7.11	9	Required	30	3.35
5	Respondent	39	3.64	10	Ordered	28	4.66

**Term and condition:** The phraseological unit 'term(s) and condition(s)' is the most frequent doublet in the corpus which follows the grammatical pattern 'noun (and) noun' (total frequency= 1412). The corpus shows two lemmas for this unit represented by the singular form 'term and condition' (frequency= 12) and the plural form 'terms and conditions' (frequency= 1400). See the following concordance lines.

. The license had a price, **term and condition** of sale clause, governed by Southern's prices,  
; that the grantee must give bond to perform' every **term and condition** of such franchise;  
under the **terms and conditions** stipulated and contracted for by the said Maison Rouge.  
. It is consequently a disaffirmance of the **terms and conditions** prescribed by the assignors  
on the part of the plaintiff of the **terms and conditions** upon which the loss is declared

Based on the concordance lines in which the doublet 'terms and conditions' is used, it is noticed that this doublet is followed by three major prepositions: 'of', 'upon', and 'on' in reference to the type of target legal document (e.g. an agreement, a lease, a franchise, a bargain, etc.), a specific article in a document (e.g. payment, termination, etc.), or a person to which these terms and conditions apply. Moreover, this doublet is often post-modified by specific lexical items such as 'prescribed', 'stipulated', 'contracted for', 'expressed', and 'contained in'. The lexical items 'terms' and 'conditions' are semantically related as they are partially synonymous as they refer to a set of arrangements, regulations and standards organizing an agreement between two entities. As far as the collocates used with 'terms and conditions', the corpus displays 84 significant collocates which occurred more than 3 times. The top-ten collocates used with this doublet are shown in table (4).

**Table 4.** Top-ten collocates used with 'terms and conditions'

No.	Collocate	Freq.	MI	No.	Collocate	Freq.	MI
1	Upon	505	3.54	6	Specified	60	4.83
2	Employment	376	6.86	7	Following	56	3.39
3	Same	197	3.05	8	Contained	40	3.76
4	Prescribe	69	6.42	9	Forth	32	3.32
5	Prescribed	66	4.81	10	Applicable	31	3.10

**Null and void:** The phraseological unit 'null and void' is the most frequent doublet in the corpus which follows the grammatical pattern 'adjective (and) adjective' (total frequency= 1727). The corpus shows only one form of the doublet as it consists of two absolute adjectives which can never be inflected. See the following examples of concordance lines.

all grants under it were declared to be null and void, by the second act.  
 . And the prayer of the bill is that the assignment may be decreed null and void,  
 the premises was constitutionally and legally impaired, and rendered null and void.  
 , pleaded in bar of the present suit, were absolutely null and void, the record not showing that  
 shall be deemed and adjudged to be utterly null and void, to all intents and purposes whatsoever.

A close reading of the concordance lines shows that the two lexical items 'null' and 'void' are used synonymously to describe a legal document which has no legal effect whatsoever. Most of the concordance lines indicate that it is used in reference to future events. The doublet is usually used with linking verbs including 'be' and 'become' to express a consequence of violating the terms of an agreement. Lexically, it is used with various verbs including 'declare', 'deem', 'adjudge', 'consider', 'prove' and 'render'. Relatedly, the corpus displays 58 collocates with frequency of 3 occurrences (MI>3). The top-ten collocates used with this doublet are shown in table (5).

**Table 5.** Top-ten collocates used with 'null and void'

No.	Collocate	Freq.	MI	No.	Collocate	Freq.	MI
1	Shall	332	4.19	6	Utterly	59	8.26
2	Declared	219	6.26	7	Become	49	4.39
3	Absolutely	124	7.76	8	Unless	42	3.36
4	Therefore	66	3.06	9	Deemed	34	4.24
5	Policy	62	3.67	10	Declaring	30	5.97

**Other frequent doublets:** The last representative examples of the doublets which are highly frequent in the corpus include 'over and above', 'any and all', and 'if and when'. These three doublets are representative of the grammatical patterns 'preposition (and) preposition', 'determiner (and) determiner', and conjunction (and) conjunction' respectively. I grouped these doublets together because they, unlike other doublets, kept the same form throughout the corpus as they are made of grammatical uninflected lexemes which are usually followed by nouns. Also, the two lexical items included

in these three doublets are near synonyms. The first doublet 'over and above' (frequency= 766) is always related to numerical values such as expenses, costs, weight, sums, rates, etc., or abstract values such as services, duties, loss, damage, etc. The second doublet 'any and all' (frequency= 1610) is used to group similar or different items together as one unit, and therefore it is mostly followed by a plural noun. However, if this doublet is followed by a singular noun, it is used in reference to all inherent aspects of the meaning that this singular noun refers to. The third doublet 'if and when' (frequency= 367) is used to express a condition. The repetition of the two conjunctions is aimed to emphasize the case.

Consider the following examples of concordance lines in which these three doublets are used.

. It also gives fifteen percent damages **over and above** these allowances.  
a very considerable amount **over and above** the debt claimed by the plaintiff in the said attachment;  
-And provided further, that **any and all** lands heretofore reserved to the United States  
may provide for issuing bonds for the purpose of funding **any and all** indebtedness now existing  
It is our intention to submit that question to the Courts, **if and when** the occasion arises.  
The special taxes or assessments here in question, **if and when** collected, will simply accrue to the special fund

The corpus displays no significant collocates for the three doublets discussed above since all of the collocates have a ratio of mutual information below 3.

Having analyzed the whole corpus quantitatively with reference to the doublets coordinated with the conjunction 'and', I would claim that this study could be used as a resource for this kind of doublets. Qualitative analysis showed how the phraseological profile of each doublet could be extracted from the corpus. These phraseological profiles would help EFL law students to improve their writing skill either autonomously or with the help of their teachers who could play an active role in this regard as the following section shall demonstrate.

## 6 Discussion

The use of corpus-based methodology in this study helped to offer a quantitative-based holistic view of the use of doublets in legal discourse in COSCO-US. Then, by means of data-driven methods, the present study offered a method for analyzing the phraseological profile of any target doublet.

The first question of this study sought to identify all the doublets whose lexical items are linked with the coordinator 'and'. These doublets totaled 84 doublets with a total frequency of 25.388 doublets. In reference to the second question, it is found that these doublets are structured in six grammatical patterns: 'verb (and) verb', 'noun (and) noun', 'adjective (and) adjective', 'preposition (and) preposition', 'determiner (and) determiners', and 'conjunction (and) conjunction'. Due to the nature of the word classes (open and closed) linked in these doublets, the grammatical patterns 'verb (and) verb', 'noun (and) noun', 'adjective (and) adjective' proved to be more frequent and productive in terms of their morphological forms (lemmas). Though the lexical items

included in these doublets can be used in different morphological forms and syntactic structures, an exploration of the concordance lines in which these doublets are used showed that legal discourse puts some constraints on the use of these doublets. In reference to the third study question which sought to reach a systematic method to analyze the phraseological profile of the target doublets, the study specified a four-step method including the identification of the grammatical pattern of the doublet, its different forms (lemmas), the semantic relationship between its component parts, and finally the most frequent collocates used along these doublets. This method would help EFL law students to learn about the formation and the use of these doublets taking into consideration the availability of the corpus, the user-friendly corpus tools offered in the website, and the possibility of comparing these doublets with regard to their pragmatic force based on contextual clues.

Based on the findings of this study, the extracted linguistic aspects of each doublet could be implemented in different activities to help EFL law students to improve their phraseological competence in writing courses based on relevant concordance lines. Some recommended exercises would target the skills of selecting the accurate lexical items used in the doublet, deciding on the proper morphological form, choosing the accurate collocate, and identifying the pragmatic force of the doublet based on the context in which it is used. Having much training on these types of exercises would hopefully minimize students' errors regarding the use of doublets on both the quantitative and qualitative levels. On the quantitative level, students' errors might include the underuse of overuse of doublets. On the qualitative level, their errors might include the form of the doublet, the appropriate preposition, the proper collocate, etc.

## **7 Conclusion**

This study aimed at exploring how the analysis of the phraseological profile of legal doublets could help EFL law students to learn about how these doublets are used in authentic legal texts. By means of qualitative and quantitative methods, it offered deep insights into the nature of legal doublets in terms of their frequency, grammatical patterns, lemmas, semantic relatedness, and collocability in order to improve writing skills. Also, the study stressed the role that online corpora could play in augmenting learners' autonomy while being engaged in free or controlled language drills.

The study offered a four-step method to maximize the benefit of using corpus-based techniques to learn about phraseology. The first step is to identify all the target phraseological units in a corpus, and then to sort the extracted phraseological units into specific grammatical patterns. The second step is to decide on all possible morphological variations. The third step is to highlight the semantic relationship between the component lexical items, and the final step is to mark the significant collocates used with such phraseological units. This method would identify all the linguistic aspects of each phraseological unit that would achieve the objectives of using the corpus. These objectives may include the exploration of specific syntactic structures, lexical relations, collocations, etc. Further future studies are recommended to test the validity of this method based on an experimental research design. Similar studies

could benefit from the methodology offered in this study to study other types of phraseological units such as lexical bundles and register-based collocations.

## 8 Acknowledgement

I take this opportunity to thank Prince Sattam Bin Abdulaziz University in Saudi Arabia alongside its Scientific Deanship, for all the technical support it has unstintingly provided towards the fulfillment of the current research project.

## 9 References

- [1] A. Boulton, "Data-driven learning: Taking the computer out of the equation," *Language Learning*, vol. 60, no. 3, pp. 534-572, 2010. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>
- [2] J. Hüttner, U. Smit, and B. Mehlmauer-Larcher, "ESP teacher education at the interface of theory and practice: Introducing a model of mediated corpus-based genre analysis," *System*, vol. 37, no. 1, pp. 99-109, 2009. <https://doi.org/10.1016/j.system.2008.06.003>
- [3] G. N. Leech, "Frequency, corpora and language learning," in *A Taste for Corpora*. In Honour of Sylviane Granger, F. Meunier, S. D. Cock, G. Gilquin, and M. Paquot, Eds. Amsterdam; Philadelphia, PA: Benjamins, 2011, pp. 7-32. <https://doi.org/10.1075/scl.45>
- [4] A. O'Keeffe, M. McCarthy, and R. Carter, *From corpus to classroom: Language use and language teaching*. Cambridge, UK: Cambridge University Press, 2007.
- [5] R. Reppen, *Using corpora in the language classroom*. Cambridge, UK: Cambridge University Press, 2010.
- [6] G. Sha, "Using Google as a super corpus to drive written language learning: A comparison with the British National Corpus," *Computer Assisted Language Learning*, vol. 23, no. 5, pp. 377-393, 2010. <https://doi.org/10.1080/09588221.2010.514576>
- [7] J. Sinclair, "New evidence, new priorities, new attitudes," in *How to use corpora in language teaching*, J. Sinclair, Ed. Amsterdam: John Benjamins, 2004, pp. 271-299. <https://doi.org/10.1075/scl.12.20sin>
- [8] N. Vyatkina and A. Boulton, "Corpora in language learning and teaching," *Language Learning & Technology*, vol. 21, no. 3, pp. 1-8., 2017.
- [9] J. Belz and N. Vyatkina, "The pedagogical mediation of a developmental learner corpus for classroom-based language instruction," *Language Learning & Technology*, vol. 12, no. 3, pp. 33-52, 2008.
- [10] S. Bernardini, *Competence, capacity, corpora. A study in corpus-aided language learning*. Bologna: CLUEB, 2000.
- [11] M. M. Jaén, "A corpus-driven design of a test for assessing the ESL collocational competence of University students," *International Journal of English Studies*, vol. 7, no. 2, pp. 127-147, 2007.
- [12] K. Kotani, T. Yoshimi, N. Hiroaki, and H. Isahara, "A Corpus of writing, pronunciation, reading, and listening by learners of English as a foreign language," *English Language Teaching*, vol. 9, no. 9, pp. 139-155, 2016. <https://doi.org/10.5539/elt.v9n9p139>
- [13] S. Phoocharoensil, "Language corpora for EFL teachers: An exploration of English grammar through concordance lines," presented at the International Educational Technology Conference. *Procedia - Social and Behavioral Sciences*, 2012. <https://doi.org/10.1016/j.sbspro.2012.11.060>

- [14] K. Ackerley, "Effects of corpus-based instruction on phraseology in learner English," *Language Learning & Technology*, vol. 21, no. 3, pp. 195–216, 2017.
- [15] L. Forti, "Learner attitude toward data-driven learning: Investigating the effect of teaching contexts," in *CALL and complexity – short papers from EUROCALL 2019*, 2019, pp. 137-143. <https://doi.org/10.14705/rpnet.2019.38.999>
- [16] J. Geluso and A. Yamaguchi, "Discovering formulaic language through data-driven learning: Student attitudes and efficacy," *ReCALL*, vol. 26, no. 2, pp. 225-242, 2014. <https://doi.org/10.1017/s0958344014000044>
- [17] S. Granger and F. Meunier, "Phraseology in language learning and teaching: Where to from here?" in *Phraseology in foreign language learning and teaching*, vol. 247-252, F. Meunier and S. Granger, Eds. Benjamins: Amsterdam & Philadelphia, 2008. <https://doi.org/10.1075/z.138.19gra>
- [18] A. Buşila, "The issue of translating legal doublets in notarial acts from English into Romanian," *Lingua Legis*, vol. 27, pp. 123-136, 2017.
- [19] E. Huertas Barros and M. Buendía Castro, "Analysing phraseological units in legal translation: Evaluation of translation errors for the English-Spanish language pair," in *Phraseology in legal and institutional settings: A corpus-based interdisciplinary perspective (Law, Language and Communication)*, S. G. Roszkowski and G. Pontrandolfo, Eds. London & New York: Routledge 2018, pp. 41-60. <https://doi.org/10.4324/9781315445724-4>
- [20] G. Pontrandolfo, "Investigating phraseology with COPSE: A contrastive corpus-based study," in *New directions in corpus-based translation studies*, C. Fantinuoli and F. Zanettin, Eds. Berlin: Language Science Press, 2015, pp. 137-160. [https://doi.org/10.26530/oa-pen\\_559833](https://doi.org/10.26530/oa-pen_559833)
- [21] K. Yunus, S. Awab, and R. Ab Rashid, "Colligations of prepositions: Essential properties of legal Pphraseology," *International Journal of Applied Linguistics & English Literature*, vol. 5, no. 6, pp. 199-208, 2016.
- [22] A. Philips, *The lawyer and society*. Ottawa: Carleton Press, 1991.
- [23] D. Cao, *Translating law*. Toronto: Multilingual Matters Ltd, 2007.
- [24] P. Tiersma, "The nature of legal language," in *Dimensions of forensic linguistics*, J. Gibbons and M. T. Turell, Eds. no. AILA Applied Linguistics Series 5) Amsterdam/Philadelphia: John Benjamins, 2008, pp. 7-26. <https://doi.org/10.1075/aals.5.03tie>
- [25] V. Bhatia, "Language of the law," *Language Teaching*, vol. 20, no. 4, pp. 227-234, 1997.
- [26] R. Gläser, "The stylistic potential of phraseological units in the light of genre analysis," in *Phraseology: Theory, analysis, and applications*, A. P. Cowie, Ed. Oxford: Oxford University Press, 1998, pp. 125-143.
- [27] S. Granger and M. Paquot, "Disentangling the phraseological web," in *Phraseology: An Interdisciplinary Perspectives*, S. Granger and F. Meunier, Eds. Amsterdam & Philadelphia: Benjamins, 2008, pp. 27-49. <https://doi.org/10.1075/z.139.07gra>
- [28] V. V. Vinogradov, *The main understanding of Russian phraseology as a linguistic discipline*. Moscow: Science, 1977.
- [29] B. Danet, *Legal discourse*. London: Academia Press, 1985.
- [30] M. Gustafsson, "The syntactic features of binomial expressions in legal English," *Text* 4, vol. 1, no. 3, pp. 123-141, 1984.
- [31] M. Gustafsson, *Binomial expressions in present-day English: A syntactic and semantic study*. Turku: Yliopisto, 1975.
- [32] Y. Malkiel, "Studies in irreversible binomials," *Lingua Legis*, vol. 8, pp. 113-160, 1959. [https://doi.org/10.1016/0024-3841\(59\)90018-x](https://doi.org/10.1016/0024-3841(59)90018-x)
- [33] T. F. Johns, "Should you be persuaded: Two examples of data-driven learning," in *Classroom concordancing*, T. F. Johns and P. King, Eds. Birmingham: ELR, 1991, pp. 1-13.

- [34] G. N. Leech, "Teaching and language corpora, a convergence," in *Teaching and Language Corpora (Applied Linguistics and Language Study)*, A. Wichmann, T. McEnery, and G. Knowles, Eds. London: Longman Publishing Group, 1997, pp. 1–23. <https://doi.org/10.4324/9781315842677-1>
- [35] E. Corino, "Bottom up specialized phraseology in CLIL teaching classes," in *Proceedings of the 12th Edition of the KONVENS Conference*, Hildesheim, 2014, pp. 68–76: Universitätsverlag Hildesheim.
- [36] U. Romer, "Corpus research and practice: What help do teachers need and what can we offer?" in *Corpora and language teaching*, K. Aijmer, Ed. Amsterdam: John Benjamins, 2009, pp. 83-98. <https://doi.org/10.1075/scl.33.09rom>
- [37] J. Smart, "The role of guided induction in paper-based data-driven learning," *ReCALL*, vol. 26, no. 2, pp. 184-201, 2014. <https://doi.org/10.1017/s0958344014000081>
- [38] S. Sonbul and N. Schmitt, "Direct teaching of vocabulary after reading: Is it worth the effort?" *ELT Journal*, vol. 64, no. 3, pp. 253-260, 2010. <https://doi.org/10.1093/elt/ccp059>
- [39] J. Y. Chang, "The use of general and specialized corpora as reference sources for academic English writing: a case study," *ReCALL*, vol. 26, no. 2, pp. 243–259, 2014. <https://doi.org/10.1017/s0958344014000056>
- [40] A. Gilmore, "Using online corpora to develop students' writing skills," *ELT Journal*, vol. 63, no. 4, pp. 363–372, 2009. <https://doi.org/10.1093/elt/ccn056>
- [41] Q. Luo and J. Zhou, "Data-driven Learning in Second Language Writing Class: A Survey of Empirical Studies," *International Journal of Emerging Technologies in Education*, vol. 12, no. 3, pp. 182-196, 2017. <https://doi.org/10.3991/ijet.v12i03.6523>
- [42] M. L. Pérez Cañado and M. B. Díez Bedmar, "Data-driven learning and awareness-raising: An effective tandem to improve grammar in written composition? ," *International Journal of Emerging Technologies in Education*, vol. 1, no. 3, pp. 1-11, 2006.
- [43] L. Zhang, "Effectiveness of college student English writing teaching Based on data-driven learning," *International Journal of Emerging Technologies in Education*, vol. 13, no. 4, pp. 23-34, 2018.
- [44] H. G. Widdowson, *Aspects of language teaching*. Oxford: Oxford University Press, 1990.
- [45] L. Flowerdew, "Applying corpus linguistics to pedagogy: A critical evaluation Applying corpus linguistics to pedagogy: A critical evaluation," *International Journal of Corpus Linguistics*, vol. 14, no. 3, pp. 393-417, 2009. <https://doi.org/10.1075/ijcl.14.3.05flo>

## 10 Author

**Waheed M. A. Altohami** is assistant professor in linguistics at the Department of English, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Saudi Arabia. Also, he worked as a lecturer in linguistics, Department of Foreign Languages, Faculty of Education, Mansoura University, Egypt. His research interests are corpus linguistics, discourse analysis, historical semantics, and translation.

Article submitted 2020-02-28. Resubmitted 2020-06-19. Final acceptance 2020-06-25. Final version published as submitted by the authors.