A dictionary of legal English collocations as an aid for mastering the legal English genre

Katarzyna Mroczyńska
Siedlce University of Natural Sciences and Humanities, Poland

Abstract

The article begins with a brief overview of collocations and their features as being central to legal language and, as such, worth studying, especially in view of the fact that legal language studies tend to be mostly interested in terminology rather than phraseology. To bridge this gap, the article offers a tool for legal English learners, i.e. the dictionary of legal English collocations based on judgments of the UK Supreme Court. Our dictionary project is aimed at analysing the corpus we created, using Sketch Engine software, a cutting-edge lexicographic tool which enables the uploading and exploration of users' own corpora. The project will focus on analysing bipartite legal English collocations appearing in the corpus. The next stage of the project will be the preparation of the final product of our research, i.e. a dictionary of legal English collocations. We believe that such a dictionary will prove a useful aid for mastering the conceptual structure of legal English.

Keywords: collocation, dictionary compiling, legal English

1. Introduction

In this article, we intend to present our ongoing project which involves compiling a dictionary of legal English collocations. This article opens with a brief summary of research findings regarding collocations, where we advocate the need for mastering collocations in order to become a proficient user of legal English. Based our observations as academic teachers working with students of Business English, including elements of legal English, we notice that students often lack sufficient exposure to phraseology which forms part and parcel of the language and consequently, they frequently struggle to find appropriate collocations in a given context. In an attempt to remedy this situation at least partly, we offer a tool for legal English learners, i.e. a dictionary of legal collocations based on judgements of the UK Supreme Court published in the period from 2009 to 2018. In the functional classification of legal texts, judicial texts belong to the group of hybrid texts, i.e. they are primarily descriptive but also prescriptive, as opposed to the other two types in the classification offered by Šarčević (1997: 11), which are purely descriptive or purely prescriptive. We believe that the Supreme Court judgements represent high-quality legal documents and, as such, may be treated as a good point of departure for a

reliable analysis of the legal English genre and may equally serve as a rich source of textual material to be used in teaching/learning. Our dictionary project[1] is aimed at analysing the corpus we created using Sketch Engine software, a cutting-edge lexicographic tool which enables uploading and exploring users' own corpora. Section 4 contains a more detailed description of functions the software offers and of the statistical measures it uses.

As legal language studies appear to concentrate mostly on terminology rather than phraseology, we would like to try and bridge this gap. To this end, the project will focus on analysing bipartite legal English collocations appearing in the corpus with the dictionary of legal English collocations, which we hope will have some pedagogic applications, to follow as the final stage of the project. Given that professional discourses such as law include a large number of formulaic expressions, multiword units and collocations, the lexical items to be presented in the dictionary will be worth integrating into a teaching curriculum so that students are exposed to these patterns, explicitly to enhance their learning outcomes.

2. Collocation explained – definitions and competing approaches

Since terminology makes an attempt at analysing, recording and describing the concepts of a specialised language, it may be of great help in providing a better and more thorough understanding of legal concepts (Bajcić 2017: 7). We tend to believe that the shift in trends regarding the development of terminology may prove useful in the field of law, which is of key interest to us. The field of law relies on the language to express legal concepts so as to produce a desired effect. However, it should be noted that the discipline of law is based on concepts which frame the professional knowledge in this area rather than words. Moreover, legal concepts need to be analysed within their extralinguistic context, as this context is of paramount importance, providing details necessary to come up with a clear and reliable interpretation of legal issues.

As Biel (2014: 42) claims, the conceptual structure and the systemic nature of legal terminology affects legal phraseology. Therefore, phraseology, and specifically collocations, act as a tool to express complex interrelations between legal concepts. Moreover, it is also capable of entering into relations with other terms. Thus, we decided to put collocations in the centre of our lexicographic project devoted to the legal English genre.

In the literature dealing with lexicological theory and lexicographic practice, we may encounter numerous definitions of the term *collocation*. To make things more complex, it seems that there is no universally accepted formal definition of collocations, not even a proposal for the definition (Mel'čuk 1998: 23).

The definitions will vary depending on the approach they have been based on. Among the three most popular approaches applied by linguists researching collocations are a frequency-

---

[1] Devising a dictionary of legal English is a project run at the Siedlce University of Natural Sciences and Humanities by Katarzyna Mroczyńska, the author of this article, and Tomasz Michta, PhD. The researchers are Assistant Professors at the Department of English Studies and Translation Studies of the Institute of Language and Literary Studies.

based approach, a semantic-oriented approach and, last but not least, a relatively new, pragmatically-driven view (Siepmann 2005: 410).

The first approach, advocated, among others, by Mel'čuk (1998), Gonzalez-Ray (2002) and Hausmann (1997), assumes the existence of a particular meaning relationship between the constituents of a collocation. As an advocate of this approach, Mel'čuk (1998: 23–24) argues that collocations form a subclass of *set phrases*, or *phrasemes*, understood as lexical units characterised by their non-compositionality; phrasemes need to be treated, stored and used as a whole. Moreover, Mel'čuk (1998: 26), in his formalist proposal, introduces two ancillary concepts for distinguishing phrasemes from free phrases, i.e. the concept of unrestrictedness ("unlimited freedom of choice among equivalent independent meanings and expressions") and of regularity ("observance of general rules in combination of meanings and expressions and is related to the concept of combination rules of language"). Accordingly, the term *phraseme* will refer to linguistic signs including a signified and a signifier which can be constructed neither unrestrictedly nor regularly.

Mel'čuk (1998: 28–29) also devises a formal definition of a collocation. He claims that collocation AB is a phrase whose signified includes the signified of one of its two constituent lexemes, e.g. A, which is freely chosen by the speaker, but the other component, a signified C, is such that the lexeme B expresses C, and it is chosen contingent on A, which means that the signifier of a collocation is not unrestrictedly constructed.

This definition may be expressed in the following formula (Mel'čuk 1998: 28):

AB = ⟨'A⊕C';/A⊕B⟩|'C' is expressed by **B** such that /A⊕B/is not constructed unrestrictedly.

Collocations, or *semi-phrasemes*, which are the focus of our attention, constitute the majority of the entire phraseme inventory. To describe them accurately and systematically, Mel'čuk (1998: 29–30) uses the Meaning-Text theory and the comprehensive concept of Lexical Functions, whose detailed description is beyond the scope of this study.

In the frequency-based view (represented by Sinclair 2004, or Kjellmer 1994, for example), on the other hand, we will examine collocations understood as a statistically significant co-occurrence of at least two or more words. We can measure the strength of a given collocation, or a syntagmatic attraction, by means of specialised tools and statistical analysis of corpus data. These statistical calculations reveal which word pairs yield significantly more co-occurrences than what would be expected by chance, taking into account the words' total frequencies in the gathered corpus material. In this manner, we are capable of establishing the most significant collocates of any given word in the language genre that the collected data represents (Sinclair 1966: 418; Hoey 1991: 6–7).

According to Lehecka, the collocation strength, i.e. the attraction between a node (W1) and its collocate (W2), is calculated based on four observed absolute frequencies in the data:

    (i)   the total number of word tokens in the corpus,
    (ii)   the number of tokens of W1 in the corpus,
    (iii) the number of tokens of W2 in the corpus, and

(iv)   the number of tokens where W1 and W2 co-occur within a specified distance (in the collocation window) from each other so that they are regarded as co-occurring (Lehecka 2015: 2).

The next step in the analysis is the comparison of the observed number of co-occurrences in the corpus and the expected number of co-occurrences, i.e. the number expected by chance given (i), (ii) and (iii). If the observed number of co-occurrences of the node and the collocate is larger than what can be ascribed to chance, then W2 is a statistically significant collocate of W1. Individual words in a language have very different frequencies. That is why the collocation strength between different word pairs needs to be done by using a statistical association measure which takes into consideration the uneven distribution of words in the data and not absolute frequencies alone, which would be insufficient for the needs of reliable collocation analysis. Nowadays, statistical tools offer over 50 different association measures (Evert 2009: 1,243), with the z-score, the t-score, MI (Mutual Information), the log likelihood ratio and Fisher's exact test being most frequently applied in the collocation analysis. Undoubtedly, the association measure a researcher chooses has a great impact on the results of the analysis, and as such, this choice certainly requires careful consideration. A thorough discussion of the advantages and the disadvantages of different associations can be found in Evert (2005); or a more concise one in Wiechmann (2008) and Evert (2004) and (2009) (Lehecka 2015).

The third approach, a younger relative of the two presented above, makes an attempt at explaining structural irregularities and non-compositionality underlying phrasemes and collocations by referring to pragmatic regularities which provide a link between the situational context and the linguistic form, which brings us back to the cognitively-based contextualisation theory offered by Fillmore (1976) (Siepmann 2005: 410–411).

All three approaches are not as divergent as they may appear at first glance, and some scholars seem to have found some common ground to build a unified view of collocations, at the same time indicating certain drawbacks of each of the individual approaches. The proposition devised by Siepmann (2005 and 2006) may be worth a more detailed presentation, as it offers a new, broader definition of *collocation* and a revised classification of the collocational spectrum.

Siepmann points out major drawbacks linked to each of the three presented approaches, i.e. the frequency-oriented approach provides us with just the raw data but does not reveal much about how this material was formed or how it should be structured; despite being appealingly intuitive, the semantically-driven view of collocations does not offer a holistic view and will only be fragmentary; and finally, the pragmatic approach, in its pure version, will fail to account for the collocations of semantic features. Therefore, he advocates an extension of the semantically-based approach "that will take account of strings of regular syntactic composition which form a sense unit with a relatively stable meaning" (Siepmann 2005: 411) based on a rigorously carried out study of the linguistic corpus. Incorporating the findings of the cognitivist camp, he also postulates loosening the definition of collocation so that it encompasses the concepts of usage norm and statistical significance, on the one hand, and the holisticity of the collocational unit, on the other hand. Collocations may be considered as self-

contained "wholes", given the fact that native speakers can ascribe meaning to general language collocations even if they lack their context (Siepmannn 2005: 438).

Siepmann (2005: 431) offers the following typology of collocation phenomena, which may cover the entire range of morpho-syntactic constructions:

a) colligation which concerns both grammatical preferences and those of longer syntagms, e.g. *far be it from me to* + INF;
b) collocation between lexemes and phrasemes, e.g. *in the end*;
c) collocation between lexemes and semantic pragmatic feature, e.g. [expectation] = *duly*;
d) collocation between semantic pragmatic feature, or long-distance collocation.

It is worth noting that instead of being based on colligational relations, some collocations are just based on semantic relations.

Although researchers seem to find it difficult to adopt a uniform, widely-accepted definition of collocation encompassing all the linguistic features of the phenomenon, a list of criteria of collocability, which will allow us to classify a given phraseological unit as collocation, can be compiled that will be basically the same across various research on collocation. These criteria include:

a) frequency of co-occurrence;
b) combinatory restrictions;
c) degree of compositionality;
d) degree of transparency;
e) span of words between node and collocate, or collocational window (Patiño 2014: 122–124).

When we adopt such a broad definition of collocation as the one offered by Siepmann, we notice that collocations are extremely widespread in the language, and as such, they seem to dominate the language use. What is more, being structurally complex units, collocations in this sense, and not just individual words, may be treated as linguistic signs on their own (Siepmann 2005: 438).

As collocations seem to dominate the language use, both general and specialist, we decided to undertake a lexicographic project aimed at analysing genre-specific collocations appearing in legal English with the assumption that they will reflect the conceptual structure of the language of the law. We were encouraged by the fact that most legal language research focuses on terminology rather than phraseology (which is relatively unknown ground). Specialist dictionaries and term bases include terms, but they do not customarily contain the collocational relations of these terms, and even if they do, collocations tend to appear in the exemplificatory material, thus making it more difficult for the user to identify them. Indeed, as dictionaries of LSP collocations are almost non-existent, we would like to bridge this gap and compile a dictionary of legal English collocations, which is the planned outcome of our research project. Although Siepmann (2006: 34) argues quite convincingly that a bilingual or multilingual dictionary has obvious advantages over a monolingual approach, its compilation inevitably is more time consuming and, in the case of a culture-bound subject field such as law, entails solving numerous problems of equivalence that stem from the asymmetry between different legal systems, a problem not to be downplayed by any lexicographer. Therefore, we decided to

compile a monolingual dictionary of collocations, hoping that in doing so we will be able to reach a wider audience (an English-Polish dictionary of legal collocations is more likely to appeal only to speakers of both English and Polish) and that our dictionary will be followed by bilingual dictionaries of legal collocations.

It is also worth mentioning that the concept of collocation does not only refer to textual statistics, but it reflects a mental representation of the lexicon, as collocations are formed by the cognitive process of priming. Hoey distinguishes three elementary types of priming: collocation, colligation and semantic preference/association, with the priming of lexical items with collocations in this psychological sense being the foundation of language structure in general (Hoey 2005: 8–9). In light of these findings, we may assume that knowing how words collocate is part of what it means to know a language or a genre of a language. Consequently, a dictionary of collocations may serve as an aid in mastering the conceptual structure of an analysed language genre.

3. Assumptions for the preparation of a dictionary of legal English collocations and applied tools

Dictionaries have accompanied people for a relatively long time, and they have served as an aid to understand a given domain of knowledge. This assistance may be particularly needed in relation to the legal domain, which affects our daily activities throughout our whole lives. Although intended to be accessible, legal regulations frequently turn out not to be understandable to non-legal professionals. It seems that this is not the language of the law as such, but the law itself that is complex (Bajčić 2017: 140). What is more, concepts used in the domain of law may lack clear-cut boundaries, which is seemingly at odds with the need for precision expected on the part of legal text users. Legislators strive for drafting precise, unambiguous rules, which, at the same time, will be able to accommodate the meaning in a way general enough to be applicable to as many situations as possible; combining the necessary precisions and the need for generalisation is one of the challenges law faces (Wagner and Gémar 2013: 179).

When it comes to the purpose of a legal dictionary, we are inclined to adopt the view offered by Bajčić, who asserts that "the main purpose of a legal dictionary is to enable users to learn about legal concepts in order to understand the law" (2017: 138).

Bearing in mind the fact that collocations reflect a language's conceptual structure, and the ability to use collocations in a correct and natural manner represents the user's mastering of the language within a given specific genre, we believe that a dictionary of English legal collocations may contribute significantly to the improvement of knowledge of the language and also of the workings of the law as such. Given that the framing of ideas in a foreign language depends on the user's linguistic competency and that we (as users) communicate by means of semi-prefabricated lexico-grammatical units rather than individual words, the learners or non-native users of a given genre need to learn a set of lexical items. Learning collocations seem to play a key role in the language acquisition process, and the set of collocations which needs to be learnt

for a person to become a competent user may be determined using the criteria of frequency, availability and learnability.[2]

Thus, the aim behind our project is to draw up a comprehensive frequency list of one hundred key legal terms, specifically nouns, and analyse their collocational behaviour. We assume that this will lead to the identification of terms and collocations that have to be mastered to be able to perform at the near-native (or lower) proficiency level.

To ensure the most reliable and up-to-date input representing examples of real-life application of the legal language, we have decided to devise our collocations dictionary on the corpus of judgements made by the UK Supreme Court, which is the final court of appeal for civil, as well as criminal, cases in England, Wales and Northern Ireland and therefore plays an important role in the development of the country's legal regulations. Moreover, the judgements are public, so their wording is published and made available on the Court's website (supremecourt.uk). To create a corpus which, as we believe, will yield reliable results and improve the depth of coverage, we decided to cover the judgements during the ten-year period from the Court's inception in 2009 to 2018. However, we are aware that the corpus material collated in this manner and only comprising the last ten years is narrowly synchronic and may display some deficiencies, e.g. it may not reflect the knowledge and experience of language users accumulated over several generations. Consequently, it is worth noting that a collocation may still exist despite the fact that it is absent from a corpus. Moreover, we are aware of the fact that that a corpus based on one source may not suffice to reflect fully the richness and complexity of the legal genre. Therefore, we see an opportunity to embark on a new more ambitious project which would rely on a fully-fledged corpus based on different types of legal texts and would provide a broader spectrum of legal phraseology.

According to Nielsen (1994: 33), while compiling a dictionary which would cover LSP communication it is of key importance to establish the target group of users and identify their lexicographic needs. To perform this task, the researchers working on the projected dictionary, who are also academic teachers specializing in LSP courses, carried out observations during their work and they noticed that, otherwise proficient language users, students often lack linguistic competence and confidence in the area of phrasemes or collocations. The researchers analysed the information regarding the use of existing LSP and general dictionaries gathered while running LSP and translation courses and it appeared that students were often unable to find a collocation they were looking for in the available dictionary resources. This gap in the resources available to students that may be bridged by the projected dictionary of legal English collocations. The dictionary is a monolingual work based on a limited set of texts and it will be used a pilot project that, if successful, may be extended into a bilingual dictionary based on a wider range of legal texts in the future.

The projected dictionary is intended to be a single-field (legal) dictionary for students of business and legal English, or LSP learner's dictionary. As it is designed to be used by students, it will contain a smaller lemma stock than LSP dictionaries for professional users.[3]

---

[2] See Siepmann (2006: 4–9).
[3] See Nielsen (1994: 41).

Undoubtedly, collocations deserve exhaustive treatment in LSP dictionaries but this need is often difficult to meet in general field dictionaries or multi-field dictionaries for reasons of limited space in the publication. Therefore, we see some room for the projected dictionary, focusing solely on the area of legal English collocations, which makes an attempt at remedying such drawbacks.

While compiling a dictionary, it is also crucial that the genuine function of a dictionary is clearly determined as it is the best criterion for (not) including particular types of information in the dictionary. According to Nielsen (1994: 44), dictionary functions may refer to text translation, composition or comprehension. As the projected dictionary is intended to provide assistance in producing texts (or utterance), its main function will be composition. To make foreign language texts or utterances, the user of a composition dictionary will need examples of language use and such real-life examples culled from the corpus accompany dictionary entries.

The determination of the target user group and the main dictionary function should be reflected in the publication macro- and microstructure. As for the former, the dictionary will have a complex macrostructure as it will contain a preface, user's guide, word list, and a cross-reference list. We decided to include these components in the dictionary as we see them as clearly and directly related to each other and adding necessary information to the scope and function of the dictionary in question. Our intention is to produce a dictionary whose all individual components will constitute a simple and coherent whole: a dictionary of legal English collocations.

We are going to follow the established rules and our dictionary will include the front matter, the word list and the back matter. The front matter will contain among other things the list of contents and the preface which gives authors' explanatory remarks concerning the dictionary function, scope and application as well the corpus on which the dictionary is based. What will also be included is a user's guide which will explain all the functional elements contained in the dictionary so that users do not have to guess themselves the meaning of various abbreviations of functional elements, for example.

As for the word list, we applied a frequency method to compile this list. The word list of the projected dictionary contains around 100 alphabetically arranged entries being legal English nouns with the highest frequency in the analysed corpus. The nouns are selected based on the frequency list generated by the statistical tools offered in SketchEngine software. It is worth noting that with this method of dictionary compilation based on a computerised corpus it is easier to find illustrative examples of the LSP terms usage. Needless to say, such examples are extremely valuable as a pedagogic aid both for students and teachers dealing with the legal English genre.

As for the back matter, the projected dictionary of legal English collocations will contain a cross-reference list giving easy access to all occurrences of a given noun in the dictionary, which as we believe is relevant to the use, scope and purpose of the dictionary.

The lemma list of a legal dictionary may be expected to feature solely specialist terms or words which have at least one specialised meaning within the field of law. Consequently, lexicographers apply the so-called pruning process where general language terms are excluded unless they have a special legal meaning differing from the normal meaning of the analysed terms (Nielsen 1994: 154).

After making some preliminary provision for the macrostructure and the word list for the projected dictionary, we also need to consider its microstructure. For the purposes of this article, we adopt the definition offered by Hausmann and Wiegand which says that "the microstructure of a dictionary article is the total set of linearly ordered information items following the lemma" (1989: 340). One of the conclusions we may draw based on this definition is that the microstructure is an ordering structure, its main function being an optimal arrangement of the article-internal information items. This optimal manner of organisation is important to a target user. Therefore, we intend to implement a microstructure which presents information in an easy to understand way and results in a user-friendly article, coherent and easy to read. The type of microstructure used in a dictionary depends on the nature of the dictionary itself, its purpose, scope, application and also its target group (Nielsen 1994: 223).

For the microstructure of the projected dictionary of legal English collocations, the following microstructure was adopted: a lemma form realisation item, i.e. a spelling form, collocations arranged by type, i.e. premodifier + TERM, TERM + noun, verb + TERM, TERM + verb, preposition + TERM, TERM + preposition, with each collocation section followed by a corpus-based example. Polysemous lemmata will have separate article sections devoted to various meanings each and following the adopted microstructure for a dictionary article.

Having established the target user group and the key function of the projected dictionary as well as having presented its macro- and microstructure, we will go on to discussing the analysis of the gathered corpus material and tools used to carry out this task. In order to facilitate the analysis of the linguistic material, Sketch Engine software,[4] a leading corpus tool widely used in lexicography, was applied. The name of the software comes from one of its functions, i.e. the wordsketch, which is a concise, yet extremely comprehensive, summary of a word's grammatical and collocational behaviour presented in a transparent one-page figure. Actually, the word sketch can be regarded as a draft dictionary entry, as the system has already found, analysed and organised all the recurring patterns for the word so that they are at the lexicographer's fingertips to be edited, reorganised and finally published (Kilgariff *et al.* 2014: 8–10).

Apart from the iconic word sketch mentioned above, the software also includes the following functions (Kilgarriff *et al.* 2014):

1. Concordance – a basic tool for any researcher working with a corpus, as it is capable of searching a corpus for a word form, a lemma, a phrase, a part of a speech tag, etc., going back to the raw data underlying any analysis. The system converts all queries into Corpus Query Language (CQL), which can be used directly.
2. Word List – the software may generate frequency lists of words, lemmas, n-grams or key words.
3. Keywords and Terms – this function enables extraction of core lexis in a corpus using a "keyness score".
4. Collocations – the tool calculates words that are statistically associated with the query term. The system uses several statistically relevant measures to find the best collocation candidates, such as T-score, MI, log likelihood, logDice, etc.

---

[4] See http://www.sketchengine.eu.

6. Word Sketch Difference – this function offers a comparison of two words based on collocations.
7. Thesaurus – creates a distributional thesaurus based on common collocation. The resulting list of words includes items in various semantic relationships.
8. Trends – this function is useful when conducting a diachronic analysis of word usage.
9. WebBootCaT – it is a set of programs that may be used to compile a user's own web corpus.

One of the advantages of the tool is the fact that apart from providing pre-loaded corpora of its own, the software also allows the user to upload, build and explore their own corpus using the WebBootCaT tool. What is more, when it is applied to seed words from a specific domain, this corpus-building procedure seems to be an extremely efficient method of discovering the terminology and phraseology of a specialist domain. Additionally, the software allows a researcher to accurately tokenise, lemmatise and tag the corpus by specific parts of speech. Tools for these processes are available for selected languages, including English (Kilgarriff *et al.* 2014).

For the purpose of analysing the material, we gathered and uploaded corpora comprising the judgments made by the UK Supreme Court in order to create our own corpus for the project. The next step was to use the Word List function to generate a frequency list, which was then used to select one hundred most frequently appearing specialist terms to be included in our legal English dictionary of collocations.

When it comes to establishing collocations, this type of software needs to use some statistical association measures that indicate the strength of association between two words. Such measures are based on various aspects of the co-occurrence of the two words in question and may be used to identify good collocation candidates, and it is basically up to the software developers as to which measures it offers its users. Until 2006, Sketch Engine used a MI-Score modified and AScore to give greater weight to the frequency of collocations. However, given the findings of more recent research, since 2006, it changed the statistic to logDice based on the Dice coefficient. In 2015, another adjustment was then made to the statistical methods applied in the software; the indices were modified again to compute the score more correctly. Currently, Sketch Engine applies a measure termed *logDice general word sketch score.* The score is computed for all possible word pairs, and the word pairs with the highest score are presented as collocation candidates.[5]

Having selected the terms, we analysed the collocational behaviour of each term individually using Word Sketch and its association measures. We have decided that each dictionary entry will include six types of collocations, i.e.

1) premodifier + TERM,
2) TERM + noun,
3) verb + TERM,
4) TERM + verb,
5) preposition + TERM,

---

[5]   For detailed discussion, see https://www.sketchengine.eu/documentation/statistics-used-in-sketch-engine/ and Rychlý (2008).

6) TERM + preposition,

and each section in every entry is arranged in this manner; whenever no significant collocations are found in a given section, then the section is omitted. Moreover, each section starts with an alphabetical list of words which form a given type of collocation with the term. Another component of an entry is a sample sentence extracted from the corpus illustrating the use of a given collocation type, and the collocation itself is in bold so that it can be easily found by the dictionary user.

A sample single section for the term 'court' is presented below:

premodifier + COURT

administrative, Appeal, appellate, appropriate, circuit, civil, commercial, County, Crown, divisional, domestic, European, federal, foreign, High, international, Lower, municipal, national, open, referring, sentencing, Supreme

The New York Times reports that "details in the article contradicted physical evidence already presented in **open court**."

It is worth noting that some words may collocate with several terms, and consequently, they will appear in the dictionary more than once. The dictionary will present the most frequent specialist terms together with their collocates in alphabetical order. We decided to set a minimum frequency threshold of five occurrences for a given collocate to be included in the main entry.

When it comes to some basic statistical data concerning the corpus created for this project, the corpus contains 9,505,800 tokens (the smallest units the corpus divides into), 8,098,719 words and 260,419 sentences culled from 636 documents representing UK Supreme Court Judgements. We hope that a corpus of such a size, created with high-quality documents from a reputable institution, will enable a reliable analysis of collocations appearing in legal English.

4. Conclusions

Despite some voices being raised as to the rationale behind a further increase in specialist dictionaries, there still seems to be a place for dictionary-like publications (although not necessarily in the form of a traditional book), as they meet a perennial human need, which is to learn a language so as to understand the world (L'Homme and Cormier 2014: 8). Thus, compiling a dictionary which will offer some insight into the concepts around which the legal system is organised may help the reader grasp the meaning of legal rules. In this article, we discussed the significance of collocations for understanding the intricacies of a specialist language and the assumptions and tools for preparation of a legal English dictionary of collocations. We believe that it is appropriate to perform further studies and that such a dictionary may serve as an aid for mastering the legal English genre. The final results, i.e. a dictionary of legal English collocations, will be presented and discussed in a separate article.

References

Bajčić, M. 2017. *New insights into semantics of legal concepts and the legal dictionary*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Biel, Ł. 2014. *Lost in the Eurofog: The textual fit of translated law*. Frankfurt: Peter Lang.

Evert, S. 2004. *Association measures*. www.collocations.de/AM/ [17/09/2019].

Evert, S. 2005. *The statistics of word cooccurrences: Word pairs and collocations*. Institutfür Maschinelle Sprachverarbeitung, Universität Stuttgart.

Evert, S. 2009. Corpora and collocations. In A. Lüdeling, and M. Kytö (eds.), *Corpus linguistics: An international handbook*, Vol. 2, 1212–1248. Berlin, New York: Mouton de Gruyter.

Fillmore, Ch. 1976. "Pragmatics and the description of discourse". In S. J. Schmidt (ed.), *Pragmatik/Pragmatics II. Grundlegungeinerexpliziten Pragmatic*, 83–104. Munich: Fink.

Gonzalez-Ray, I. 2002. *La phraséologie du français*. Toulouse: Presses Universitaires du Mirail.

Hausmann, F. J., and H. E. Wiegand. 1989. Components part and structures of general monolingual dictionaries: A survey. In F. J. Hausmann, O. Reichmann, H. E. Wiegand, and L. Zgusta (eds.) *Wörtebücher. Ein internationals Hanbuch zur Lexikographie*, vol. 1, 328–360. Berlin/New York: de Gruyter.

Hausmann, F. J. 1997. Tout est idiomatique dans les langues. In M. Martins-Baltar (ed.), *La locution entre langue et usage*s, 227–290. Fontanay/Saint-Claud: ENS Éditions.

Hoey, M. 1991. *Patterns of lexis in text*. Oxford: Oxford University Press.

Kilgarriff A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, and V. Suchomel. 2014. The sketch engine: Ten years on. *Lexicography*, 1: 7–36.

Kjellmer, G. 1994. *A dictionary of English collocations*. Oxford: Clarendon Press.

Lehecka, T. 2015. Collocation and colligation. In J.-O. Östman, and J. Verschueren (eds.), *Handbook of pragmatics*, 1–20. John Benjamins Publishing.

L'Homme, M., and M. C. Cormier. 2014. Dictionaries and the digital revolution: A focus on users and lexical databases. *International Journal of Lexicography* 2014/ doi:10.1093/ijl/ecu023.

Mel'čuk, I. 1998. Collocations and lexical functions. In E. P. Cowie (ed.), *Phraseology, theory, analysis and application*, 23–53. Oxford: Clarendon Press.

Nielsen, S. 1994. *The bilingual LSP dictionary. Principles and practice for legal language*. Tübingen: Gunter Narr Verlag.

Patiño, P. 2014. Towards a definition of specialized collocation. In G. Quiroz Herrera, and P. Patiño Garcia (eds.), *LSP in Colombia. Advances and challenges*, 119–133. Brno: Peter Lang Publishing.

Rychlý, P. 2008. A lexicographer-friendly association score. *Proceedings of 2nd Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN*, 2: 6–9.

Šarčević, S. 1997. *New approach to legal translation*. The Hague/London/Boston: Kluwer Law International.

Siepmann, D. 2005. Collocation, colligation and encoding dictionaries. Part I: Lexicological aspects. *International Journal of Lexicography* 18(4): 409–443.

Siepmann, D. 2006. Collocation, colligation and encoding dictionaries. Part II: Lexicographic aspects. *International Journal of Lexicography* 19(1): 1–39.

Sinclair, J. 1966. Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds.), *In memory of J. R. Firth*, 410–430. London: Longman.

Sinclair, J. 2004. *Trust the text. Language, corpus and discourse*. New York/London: Routledge.

Wagner, A., and J.-C. Gémar. 2013. Materializing notions, concepts and language into another linguistic framework. *International Journal for the Semiotics of Law* 26: 731–745.

Wiechmann, D. 2008. On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory* 4(2): 253–290.

Online sources
supremecourt.uk
www.sketchengine.eu