

The Use of Software for the Analysis of Lexical Properties of Legal Discourse

Daiva Macko

crossref <http://dx.doi.org/10.5755/j01.sal.0.20.1190>

Abstract. The use of computational tools in linguistic research is at the core of corpus linguistics. Currently, specialised lexical software contains elaborate statistical measures that enable a detailed quantitative analysis of corpus data. This paper analyses typical collocations frequently used in the appellate judgments of the European Court of Justice (ECJ). Right verbal collocates of *Court* are analysed in terms of frequency, statistical significance and characteristic semantic patterns. The *WordSmith Tools program*, Version 5.0 was used to measure the frequency and significance of the collocations; specialised computational tools were also used to compare the use of selected collocations with the use of corresponding collocations in the British National Corpus, which was used as the source of general English. The research results show that typical collocations used in the appellate judgments of the ECJ differ from the general English language in terms of frequency and statistical significance and exhibit unique semantic characteristics, therefore suggesting that there are considerable lexical differences between legal and general English that should be taken into account in teaching and learning.

Keywords: *collocation, corpus, frequency, negative stance, statistical significance.*

Aim

The present research aims to illustrate the use of lexical software for the analysis of lexical properties of legal discourse. It is supposed here that the most frequent uses are more likely to be characteristic of the language variety analysed and therefore information on the frequency and statistical significance of lexical items in the specific genre is of great value in characterising the specificity of the discourse. The differences in the use of selected collocations between the general and specialised English also imply that specific collocational competence should be involved in teaching and learning specialised English in general and legal English in particular.

Previous Research

The use of software for linguistic research is of great value. Firstly, it provides a linguist with a novel type of data. For instance, wordlists and concordances are products generated from the corpora by the use of specialised software. These products are available due to computer technologies exclusively and are therefore unique. In addition, as the capacity of computers grows, it is possible to store increasingly larger amounts of data. The specialised linguistic software allows generating frequency lists from large corpora within a few minutes, which would otherwise be hardly feasible at all. As Biber et al. (2004, p.21–22) note, not only are such data more precise and complete, but they are also more representative of the language variety under investigation. In addition, research in the field of collocational studies has shown that the use of computational tools provides data that are not accessible by intuition, suggesting that the users of language are to some extent unaware of their own collocational competence and the patterns that they produce (Widdowson, 2000, p.6). This objectivity distinguishes corpus linguistics as a valuable quantitative method.

Sinclair has often emphasized the importance of objective observance of language in use in order to find *evidence*, or

facts about language and its regularities (for example, see Sinclair, 1991, p.39). The empirical nature of this methodological approach relies on elaborate quantitative analysis of the corpus data. As McEnery & Wilson (2001, p.77) point out, proper and valid sampling and significance techniques provide not only a precise information on the frequency of certain linguistic phenomena, but they also enable comparisons between different corpora.

The earliest empirical research into collocation involving the use of computers was the OSTI report conducted by Sinclair et al. Since then, many modern statistical tools have been developed to analyse collocations. However, the use of statistical measures needs to be balanced with qualitative analysis (McEnery & Wilson, 2001, p.76–77). Probably the main benefit of applying computational techniques in corpus work is proving that the use of language includes recurrent prefabricated constructions (Kennedy, 1998, p.270). Yet, as Sinclair (2004) has summed up,

the point is that nobody believes that language occurs by chance. <...> Statistics, however, only tells us that co-occurrence of two (or more) items is probably not accidental.

It is generally agreed that the origins of the concept of collocation in linguistics lie in Firth's definition of the phenomenon as *actual words in habitual company* (Firth, 1957, p.14 quoted in Kennedy, 1998, p.108), or "the company words keep" (Firth quoted in Hill, 2000, p.48). In the current research, a statistical approach to collocation is followed rather than a semantically-based approach. A statistically-based concept of collocation relies on the application of computational tools to large corpora and extraction of recurrent patterns of words (Siepmann, 2005, p.410–411). The statistical approach was advocated and developed by Sinclair (Crowther et al, 2002, p.58). The frequency criterion seems to be acceptable to many linguists and thus can be stated to lie at the heart of the statistically-based concept of collocation (see Bartsch, 2004, p.59–60; Otani, 2005, p.5; Hanks, 2008, p.222;

Lewis, 2000, p.127, etc.). For example, Biber et al. (1999, p.988) define collocations as statistical associations of words that often co-occur together. In principle, the statistical approach to collocation implies that the validity of results obtained is directly dependent upon the number of recurrent patterns in a large corpus. In other words, the greater the co-occurrence of certain words in the same corpus, the more likely they will collocate with each other than appear separately. Although it sounds reasonable, this point is criticised by Siepmann (2005, p.411), who notices that it remains unclear at which point frequency becomes significant enough. As a result, it is often considered that there are no clear boundaries to mark the significance of collocates (Otani, 2005, p.5; Kennedy, 1998, p.117).

The predominant semantic properties of collocation stem from its contextual origins and the importance of repetition in a text. It is considered that through constant repetition and repeated co-occurrences textual and intertextual meaning is formed (Siepmann, 2005, p.409; Stubbs, 2001, p.157). It is worth remembering that Firth also undermined the repetition in language as a source of typicality (*typical, recurrent and repeatedly observable*, Firth, 1957, p.35 quoted in Tognini-Bonelli, 2001, p.164). According to Tognini-Bonelli, the very concept of collocation arises from the above mentioned theoretical premises. She also notes that the Firthian theory postulates the priority of lexis over grammar (collocation should be observed first and colligation inferred after, as it is a more abstract feature), which obviously has implications for language teaching by shifting the focus to lexis rather than grammar¹.

As regards the form and content of collocation (i.e. the number and nature of elements that constitute it), collocations composed of the so-called content words are generally referred to as lexical (Wei, 1999, p.8; Lewis, 2000, p.134), distinct from grammatical collocations involving a grammatical structure or containing prepositions. The latter are usually referred to as colligation (see Siepmann, 2005, p.411–419; Sinclair, 2000, p.200; Hanks, 2008, p.222; Hoey and Brook, 2008, p. 294; Hoey, 2000, p.234). The binomial structure of collocation is said to be grounded in the statistical concept of collocation that focuses on the lexical connection between words (Marcinkevičienė, 2010, p.140). However, in use these combinations of words are almost always embedded in certain grammatical structures, thus the number of the constituents of collocation is actually more than two items (Lewis, 2000, p.134).

In the field of collocational studies, corpus linguistics is chosen by many researchers as a methodological guideline. Probably, this choice is largely motivated by the importance of frequency criterion. There are various kinds of corpora distinguished, but the most relevant distinction in this research is that between general and specialized corpora. The former contain texts from different genres and often include spoken and written language, while specialized corpora are designed for specific research and are confined to language used only in particular kinds of texts or

situations (Kennedy, 1998, p.19–20; Paltridge, 2006, p.156-157). It is stated that the corpus size interacts with the reliability of the analysis: smaller corpora are suitable for analysing more frequent items, while rarer features require larger samples (see McEnery & Wilson, 2001, p.80).

As a result of investigation in the field, a fairly novel concept of *collocational competence* has emerged. It is often emphasized as a vital skill for adequate knowledge of language (see, for example, Hill, 2000, p.49; Juknevičienė, 2008, p.119). Consequently, mastery of special languages can also be regarded as largely dependent on collocational competence, since lexis is an important attribute of a genre. Therefore, a lot of research focuses on typical collocations in special languages as opposed to common collocations in general language.

However, it seems that little research has been done on typical collocations in the legal language in general and in the legal language of the EU². As regards the legal English of the Commonwealth countries, Bhatia (1993) and Maley (1994) have discussed its genre-specific traits in detail. Bhatia defines judgments as one of the most conventionally standardized disciplinary genres (together with legislation and case-law) in law (Bhatia, 2000, p.82). Nevertheless, Bhatia notes that legal English in Europe differs considerably from the countries of the common law system (Bhatia, 1993, p.139), while Maley also emphasizes that particularly at the appellate level structural differences of continental and common law systems are evident (Maley, 1994, p.44). The nature of the EU law system is distinct both from continental and common law. Yet, to the best of my knowledge, so far the discursal and linguistic characteristics of the EU legal English have not been systematized yet. Collocational studies in this field could provide valuable insights into its lexical characteristics.

Data and Methods

The research is based on the analysis of a corpus composed of 50 judgments on appeal of the European Court of Justice. The size of the corpus is 528 073 words. The judgments are available on the website <http://eur-lex.europa.eu/> in the Internet. The chosen judgments were retrieved from this database in the following sequence: filtering the data by specifying a file category (case-law); narrowing the selection to the documents issued by the Court of Justice; and filtering files according to the type of procedure — choosing the appeal procedure. As a result of this search, 432 judgments were available at the moment of selection (3 November 2009). The time span of these judgments begins from 1 October 1991; the database is frequently updated. In order to compare the results with the data from a larger, general corpus, the British National Corpus (BNC, 100 million words) was used.

The corpus was composed with the aim to analyse the most recent available data, as it was expected that this material would be the most representative of the current use of the legal English language of the European Union (EU)

¹ For research based on the so-called lexical approach see, e.g., Lewis, 2000.

² Although there is literature available on general lexical and syntactic properties of the language of the law (see Vystrčilová, 2000, Gibbons, 1994; Cao, 2007; Ingels, 2006).

institutions (representativeness is commonly distinguished as one of the key characteristics of a corpus, see Tognini-Bonelli, 2001, p.52-62; McEnery & Wilson, 2001, p.30). Therefore, the most recent judgments were chosen, covering the time span from 21 February 2008 until 10 September 2009. The authorship of the selected texts is attributable to groups of persons rather than a single author, because a judgment is arrived at by a Chamber composed of several judges.

The program WordSmith Tools (WS), Version 5.0 was used to extract collocations and calculate their frequency and statistical significance scores. Computational tools available on the Internet were also applied. The qualitative part of the research was combined with computational analysis and involved manual scrutiny of relevant (i.e. statistically significant and most frequent) collocations. The focus was centred upon the classification of data into semantic patterns.

N Concordance	
24	been annexed to that statement, the Court held that ADM could not properly
25	279 of the judgment under appeal, the Court held that Schneider could not
26	on Adams v Commission, in which the Court held that the expiry of a limitation
27	A. In paragraph 57 of that judgment, the Court held that, in this case, it had not
28	56 In reaching this conclusion, the Court held in particular, in paragraphs
29	the application for annulment 18 The Court held that there was no need to rule
30	nor the exercise of those rights. The Court held that, contrary to the
31	Pak v Commission, paragraph 41, the Court held , first, that prices below
32	by the applicants. 55 Finally, the Court held that the damage incurred by

Fig. 1. A Sample of a Machine Generated Concordance.

Node-word — “the word that appears in the middle of the screen in a list of concordances” (Concordancing glossary). Sinclair refers to a node as “the word that is being studied” or “a central word” in a “machine-generated concordance” (Sinclair, 1991, p.105–116).

In the above given example the node-word is *Court*.

Collocate — “any word that occurs in the specified environment of a node” (Sinclair, 1991, p.115). The word *held* in the example above stands for a collocate. The *Concordancing glossary* provides a common definition based on the frequency criterion: “a word that appears with another word more often than simple chance would suggest”.

Firstly, the WS program was used to generate a wordlist³ in order to find out the most frequent content words in the corpus. It turned out that *Court* was the most frequent lexical item used (5386 occurrences), therefore it was chosen as a node-word. Afterwards, the concordance of the chosen node-word was generated in order to find out its collocates. Following Sinclair’s recommendations (see Sinclair, 1991, p.106), the WS program was set to count collocates within a span of ten words, i.e. five words to the left and five words to the right of the node-word. In total, 710 collocates of the chosen node-word were extracted.

Results and Discussion

The concepts underlying the **quantitative analysis** are the following.

Concordance — “a comprehensive listing of a given item in a corpus (most often a word or a phrase), also showing its immediate context” (McEnery & Wilson, 2001, p.197). Technically, it can also be defined as

a list of all the words, or a certain word, used in a text or in a body of texts, together with a context in which the words appear. This context is usually no more than 7 or 8 words to the left and right of the node word (Concordancing glossary).

The above mentioned context is usually referred to as a *span* (ibid). Sinclair refers to concordance as a “first stage in examination of an item as a node” (Sinclair et al., 2004, p.71). Consider the following example of a machine generated concordance:

The scope of the current research was limited to the right verbal collocates of the node-word analysed (syntactically, the node-word *Court* takes the role of the subject in a sentence; therefore, this word has the majority of its verbal collocates on the right). The focus was on lexical verbal collocates, which excludes the so-called form-words from the scope of the research. Thus such auxiliaries and modals as *is, was, did, should* were ignored, just as the so-called delexicalised verbs (see Juknevičienė, 2008, p.120), e.g. *have, take, make, give* etc., unless they were used in a uniform sense (e.g. *to give a judgment; to make a decision*). With the purpose of restricting the number of analysed instances, only collocates occurring not less than 5 times within the same grammatical pattern were selected.

64 right verbal collocates of the chosen node-word matching the above given criteria were selected. The following list presents ten most frequent right verbal collocates of *Court* in the corpus:

- 1) ***Court held that*** (320);
- 2) ***Court set aside sth.*** (95): *the Court should set aside the judgment;*
- 3) ***Court found that*** (116);
- 4) ***Court stated that*** (112);
- 5) ***Court erred in sth.*** (105): ***Court erred in law;***
- 6) ***Court rejected sth.*** (85): ***Court rejected appellant’s arguments/ a plea/ an action etc.;***

³ A wordlist is a list of words automatically generated in both alphabetical and frequency order (WordSmith Tools Version 5.0).

- 7) **Court dismissed sth.** (61): **Court dismissed** an action/ an application/ an appeal;
- 8) **Court ... dismiss sth.** (48): **Court should dismiss** an appeal;
- 9) **Court infringed sth.** (59): **Court infringed** an Article/ an obligation/ a principle/ a provision etc.;
- 10) **Court failed to do sth.** (58): **Court failed** to explain/ to respond to arguments/ to state adequate reasons for finding etc.

The most frequent collocations of the chosen syntactic pattern were almost always subject to uniform colligational patterns indicated in the above given list. In addition, in some cases the lexical context surrounding the collocations was also rather uniform.

As pointed out by McEnery & Wilson (2001, p.86), raw frequency numbers are not comparable with data from other corpora due to differences in size. In order to compare the frequency of the collocations extracted with the data from the BNC, certain computational tools were applied. The **log-likelihood** calculator was used to compare the relative frequencies between the two corpora. The results of the log-likelihood test are given in Appendix 1. A '+' sign indicates that the frequency of a certain collocation in the corpus outnumbers the frequency of the same collocation in the BNC. The higher the value, the more significant the difference between two frequency scores is and the lower the probability that the statistical difference is accidental.

In comparison with the BNC, the numerical log-likelihood values obtained show that the use of the selected collocations in the corpus of the current research is much more frequent than in the BNC, i.e. the frequency of the collocations selected is significantly higher in my corpus than in the BNC. Consequently, the results suggest that typical collocations extracted serve as generic markers, i.e. they distinguish this legal subgenre from the general language. It is worth noting that the lowest score is higher than 15.13, which means that the chances of unreliability of the calculations amount to only 0.01 per cent.

In addition, the statistical **MI** (mutual information) test was applied to the selected collocations. The importance of this test is reinforced by theoretical doubt that the frequency alone does not necessarily constitute a collocation and the co-occurrence might be accidental. Thus, MI score is a test designed to measure the statistical significance of collocation. It

'compares the probability that the two items occur together as a joint event ... with the probability that they occur individually', i.e. by chance. The higher the MI score, the more significant the collocation is, whereas the values below zero show that words co-occurred by chance (McEnery & Wilson, 2001, p.86).

The MI test was applied to the selected data using the WS program, while the BNC provides an option of displaying MI scores together with the collocations extracted. To assure that results are reliable, some linguists⁴ recommend

to set the cut off point for MI values at 3 and exclude the values below this point. This recommendation was followed.

The results of the MI test illustrate that the most frequent collocations are not necessarily the most significant, and vice versa. For example, the collocation *Court quash* (5 occurrences) ranked as the most significant, while the most frequent collocation *Court held* (326 occurrences) ranked the seventh in the statistical significance list (see Appendix 2). The collocations selected from the analysed judgments proved to be statistically significant, while their significance in the BNC was not as high (except for *ruled* and *upheld*, which had a higher MI score).

The majority of the most significant collocations shared the same semantic property: they seemed to be marked for negative stance. This observation prompted a qualitative analysis of the instances selected.

It turned out that in addition to lexico-syntactic patterns, the collocations analysed were also subject to certain patterns of attitudinal meaning. In linguistics, communication of "feelings, attitudes, value judgments, or assessments" are termed stance (Biber et al., 1999, p.966). While the expression of feelings, naturally, is not welcome in juridical settings, the expression of attitudes and assessments seems to be involved in the argumentation of the parties to the proceedings.

The genre of appellate judgments per se presupposes a negative evaluation of the court of first instance's decisions and consists of numerous indications of errors in the judicial reasoning. This negative stance is expressed in verbs that have an element of negative evaluation, for example, *to err*, *to infringe*, *to distort*, *to misconstrue* etc. Consider the list of collocations marked for negative stance within the most typical context of the corpus:

- 11) **Court erred in sth.** (105): **Court erred** in law (85)
- 12) **Court infringed sth.** (59): **Court infringed** an Article/ an obligation/ a principle/ a provision etc.
- 13) **Court failed to do sth.** (58): **Court failed** to explain/ to respond to arguments/ to state adequate reasons for finding etc.
- 14) **Court distorted sth.** (41): **Court distorted** the facts/ the evidence etc.
- 15) **Court misconstrued sth.** (10): **Court misconstrued** the legal criteria/ the provision etc.
- 16) **Court misinterpreted sth.** (8): **Court misinterpreted** an Article/ a concept etc.
- 17) **Court misapplied sth.** (6): **Court misapplied** the article/ the provisions/ the test etc.
- 18) **Court disregarded sth.** (8): **Court disregarded** the fact/ the Article etc.

Such collocations were always followed by evaluative phrases in the analysed judgments. For example:

- 19) *By the second part of this ground of appeal, the appellant claims that the Court of First Instance failed to state adequate reasons for rejecting the appellant's arguments ...*

⁴ See Martin Weisser's website.

- 20) *By her second plea, the **appellant submits** that the **Court of First Instance infringed** the principle of equal treatment.*
- 21) *By the third branch of their fourth ground of appeal, the **appellants claim** that the **Court of First Instance erred** in law in the application of the principle of non-discrimination.*
- 22) *... the **appellant claims** that the **Court of First Instance distorted** the evidence on which it based its analysis ...*
- 23) *By the first part, the **appellants submit** that the **Court of First Instance misconstrued** that provision by not properly verifying ...*
- 24) *In the first part, **Bolloré maintains** that the **Court of First Instance infringed** the rights of the defence in refusing to endorse ...*
- 25) ***Divipa argues** that the **Court of First Instance distorted** the clear sense of the evidence by ...*

Interestingly, after examining the linguistic context of such collocates, it turned out that the appellant's arguments account for a large portion of the right verbal collocates of *Court* marked for negative stance (e.g. consider the above given instances). The regularities observed stem from the peculiarities of juridical settings. At the appellate instance, the appellant is to point out his reasons for appealing. Specifically, he is supposed to give legitimate reasons for his dissatisfaction with the court's decision. Therefore, he has to continuously refer to the court of first instance's arguments or actions that he considers to be erroneous in some respects. Naturally, the appellant's argumentation involves evaluative aspects, thus accounting for the verbal collocates of *Court* marked for negative stance, while linguistically, that-clauses serve as a convenient syntactic pattern to structure and present the appellant's claims. The figure below provides a schematic illustration of the above discussed observations.

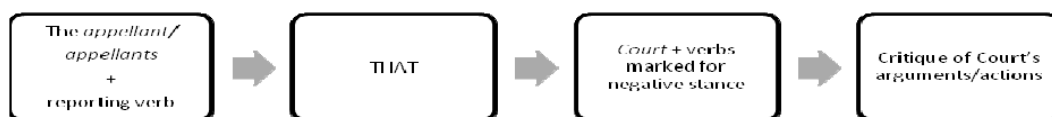


Fig. 2. *The Structure of Appellant's Arguments.*

As regards the attitudinal patterning, the concept of semantic prosody is common in the field of collocational studies. It is understood as “an attitudinal and pragmatic meaning” opposed to referential meaning (Sinclair, 2000, p.200). Jackson understands it as “particular negative or positive connotations” (Jackson, 2002, p.16). It seems that this term has a particular purpose in studying collocation and relates to the special peculiarity of words to collocate seemingly unpredictably and also for one collocate to enhance certain semantic aspects of the other collocate. Sometimes, these aspects can be unexpected, thus referred to as “latent categories of meaning” by Sinclair (2000, p.198). For example, Hoey states that the verb *happen* is more likely to associate with unpleasant events (2000, p.232).

In the instances discussed above, right verbal collocates of *Court* convey negative stance explicitly. The negativity is either encoded morphologically (in the prefixes *mis-*, *dis-* in words *misconstrued*, *misinterpreted*, *misapplied*, *disregarded*) or lexically (*erred*, *failed*, *infringed*, *distorted*). Yet another group of collocates could be distinguished consisting of verbs that do not denote negative actions per se, yet they occur solely in the semantic environment of evaluation or of claiming something to be erroneous or illegal, for example, *Court committed an error of law / a manifest error* (this phrasing is manifest in as many as 36 instances out of 37 co-occurrences). The latter collocations could be regarded as having a negative semantic prosody. Similarly, the collocation *Court applied*, although itself rather neutral, can be said to have an evaluative semantic prosody, since it is used in the statements of evaluative nature: it is frequently surrounded by words *correctly/*

incorrectly and similar expressions. Consider the examples of negative and evaluative semantic prosody:

- 26) *The appellants also maintain in that regard that the **Court of First Instance committed an error of law** in failing to recognise that ...*
- 27) *The appellants claim that the **Court of First Instance committed an error of law** in using the statement of objections as a benchmark ...*
- 28) *The appellant claims that the **Court of First Instance committed a manifest error** in its appraisal of the facts relied on in the determination of the injury.*
- 29) *Moreover, in its examination of the contested decision, the **Court of First Instance itself committed manifest errors** of assessment and fundamentally misconstrued the evidence ...*
- 30) *In the alternative, the appellants submit that the **Court of First Instance applied an excessively high standard of reasoning** that is inconsistent with settled case-law and fails to take account of the particular context and nature of proceedings for the control of concentrations.*
- 31) *Finally, as regards the third part, the appellant maintains that the **Court of First Instance incorrectly applied** the principles identified in the Baby-Dry judgment.*
- 32) *In the alternative, Impala submits that the **Court of First Instance applied the correct test** for establishing market transparency in the judgment under appeal, namely ...*
- 33) *The Commission submits that the **Court of First Instance erroneously applied** the case-law mentioned in paragraph 22 ...*

- 34) *In this ground of appeal, the Kingdom of Belgium alleges that the Court of First Instance wrongly applied the principle of proportionality in considering that ...*

Interestingly, statistical measures applied showed that the collocates marked for negative stance proved to be the most significant (see MI score results in Appendix 2). These collocates can therefore be regarded as typical to the sub-genre under investigation, i.e. they can be viewed as the most characteristic lexical elements of the appellate judgments of the ECJ. As such, they are specific and therefore do not occur in general English as frequently; in fact, most of them (*infringed; distorted; misconstrued; misinterpreted; misapplied*) were not available in the BNC at all. These findings also suggest that the narrower the genre, the more specific collocations it tends to exhibit in relation to general language.

Other collocations, in contrast, seem rather neutral in terms of stance and are used to express the reasoning of the parties to the proceedings. The latter verbal collocates resemble what Biber et al. classify as mental and communication verbs (Biber et al., 1999, p.362–363). Just as Biber et al. suggest (*ibid*, p.660–670), mental and communication verbal collocates were frequently subject to colligational patterns with that-clauses which introduced reported statements (e.g. *Court stated that, Court considered that, Court concluded that, Court noted that* etc.).

Some verbal collocates indicate the court's actions rather than its reasoning and argumentation processes; they are used to name certain procedural steps in decision-making process and relate to the court as a procedural body, for example: *Court dismissed an appeal, Court ruled on a plea, Court examined a dispute* etc. Following Biber's classification, these verbal collocates would fall under the heading of activity verbs (*ibid*, p.362). Collocations with these verbs can also be regarded as neutral in terms of stance.

Conclusions

The current study showed that the language of the judgments on appeal of the European Court of Justice is significantly different from the general English language. The results of statistical analysis carried out using specialised lexical software prove that in terms of frequency and statistical significance of the analysed collocations the language of the appellate judgments in the EU law is remarkably different from general English. The qualitative analysis revealed that collocations analysed exhibit semantic properties that allow classifying them into attitudinal patterns.

The semantic analysis suggests that the genre of appellate judgments is unique because it provides collocations that express numerous ways of saying that the court was wrong. The results obtained from statistical analysis show that the right verbal collocates of *Court* marked for negative stance accounted for the most significant collocations throughout the corpus and in comparison to the BNC, thus proving that collocations serve as a source of typicality and suggesting that perhaps some ways of expressing the wrongfulness of court's actions are specific to the English of the EU law only.

The results suggest that a specific collocational competence should be involved in producing the language of the EU law. The current research reaffirms that the study of legal English should be specialised, i.e. it should differ from teaching general English language.

Since the research was limited in various respects, suggestions for further research arise. The size of the corpus could be expanded in order to achieve more valid results, as it is supposed that larger corpora provide rarer uses (see McEnery & Wilson, 2001, p.80). For example, the judgments could cover a larger time span. Having in mind the variety of subgenres of legal discourse, the appellate judgments analysed cover a relatively narrow area of the language of the EU law. In comparison, different subgenres could be considered. A parallel Lithuanian corpus could be composed for further contrastive studies of the genre of appellate judgments of the ECJ, since all of the selected judgments are translated into Lithuanian. It would be interesting to investigate the English language of appellate judgments in the countries of common law in order to compare it with the legal English of the EU law.

References

1. Bartsch, S., 2004. Structural and Functional Properties of Collocation in English. Available at: <http://books.google.com/> [Accessed on 17 August 2009].
2. Biber, D., Conrad, S. Reppen, R., 2004. Corpus Linguistics: Investigating Language Structure and Use. Cambridge: Cambridge University Press.
3. Biber, D., et al, 1999. Longman Grammar of Spoken and Written English. Harlow: Pearson Education Limited.
4. Bhatia, V. K., 1993. Legal Discourse in Professional Settings. *In: Language Use in Professional Settings*. London: Longman. pp.101–144.
5. Bhatia, V. K., 2000. Discourses of Business and Law: is there a Common Core? *In: Dvorakova, J. (ed.) Proceedings of the LSP forum '99, Prague, Czech Republic, September 17–19, 1999*. Prague: Asociace učitelů angličtiny, pp.81–86.
6. Cao, D., 2007. Translating Law (Topics in Translation). Great Britain: Multilingual Matters LTD.
7. Concordancing Glossary. Available at: <http://www.nsknet.or.jp/~peterr-s/index.html> [Accessed on 6 March 2010].
8. Crowther, J., Dignen, S., Lea, D., 2002. Oxford Collocations Dictionary for Students of English. *International Journal of Lexicography*, no. 16 (1), pp.57–61.
9. Gibbons, J., 1994. Language Constructing Law. *In: Gibbons, J. (ed.) Language and the Law*. London and New York: Longman, pp.3–10.
10. Hanks, P., 2008. The Lexicographical Legacy of John Sinclair. *International Journal of Lexicography*, no. 21 (3), pp.219–229.
11. Hill, J., 2000. Revising Priorities: from Grammatical Failure to Collocational Success. *In: Lewis, M. (ed.) Teaching Collocation*. England: Language Teaching Publications, pp.47–68.
12. Hoey, M., 2000. A World Beyond Collocation: New Perspectives on Vocabulary Teaching. *In: Lewis, M. (ed.) Teaching Collocation*. England: Language Teaching Publications, pp.224–243.
13. Hoey, M. & Brook, M., 2008. Lexicography, Grammar and Textual Position. *International Journal of Lexicography*, no. 21 (3), pp.293–309.
14. Ingels, M. B., 2006. Specific Characteristics of Legal English Writing. *Legal English Communication Skills*. Acco Leuven/Veerburg, pp.54–63.
15. Jackson, H., 2002. Lexicography. An Introduction. London and New York: Routledge. <http://dx.doi.org/10.4324/9780203467282>
16. Juknevičienė, R., 2008. Collocations with High-frequency Verbs in Learner English: Lithuanian Learners vs Native Speakers. *Kalbotyra*, no. 59 (3), pp.119–126.
17. Kennedy, G., 1998. An Introduction to Corpus Linguistics. London and New York: Longman.

18. Lewis, M., 2000. Language in the Lexical Approach. *In*: Lewis, M. (ed.) Teaching Collocation. England: Language Teaching Publications, pp.126–154.
19. Maley, Y., 1994. The Language of the Law. *In*: Gibbons, J. (ed.) Language and the Law. London and New York: Longman, pp.11–49.
20. Marcinkevičienė, R., 2010. Lietuvių kalbos kolokacijos. Kaunas: Vytauto Didžiojo universiteto leidykla.
21. McEnery, T., Wilson, A., 2001. Corpus Linguistics. Edinburgh: Edinburgh University Press.
22. Otani, H., 2005. Investigating Intercollocations – Towards an Archaeology of Text. *International Journal of Lexicography*, no. 18 (1), pp.1–24.
23. Paltridge, B., 2006. Discourse Analysis. An Introduction. London, New York: Continuum.
24. Siepmann, D., 2005. Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography*, no. 18 (4), pp.409–443.
25. Sinclair, J., 1991. Corpus, Concordance, Collocation. Oxford: Oxford University Press.
26. Sinclair, J., 2000. Lexical Grammar. *Darbai ir dienos*, no. 24, pp.191–203.
27. Sinclair, J. et al., 2004. English Collocation Studies: The OSTI Report. *In*: Krishnamurthy, R. (ed.) London, New York: Continuum.
28. Stubbs, M., 2001. Texts, Corpora and Problems of Interpretation: A Response to Widdowson. *Applied Linguistics*, no. 22 (2), pp.149–172.
29. Tognini-Bonelli, E., 2001. Corpus Linguistics at Work. Amsterdam/Philadelphia: John Benjamins Publishing Company.
30. Vystrčilova, R., 2000. Legal English. *Philosophica*, no. 73, pp.91–96.
31. Wei, Y., 1999. Teaching Collocations for Productive Vocabulary Development, pp. 1–8. Available at: <http://www.eric.ed.gov> [Accessed on 12 May 2009].
32. Weisser, M., 2007. N-grams, Collocations and Idiomatic Constructions. Available at: <http://ell.phil.tu-chemnitz.de/analysis/collocations.html> [Accessed on 02 May 2009].
33. Widdowson, H. G., 2000. On the Limitations of Linguistics Applied. *Applied Linguistics*, no. 21 (1), pp.3–25.

Sources

1. Judgments on appeal of the European Court of Justice: Available at: <http://eur-lex.europa.eu> [Accessed in May 2009].
2. BNC: Available at: <http://corpus.byu.edu/bnc/> [Accessed in May 2009].

Computational Tools

1. Scott, M., 2009. WordSmith Tools Version 5.0. Liverpool: Lexical Analysis Software.
2. Log-likelihood calculator. Available at: <http://ucrel.lancs.ac.uk/llwizard.html> [Accessed in May 2009].

Daiva Macko

Programinės įrangos naudojimas analizuojant leksines teisinio diskurso ypatybes

Santrauka

Kompiuterinės technologijos yra tiek tekstynų lingvistikos susiformavimo sąlyga, tiek svarbiausias metodologinis įrankis. Dėl specialios programinės įrangos panaudojimo lingvistikoje atsirado naujas kalbos tyrimo objektas – tekstynai ir buvo sukurti vertingi kiekybiniai jo analizės metodai. Kolokacijų tyrinėjimai yra vieni produktyviausių tekstynų lingvistikoje. Šio tyrimo tikslas – pademonstruoti kelias programinės įrangos panaudojimo galimybes pasirinkto tipo kolokacijoms tirti. Tyrimo objektu pasirinkti apeliaciniuose Europos Teisingumo Teismo sprendimuose vartojami dešinieji veiksmažodiniai žodžio court (teismas) kolokatai. Atrinktų kolokacijų vartoseną buvo lyginama su bendrinės anglų kalbos ypatybėmis vartojimo dažnumo ir statistinio reikšmingumo aspektais. Bendrinės anglų kalbos šaltiniu pasirinktas Britų nacionalinis tekstynas (BNC). Kiekybinė tyrimo dalis grįsta statistine duomenų analize ir buvo atlikta naudojantis programa „WordSmith Tools“ (WS), 5-tąja versija (Scott 2008). Naudojantis logaritminio tikėtumo santykio skaičiuokle buvo palygintas santykinis atrinktų kolokacijų pavartojimo dažnumas su tų pačių kolokacijų dažnumu BNC. Kokybinėje tyrimo dalyje atrinktos kolokacijos klasifikuotos semantiniu aspektu. Tyrimo rezultatai patvirtina, kad kolokacijų vartoseną Europos Teisingumo Teismo apeliacinių sprendimų anglų kalboje skiriasi nuo tų pačių kolokacijų vartosenos bendrinėje anglų kalboje, tai parodo, kad šis teisinio ES institucijų diskurso tipas pasižymi specifinėmis leksinėmis ypatybėmis, į kurias mokant(is) specializuotos teisinės anglų kalbos yra labai svarbu atsižvelgti.

Straipsnis įteiktas 2012 01
Parengtas spaudai 2012 05

About the author

Daiva Macko, Acting Chief Specialist (Translator), *Statistics Lithuania*.

Research interests: corpus linguistics, translation studies.

Address: *Statistics Lithuania*, Gedimino Ave. 29, LT-01500 Vilnius, Lithuania.

E-mail: daiva.macko@stat.gov.lt

APPENDIX

Appendix 1. Right Verbal Collocates of *Court* ($p < 0.0001$; critical value = 15.13)

<i>Right verbal collocates of Court</i>	<i>Occurrences in the chosen judgments</i>	<i>Occurrences in the BNC</i>	<i>Log likelihood</i>
HELD	326	494	+2325.43 (overuse)
SET	119	46	+1054.44
FOUND	116	150	+854.94
FIND	6	65	+22.54
FINDS	11	18	+77.17
STATED	112	39	+1003.66
STATES	8	1	+77.71
ERRED	108	8	+1075.64
REJECTED	85	79	+666.02
REJECT	7	0	
DISMISSED	61	48	+491.33
DISMISS	48	10	+450.68
INFRINGED	59	0	
FAILED	58	23	+512.46
CONCLUDED	49	34	+402.42
CONCLUDE	9	13	+64.85
CONSIDERED	46	66	+331.93
CONSIDERS	18	33	+123.09
POINTED OUT	46	0	
DISTORTED	41	0	
NOTED	37	9	+343.04
NOTES	8	1	+77.71
COMMITTED	37	17	+321.33
APPLIED	36	15	+316.29
ACCEPTED	32	62	+216.02
REFERRED	30	31	+230.72
REFERS	5	7	+36.26
EXAMINED	30	7	+279.12
RULED	28	267	+111.63
RULE	15	29	+101.31
OBSERVED	26	10	+230.51
OBSERVES	5	0	
RELIED	26	10	+230.51
ANNULLED	22	3	+212.64
DECIDED	20	132	+92.98
BASED	18	21	+135.35
GIVE	18	104	+87.96
GIVES	11	23	+72.91
ORDER	17	148	+70.57
ORDERED	6	117	+16.27
EXERCISE	16	41	+100.73
ADDED	14	11	+112.79
MAKE	14	226	+46.62
RECALLED	12	2	+114.51
ASSESSED	12	1	+118.93
ASSESS	7	14	+46.90
INFERRED	12	0	
DECLARE	11	16	+79.15
DECLARED	8	24	+48.25
INTERPRETED	11	3	+101.96
MISCONSTRUED	10	0	
RECOGNISED	8	16	+53.60
DETERMINE	8	36	+42.64
DETERMINED	5	8	+35.25
MISINTERPRETED	8	0	
SUBSTITUTED	8	3	+71.12
DISREGARDED	8	2	+74.00
ACKNOWLEDGED	7	5	+57.24
MISAPPLIED	6	0	
GRANTED	6	55	+24.35
UPHELD	6	96	+18.36
ADJUDICATE	6	4	+49.57
IMPOSED	5	20	+27.68
QUASH	5	13	+31.36

Appendix 2. MI Score.

<i>Right verbal collocates of Court</i>	<i>Joint frequency</i>	<i>MI</i>	<i>Joint frequency in the BNC</i>	<i>MI in the BNC</i>
QUASH	5	6.62	13	6.58
ERRED	108	6.39	8	5.57
MISINTERPRETED	8	6.29		
DISMISS	48	5.95	10	3.07
MISCONSTRUED	10	5.85		
RULED	28	5.69	267	6.24
HELD	326	5.69	494	3.63
MISAPPLIED	6	5.62		
SUBSTITUTED	8	5.37		
FOUND	116	5.35		
INFRINGED	59	5.24		
DISTORTED	41	5.20		
COMMITTED	37	5.17		
ADDED	14	5.14		
DECLARE	11	5.07	16	3.53
FAILED	58	5.03		
STATED	112	5.03		
INFERRED	12	4.95		
DISREGARDED	8	4.91		
POINTED	53	4.89		
TOOK	42	4.89		
ACKNOWLEDGED	7	4.84		
CONCLUDED	49	4.83		
ANNULLED	22	4.79	3	4.40
DISMISSED	61	4.70	48	3.47
EXPLAINED	10	4.69		
RECALLED	12	4.65		
ACCEPTED	32	4.65		
REJECTED	85	4.53	79	3.76
NOTES	8	4.53		
NOTED	37	4.50		
CONSIDERS	18	4.43		
DECIDED	20	4.32		
OBSERVED	26	4.30		
OBSERVES	5	4.29		
EXAMINED	30	4.26		
SET	119	4.24		
CONSIDERED	46	4.21		
RECOGNISED	8	4.16		
FINDS	11	4.01		
CONCLUDE	9	3.98		
INTERPRETED	11	3.75		
ASSESSED	12	3.66		
RELIED	26	3.54		
REFERRED	30	3.34		
APPLIED	36	3.25		
UPHELD	6	3.09	96	6.79