



Corpora and Computation in Teaching Law and Language

*Ruth Breeze**

Abstract

Corpus use has revolutionised the teaching of languages for specific purposes. In this article, I review some of the ways in which corpus studies can enhance our understanding of the regularities in legal language, looking particularly at formulaic language in four different areas of legal language: academic law articles, case law (judgments and opinions), documents (contracts, merger agreements, etc.) and legislation. After a brief overview of lexical issues in legal language, I look in greater detail at 4- to 8-word bundles in legal texts. After some consideration of legal modality and recurring syntactic structures, I show how these aspects come together with the phenomenon of bundles and formulaicity. I then provide some examples of how the kind of information provided here by specialised corpora can be exploited for teaching purposes.

Keywords

Legal English, corpus linguistics, multimodality, formulaic language, bundles, modal verbs

Submitted: 22 December 2016, accepted: 29 May 2017, published online: 28 June 2017

* *Ruth Breeze*: University of Navarra, Spain, rbreeze@unav.es. The author would like to thank the Instituto Cultural y Sociedad and the School of Law at the University of Navarra for their support.

1. Introduction

The construction of large corpora and the availability of increasingly sophisticated computer tools to process language information have profoundly changed the way we understand and teach languages.

On the one hand, it is now easier to search for words in context, both to gain a deeper understanding of the ways a particular word is used, its collocates, associated prosody, local grammar, etc., and to study lexical patterns that run through a range of material from different sources. On the other hand, we can also start from the text, and use corpus affordances to find out what makes a particular text or genre different from others, learn more about the contrast between spoken and written language, or gain insights into generic patterning that is not apparent to the naked eye. The insights from this not only help us to explain language phenomena more clearly to our students, but they also allow us to construct better didactic tasks and provide richer, more varied examples to use in the classroom.

All of this is true for language teaching in general, but it is even more important in the area of languages for specific purposes. Corpus use facilitates the creation of subject-specific wordlists, enhances the investigation of professional genres, and provides a wealth of insights into the socio-cultural phenomenon of specialised language. This article is intended to provide a selection of different ways in which corpora and computation can help us approach the teaching of legal English. Here, I describe how our understanding of aspects such as word frequency, keywords and bundles can be operationalised in preparing course material. This article considers the particular texture of legal texts in different genres, and the way in which formulaic language serves both to constitute the frames and fill the slots in legal discourse, particularly in the most highly conventionalised genres such as documents and legislation. My discussion then points to ways in which teachers can use corpora to gain deeper knowledge of complex text structures and formulaic expressions, in order to scaffold student learning.

2. Insights into specificity: why is legal English different?

Among specialised professional languages, legal English has the reputation of being one of the most difficult for the layperson. It presents challenges on many levels: in lexical areas (specialised terminology, often of Latin or Norman French origin, sometimes involving archaisms or redundant expressions), discourse organisation (very long sentences with many embedded clauses, the persistence of features such as compound reference words, such as “hereinafter”), interpersonality (performative speech acts, highly formal register, third-person address), grammar (frequency of conditional structures, characteristic modal system based on “shall” and “may”), and so on (Alcaraz

& Hughes, 2002). Despite initiatives such as the UK “Civil Procedure Rules” (1998) or the US “Plain Writing Act” (2010), much legal language remains inaccessible to non-specialists. Although studies based on close analysis of texts and interactions, or the diachronic development of particular genres, are valuable for understanding what is special about specialised language, computer-based investigation also offers a useful way of bringing out the unique nature of particular types of legal discourse. Corpus studies can potentially answer the question whether legal language is truly “different” from other kinds of English, that is, whether a “legal register” exists that runs through different genres, and what it might contain. They can also help us to show our students what variations occur from one legal genre to another, or even within one particular kind of text.

Legal English is, of course, a vast area containing many sub-domains which vary in terms of vocabulary, structures and genres. However, the huge expansion in international trade over the last twenty years has meant that the field of commercial law can be identified as particularly important for legal practitioners outside the English-speaking world. This means that law students taking degrees and LL.M.s are likely to benefit most from language support in this specific field. In order to approach this area of legal English using corpus linguistic tools, I gathered two million words from the area of commercial law, divided into four corpora of approximately 500,000 words each from: academic law articles on commercial and corporate law, case law (judgments and court opinions), legislation (Companies Acts) and legal documents (contracts, commercial lease agreements, merger agreements and so on) (see Breeze, 2013, for a more detailed description). WordSmith 6 and SketchEngine were used to perform the various quantitative tests used below.

Table 1: Comparative data in four legal corpora.

	Academic	Cases	Documents	Legislation
Lexical difference	3.69	3.55	5.69	5.73
Type/token ratio	38.52	35.52	29.43	24.16
Mean word length (in letters)	4.99	4.74	4.99	4.67
Mean sentence length (in words)	20.22	23.5	51.59	45.66

Note: *Lexical difference* is calculated by “compare corpus”, using EnTenTen13 as a reference corpus. Higher numbers indicate greater differences.

Table 1 represents a starting point for the quantitative study of legal English. The table shows that the documents and legislation corpora differed more sharply from “general language” (represented by the EnTenTen13 corpus of general English) than the academic or case law corpora. Conversely, the type/token ratio was higher in academic and case law, and lower in legislation and documents, which is reasonable, given the technical nature of the lexis in legislation and documents: the same words are likely to be repeated for the sake of clarity, or because similar formulae are being used. The

mean sentence length was much greater in legislation and documents: The conventions governing these genres are quite different from those that characterise academic writing, since it is usual for a considerable amount of information to be included in one sentence, and it is possible for a single sentence to extend over several paragraphs or sections of the text.

3. Lexical issues

Like most professional areas, the law is rich in specialised terminology. Technical terms are an essential feature of specialised areas. Moreover, it is also likely that certain types of written document (instructions, technical reports, etc.) will rely more heavily on technical terminology than, say, promotional websites or letters to clients. As we saw in Table 1, although all four corpora presented a substantial degree of lexical difference from the general English reference corpus (EnTenTen13), the Documents and Legislation corpora contrasted more dramatically with the reference corpus, which indicates that the vocabulary of these corpora is much more specialised.

The measure of keyness allows us to find out which words are particularly frequent in the corpus in question, in comparison with the larger reference corpus (in this case, EnTenTen13). To show how this can be used, Table 2 displays the eight verbs in each corpus with the highest keyness scores.

Table 2: Verbs with highest keyness score (reference corpus: EnTenTen13) in each corpus.

Academic		Cases		Documents		Legislation	
Arbitrate	701	Dismiss	64	Contribute	304	Authorise	107
See	101	See	61	Indemnify	258	Allot	69
Litigate	65	Allege	54	Affiliate	159	Restate	59
Pre-empt	36	Abet	45	Assume	106	Contravene	55
Enforce	32	Imply	37	Exclude	105	Specify	52
Liquidate	29	Litigate	34	Contemplate	94	Confer	49
Preclude	29	Subrogate	32	Amend	87	Comply	45

As Table 2 shows, many verbs with a highly technical meaning have a high keyness score, which means that technical words are very frequent and so learners will need to be familiar with them in order to make progress in their comprehension of legal texts. Similar data could be presented for nouns or adjectives, shedding light on the need to emphasise technical lexis when teaching legal language. The use of corpora also makes it possible to zoom in on particular specialised areas within one field. So, for example, if we compare a “minicorpus” of contracts of sale with the main Documents corpus, we

can use the keywords function to find out what lexical items are going to be particularly frequent when we are looking at contracts of sale. As Table 3 illustrates, some of these words are predictable, while others might come as a surprise. When compiling course material, use of such procedures can help teachers to ensure that students have adequate vocabulary coverage from the type of document they are likely to encounter, or from a specific range of documents.

Table 3: Keywords in subcorpus of contracts of sale (keyness >45).

buyer	closing	allocated	preferential
defect	assets	intangible	records
seller	title	escrow	affected
purchase	knowledge	past	transaction

Information about the vocabulary that is specific to each area can be used, in combination with our knowledge of the lexis of legal texts in general, to generate practice and revision exercises such as Exercise One. Such exercises at first appear difficult, because of the clustering of unfamiliar vocabulary within a complex sentence. However, when students learn to approach the task systematically, they soon find that their understanding of the legal background and the interactional character of the clause enables them to solve the problem easily and build up confidence to tackle longer texts.

Exercise One

Put the words in bold into the correct gap in this extract from a contract of sale:

Escrow	Buyer	Sellers	Claims
Promptly upon the expiration of the Claims Period, (<i>Answer: Buyer</i>) shall pay to (<i>Answer: Sellers</i>) an aggregate amount equal to that portion of the (<i>Answer: Escrow</i>) Amount that has not been used to satisfy Buyer's indemnification (<i>Answer: Claims</i>).			

Nonetheless, since legal English textbooks often have a lexical orientation (cf. Krois-Lindner, 2006; Brown & Rice, 2007; Reinhart, 2007; see also Breeze, 2015), and students are generally extremely aware of the need to acquire a large specialised vocabulary, this will not form the main object of the present paper.

4. Exploring formulaic language

Moving on from simple word frequencies and keyness to lexical patterning, the first feature that strikes many people when they read certain types of legal text is its highly formulaic nature. To examine formulaicity, I worked with the concept of the “lexical bundle”, first applied in Biber et al. (1999), which is specifically used to mean frequently recurring sequences of words regardless of the nature of the kind of links that might exist between them. In other words, such “bundles” may not be collocations or set

phrases, but the fact that they recur frequently may have some significance for our understanding of specialised language (Biber & Conrad, 1999). Biber (2006) brought to light patterns that emerge from seemingly fragmentary bundles, revealing a certain degree of regularity within fragmentation. For example, he documented the presence of large numbers of stance expression fragments, discourse organising fragments, and referential expressions, as well as a certain number of set phrases. Other authors working on spoken academic discourse have shown that such bundles often have discourse organising functions (Csomay, 2004; Nesi & Basturkmen, 2006). Studies of academic written language, on the other hand, have shown that bundle use varies across disciplines (Hyland, 2008): research-oriented bundles used in the sciences prioritised empirical methods and findings, while text-oriented bundles in humanities and social science disciplines reflected the value accorded to coherent argument (Hyland, 2008: 16). This section will show how examination of bundles sheds light on legal discourses and provides material that can be exploited pedagogically.

When the four corpora in this study were examined using WordSmith to identify frequent bundles, the corpus with the greatest number of repeated combinations of 4 to 8 words was the legislation corpus, followed by the documents corpus. The academic and cases corpora made use of fewer long bundles, although they did have more 4- and 5-word bundles than would be expected in, say, fiction or media texts.

Table 4: Frequency of different 4- to 8-word bundles in the four corpora (from Breeze, 2013).

	Academic	Cases	Documents	Legislation
8-word bundles	0	1	19	54
7-word bundles	1	3	38	75
6-word bundles	2	5	80	115
5-word bundles	8	18	171	284
4-word bundles	53	76	384	564

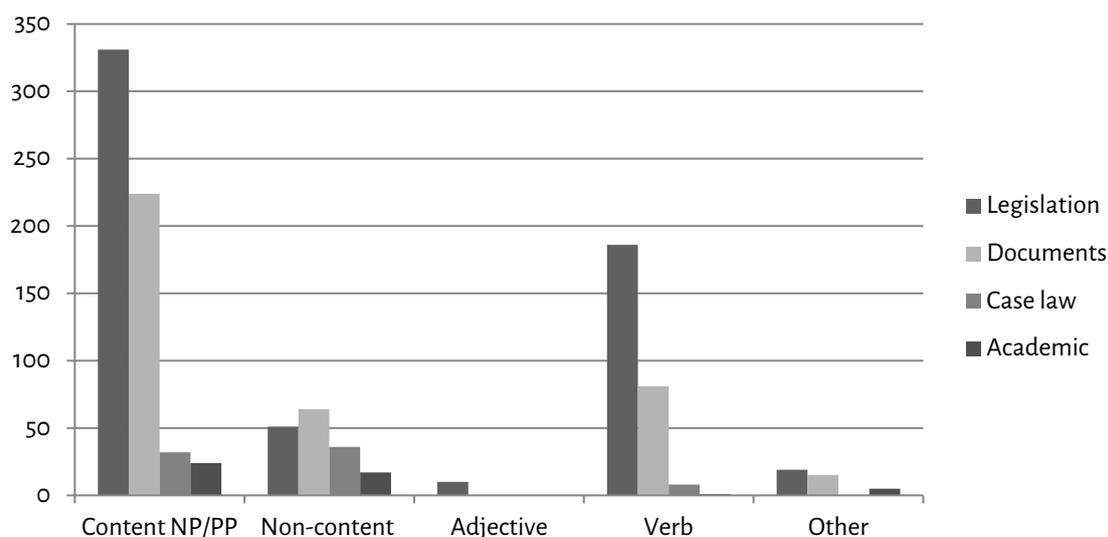
Note: Numbers denote raw (absolute) frequencies in each corpus of approximately 0.5 million words.

As Table 4 suggests, the most striking bundles were those of 5, 6, 7 and 8 words found in the Documents and Legislation corpora. These tended to be either heavy noun phrases such as “officer or secretary of the board of directors”, or verb phrases such as “shall have the meaning set forth in the” which reflect formulae used in documents. Although such phrases might not seem particularly attractive pedagogically, we should bear in mind that speed of comprehension (as well as production) generally improves when students learn to recognize (or produce) fixed or semi-fixed lexical chunks (Nattinger & DeCarrico, 1992; Wray, 2000). Since many legal documents consist of sentences like example 1 (below), the ability to recognize and process fixed formulae is a skill that students should acquire. Focusing students’ attention on how to divide the sentence into its component chunks is likely to be useful for comprehension purposes, and absolutely essential if translation forms part of the curriculum.

(1) No waiver by either party hereto / of any breach of this Agreement / shall be deemed to be / a waiver / of any preceding or succeeding breach / of the same or any other provision hereof.

Following on with the topic of bundles, since 4-word bundles were frequent, we focused on classifying these, using a procedure based on Biber (2006). Four main categories emerged: content noun phrases, non-content phrases, verb phrases, and instructions. Around 4% of the 4-word-bundles had to be discarded because it was not clear which category they might belong to.

Figure 1: Bundle types in the four corpora.



From the information displayed in Figure 1, it is evident that the category of “non-content” bundles accounted for a significant proportion of these bundles. On closer inspection, many of these turn out to be complex prepositional phrases such as “in the context of”, “on behalf of the”, “at the time of” or “in the event of”. As in the case of the longer bundles, familiarity with these patterns should help students to gain reading speed and improve their accuracy.

Table 5: Ten most frequent 4-word prepositional phrases in Documents and Legislation corpora.

Rank	Documents	Legislation
1.	In accordance with the	In the case of
2.	On behalf of the	For the purpose of
3.	With respect to the	In accordance with the
4.	In connection with the	In respect of the
5.	In the case of	With respect to the
6.	In respect of the	On behalf of the
7.	As a result of	Within the meaning of
8.	To the extent that	To the extent that
9.	To the knowledge of	As a result of
10.	In the event of	In the event of

Since 4-word bundles are particularly frequent in the Documents and Legislation corpora, Table 5 shows the ones which occur most in each corpus. The most common prepositional phrase bundles appear in the context of the need for inter- and intratextual reference in legal texts (i.e. “in accordance with the”), the need for delimitation and precision (i.e. “to the extent that”), and a process of nominalisation of causal and conditional relations (i.e. “as a result of” to replace “because”, and “in the event of” to replace “if”), which has been discussed elsewhere as a typical feature arising from the need to assign technical legal values to actions or utterances (Vázquez Orta, 2010: 273). In this context, exercises of the following type can be used to raise students’ awareness of this type of bundle.

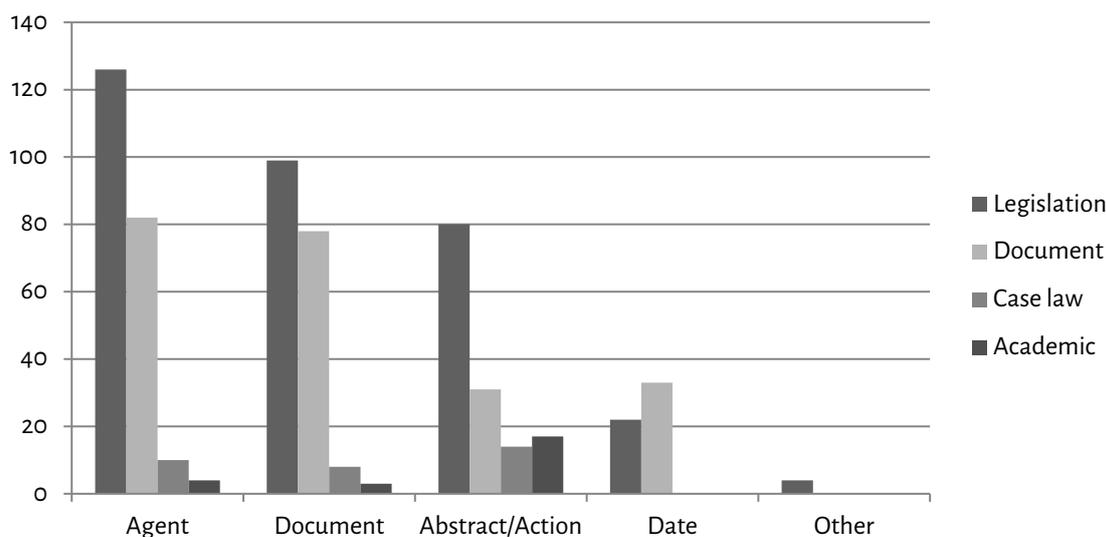
Exercise Two

Complete these phrases from legal documents and laws using “in”, “on”, “within” or “of”:

- A. (..... *Answer: In*) the event (..... *Answer: of*) a breach or threatened breach of the terms of this Agreement by Consultant, the parties hereto agree that monetary damages would be an inadequate remedy for said breach.
- B. Any person who is a worker (..... *Answer: in*) the meaning (..... *Answer: of*) the Act and is over 16 years of age may join a trade union.
- C. When an arbitrator considers that a statement of claim made (..... *Answer: on*) behalf (..... *Answer: of*) the claimant should be the subject of two or more separate arbitrations, he may refuse to deal with multiple claims in a single reference.

Since content bundles are frequent, it is also interesting to examine which type of content bundle is more frequent in the different corpora. Figure 2 below shows the proportion of 4-word bundles belonging to the category of “content” that could be classified as representing agents (people or institutions), documents (laws, contracts, etc.), and abstract concepts or actions.

Figure 2: Bundle categories in the four corpora.



As we can see from Figure 2, the names of documents and abstract concepts figured largely in these texts. Documents ranged from legislation (“Model Business Incorporation Act”) to everyday documents in the life of a company (“this memorandum of association”). Abstracts ranged from theoretical entities such as “contractual choice of law” or “the corporate law market” in academic texts, to aspects of corporate practice (“ordinary course of business”, “all liabilities and obligations”) in the documents corpus. Again, the frequent bundles can be identified and used in the classroom in order to familiarize students with the texture of legal texts.

Exercise Three provides a scaffolded approach to understanding a dense clause concerning the bundle “a Material Adverse Effect”, embedded within a complex grammatical structure. The twofold difficulty (heavy noun phrase and conditional passive of “expect”) is disentangled stage by stage, as students are invited to try to express the same legal concept in everyday language.

Exercise Three

Read the following clause from a merger agreement. You are going to explain this clause to a client who is not a legal specialist. Make some notes to help you give your explanation.

Absence of Certain Changes. Since December 31, 2007 until the date hereof, (1) the Company and the Company Subsidiaries have conducted their respective businesses in all material respects in the ordinary course, consistent with prior practice, (2) except for publicly disclosed ordinary dividends on the Common Stock and outstanding Company Preferred Stock, the Company has not made or declared any distribution in cash or in kind to its shareholders or issued or repurchased any shares of its capital stock or other equity interests and (3) no event or events have occurred that has had or would reasonably be expected to have a Material Adverse Effect.

Before you give your explanation, answer the following comprehension questions:

1. What is a Material Adverse Effect? (*Answer: Something that has happened in the company that would make it less attractive to buy.*)
2. Re-read the final phrase: “no event or events have occurred that has had or would reasonably be expected to have a Material Adverse Effect”. Try to express this without using the passive. (*Answer: Nothing has happened in the company that would make people think that it is less attractive to buy.*)
3. If you have problems with the last sentence, check the way the passive is used here. We can say that “we expect that an event will have a Material Adverse Effect”. How can we express this idea impersonally, using “is expected to”? (*Answer: An event is expected to have an MAE.*) How can we express this to make it sound as if we are not certain? (*Answer: An event could be expected to have an MAE.*) How can we express this to make it sound as if people would reasonably expect the event to have an MAE? (*Answer: An event would be reasonably expected to have an MAE.*)

Now work in pairs, taking turns to be the client who does not understand the clause and lawyer who has to explain the clause.

5. Grammatical patterns

As the examples in the previous section show, the degree of grammatical complexity in legal texts is often very high, and this phenomenon presents a considerable challenge

for teachers. One approach to this might be to identify some of the reasons why the sentences in Documents and Legislation (see Table 1) are so long, in terms of the pragmatic functions that are being fulfilled and the conventions associated with their realisation. One particular function that has been associated with legal texts since Babylonian times (see Breeze, 2013) is that of seeking to define the consequences of actions, or the conditions in which something may or must be done. The use of “if” is frequent in this context, even though other mechanisms also exist (such as “in the case of” or “in the event of” used with nominalisations, as mentioned above).

Table 6: Frequency per 10,000 words of “if” and “if”/“had” in the four corpora.

	Academic	Cases	Documents	Legislation
If	23.1	44.0	25.4	43.9
If + (1-4) + had	0.5	1.3	0.3	1.3

Table 6 shows the frequency of “if” (as a crude measure of conditional structures in which verbs are used) and “if” plus “had” (within five words either way, calculated using the “filter” option on SketchEngine) (as a crude measure of counterfactual conditionals) in the four corpora. Interestingly, although the frequency of “if” in the Academic and Documents corpus is similar to that found in BNC (22 per 10,000) and EnTenTen13 (24.6), the frequency in Cases and Legislation is much higher. This is logical, in that conditional-type structures are strongly associated with proceedings and texts in which different courses of action and their consequences are laid out. It is slightly more difficult to see why such structures are less frequent in Documents, but the answer would appear to lie in the point that such documents (mainly contracts of different kinds) exist to set out what must and must not be done in a particular situation, rather than to allow for many different contingencies (as in the case of legislation) or to determine whether or not something was actually done, and whether or not that action violated a particular rule or principle (as in case law).

The following simple exercise based on an extract from the Legislation corpus illustrates the way students can be encouraged to develop an awareness of the characteristic “if” structures in legal texts. When used with law students, this type of exercise serves to draw students’ attention to specific features of the language of the text, as well as to encourage close reading. It is likely that the same type of exercise would also have potential for use with, say, students of translation, but in this case, the analysis could be directed towards linguistic aspects, such as the difference between “limited by” and “limited to”, or the reasons motivating the use of repetitions in legal genres.

Exercise Four

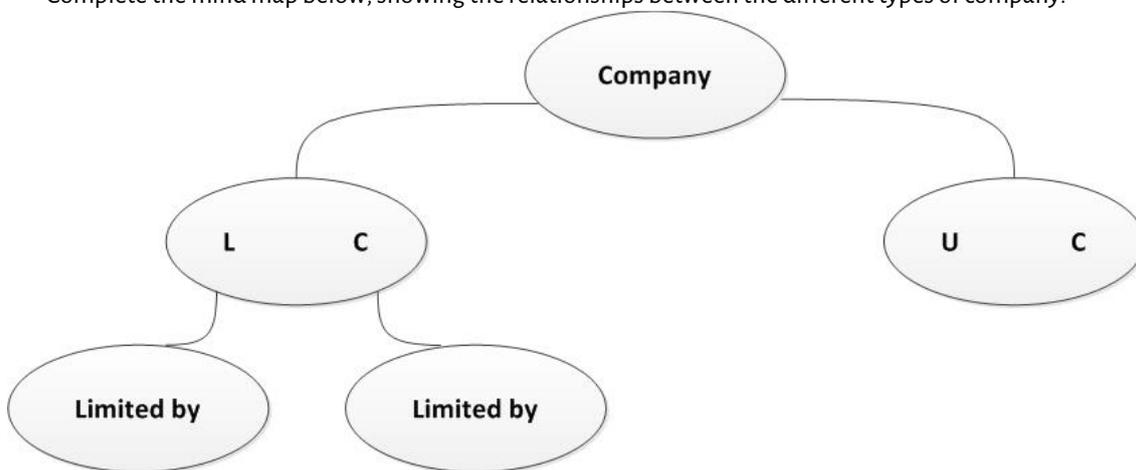
Read the following extract about different types of company:

A company is a “limited company” if the liability of its members is limited by its constitution. It may be limited by shares or limited by guarantee. If their liability is limited to the amount, if any, unpaid on the shares held by them, the company is

“limited by shares”. If their liability is limited to such amount as the members undertake to contribute to the assets of the company in the event of its being wound up, the company is “limited by guarantee”. If there is no limit on the liability of its members, the company is an “unlimited company”.

The word “if” occurs five times in the above extract, and one of the uses is different from all the others. Which one is different?

Complete the mind map below, showing the relationships between the different types of company.



6. Modality

It is well known that the system of modality in English-language legal documentation traditionally finds its central axis in “shall” (obligation) and “may” (permission) (Trosborg, 1997). Here, we find that “shall” is the 9th most frequent word in the Documents corpus, and is also common in legislation, while in the Academic and Case law corpora the frequency of “shall” is much closer to that found in general English. On the other hand, “may” is particularly frequent in Legislation (seemingly, legislators are more interested in granting permission than in prescribing). All four legal corpora have higher frequencies of “may” than the general corpus (see Breeze, 2014, for an investigation of the use of “may” in legal correspondence). Regarding the other modal verbs quantified here, “will” is much more frequent in the general corpus than in the legal corpora, while “should” occurs more in the Academic and Case law corpora, and hardly at all in Documents and Legislation (which are not associated with discourse functions such as advising, recommending or predicting commonly fulfilled by “should”).

Table 7: Frequency per 10,000 words of modal verbs of obligation and permission across corpora.

	Academic	Cases	Documents	Legislation	EnTenTen13
Shall	3.2	4.4	103.5	39.6	1.0
May	27.1	13.4	28.7	50	11.0
Will	19.8	10.9	17.8	28.9	38.8

	Academic	Cases	Documents	Legislation	EnTenTen13
Must	7.4	8.8	2.5	29.9	4.7
Should	12.8	11.8	1.2	1.4	8.9

If we combine the information we have about the frequency of “shall” with what we know about verb phrase bundles (see above), we find that “shall” commonly occurs in bundles such as “shall be read as” and “shall be deemed to” in Legislation, and “shall be governed by”, “shall have the right”, and “shall be deemed to” in Documents. Although the use of “must” rather than “shall” is preferred by some advocates of plain English (Garner, 2001), change has been slow outside the United States, particularly among drafters of legislation and legal documents (Williams, 2011): for example, Williams (2013) shows that the frequency of “shall” remained stable among EU drafters between 1973 and 2010. In fact, although use of “shall” for obligation is an archaism, its use in technical legal contexts rarely presents difficulties of comprehension once the reader is aware of this convention. Exercise Five below is designed to encourage students to focus on the ways in which one frequent bundle is used in legal documents. Such exercises should enable law students to become more familiar with the language of documents and learn to take advantage of its formulaic nature in order to read more efficiently. Exercises in chunking are also likely to bring benefits for translation students.

Exercise Five

In legal documents, the technical phrase “shall be deemed to” has a special role in spelling out the terms of an agreement or understanding in order to ensure shared comprehension. Look at the following extracts from legal documents, and complete the table below. Then answer questions A to C.

- 1. If any notice of a proposed sale of Guaranty Collateral shall be required by law, such notice shall be deemed reasonable and proper if given at least ten (10) days before such sale or other Disposition.*
- 2. If any term or provision of this Guaranty shall be deemed prohibited by or invalid under any applicable law, such provision shall be invalidated without affecting the remaining provisions of this Guaranty.*
- 3. Nothing in this Agreement shall be deemed to create or constitute a partnership, agency, representative or other relationship between the Parties.*
- 4. No failure on the part of a Party to exercise or delay in exercising any right hereunder will be deemed a waiver thereof, nor will any single or partial exercise preclude any further or other exercise of such or any other right.*
- 5. Any notice, request, instruction or other document to be given hereunder by any party to the other will be in writing and will be deemed to have been duly given (a) on the date of delivery if delivered personally or by telecopy or facsimile, upon confirmation of receipt, (b) on the first business day following the date of dispatch if delivered by a recognized next-day courier service, or (c) on the third business day following the date of mailing if delivered by registered or certified mail, return receipt requested, postage prepaid.*

Now fill in the table below about how the verb “to deem” is used in legal documents, using examples from the box above:

- With infinitive: (Answer: shall be deemed to create)
 With past infinitive: (Answer: will be deemed to have been duly given)
 With adjective: (Answer: shall be deemed reasonable and proper)
 With past participle: (Answer: shall be deemed prohibited)
 With noun: (Answer: will be deemed a waiver)

Read the instructions and complete the sentences in each case:

A. You want to define the term “employee” so that it excludes members of the board of directors. Use “shall be deemed” plus infinitive to complete the sentence from the articles of association.

For purposes of this Article III (and the definition of “Maximum Option Number”) only, the term “Employee” (Answer: shall be deemed to exclude) members of the Board of Directors of the Company.

B. You need to say that nothing in this agreement is supposed to create third party rights.

Nothing in this Agreement (Answer: shall be deemed to create) any third party beneficiary rights in any Person or entity not a party to this Agreement.

C. You need to say that stockholders who do not make a proper stock election in the correct way will be understood not to have made a stock election.

Any holder of Company Common Stock who does not properly make a Common Stock Election prior to 5:00 p.m., New York City time, on the Election Date, (Answer: shall be deemed not to have made / shall not be deemed to have made) a Common Stock Election, and all of such holder’s Company Common Shares shall be converted into the right to receive the cash merger consideration as set forth in Section 3.2(c)(i), subject to Section 3.2(c)(iv).

7. Verb-preposition combinations

Given the difficulty of legal English for students, it is perhaps interesting to look briefly at an area where legal language may actually present learners with fewer problems than general English does. Studies of legal English have generally paid little attention to verb/preposition combinations, and where textbooks have done so, the focus is usually on verbs with dependent prepositions (e.g. Krois-Lindner, 2006, contains many useful exercises involving typical combinations of verb and preposition), although phrasal verbs are also found (e.g. carry out, spin off). Table 8 shows the main verb-plus-preposition combinations found in the four corpora, obtained using corpus query language in SketchEngine.

Table 8: Ten most frequent verb-preposition combinations in the four corpora (raw frequencies).

Rank	Academic	Cases	Documents	Legislation
1.	Base on 188	Deal with 51	Comply with 119	Comply with 197
2.	Apply to 141	Apply to 35	Enter into 73	Apply to 195
3.	Relate to 115	Enter into 32	Result in 51	Deliver to 56
4.	Enter into 115	Engage in 27	Deliver to 51	Vote on 47
5.	Depend on 90	Dispose of 27	Apply to 44	Provide for 36
6.	Refer to 69	Comply with 27	Participate in 37	Subscribe for 26
7.	Assign to 63	Result in 25	Pay to 26	Participate in 25
8.	Provide for 60	Rely on 24	Inure to 25	Dispose of 22
9.	Deal with 51	Refer to 22	Cooperate with 24	Carry on 22
10.	Involve in 48	Account for 21	Interfere with 23	Apply for 18

Knowledge of the verb-preposition combinations that students are most likely to encounter in different genres enables teachers to compile exercises that focus attention on particular issues that are likely to be problematic, such as the choice of preposition. Exercise Six was built from the Documents corpus, and represents a model for centering students' attention on frequent verb-preposition combinations. Such exercises allow students to practise the notoriously difficult area of dependent prepositions in a specifically legal context.

Exercise Six

Choose an appropriate preposition from the list in bold to fill the gaps in the following extracts from legal documents:

with **to** **in** **into** **with**

1. Such Investor has the corporate or other power and authority to enter (*Answer: into*) this Agreement.
2. Tenant shall comply (*Answer: with*) any reasonable regulations made by the Landlord regarding the use and occupation of the Premises.
3. The Parties shall not do anything that might interfere (*Answer: with*), obstruct or delay the satisfaction of all or any of the Conditions.
4. The provisions of this Agreement shall be binding upon and inure (*Answer: to*) the benefit of the Parties and their respective successors and permitted assigns.
5. The Company shall use its reasonable efforts to structure the Transactions in a manner that does not result (*Answer: in*) any material tax to the Executive.

8. Textures and trends

Returning to our general overview of formulaic language and typical patterns above, it is possible to put together some ideas about how bundles and syntactic patterning work in the different corpora, and therefore in the different genres. Of course, all language may be underpinned by a restricted range of structural possibilities which offer slots to a vast range of lexical options in order to generate an almost infinite series of meanings. However, what makes specialised language particularly fascinating is that the structures, slot-fillers and meanings are all much more closely circumscribed.

Legal language offers a range of “textures” in this sense, going from the more varied, more loosely structured organisation of academic discourse, to the highly formulaic, tightly structured language of documents and legislation. In what follows, I outline the main findings from the four corpora in terms of the way formulaic language works in each, thinking particularly of the different functions which bundles appear to have in the different types of text.

Academic legal texts contain abstract conceptual noun phrases which act as placeholders, but also non-content prepositional phrases used for referential framing:

(2) The illogic in preserving a distinction between *void* and *voidable* contracts can be illustrated in the context of a contract that is void ab initio for illegality.

Case law contains many verb phrases representing speech acts or indicating textual orientation. Place-holders are often noun phrases, which can be concrete (actors, documents) or abstract (concepts). Non-content prepositional phrases are used for referential framing:

(3) The Claimant was then to divide up the money *in accordance with the other terms of the partnership* to which I shall refer below.

Documents contain many heavy noun phrases (actors, documents, concepts) as place-holders (4), and many non-content prepositional phrases for referential framing, as well as post-modifiers that are also used for intra- and intertextual reference (5):

(4) A shareholder may also take action against *another shareholder or director* pursuant to *these Articles of Association*.

(5) If such disclosure is made *in accordance with the confidentiality obligations set forth in this Agreement*.

Legislation abounds in prepositional phrases which orient the reader within the text or towards other documents, as well as performing functions related to referential framing. It also contains many deontic/regulatory phrases used to connect concepts together. On the other hand, heavy noun phrases (actors, documents and concepts) act as place-holders.

(6) Nothing in the preceding provisions of this section / shall be construed as preventing the use of a registered trademark by any person for the purpose of identifying goods and services.

If we put these ideas together in terms of frames (discourse structures) and slots (places for noun phrases, such as actors, documents or concepts), extract (7), from the articles of association of a company, can be chunked in various ways, and could be presented either in the form of a frame awaiting slots (8), or in the form of place-holders to be joined together by a frame (9).

(7) The board of directors' resolutions / in respect of / all other matters / may be passed by / the affirmative vote of / a simple majority of the directors.

(8) in respect of may be passed by

(9) The board of directors' resolutions all other matters the affirmative vote of / a simple majority of the directors.

Similarly, the legislation clause below (10) could be divided as proposed here, and then presented as a frame (11) or as place-holders (12).

(10) A person guilty of an offence / is liable on summary conviction to a fine / not exceeding level 3 / on the standard scale / and, / for continued contravention, / a daily default fine / not exceeding one tenth of level 3 / on the standard scale.

(11) is liable on summary conviction to and, for continued contravention,

(12) A person guilty of an offence a fine not exceeding level 3 on the standard scale
..... a daily default fine not exceeding one tenth of level 3 on the standard scale.

As teachers, it is extremely important for us to show students how this type of slot-frame interaction occurs in the texts we use. With some support, law students will be able to draw on their familiarity with the interactional framework of legal discourse to decode the text satisfactorily. For translation students, on the other hand, guided work with structures of this kind will help them build an awareness of the law as a system, with its actors, actions, eventualities and consequences, and to gain a feeling for the special pragmatics which underpins the language of the law.

9. Concluding reflections

For people who teach legal English, it is essential to make students aware not only of specific high-frequency terminology, but also of the typical formulaic language that they will encounter. This can be done by consciousness-raising exercises, and by setting tasks such as those exemplified in this article. On a basic level, students should then also be encouraged to develop chunking skills, so that they can read and interpret legal documents or legislation more easily. On a more advanced level, it is also important for student to gain hands-on user knowledge of the typical framework structures that sustain particular legal genres, particularly different types of contract clause. For example, students can be given model clauses or templates that have to be completed using specific information, or clauses that have to be corrected or adapted to new situations.

Looking ahead, we need further research based on larger and wider corpora in order to examine how the type of formulaic language identified here behaves in other legal genres and domains. Moreover, since our current knowledge is based mainly on written evidence, it would be stimulating to examine how formulaic language operates in multimodal terms, looking at spoken legal language across a range of contexts. This would enable us to overcome the distortions imposed by the exclusive focus on textual evidence, and develop a more ecologically valid understanding of legal language as a spoken system with pragmatic, interactional and even performative dimensions. We also need to work on how to exploit corpora in the classroom, with consideration of how the kind of information provided by corpora can be accessed and used productively by students themselves. New technological affordances, such as searchable learner-friendly corpora or multimodal corpora, are currently opening up exciting new possibilities for research and teaching in this area.

References

- Alcaraz, Enrique & Hughes, Brian (2002). *Legal translation explained*. Manchester: St Jerome Publishing.
- Biber, Douglas (2006). *University language: a corpus-based study*. Amsterdam: John Benjamins.
- Biber, Douglas & Conrad, Susan (1999). Lexical bundles in conversation and academic prose. In Haselgård & Oksefell (Eds.), *Out of Corpora: Studies in Honour of Stig Johansson* (pp. 181–189). Amsterdam: Rodopi.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Breeze, Ruth (2013). Lexical bundles in four legal genres. *International Journal of Corpus Linguistics*, 18(2), 229–253. DOI: [10.1075/ijcl.18.2.03bre](https://doi.org/10.1075/ijcl.18.2.03bre).
- Breeze, Ruth (2014). The discursive construction of professional relationships through the legal letter of advice. In Breeze et al. (Eds.), *Interpersonality in legal genres* (pp. 281–302). Bern: Peter Lang.
- Breeze, Ruth (2015). Teaching the vocabulary of legal documents: a corpus-driven approach. *ESP Today* 3(1), 44–63. Available at www.esptodayjournal.org/esp_today_back_issues_vol4.html.
- Brown, Gillian & Rice, Sally (2007). *Professional English in Use: Law*. Cambridge: Cambridge University Press.
- Csomay, Eniko (2004). Linguistic variation within university classroom talk: A corpus-based perspective. *Linguistics and Education*, 15(3), 243–274. DOI: [10.1016/j.linged.2005.03.001](https://doi.org/10.1016/j.linged.2005.03.001).
- Garner, Brian (2001). *Legal writing in plain English* (2nd ed.). Chicago: University of Chicago Press.
- Hyland, Ken (2008). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. DOI: [10.1016/j.esp.2007.06.001](https://doi.org/10.1016/j.esp.2007.06.001).
- Krois-Lindner, Amy (2006). *International Legal English*. Cambridge: Cambridge University Press.
- Nattinger, James & DeCarrico, Jeanette (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nesi, Hilary & Basturkmen, Helen (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11(3), 283–304. DOI: [10.1075/bct.17.03nes](https://doi.org/10.1075/bct.17.03nes).
- Reinhart, Susan (2007). *Strategies for legal case reading and vocabulary development*. Ann Arbor: University of Michigan Press.
- Trosborg, Anna (1997). Text Typology: Register, Genre and Text Type. In Trosborg (Ed.), *Text Typology and Translation* (pp. 3–23). Amsterdam/Philadelphia: Benjamins. DOI: [10.1075/btl.26.03tro](https://doi.org/10.1075/btl.26.03tro).
- Vázquez Orta, Ignacio (2010). A genre-based view of judgments of appellate courts in the common law system. In Gotti & Williams (Eds.), *Legal discourse across languages and cultures* (pp. 263–284). Bern: Peter Lang.
- Williams, Christopher (2011). Legal English and Plain language: an update. *ESP Across Cultures*, 8, 139–151.
- Williams, Christopher (2013). Changes in the verb phrase in legislative language in English. In Aarts et al. (Eds.), *The verb phrase in English: Investigating recent language change with corpora* (pp. 353–371). Cambridge: Cambridge University Press. DOI: [10.1017/CBO9781139060998.015](https://doi.org/10.1017/CBO9781139060998.015).
- Wray, Alison (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics*, 21(4), 463–489. DOI: [10.1093/applin/21.4.463](https://doi.org/10.1093/applin/21.4.463).

Note: JLL and its contents are Open Access publications under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).



Copyright remains with the authors. You are free to share and adapt for any purpose if you give appropriate credit, include a link to the license, and indicate if changes were made.

Publishing Open Access is free, supports a greater global exchange of knowledge and improves your visibility.