

STANISŁAW GOŹDŹ-ROSZKOWSKI

## DISCOVERING PATTERNS AND MEANINGS: CORPUS PERSPECTIVES ON PHRASEOLOGY IN LEGAL DISCOURSE

### Abstract

This article attempts to demonstrate different perspectives on investigating legal terminology by focusing on various types of textual recurrence. It discusses the results of a corpus-based and corpus-driven analysis of *discovery*, a term central to the US trial practice and criminal proceedings. The analysis starts from investigating traditional collocational patterns, both nominal and verbal. It describes meanings resulting from the emerging co-occurrence patterns between the term and various word forms identified in its co-text. Then the paper proceeds to explore computer-generated clusters formed around *discovery*. Finally, this article refers to the concepts of semantic preference and semantic sequence (Hunston, 2008) to show how *discovery* tends to be found with a range of different word forms albeit all belonging to the meaning group of 'limitation' or 'restriction'. The connotational value of the term is presented by providing some textual evidence of *discovery* found in co-texts where writers express their unfavourable evaluation towards this legal concept. By drawing on Hunston's (2008) concept of semantic sequence, the analysis illustrates how corpus linguistics can complement more traditional approaches to terminology description. It is argued that only by combining various approaches to the study of word combinations is it possible to gain important insights into the phraseological behavior of legal terms.

**Keywords:** phraseology, legal terminology, corpus linguistics, legal discourse.

## 1. INTRODUCTION

Despite the growing interest in phraseology and its recognized pervasiveness in all language fields (e.g. Granger & Meunier, 2008; Meunier & Granger, 2008; Römer & Schulze, 2009), research into phraseological items and patterns in specialist discourse domains seems to be lagging behind. Legal discourse is one of those domains, with very few studies addressing explicitly the nature and role of word combinations in its different textual manifestations (see however Goźdź-Roszkowski, 2011). The significance of word combinations in legal language was noticed and discussed in Kjaer (1990). In her analysis of phraseology in German legal language, Kjaer points to the crucial link between formulaicity of legal instruments and the corresponding legal effect by noting that failure to employ prefabricated word combinations which are directly prescribed by law “[will] result in the *invalidation* (my emphasis) of the whole text of which they form a part” (1990: 28). Yet, multi-word units in legal language have been usually examined as a case of lexical doubling or binomials (e.g. Gustaffson, 1984) and discussed in the context of distinctive features of legal language (e.g. Tiersma 1999) or redundancy in specialized discourse (e.g. Crystal and Davy, 1969; Gotti 2003: 55-52). Unquestionably, corpus linguistics and the emergence of specialized computerized resources have given fresh impetus to the study of phraseology. This has resulted in the emergence of the distributional (Evert, 2004), frequency-based (Nesselhauf, 2004) approach which adopts a bottom-up corpus-driven approach to identify lexical co-occurrences (Sinclair, 1987). In consequence, phraseology research carried out from this perspective generates various word combinations which defy the traditional predefined linguistic categories. As Sinclair (2004: 19) points out “a huge area of syntagmatic prospections” has become available. This recent distributional approach to phraseology has obviously not superseded more traditional corpus-based treatments of word combinations in legal texts. For example, Bhatia et al. (2004: 212) despite some skepticism, acknowledge the significance of corpora in “identifying and exploring the use of specifically favoured expressions”. They then move on to explore the use of noun-verb collocations in a corpus of law cases in order to determine the extent and range of variation in different business-related discourses of management, marketing, accountancy and economics. The use of bilingual corpora has led to some comparative collocational studies as in Biel (e.g. 2012) who analyzes key terms and their collocations in English and Polish company law based on her corpus data. Pontrandolfo (2012) scrutinizes collocational patterns in Italian

and Spanish criminal judgments. This type of studies approaches word combinatorics in terms of syntagmatic co-occurrence involving two lexemes, where at least “one component of the collocation must be a term for which a conceptual description is (or at least may be) available” (Heid, 2001: 788).

The distributional approach to word combinations in legal texts has been largely associated with a computer-aided identification of uninterrupted sequences of word forms of a specified length (e.g. 3-word, 4-word or 5-word sequences) identified solely on the basis of their frequency. It thus represents a corpus-driven rather than corpus-based methodology. Tognini-Bonelli (2001: 84-87), who is widely credited with introducing such distinction, maintains that ‘corpus-based’ research rests on the validity of linguistic structures derived from linguistic theory. The research goal in this approach consists in examining the use of pre-defined linguistic features. In the case of multi-word units, this involves pre-selecting such expressions and then analyzing the corpus to determine how they are used (e.g. Moon, 1998). In contrast, ‘corpus-driven’ approach is more inductive in that corpus analysis alone leads to the emergence of linguistic constructs. This approach makes minimal a priori assumptions about the linguistic constructs employed in the analysis.

Such linguistic constructs have been variously defined and referred to as lexical bundles (Biber et al 1999), clusters (Scott 2008), n-grams, routines, formulas, prefabricated patterns, etc. Goźdz-Roszkowski in a series of studies (2006a; 2006b; 2007; 2011: 109-143) documents the patterns of use of recurrent multi-word sequences (lexical bundles) both in individual legal genres (e.g. contracts or judgments) and across a range of legal genres. The findings contained in these studies show that recurrent multi-word expressions, due to their sheer frequency, play a significant role in constructing legal texts, albeit to a varying degree. It emerges that all legal genres use lexical bundles. However, the genres use bundles to differing extents, and for different functions. For example, lexical bundles in legislation and contracts account for as much as 10% of the language found in legislative and contractual texts. Indeed, it turns out that multi-word combinations of this type can be effectively used to discriminate between different legal genres. Apart from the discriminating function, the lexical bundles methodology can be extremely useful in identifying terminological phrases (see for example Goźdz-Roszkowski, 2006b). Ultimately, it is argued that legal genres can be described and differentiated in terms of their preferred phraseologies and that the phraseological preferences correlate strongly with the different communicative priorities and epistemological precepts of the legal genres.

Yet another recent approach to phraseology concerns recurrent co-occurrence patterns of meaning elements rather than specific, formally defined linguistics units as reflected in the recently proposed concept of ‘semantic sequence’ (Hunston 2008). This perspective on recurrent patterns is discussed in detail in Section 2 below. Despite the paucity of collocational and phraseological studies in English legal language, it can be already suggested that combining different perspectives and methodologies is likely to result in a more comprehensive and fairly exhaustive phraseological descriptions of specialized discourses, and especially in the descriptions of domain-specific terminology. The aim of this paper is thus to demonstrate how the three different, albeit related approaches (which can be, tentatively, referred to as *collocational*, *cluster-* and *meaning-based*) to phraseology can be effectively applied to build a phraseological profile of legal terms. In doing so, I will focus on the term *discovery* which refers to a crucial concept in US trial practice and criminal proceedings.

## 2. THE MANY FACES OF PHRASEOLOGY

In view of the highly variable and wide-ranging scope of phraseology, it seems necessary to revisit the very notion of what constitutes phraseology. The well-known classic definition according to which phraseology concerns “the study of the structure, meaning and use of word combinations” (Cowie 1994: 3168) may turn out to be not very useful, especially in the face of rapid developments taking place in corpus linguistics. Extensive corpus research into multi-word sequences has led to the growing recognition of the central importance of phraseology for language description (e.g. Sinclair, 1991; Hunston and Francis, 1999; Römer and Schulze, 2009). For example, there have been several studies relying on multi-word sequences to discriminate between different registers (e.g. Biber et al., 1999; Biber, Conrad and Cortes, 2003), text types (e.g. Stubbs and Barth, 2003) and disciplines (e.g. Groom, 2005; Hyland, 2008). Methodological concerns resulting in the identification and categorization of phraseologisms have become of central importance for researchers who would like to share their assumptions, concepts and findings. Gries (2009: 4) notes that the notion of phraseology has become so widespread that many scholars tend to use it without providing clear definitions thus precluding wider applicability and comparability of their work. Indeed, a crucial question should be posed concerning the status of multi-word units as phraseologisms. Gries (2009: 4) pleads for “a rigo-

rous definition of co-occurrence phenomena in general, and phraseology in particular”. He offers six criteria useful in defining phraseology: (1) the *nature* of the elements involved in a phraseologism; (2) the *number* of elements involved in a phraseologism; (3) the *number of times* an expression must be observed before it counts as a phraseologism; (4) the permissible *distance* between the elements involved in a phraseologism; (5) the degree of *lexical and syntactic flexibility* of the elements involved; and finally (6) the role that *semantic unity* and *semantic non-compositionality / non-predictability* play in the definition (2008: 4). After discussing each criterion, Gries (2008: 6) proposes what could be considered as, probably, the broadest definition of a phraseologism: “the co-occurrence of a form or a lemma of a lexical item and one more or additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance”. This definition is worth noting for being explicit with regard to each of the above-mentioned parametres and for extending the range of phenomena regarded as phraseologism. As a consequence, it encourages researchers to define carefully the level at which they examine a potential phraseologism and to decide how many elements a phraseologism is supposed to comprise. It also prompts one to consider many types of multi-word expressions as phraseologisms, especially those computer-generated.

There is yet another concept useful in the description of recurrence or regularity in texts. In 2008, in her article “Starting with the small words. Patterns, lexis and semantic sequences”, Susan Hunston proposes the concept of *semantic sequence* defined as “recurring sequences of words and phrases that may be very diverse in form and which are therefore more usefully characterized as sequences of meaning elements rather than as formal sequences” (2008: 271). In addition, the semantic sequence consists of “the core word, item’, the complementation pattern or patterns associated with that word (such as a *that*-clause, *wh*-clause, or a prepositional phrase with a specific preposition) and a number of phrase types occurring before the core word which are, in spite of being diverse in form, consistent in terms of meaning” (2008: 272). The concept of semantic sequence looks particularly promising for analyzing disciplinary discourses. Studies done by Charles (2004), Gledhill (2000) and Groom (2007) provide important evidence that disciplines may be characterized by their unique phraseological profiles or templates. It appears that there are various ways of identifying semantic sequences. Both Gledhill and Groom argue that grammatical words can be treated as a useful starting point for identifying semantic sequences in

domain-specific corpora. For them, prepositions represent the ‘core item’ in the sequences they identify. On the other hand, as Hunston convincingly demonstrates, semantic sequences can also be captured by starting with a lexical word, phrase or pattern. Irrespective of the initially selected ‘core item’, the next analytical stage involves examining relevant co-texts of a lexical item, grammatical words (e.g. prepositions) or a grammatical pattern. In Goźdź-Roszkowski (2012) I demonstrate the applicability of this concept to investigate the construction of epistemology in legal academic discourse. Starting with the pattern *the* + Noun + (appositive) *that*-clause, I examine two nouns *idea* and *notion*. It is generally accepted that status nouns found in this pattern are particularly important in the construction of knowledge (Halliday and Matthiesen 2004) because they indicate the epistemic status of the proposition expressed in the *that*-clause. Projected clause of this type are thus important to disciplinary epistemology. It turns out in my study that both *idea* and *notion* co-occur with highly evaluative meaning elements manifested through diverse language forms. In the case of *the notion that*, it is possible to propose a frequently-occurring semantic sequence consisting of ‘institution’ + ‘accept or reject’ meaning group + *the notion* + *that*-clause. As will be demonstrated below, this perspective can be also used effectively to uncover recurrent semantic patterns of a highly technical legal term.

### 3. CORPUS LINGUISTICS, TERMINOLOGY AND SPECIALIZED DISCOURSES

As indicated above, this paper aims to discuss the results of a corpus-based analysis of a selected legal term, i.e. *discovery*, with a view to underscoring the importance of contextual and phraseological information in terminology description in reliance on certain concepts proposed in corpus linguistics research to inform terminology practice. It reflects the recent trend which attempts to bring the study of terminology back to the study of real language usage (cf. Temmerman, 2000). Thus, a more descriptive approach is advocated., whereby legal terminology is studied in texts, which implies the study of *parole*, i.e. real language usage. Further, it is argued that any descriptive terminological work should widen its scope by examining the co-texts in which a term is found with a view to uncovering its relevant phraseological patterns. The benefits of corpus-based methodology for terminology research and practice can hardly be overestimated. Almost two decades ago Bowker (1996) was one of the first scholars to advocate this

approach and to pointing out its three major advantages. First, machine-readable corpora combined with the requisite tools enable one to reduce the time needed to research candidate terms. Large quantities of data can be processed rapidly exposing terminologists to a large number of conceptual descriptions. Having recourse to well-annotated corpora allows terminologists to focus on those parts of the texts that are terminologically relevant (Bowker, 1996: 31-32). Thus, with the right corpus and tools, it is possible to gain access to knowledge-rich contexts (Bowker, 1996) useful for conceptual analysis. Second, corpus data facilitates the examination of syntactic and semantic information encoded in linguistic patterns which are difficult to intuit or observe when scanning texts manually. This type of information can be retrieved through the use of concordances, also known as key words in context (KWIC). Concordances can provide information on the combinatorics of terms. Third, unlike in the case of conventional term banks, corpora provide a wealth of examples. Bowker (1996: 32-33) rightly observes that corpora present a variety of contexts as well as more extensive contexts. The accumulation of contexts is essential if one aims to examine whether a particular term is found in recurrent regular patterns of co-occurring items. In a similar vein, Pearson (1998) in her seminal book *Terms in Context* demonstrates the advantages of integrating terminology and corpus linguistics in order to develop an effective and efficient methodology for retrieving information about terms from corpora.

At the same time, there is a growing awareness of the role of language patterns as one of the basic constructs employed by institutions to build up our linguistic, conceptual and ideological view of the world which may convey messages implicitly without the reader being intuitively or consciously aware. Increasingly, this type of language constructs have been identified in specialized discourse. Schulze and Römer (2008) in their introduction to the special issue of *International Journal of Corpus Linguistics: Patterns, Meaningful Units and Specialised Discourses* emphasize the inextricable link between the performativity in language and the way “speakers and writers do things by predominantly and unconsciously employing patterns or phraseological items” (2008: 1). Studies of specialised language collected in that issue demonstrate how specialised meanings are expressed and encoded by means of phraseological patterns and how they are linked to the particular domains in which they were used. The present paper aims to contribute to this growing body of research by investigating recurrent patterns of the legal term *discovery*.

In what follows, I turn to present the data and methodology adopted in this study (Section 4) and then discuss the findings (Section 5).

#### 4. DATA

The data used in the analysis consists of 114 different opinions given by the Supreme Court of the United States of America. These were collected at random from the period between 2000 and 2007<sup>1</sup>. The corpus totals 1,182, 246 words. Under the largely judge-made law system, the Supreme Court opinions represent the primary vehicle through which American law develops (Lee, Hall and Hurley, 1999). The Supreme Court is regarded as the national symbol of justice. It is the highest court in the United States consisting of the Chief Justice and eight Associate Justices. The U.S. Supreme Court has both original and appellate jurisdiction, but it exercises the former only in rare instances. The primary task of the Court is thus appellate. In that capacity it serves as the final arbiter in the construction of the Constitution of the United States and it provides a uniform interpretation of the law. However, to a large extent, it attempts to adhere to precedent, the well-known doctrine of *stare decisis*, i.e. “let the decision stand”. In this way, the precedents are given the authority of established law. It is thus little wonder that learning to read and understand cases, especially those given by the US Supreme Court, is a fundamental task that is mastered in all law schools. Worth bearing in mind is that the term *opinion* can be used in at least two somewhat different senses. First, more generally, an opinion may refer to the official decision of a court of justice and it is then interchangeable with the term *judgment*, which is defined as “the official and authentic decision of a court of justice upon the respective rights and claims of the parties to an action or suit therein litigated and submitted to its determination. The final decision of the court resolving the dispute and determining the rights and obligations of the parties” (Black, 1990). Second, the term *opinion* is also used to denote the reason which the court gives for its decision. It then refers to the statement made by a judge or court of the decision reached in connection with a case heard before them. Such a statement explains the law as applied to the case and provides the reason on the basis of which the judgment is made. The analysis presented below is based on opinions in the second sense, i.e. strictly judicial discourse providing legal argumentation for the decision reached. This study focuses on the legal term *discovery*, which denotes a significant concept in American legal procedure. The authoritative *Black’s Law Dictionary* clarifies this term

---

<sup>1</sup> In fact, the data set is part of a larger American Law Corpus compiled by the author and described in detail in Goźdź-Roszkowski (2011: 27-33).



in the following ways: First, in the general sense, *discovery* means “the ascertainment of that which was previously unknown; the disclosure or coming to light of what was previously hidden”. More specifically, it is employed as a fundamental concept in the trial practice where it is attributed the sense of “Disclosure by defendant of facts, deeds, documents or other things which are in his exclusive knowledge or possession and which are necessary to party seeking discovery as a part of a cause of action pending”. In this sense, *discovery* is aided by the following tools: “depositions, written interrogatories, productions of documents or things, permission to enter upon land, etc.” (Black’s Law Dictionary). As can be seen, *discovery* is a highly technical term used to denote a fact-finding process based on the assumption that a free exchange of information regarding a particular case is beneficial to the parties in that it helps to uncover the truth. In what follows we examine the contexts in which *discovery* is found in the hope of identifying prevalent patterns of using this term in judicial practice. The data was handled using the popular suite of text-processing programmes known as the *WordSmith Tools* (published by Oxford University Press since 1996, now at version 5.0 <http://www.lexically.net/wordsmith/>).

## 5. METHODOLOGY AND RESULTS

In what follows, this paper provides methodological background and it discusses the results of analysis carried out at three levels: 1) collocational (co-occurrence between the node word (i.e. the term *discovery*) and its collocates; 2) clusters; and (3) semantic sequence.

### 5.1. COLLOCATIONAL ANALYSIS

The collocational analysis was carried out using the in-built collocate feature of the Concord Tool. The collocate horizon was set at 5, which means that the programme searched for potential collocates within five words to the right and left of the node word, i.e. the term under investigation. The Concord search of *discovery* has returned 205 instances of this word found in the entire corpus. After enabling the ‘collocates’ function and analyzing its output, the following collocates can be identified. First, there are several nominal collocates identified within five words of the node word, where *discovery* serves as a modifier:

- discovery <requests (26), order (s) (29), plan (13), process (7)>

The figures in ( ) brackets denote the raw frequency with which a given word collocates with the term *discovery*. It appears that *discovery* is frequently found in two-word nominal phrases closely related to the procedure as for example in (1) and (2):

- (1) When petitioner, an attorney representing a plaintiff, failed to comply with certain **discovery orders**, the Magistrate Judge granted the respondent's motion for sanctions against petitioner under Federal Rule of Civil Procedure 37(a)(4).
- (2) Some of the sanctions in this case were based on the fact that petitioner provided partial responses and objections to some of the defendants' **discovery requests**.

Left-sided nominal collocates include:

- discovery <court (6), objections (4) found three words to the left of the node word and scope (8) located at three and two words left of the node word:

The co-occurrence of *court* and *discovery* shows the former as the primary institutional interactant involved in controlling what is admitted or denied in the fact-finding process:

- (3) In due course the District **Court** approved respondents' discovery plan (...).

The collocate *objections* also signals certain procedural properties associated with the concept of *discovery* when one of the litigants can disagree with the way the process is handled:

- (4) Government stated: "We did not choose to offer written **objections** to [the discovery plan]"

The examination of concordances generated for *discovery* reveals that apart from *objections*, two related verbal forms *object* and *objecting* can also be found co-occurring with the term, albeit less frequently (see example (5)). This observation corroborates the fact that the way discovery is conducted can be open to criticism from the litigants and it suggests the rather obvious need for studying co-occurrence patterns both at the level of word forms and lemma.

- (5) Under these circumstances, instead of requiring petitioners to **object** to particular discovery requests, the District Court (...)

Eight occurrences of *scope* found in the co-texts of *discovery* seems to indicate that the extent to which discovery process tends to be conducted is indeed perceived as an issue in judicial discourse:

- (6) A district court is not subject to criticism if it awaits a party's motion before tightening the **scope** of discovery.
- (7) Defendants object to the **scope** of plaintiffs' discovery requests and to the undue burden imposed by them

Example (7) is particularly noteworthy as it shows that the lemma *object* and *scope* are not only collocates of *discovery* but they may collocate with each other. In other words, litigants may object to the extent and the manner of fact-finding process. These collocates begin to paint the picture of this concept as flexible and shaped by the litigants under judicial supervision.

Apart from the noun + noun combinations, there are several important verbal collocates preceding *discovery* as represented by the V + N pattern:

- Discovery<*allow, object, conduct, narrow, obtain, reject, order, seek, submit, vacate*>

In terms of frequency, the verbal collocations are weaker than the nominal ones (on average three occurrences) but they display a wider range of lexical choice.

- (8) Moreover, even if this Court affirms the decision below and **allows discovery** to proceed in the District Court, the issue that would ultimately present itself still would have no bearing upon the reputation and integrity of Richard Cheney.
- (9) The District Court **ordered discovery** here, not to remedy known statutory violations, but to ascertain whether FACA's disclosure requirements even apply to the NEPDG in the first place.

The sentence examples (8) and (9) illustrate how courts operate with respect to *discovery*. They are active agents overseeing the process of *discovery* ensuring that it complies with the requisite law. Obviously, parties involved in a litigation can also influence this process, as illustrated by the following examples:

- (10) Banks attached affidavits from Farr and Cook to a February 1999 motion **seeking discovery** and an evidentiary hearing
- (11) The Government, however, did in fact **object to the scope of discovery** and asked the District Court to **narrow** it in some way.

Excerpt (11) corroborates the observation made above recognizing the degree to which *discovery* is perceived as a key issue involved in the way the concept of *discovery* is discoursed in legal opinions. Indeed, it can be clearly seen that *discovery* can be a contentious issue between the parties to litigation or a litigant and the overseeing court. Undoubtedly, there are conflicting views as to the extent to which the process of *discovery* should be carried out, with one party often insisting on limiting it to some degree. The issue of the scope of discovery tends to play a role upon the examination of adjectives preceding this term.

- **discovery** <*broad* (8); *civil* (4); *pretrial* (3); *proposed* (3) >

The limited range of attributive adjectives includes two technical collocates. Example (14) shows *civil* used as a technical way of distinguishing the legal area involving a particular case heard by the court, while excerpt (15) indicates the stage of the legal proceedings where discovery was needed. There are eight instances when the term is immediately preceded by the word *broad*, as for example, in excerpts (12) and (13):

- (12) Thus, granting **broad discovery** in this case effectively prejudged the merits of respondents' claim for mandamus relief--an outcome entirely inconsistent with the extraordinary nature of the writ.
- (13) **Broad discovery** should be encouraged when it serves the salutary purpose of facilitating the prompt and fair resolution of concrete disputes.
- (14) There are no checks in **civil discovery** analogous to the constraints imposed in the criminal justice system to filter out insubstantial legal claims.
- (15) Most of the **pretrial discovery** related to the question of whether the 1995 Amendment to the IOLTA Rules had indirectly lessened the earnings of LPOs

Excerpts (12) and (13) add to the emerging semantic pattern whereby *discovery* is discoursed in terms of its scope. These two examples highlight the advantages of broad discovery. These excerpts contain a strong evaluative element in the judicial argumentation pointing either towards the advantages or disadvantages of broad discovery in pretrial proceedings.

Interestingly, *broad* is not restricted solely to the one sentential position. It can be more freely distributed as in the following example, where it is found within five words to the right of the node word:

- (16) Although acknowledging that **the scope of** respondents' **discovery** requests was **overly broad**, the appeals court nonetheless agreed with the District Court that petitioners should bear the burden of invoking executive privilege and of objecting to the discovery orders with detailed precision.

Worth noting is the presence of the adverb *overly* which clearly carries negative evaluation of the scope of such discovery. The data shows that *discovery* also collocates with the lemma *narrow*. There are several attested instances of limiting the scope of *discovery* by using this lemma. Before we move on to explore semantic aspects involved in lexical co-occurrences of the term *discovery*, let us examine the results of a cluster analysis.

## 5.2. CLUSTER ANALYSIS

The Concord tool in the *WordSmith Tools* software contains a feature which enables one to identify patterns of repeated phraseology understood as pre-defined sequences of word forms. Mike Scott, the author of the *WordSmith Tools*, refers to such constructs as *clusters* (Scott, 2000). Other related terms include lexical bundles (Biber et al., 1999) or n-grams (Fletcher, 2006). In short, a cluster, n-gram or lexical bundle is an unbroken sequence of n words. For example, a 4-gram is simply four consecutive words identified by a programme in a text. It should be noted that lexical bundles are somewhat more sophisticated than n-grams or clusters chiefly because of the multiple text requirement, i.e. a sequence is considered lexical bundle if it is found, for example in five different texts. In this way, idiosyncratic local repetitions can be excluded. Additionally, lexical bundles tend to have very high frequency cut-off points of at least 10 times per million words (Biber et al., 1999). Strictly, the analysis done in this study concerns clusters, not lexical bundles. This study is concerned with exploring the extended collocations of a single term rather than characterizing text types of genres. Different, more relaxed criteria can be thus applied. Predictably, clusters centered around a term will be far less numerous. Clusters are calculated using the existed concordance lines. Clusters settings in this study include length determined at between 3 and 5, horizons set at 5 words to the left and right of the node word and the minimum cluster frequency of 3 calculated using the LogLikelihood relation. Adopting these parameters resulted in identifying 80 clusters.

### 5.2.1. *Discovery* clusters

To a large extent, the cluster analysis reiterates some of the findings already obtained when examining the collocates of *discovery*. There are however some additional advantages. First, clusters highlight some co-occurrences not immediately visible in collocates. For example, the term *evidentiary hearing* tends to be found in the vicinity of *discovery* and their co-occurrence indicates that these two terms are part of the more general *rules of evidence*. The same observation applies to another important term – the *mandamus petition*. These clusters represent potential legal procedure-related phrases. Second, cluster may show some consistency in function expressed though diverse language form. This is illustrated, for example in the two different clusters *objections to the discovery*, *objecting to the discovery* and the other negatively-charged clusters such as *overly broad discovery requests*, or *no discovery is appropriate*. They all contribute to the sense of negative evaluation inherent in the concept of *discovery*. There are several potential terminological phrases with *discovery* being part of them; some are phrasal:

*the discovery orders, the discovery requests, the discovery process, the discovery plan, the scope of discovery, the civil discovery, the mandamus petition, an evidentiary hearing, discovery and an evidentiary hearing*

while others tend to be clausal:

*to vacate the discovery, vacate the discovery requests, vacate the discovery orders, to narrow discovery, to dispute the discovery, dispute the discovery requests,*

These clusters show that actions related to *discovery*-centred terms such as *discovery requests* or *discovery orders* focus on cancelling or at least reducing the process. Some repetition can be noticed among the clusters due to the fact that certain clusters are in fact part of larger sequences. Compare, for example, *to dispute the discovery* and *dispute the discovery requests*.

Certain clusters, even if not obviously terminological, are overtly evaluative and provide an important insight into the nature of this concept, especially the way in which *discovery* is discussed in judicial discourse:

*objections to the discovery, objecting to the discovery, overly broad discovery requests, a routine discovery dispute, no discovery is appropriate, this is not a,*

Apart from the value-laden lexical words like *objecting*, *overly broad* or *appropriate*, the negation markers *not* and *no* may signal a negative discourse

prosody related to the term *discovery*. Finally, a large proportion of clusters are obviously part of still larger sequences deemed irrelevant at this stage: *the discovery requests are, of the discovery orders, with all discovery to, proposed discovery plan, allowing discovery to*

Worth bearing in mind is that clusters represent merely *potential* or *candidate* term phrases. More work is usually needed to validate their terminological status. More generally, n-grams cannot be equated with units of meaning, in other words they have no linguistic validity. They should be treated as a helpful starting point in search of multi-word units. This is clear when we look at the cluster output. Some of the sequences are well-formed while others cannot function on their own and should be classified and discarded as ‘noise’. The collocational and cluster analyses carried out so far have revealed important lexical co-occurrences showing the term *discovery* as forming part of compound nominal terminological phrases as well as co-occurring with procedure-related verbs. What we have also noted is that there might be slightly different types of co-occurrences or relations where *discovery* is involved. This term tends to co-occur with words like *scope*, *broad* or *narrow*, i.e. words referring to the extent or degree to which *discovery* should be conducted. We also observed that, especially *scope* of the discovery is perceived as an issue to which litigants can object. It thus emerges that there is a different type of relation not only involving collocates understood as individual word forms or even lemma but ‘semantic preference’ and ‘semantic prosody’. These two concepts will be explored in the next section.

### 5.3. SEMANTIC PREFERENCE AND SEMANTIC SEQUENCE

As already noted, it is often more effective to identify phraseology attached to lemma rather than individual word forms. In the present study, this observation is suitably illustrated by examining the lemma NARROW. Examples (17) through (20) show that the co-occurrence between *discovery* and *narrow* can be expressed through different word forms found in various textual positions.

- **discovery <narrow + N (12)>**

- (17) Mindful of “the judiciary’s responsibility to police the separation of powers in litigation involving the executive,” the Court of Appeals expressed confidence that the District Court would “respond to petitioners’ concern and **narrow discovery** to ensure that [respondents] obtain no more than they need to prove their case.” *Id.*, at 1106.

- (18) The District Court, moreover, did not err in failing to **narrow** discovery on its own initiative;
- (19) The Government, however, did in fact object to the scope of discovery and asked the District Court to **narrow it** in some way;
- (20) Although the Court cites *United States v. Poindexter*, 727 F. Supp. 1501 (DC 1989), as "sound precedent" for district-court **narrowing of discovery**, see ante, at 19-20, the target of the subpoena in that case, former President Reagan, unlike petitioners in this case, affirmatively requested such **narrowing**, 727 F. Supp., at 1503.

If we recall that *discovery* also collocates with *broad*, as shown in Section 5.1, it appears that, perhaps, the *broad v. narrow* opposition might be to some extent indicative of the way in which *discovery* is conceptualized in the judicial discourse. In other words, there might exist some semantic relation between *discovery* and certain co-occurring items. In order to validate this claim, more evidence of such items would be needed. It should be stressed at this point that the traditional collocational analysis which looks at the strength of collocational relations between the node word and an individual word form might offer only limited information on the phraseological behavior of the term in question. The next step is therefore to study each and every concordance line in the hope of identifying more evidence of such semantic relation. Indeed, the search has proved to be extremely rewarding showing that there are several words which could be subsumed under the semantic category of ‘determining the scope of discovery’:

- (21) First, as the Court of Appeals recognized, see supra, at 8-9; infra, at 11, should the Government so move, the District Court could **contain discovery** so that it would not be “**wide-ranging**.”
- (22) Indeed, the appeals court has already suggested **tailored discovery** that would avoid “effectively prejudg[ing] the merits of respondents’ claim;
- (23) Without taking up the District Court’s suggestion of that court’s readiness to **rein in discovery**, see 219 F. Supp. 2d, at 54, the Government, on behalf of the Vice President, moved, unsuccessfully, for a protective order and for certification of an interlocutory appeal pursuant to 28 U. S. C. § 1292(b).
- (24) A district court is not subject to criticism if it awaits a party’s motion before **tightening the scope of discovery**;



- (25) On **limiting discovery** to the issue of membership, the Court of Appeals indicated its agreement;
- (26) the District Court approved respondents' **extensive discovery** plan, which included detailed and far-ranging interrogatories and sweeping requests for production of documents.

Excerpts (21) through (25) show a range of different words expressing the idea of determining the scope of *discovery*, usually by limiting it. These are verbs, such as *contain*, *rein in*, *tighten* or *limit* or adjectival forms like *tailored* and *wide-ranging*. Note that even *wide-ranging* is used in the negative context, where it is indicated that it is not a desirable situation to have 'a wide-ranging discovery.' The opposite seems to prevail, that is, discovery should be clearly kept in check. Quantitatively, a particular word form may be relatively infrequent. Some forms are actually found only once, but taken together, they represent a sizeable group. Other related items belonging to the same lexical field include *a tightly-reined discovery*, *reined-in discovery*, *discovery-tailoring measures*, *limited discovery*, *keep discovery within appropriate limits*, *reduction of the District Court's discovery*, etc.

The findings provided above can be viewed in terms of the concept of semantic preference. This concept is understood here as "the relation between the node word and a lexical field which signals frequent topics in the immediate co-text". (Stubbs, 2009: 22). Thus, it is argued here that there is a relation of semantic preference between the term *discovery* and the lexical field of 'limiting the scope'. Strictly, there are also words indicating the expansion of the scope of *discovery* like *wide-ranging* or *extensive* but they are not as numerous and frequent as those expressing the sense of limitation. The linguistic material analyzed in this study seems to fit in with the fundamental characteristics of a semantic sequence, i.e. the co-occurrence between the word or term and a range of various word forms referring to the same semantic field. As can be seen in the examples provided above, *discovery* tends to be found with different word forms, however it remains linked to the semantic value of limitation or restriction. This somewhat looser (than in the case of traditional collocational analysis) association can be generalized as a semantic sequence (as shown in Table 1).

---

**legal or natural person** (*court/government/defendants/plaintiffs*) +  
 'delimited scope' vocabulary [e.g. *narrow*, *broad*, *tailored*, *contain*, *wide-ranging*, etc.] + **discovery**

---

Table 1. Proposed semantic sequence for the term *discovery*

Quantitatively, this semantic sequence accounts for approximately 30% of all instances where *discovery* is found. In subject positions, there is usually a legal or natural person, such as *court* or *plaintiff*, co-occurring with the lexical set belonging to ‘limiting the scope’ field and the term *discovery*. It should be stressed that the sequence is by no means fixed in that it does not follow any specific formal structure. Diverse word forms from different part-of-speech categories are found with varying frequency both to the left and right of *discovery*. This loose form makes it an elusive target for the traditional collocational analysis.

The description of meaning in terminology tends to focus on the denotational value admitting of little if any connotational value. In fact, the connotational significance of lexis has not always been widely recognized.<sup>2</sup> The term *connotation* has been defined, for example, as “the vaguer associations of a word for a group or individual” (Cook, 1992: 8) or in terms of “secondary implication” of an item (Lyons, 1977). Overall, connotation was perceived as “incidental to language rather than an essential part of it (Leech, 1974: 15). The potential connotations of terms appear to be beyond the realm of traditional terminology, which tends to focus on terms alone without paying much attention to the context of their use (cf. Wüster, 1991). In the case of the term *discovery*, connotation, especially expressive connotation, seems to play an important role in understanding its underlying concept. In what follows, the term *discovery* has been examined in order to find out whether its use involves favourable or unfavourable evaluation by the writer towards what they describe. In Section 5.1, we noted that certain collocates may carry negative evaluations of the discovery process. Apart from the lemma *object*, we noted that, for instance the phrase *overly broad*, as well as some of the clusters in Section 5.2 signal negatively-charged discourse. Also, the lexical field of ‘limiting the scope’ located within the semantic sequence provided above contains items that are highly evaluative, such as, for example, *a tightly-reined discovery*, *reined-in discovery* or *keep discovery within appropriate limits*. The analysis of the relevant co-texts has revealed that in fact in the overwhelming majority of cases, *discovery* is found in other negative contexts. For example, *discovery* is often mentioned in the context of abusing its original purpose, i.e. ascertaining the facts of the case (excerpts 27 and 28).

---

<sup>2</sup> Partington (1998: 65) provides a brief but useful overview of how connotation has been defined and perceived in traditional semantics.

- (27) Allowing a plaintiff to forgo giving any indication of the economic loss and proximate cause would bring about the very sort of harm the securities statutes seek to avoid, namely the **abusive practice** of filing lawsuits with only a faint hope that **discovery** might lead to some plausible cause of action.
- (28) In 1970, the prerequisites for imposing sanctions were redesigned “to encourage judges to be more alert to **abuses** occurring in the **discovery** process.”

In addition, *discovery* tends to be perceived as a potential threat to the smooth, speedy and effective trial practice. The judiciary appears to be concerned that discovery procedure could be manipulated by unscrupulous litigants and their counsel. Protecting the interests of the parties is a recurring theme in such contexts (excerpts 29 and 30) and it can be seen particularly well in excerpt (30), which provides justification for the disposition handed down by a judge.

- (29) To permit an immediate appeal from such a sanctions order would undermine the very purposes of Rule 37(a), which was designed to protect courts and opposing parties from **delaying or harassing tactics** during the **discovery** process.
- (30) In accord with the Court of Appeals, I am "confident that [were it moved to do so] the district court here [would] protect petitioners' legitimate interests and **keep discovery within appropriate limits.**" 334 F. 3d, at 1107.9 I would therefore affirm the judgment of the Court of Appeals.

There are many more instances of negative discourse associated with the term *discovery*. A list of selected examples can be found in Table 2:

- 
- the contention that *discovery* would be unduly **burdensome**;
  - reined-in *discovery*—is surely a **doubtful proposition**
  - the Government mentioned “**excessive discovery**” in support of its plea to be **shielded from any discovery**;
  - There are **no checks** in civil *discovery* analogous to the constraints imposed in the criminal justice system to filter out insubstantial legal claims;
  - Pretrial motions and *discovery* **consumed three years**;
- 

Table 2 Examples of negative evaluation of the term *discovery*

The negative evaluation of the contexts in which *discovery* tends to be found is not related to a single expression or pattern. Instead, we note the presence of a range of diverse lexical items which signal a negative stance on the part of the writers. In some cases, the negative evaluation is more implicit and it should be inferred, as in the last two examples shown in Table 2, where ‘the period of three months’ suggests a protracted procedure and the absence of ‘checks in the civil discovery’ is viewed as a serious drawback. At the same time, these two examples also highlight the link between ‘limiting the scope’ phraseology and the expression of evaluation. The semantic sequence presented above could be thus extended, at least in some cases, to include semantic prosody understood as “the discourse function of the unit” and indicating “the speaker’s communicative purpose” (Stubbs 2009: 22). It seems clear that determining the scope of discovery is in many cases strongly associated with negative evaluation.

#### 4. CONCLUSIONS

This study has attempted to demonstrate different perspectives on investigating term phraseology in judicial discourse by focusing on different types of textual recurrence. First, I documented the most common nominal, adjectival and verbal combinations in which the term *discovery* is found. This basic collocational analysis led to the identification of several important terminological co-occurrences. These were corroborated and extended through cluster analysis. Cluster analysis was noted to offer a quick and efficient way of identifying potential terminological phrases. Further investigation has revealed that a significant proportion of the co-occurring items belong to the lexical field which refers to the idea of limiting or restricting. In this particular case, the lexical items refer to the imposition of constraints on the scope of *discovery*. These lexical items cannot be subsumed under any specific formal and fixed structure. Instead, they represent diverse grammatical categories used in various syntactic positions. Apart from the shared lexical field, the only other common feature is related to the subject or agent, which is a legal or natural person involved in the legal proceedings. Traditional collocational and cluster analyses prove to be of limited use when it comes to this type of textual recurrence. It turned out that the term *discovery* is found in many concordance lines which do not contain collocates but they may contain patterns. For example, words like *reined-in* or *discovery-tailoring measures* belong to the lexical field of ‘restrictive vocabulary’ used

to determine the scope of *discovery* and they represent evaluative language, however they do not occur often enough to be collocates. In order to account for this regularity, I found it useful to invoke certain concepts proposed within the realm of corpus linguistics, such as semantic preference, semantic prosody and semantic sequence. I argue that in order to account for the broad phraseological profile a legal term it is advisable to refer to Sinclair's model of extended lexical units (Sinclair 1998, 2005). Terms should be considered as extended units of meaning including collocation, colligation, semantic preference and semantic prosody. Such extended units of meaning should draw upon both their immediate co-texts as well the wider context of the type of legal discourse in which they are found and the professional community which uses them. It is very instructive to see how phraseological analysis can identify the most salient discourses and overtly evaluative ways of referring to a highly technical legal term. This type of knowledge, not readily available in traditional lexicographic and terminographic resources seems indispensable for LSP users, especially translators and terminologists.

#### BIBLIOGRAPHY

- Bhatia Vijay K., Langton Nicola M., Lung Jane, 2004, «Legal discourse: Opportunities and threats for corpus linguistics» [in:] Ulla Connor *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing, pp. 203-231
- Biber Douglas., Stig Johansson, Geoffrey Leech, Susan Conrad, Finegan Edward, 1999, *The Longman Grammar of Spoken and Written English*, London: Longman.
- Biber D., Conrad S., Cortes V., 2003, «Lexical Bundles in Speech and Writing: an Initial Taxonomy» [in:] Andrew Wilson, Paul Rayson, Tony McEnery, *Corpus Linguistics by the Lune*, Peter Lang.
- Biel Łucja, 2012, «Areas of similarity and difference in legal phraseology: collocations of key terms in UK and Polish company law» [in:] Antonio Pamies, José Manuel Pazos Breña, Lucía Luque Nadal (eds.) *Phraseology and Discourse: Cross-Linguistic and Corpus-based Approaches*, Baltmannsweiler: Schneider Verlag, pp. 225-233.
- Black Henry Luce, Black Henry Campbell, 1990, *Black's Law Dictionary with Pronunciations*, New York: West Publishing Company.
- Bowker Lynne, 1996, «Towards a Corpus-Based Approach to Terminography» [in:] *Terminology*, 3.1: 27-52.
- Charles M., 2004, *The construction of stance: a corpus-based investigation of two contrasting disciplines*, Unpublished PhD thesis, University of Birmingham.
- Cook Guy, 1992, *The Discourse of Advertising*, London: Routledge.
- Cowie Anthony Paul, 1994, «Phraseology» [in:] Asher, R.E. (ed.) *The Encyclopedia of Language and Linguistics*, 3168-3171.
- Crystal David, Davy Derek, 1969, *Investigating English Style*, London: Longman.

- Evert Stefan, 2004, *The Statistics of Word Cooccurrences: Word Pairs and Collocations*, PhD dissertation, University of Stuttgart.
- Fletcher William, 2006, «Concordancing the web: promise and problems, tools and techniques» [in:] Marianne Hundt, Nadja Nesselhauf, Carolin Biewer, *Language and computers. Corpus Linguistics and the Web*. Pp. 25-45 (2). Rodopi.
- Gledhill Christopher, 2000, «The discourse function of collocation in research article introductions» [in:] *English for Specific Purposes*, 19, 115-135.
- Gotti Maurizio, 2003, *Specialised Discourse. Linguistic Features and Changing Conventions*, Bern: Peter Lang.
- Goźdź-Roszkowski Stanisław, 2011, *Patterns of Linguistic Variation in American Legal English. A Corpus-Based Study*. Peter Lang.
- Goźdź-Roszkowski Stanisław, 2007, «Pattern and meaning in judicial argumentation. An exploratory study » [in:] *Corpora and ICT in Language Studies* (Walinski et al. ed.), Łódź Studies in Language, Frankfurt/Mein Peter Lang.
- Goźdź-Roszkowski Stanisław, 2006a, «Recurrent Word Combinations in Judicial Argumentation. A Corpus-Based Study» [in:] *Langue, Droit, Société. Language, Law, Society* (Danuta Bartol, Anna Duszak, Hubert Izdebski, Jean-Marie Pierrel (eds), Nancy, 139-152.
- Goźdź-Roszkowski Stanisław, 2006b, «Frequent phraseology in contractual instruments. A corpus-based study» [in:] *New Trends in Specialized Discourse Analysis*, (Linguistics Insights. Studies in Language and Communication) Maurizio Gotti, David Giannoni (eds.), Frankfurt/Mein Peter Lang.
- Granger Sylviane, Meunier Funny (eds.), 2008, *Phraseology. An interdisciplinary perspective*, Amsterdam/Philadelphia: John Benjamins Publishing.
- Gries Stefan Thomas, 2008, «Phraseology and Linguistic Theory: A brief Survey» [in:] S. Granger, F. Meunier (eds.) *Phraseology: An Interdisciplinary Perspective*, Amsterdam/Philadelphia: John Benjamins, 3-25.
- Groom Nicholas, 2007, «Phraseology and epistemology in Humanities writing», Unpublished PhD thesis, University of Birmingham.
- Groom Nicholas, 2005, «Pattern and meaning across genres and disciplines: An exploratory study». *Journal of English for Academic Purposes*, 4, 257-277.
- Gustafsson Marita, 1984, «The Syntactic Features of Binomial Expressions in Legal English» [in:] 4 Text 123-132.
- Halliday Michael Alexander Kirkwood, Matthiesen Christian, 2004, *An Introduction to Functional Grammar*, 3<sup>rd</sup> edition. London: Arnold.
- Heid Ulrich, 2001, «Collocations in Sublanguage Texts: Extraction from Corpora» [in:] *Handbook of Terminology Management*, vol.2, Sue Ellen Wright, Gerhard Budin (eds.), Amsterdam/Philadelphia: John Benjamins Publishing.
- Hunston Susan, 2008, «Starting with the small words. Patterns, lexis and semantic sequences» [in:] *International Journal of Corpus Linguistics*, 13:3, 271-295.
- Hunston Susan, Gill Francis, 1999, *Pattern Grammar: a Corpus-driven Approach to the Lexical Grammar of English*, Amsterdam, Benjamins.
- Hyland Ken, 2008, «As can be seen: lexical bundles and disciplinary variation» [in:] *English for Specific Purposes*, 27 (2008) 4-21.
- Kjaer Anna, 1990, «Context-conditioned word combinations in legal language» [in:] *Terminology Science and Research*, vol 1, no.1-2: 21-32.
- Leech Geoffrey, 1974, *Semantics*, Harmondsworth: Penguin.
- Lee Debra S., Hall Charles, Hurley Marsha, 1999, *American Legal English. Using Language in Legal Contexts*, Ann Arbor: The University of Michigan Press.
- Lyons John, 1977, *Semantics*, Cambridge: Cambridge University Press.

- Meunier Fanny, Granger Susan (eds.), 2008, *Phraseology in Foreign Language Learning and Teaching*, Amsterdam/Philadelphia: John Benjamins Publishing.
- Moon Rosamund, 1998, *Fixed Expressions and Idioms in English: A Corpus-Based Approach*, Oxford: Clarendon Press.
- Nesselhauf Nadja, 2004, «What are collocations?» [in:] Allerton, D., N. Nesselhauf, P. Skandera (eds.) *Phraseological Units: Basic Concepts and their Application*, 1-21. Basel: Schwabe.
- Partington Alan, 1998, *Patterns and Meanings: using corpora for English language research and teaching*, Amsterdam: Benjamins.
- Pontrandolfo Gianluca, 2012, «Collocational Patterns in Translated vs. Non-Translated Criminal Judgments: A Case Study» [in:] Lew N. Zybatow, Alena Petrova, Michael Ustaszewski (eds.): *Translation Studies: Old and New Types of Translation in Theory and Practice*, (=Forum Translationswissenschaft 16). Frankfurt am Main u.a.: Peter Lang. 111-116.
- Renouf Antoinette, Sinclair John, 1991, «Collocational frameworks in English» [in:] Aijmer K, Altenberg B. (eds.) *English corpus linguistics: studies in honour of Jan Svartvik*, London: Longman.
- Römer Ute, Schulze Rainer (eds.), 2009, *Exploring the Lexis-Grammar Interface*, Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Schulze Rainer, Römer Ute, 2008, «Patterns, meaningful units and specialized discourses» [in:] *Intentional Journal of Corpus Linguistics*, 13:3, 265 – 270.
- Scott Mike R., 2008, *WordSmith Tools Help Manual*, Version 5.0 Liverpool: Lexical Analysis Software.
- Sinclair John, 2004, «Trust the text » (in:) J. Sinclair, R. Carter (eds.) *Trust the Text – Language, Corpus and Discourse*, 9-23. London: Routledge.
- Sinclair John, 1991, *Corpus, concordance, collocation*, Oxford, Oxford University Press.
- Sinclair John, 1987, *Looking Up: An Account of the COBUILD Project in Lexical Computing*, London: Collins.
- Stubbs Michael, 2009, «Technology and phraseology. With notes on the history of corpus linguistics» [in:] Ute Römer, Rainer Schulze (eds.) *Exploring the Lexis-Grammar Interface*. John Benjamins Publishing Company: Amsterdam/Philadelphia.
- Stubbs Michael, Barth Isabelle, 2003, «Using recurrent phrases as text-type discriminators. A quantitative method and some findings» [in:] *Functions of Language* 10.1, 61-104.
- Temmerman Rita, 2000, *Towards New Ways of Terminology Description. The Sociocognitive Approach*, John Benjamins Publishing Company: Amsterdam/Philadelphia.
- Tiersma Peter, 1999, *Legal Language*, Chicago: University of Chicago Press.
- Tognini-Bonelli Elena, 2001, *Corpus Linguistics at Work*, Amsterdam: John Benjamins.
- Wray Alison, 2002, *Formulaic Language and the Lexicon*, Cambridge: Cambridge University Press.
- Wüster Eugene, 1991, *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*, 3. Bonn: Romanistischer Verlag.

W POSZUKIWANIU WZORCÓW I ZNACZEŃ.  
METODY KORPUSOWE W ANALIZIE FRAZEOLÓGICZNEJ DISKURSU PRAWA

Streszczenie

Praca prezentuje różne podejścia w badaniu frazeologii dyskursu prawa w ujęciu korpusowym poprzez analizę angielskiego terminu *discovery*. Termin ten, niezwykle istotny dla amerykańskiej procedury sądowej, oznacza przedprocesowe przedstawienie materiału dowodowego. Analiza koncentruje się na badaniu kolokacji, zbitek wieloelementowych (n-gramów) oraz tzw. sekwencji semantycznej (*semantic sequence*) powiązanych z terminem *discovery*. Wyniki badania sugerują, że w miarę wyczerpujący i spójny profil frazeologiczny terminu można uzyskać tylko dzięki zintegrowaniu zastosowaniu różnych metod analitycznych.

**Słowa kluczowe:** frazeologia, terminologia prawnicza, językoznawstwo korpusowe, dyskurs prawa.