AGNIESZKA RZEPKOWSKA[1]
University of Siedlce, Poland
https://orcid.org/0000-0003-3381-0870

Employee, *worker*, *jobholder*, *agent*, *staff* and *workforce* in UK employment legislation: A genre-specific corpus study on synonymy, collocations and meaning

Abstract. In legal texts, synonymy may lead to confusion, especially if the synonymous words are terms which, by definition, should be unambiguous. This paper addresses the issue of synonyms in legal language through a genre-specific corpus study of *employee*, *worker*, *jobholder, agent*, *staff* and *workforce* – legal terms that appear similar in meaning – in the corpus of UK employment legislation. Specifically, the study looks at (a) the distribution of the terms in the corpus to determine the areas of law in which they are used, (b) the definitions of these terms in legal dictionaries, as well as general and business English dictionaries if the legal dictionaries fail to provide definitions, along with legal definitions from the 12 legislative documents constituting the corpus, (c) the immediate context of use (the co-text) to identify the most typical word combinations with the terms (candidate collocates), and (d) the differences between the terms based on the definitions and the collocational profile[2] of the terms. The findings suggest that, to some extent, the meanings of the terms overlap, indicating that they function as synonyms. However, they are not interchangeable in legislative acts as indicated by both their distribution in the corpus and their immediate context. Additionally, the study identified not only candidate collocations but also several multi-word terms defined within the legal acts.

Keywords: collocation, corpus, legal discourse, legal language, multi-word term, synonym, word combination

---

1   Address for correspondence: Instytut Językoznawstwa i Literaturoznawstwa, Uniwersytet w Siedlcach, ul. Żytnia 39, 08-110 Siedlce, Poland. E-mail: agnieszka.rzepkowska@uws.edu.pl

2   A collocational profile is understood here as the immediate linguistic context (or co-text) in which the words tend to appear in the corpus. This profile includes the most frequent word combinations of the words subject to rules determined further in the paper ("The study").

1. Introduction

Synonymy is a phenomenon generally not welcome in legal discourse, particularly where legislation is concerned. Lawyers typically associate each term with a distinct meaning and tend to avoid using synonyms. However, despite being frowned upon, synonyms, or near-synonyms, have nonetheless found their way into legal terminology and settled quite comfortably in legal texts (Klabal, 2022; Matulewska, 2016; Goźdź-Roszkowski, 2013).

From a pragmatic point of view, synonymy can be seen as a relation between two words in use – words that "map to the same meaning or concept" (Murphy, 2003, p. 145). Therefore, studying the context in which they appear may shed light on the differences between their meanings – sometimes subtle but still present and recognised by experts in a given field. The differences may not be obvious to non-experts for whom a dictionary definition, if available, may not be enough to illustrate the discrepancies in meaning. However, what can help illuminate these distinctions is seeing the words in their 'natural environment' – the texts in which they were originally used. In legal contexts, such texts are often legislative documents.

One avenue which can be taken to investigate the synonymous relations between words is to examine the other words that cooccur with the lexical units being studied. The nature of these cooccurrences may vary, depending on the strength of the connection between words and their mutual expectancy. A text can reveal a spectrum of different semantic relations between cooccurring lexical items, some loose, others more fixed – ranging from free combinations to collocations and ending with idioms (the most fixed of all three). Collocations, in particular, can provide valuable insights into the actual usage of words in both general and special-purpose languages. The fact that a word or a word combination collocates with a certain lexical item and not with others can illustrate part of its meaning. Legal language is no exception to this phenomenon.

Corpus tools are useful for identifying word combinations. They provide simple calculations of word frequencies and more complex statistical measures that make use of frequency counts. Some of the measures, such as MI-score, MI2-score, Delta P, Cohen's d and logDice – the latter used by default in Sketch Engine (Brezina, 2018, p. 70) – calculate the association measure between words in a corpus and are thus used to identify candidate collocations. Corpus analytics can also be used to compare words based on the lexical items they tend to co-occur with, helping to distinguish the use of synonyms in specific types of texts.

The paper explores whether genre-specific corpus studies could help answer questions about the use of near-synonymous legal terms in legal contexts using the example of *employee, worker, jobholder, agent, staff* and *workforce*. These six terms are semantically related and may be treated as synonyms or near-synonymous in the context of UK employment law. Specifically, the study investigates:

— whether genre-specific corpus studies can provide insights into the specific area of law in which each of the terms is used;

— whether the terms are synonyms and, if yes, how closely related they are; and

— whether a list of the most frequent word combinations with the terms (candidate collocates) compiled using corpus tools can reveal, or contribute to revealing differences between them.

Following a literature review on synonymy in legal language, legal discourse and collocations, this paper proceeds with the four steps listed below:

1. examining the distribution of the terms in the corpus to identify the areas of employment law where they are most frequently used,

2. finding and comparing dictionary definitions of the terms with their legal meanings as established in UK employment legislation;

3. analysing the immediate context in which the terms appear in the corpus by identifying the strongest word combinations (with candidate collocates), using the logDice score calculated by Sketch Engine; and

4. differentiating between candidate legal collocations and multi-word terms, and interpreting the meaning of the terms based on the word combinations found in the corpus.

2. Legal language and legislative documents

There is an inextricable connection between law and language. Everything that law encompasses needs to be expressed in language and vice versa: language is the primary means of understanding law. As law applies to and is applied in various walks of life, its language naturally varies in different contexts. Goźdź-Roszkowski aptly observes that "what is routinely referred to as 'legal language', represents an extremely complex type of discourse embedded in the highly varied institutional space of different legal systems and cultures" (Goźdź-Roszkowski, 2012, p. 1). He also elaborates on a range of categories of legal language discussed in the literature, including:

— frozen, formal, consultative and casual types of written texts (Danet, 1980, p. 371);

— judicial discourse, courtroom discourse, language of legal documents and the discourse of legal consultation (Maley, 1994, p. 16);

— the language of law (the language of legislation and contracts) and other uses of "legal language" (Trosborg, 1995, p. 31).

Similarly to Trosborg, Wróblewski distinguishes between two types of legal language: the language of the sources of law, that is legislature and contracts (in Polish *język prawny*), and the metalanguage of law, the language used, among others, by lawyers to talk about law (in Polish *język prawniczy*) (Wróblewski, 1948). Bhatia (2006) differentiates between the primary legal genre, which includes primary sources of law, and the secondary legal genre, which comprises a reproduction of legislation that exhibits a high degree of intertextuality with the primary sources[3]. De Groot proposes a similar division, focusing not on legal language in general but on legal vocabulary: (i) the vocabulary employed by legislators in acts of law, (ii) the vocabulary used by

---

3 Klabal (2019, p. 167) relates to them as higher-order genre and lower-order genre, respectively.

lawyers of a legal system and in commentaries on that legal system, and (iii) terminology found in general publications concerning that legal system (De Groot, 1996, p. 378). In light of this, we can classify legislative documents as written texts that use the language of law as defined by Trosborg and Wróblewski, featuring vocabulary specifically used by legislators.

Legal language is noted for its precision, indeterminacy, specialization, complexity and conservatism (Goźdź-Roszkowski, 2012). These features characterise both the primary and secondary legal genres. Precision is exemplified, among others, by the use of legal terms: words or phrases of specific meaning outlined in a given legal act. Indeterminacy allows lawyers to adjust the interpretation of a legal act to current circumstances. Specialization indicates that a text employs a language designed for special purposes. Complexity pertains, among others, to the intertextuality of legal texts in the continental systems of law or the all-inclusiveness of documents in the common law system. In English legal language, conservatism is manifested through traditional grammar, which includes old-fashioned words and phrases, such as modals and semi modals like *shall* and *may*, formal adverbs such as *whereby, therein, hereby,* etc., and the use of historically-formed expressions like doublets and triplets.

All those features make legal discourse very unique. On the one hand, precision is required, on the other, indeterminacy is also present. In the context of this paper, these two features deserve special attention. Seen, among other things, in a consistent use of legal terminology, precision appears to exclude the possibility of using absolute synonyms in a legal text, particularly in legislative documents. On the other hand, it is this same precision that paves the way for near-synonyms, or terms that semantically seem very close but differ contextually. The context here may refer to a given branch of law, legal genre, particular enactment, its section, paragraph or a single sentence. Indeterminacy may be employed by the legislator on purpose to allow a great deal of freedom in interpreting certain provisions when generalisation and flexibility of solutions are needed, and when neither precision nor determinacy is necessary or recommended (Kaczmarek, 2013, pp. 55–56).

There is no discussion about legal language without mentioning its core element: legal terms. In general, a term is a word or a word combination that represents a specific concept in a specific terminological system (Lukszyn, 2001, pp. 9–14). The main features of terms include:
— specialisation, which is evident in its use by specific users in specific situations in reference to specific objects;
— conventionality, which arises from the fact that terms are not formed in the course of a natural linguistic process but are the outcome of a purposeful action by a specific group of professionals;
— system-based meaning, which means that each term is part of a specific terminological system;
— accuracy and explicitness, which means that each term is defined in a specific way within a given professional context; and
— neutrality in terms of emotive and stylistic features (Kornacka, 2005, p. 131).

The status of a term in law-making acts comes from those documents where it is either directly or indirectly defined. The context of the legislative document can help to reconstruct the meaning, or the co-text of the term may indicate that the word or phrase has a specific system-based meaning (Rzepkowska, 2021, pp. 20–21). In addition to collocations or free combinations, multi-word terms are word combinations used in legal documents and their use is rarely incidental. However, distinguishing between a term and a non-term can be problematic as the difference between a term and a phraseme in legal texts may become obscure (Biel, 2012, p. 227) and requires an in-depth systemic and contextual analysis taking into account the features listed above.

An analysis of word combinations comprising terms must consider the phenomenon of nested terms. These are one-word or two-word terms that form part of longer terms (Marciniak & Mykowiecka, 2014a; Marciniak & Mykowiecka, 2014b) like in *fixed-term employee*, where both *employee* and *fixed-term employee* are legal terms in the relevant statutes[4]. Therefore, it is important to bear in mind that a term found in a corpus may also be part of larger multi-word terms, thus going beyond a typical collocation. It is possible to initially verify with special algorithms whether a given word combination is a term. For instance, when discussing nested terms, Marciniak and Mykowiecka suggest taking into account the frequency of occurrence of terms in isolation and nested, as well as the number of different contexts in which the whole nested phrase appears (Marciniak & Mykowiecka, 2014b, pp. 2–3). However, the final decision should be based on the context of use. In legal texts, the context often provides a definition of a term, offering a definite proof that a given phrase is a term. However, often does not mean always. The issue becomes more complicated when a legal term is not defined in a given statute or a number of statutes regulating a specific matter, but is still regarded by lawyers as a legal term. As Rycak observes in relation to Polish law on working time regulations, there are no definitions of a number of terms that the legislator uses when regulating working time. Additionally, the legislator uses certain terms illogically. All of that leads to numerous disputes on fundamental issues in legal academic writings (Rycak, 2008, p. 15). In such cases, determining whether a lexical item is a term requires a thorough analysis of its usage context, supported by in-depth, expert knowledge of the subject matter.

3. Synonymy

Synonymy has a long history of research in linguistics and has been characterised in various ways. Lyons distinguishes between complete synonymy and absolute synonymy. The former is defined as two items sharing descriptive, expressive and social meaning, and the latter as two items that feature "the same distribution and [being] completely synonymous in all their meanings and in all their contexts of occurrence" (Lyons, 1981, p. 148). While complete synonymy is rare, absolute synonymy, as defined in this way, is considered nearly non-existent in natural language

---

4 The example of *fixed-term employee* has been taken from the corpus under analysis.

as words are rarely fully interchangeable in all contexts. Clark argues that language eliminates absolute synonyms because, over time, one word in a synonymous pair tends to fall out of use or take on a slightly new meaning (Clark, 1992, p. 177; see also Ullmann, 1972, p. 141). Scholars also note a scale of synonymy, with some words being very close in meaning to one another and others being more distant (Cruse, 1986, pp. 265–266; Goźdź-Roszkowski, 2013, pp. 96–97).

English legal language provides us with a specific type of synonymy, which manifests in the phenomenon of doublets, triplets and even quadruplets (Crystal, 2006) in legal texts[5]. These are expressions representing very similar or complementary concepts, belonging to the same grammatical category and usually joined by *and* or *or* (based on Carvalho, 2008, p. 1; Bhatia, 1993, p. 108). Vázquez y del Árbol (2006, 2014) notes that they tend to appear particularly in notarial documents and testaments, though they are present in other types of texts as well, for instance contracts (Carvalho, 2008). Discussing doublets, Buşilă observes that "[e]ven though these constructions are etymologically validated, semantically – they become tautologies or pleonastic phrases: *ideas and opinions*, *null and void*, *defamatory or untrue*, *relevant and sufficient*, *unreasonably or arbitrarily*, *final and unappealable*, *costs and expenses*, etc." (Buşilă, 2016, p. 190) and they are used "more as an incantation than for any legal reason" (Buşilă, 2016, p. 190). Adams shares this view, adding that the use of certain strings of synonyms or near-synonyms like *sell, convey, assign, transfer, and deliver* in a purchase agreement may be an example of improvisation and result from "[finessing] the often-awkward task of selecting the best word for a given provision" (Adams, 2004, p. 204).

This phenomenon of "redundant synonyms" (Adams, 2004, p. 204) may serve as an example of synonymy rooted in legal practice and contrasts with the general opinion that synonymy in its extreme form is generally undesirable in legal practice. For example, Jopek-Bosiacka observes that:

> [i]n order for the law to function, the principle of semantic accuracy or language consistency must be observed. Once a technical term was selected, it must be repeated over and over again instead of using synonyms. The use of synonyms is discouraged in legal texts because the user might think that reference is being made to a different concept. (Jopek-Bosiacka, 2011, p. 16)

However, a number of studies, often carried out with corpus-analysis tools, show that synonymy of varying degree exists in legal contexts (Goźdź-Roszkowski, 2013; Matulewska, 2016; Cao, 2007). Researchers' interest in legal synonyms spans different areas of law, such as contract law (Biel, 2012) and competition law (Biel & Koźbiał, 2020); different legal systems, such as the American legal system (Goźdź-Roszkowski, 2013) and the EU legal system (Biel & Koźbiał, 2020); different areas in which synonyms are an issue, such as translation (Cao, 2007; Chroma, 2011; Jopek-Bosiacka, 2011) and intralingual synonymy (Matulewska, 2016); and different legal texts,

---

5   Bhatia (1993) refers to them as binomial and multinomial expressions.

such as notary acts (Vázquez y del Árbol, 2014; Buşilă, 2016) or agreements (Calvalho, 2008). The corpora used by researchers often consist of certain types of texts, such as legislative documents, or a collection of texts of different legal genres.

The phenomenon of synonymy in legal texts is usually defined in a very broad sense. Discussing legal language, Matilla explains that synonymy is when "two or several terms express the same concept" (Matilla, 2006, p. 144). In her study on the semantic relations between legal terms, Matulewska defines synonymy as "a semantic relation that binds two terms with the same referential meaning – but not necessarily the same pragmatic meaning – which belong to the same part of speech and differ in spelling" (2016, p. 163).

As synonymy can be viewed across a broad spectrum, there are numerous ways of classifying it in the literature. Matulewska distinguishes between interlingual synonyms, corresponding to equivalence as addressed in translation studies, and intralingual synonyms, which are words with similar meanings in the same language. She also lists different situations in which synonymous terms occur: (1) between different languages: (a) in vernacular and legal languages, (b) in legal and other special purpose languages; and (2) within legal language: (a) due to a lack of terminological consistency, (b) due to the passage of time (synonymous terms in diachronic perspective), (c) due to different text genre (legal-genre dependent synonymy), and (d) due to different branches of law (branch-of-law-dependent synonymy) (for examples of each type of synonyms see Matulewska, 2016, pp. 164–170). Basing her classification on Murphy (2003), Chromá discriminates between lexical synonymy (relations between lexical units) and propositional synonymy (relations between syntactic units) also referred to as paraphrase (Chroma, 2011, pp. 39–40). Klabal makes a distinction between synonyms across parts of speech and categorises synonyms into: adjectival synonyms, verbal synonyms, nominal synonyms and prepositional synonyms (Klabal, 2022, pp. 72–73).

A concept of synonymous relations often referred to by scholars is plesionymy. The term relates to situations where words are not exchangeable in all contexts; they are very similar but not identical in meaning, with their denotation, connotation, implicature, emphasis and register possibly varying (Edmonds & Hirsty, 2002, p. 107; Yevchuk, 2021, p. 204; Austin, 1962). Cruse (1986, pp. 285–286) categorizes plesionyms alongside absolute synonyms and cognitive synonyms, noting that among these three synonym categories, plesionyms are the least similar on the scale of similarity. In fact, plesionymy is useful for capturing various relations, like hypernymy and hyponymy, between different legal terms and ordinary words. Jopek-Bosiacka, writing about legal terms that have acquired their general meaning from ordinary words, gives such examples as *assault*, *battery*, *purchase*, or *domicile* (Jopek-Bosiacka, 2011, p. 11), showing that the way they are understood in legal language differs from their general use. Yet, they do remain synonymous to some extent, and thus can be treated as examples of hypernyms and hyponyms.

Legal synonyms are a subject of research in translation practice. Cao provides a number of examples of legal synonyms, and points out that they are not identical because what makes them different is their connotations (e.g. in the case of *encumbrance*, *mortgage*, *charge*, *pledge*,

lien), the type of legal writing they are used in and the area of law they pertain to (2007, pp. 71–73). In discussing legal translation practice, she illustrates that identifying and ascertaining the legal meaning of a word in relation to its general meaning takes place through an analysis of the context the word occurs in. "This includes both the wider legal context, such as a particular area of law, and the immediate linguistic context such as a sentence, the paragraph and the entire text in which the word is used" (Cao, 2007, p. 70). Examining the context is particularly important as there are terms in legal texts that originate from general language "but are assigned a special legal meaning by each legal system" (Jopek-Bosiacka, 2011, p. 10). Legal synonymy in the context of translation is also studied by Biel and Koźbiał, who explore near-synonymous legal terms in EU English-Polish competition law (Biel & Koźbiał, 2020). They investigate source-text synonymy in translation and conclude that "source-text synonymy causes variation and that, due to asymmetries between languages, it is difficult to control synonymy and standardise variants in translation" (Biel & Koźbiał, 2020, p. 87).

This short literature review brings us to a conclusion that synonymy is a gradable phenomenon, where instances of absolute synonymy are rare, particularly in legal language. More frequent are near-synonyms or plesionyms (as some scholars call them). Also, learning about synonymy entails investigating the context in which two or more lexical units appear. Lastly, despite being omnipresent, synonymy, particularly in its extreme form, is generally viewed as an undesirable phenomenon in legal discourse, and rarely allows for full interchangeability.

4. Collocations

So natural to native speakers and at the same time so hard to acquire in foreign-language learning, collocations are an essential component of every language. The term collocation is credited to Firth (1957) who first used it in his linguistic theory. Burkhanov states that the concept is used to "account for characteristic word combinations which have acquired an idiomatic, or rather semi-idiomatic, semantic relationship because of their frequent cooccurrence in the *context* (1), as *dog* and *bark*, *dark* and *night*" (Burkhanov, 1998, p. 39). Cruse, on the other hand, refers to collocations as habitually cooccurring "sequences of lexical items" (Cruse, 1986, p. 40), in which each constituting lexeme is semantically transparent as opposed to an idiom, the meaning of which is not a direct derivative of its constituents. The *Dictionary of Lexicography* defines collocation as "the semantic compatibility of grammatically adjacent words" (Hartmann & James, 1998, pp. 22–23). This definition pays attention to the patterns of cooccurrence of words such as adjective-noun, noun-verb and verb-preposition. Hartmann and James suggest that collocations should be viewed in opposition to idioms and free combinations as they are less fixed than the former and more fixed than the latter. Cowie takes a similar approach placing collocations next to idioms and quasi-idioms under the category of semantic phrasemes, treating them as the least fixed of the three types (Cowie, 1998, p. 30). The dividing lines between free combinations and collocations, and between collocations and idioms are sometimes blurry, with many phrases existing on the borderline between these categories.

Frequency as a means of assessing collocations was anticipated by Firth (1957/1968, p. 180) and Halliday (1961, p. 277). The increasing accessibility of corpus analytics offering various quantitative formulas in which frequency is the core element has recently made the frequency-based approach a feasible tool for finding and evaluating collocations. The calculation of the probability of cooccurrence and the resulting association measure is done with the use of such data as the number of tokens in the corpus, the frequency of the node, the collocate and the collocation as a whole, i.e. the node and collocate, in the whole corpus and the collocation window size (Brezina, 2018, p. 70). On the other hand, the phraseological approach views collocation as an association of lexemes that tend to occur in certain grammatical constructions. The meaning of such word combinations should be transparent (derivable from the meanings of the constituting lexemes), which distinguishes them from idioms. They are lexically variable but the selection of individual components, that is the collocates of the node, can be restricted at one or more points (Cowie, 1994, pp. 115–116; Sinclair, 2005), which makes them also different from free combinations. Mitigating the dispute concerning which approach to collocations is best, Michta and Mroczyńska (2022, p. 14) justly note that each of the approaches answers different questions and thus they should not be perceived as two opposing methods but as partners walking hand in hand. Hence neither should be called "empty" (Siepmann, 2005, p. 411) because if one of them is, the other may be as well.

5. The study

To shed some light on the information that can be reconstructed from the collocational profile of legal terms, this analysis will focus on six words found in UK employment legislation: *employee*, *worker*, *jobholder, agent*, *staff* and *workforce*. These terms will be examined alongside their synonymous relationships based on definitions from both dictionaries and the legal contexts in which they appear. As the selected lexical units are nouns, they will be studied from the viewpoint of lexical synonymy, particularly nominal synonymy. Nominal synonyms in legal language are "often terminological in nature and usually cannot be used interchangeably, or may also be a case of collocations or jurisdictional variation" (Klabal, 2022, p. 72). Therefore, I do not expect to find examples of absolute synonymy (Lyons, 1981) in this analysis. The question is whether there are instances of complete synonymy or near-synonymy among these terms.

Considering the type of texts in the corpus, comprising 12 UK statutes on various employment-related issues, the analysis is also expected to yield findings on legal-genre-dependent synonymy. That is in line with Lyons' assertion that "frequent use of a word or phrase in one range of contexts rather than another tends to create a set of associations between that word or phrase and whatever is distinctive about its typical contexts of cooccurrence" (Lyons, 1981, p. 150). In this research, the context is understood in two ways: as the whole text in which the terms are found, and the very close environment in which the terms appear and where their collocates are found – also referred to by scholars as co-text (Halliday, 1999, p. 3). A review of word combinations – ranging from free combinations to relatively fixed specialised collocations

and multi-word terms – serves as a means to understand their typical context as employed by the legislator.

5.1. Data and methods

5.1.1. The corpus

This paper uses the term "corpus" in the sense proposed by Sinclair as "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (Sinclair, 2005, p. 16). The corpus tool selected for the analysis is Sketch Engine[6] which processes texts to present results in the form of Word Sketches, concordances and word lists, among others. This research takes advantage of all these functions to learn about the collocations of words under analysis and unlocks the potential of context as a medium for filling gaps in understanding differences between near-synonyms.

The study relies on the author's corpus of legal texts compiled according to the predefined criteria, including the legal system, legal domain and type of texts. The corpus comprises legislative documents governing employment relations in the United Kingdom. These documents were selected based on the UK government websites[7] and commercial legal websites[8] offering information and advice on employment regulations in the UK. The laws referred to by experts as the most important where subsequently found on the official legislation website, which includes all enacted legislation for the UK, Scotland, Wales and Northern Ireland[9]. The final corpus consists of 12 documents, 1.2 million tokens[10] and over 760 thousand words[11] (Table 1).

Table 1. UK Employment Law corpus composition

Legislative document	Tokens	Percentage of the corpus
Trade Union and Labour Relations (Consolidation) Act 1992	303,082	25%
Employment Rights Act 1996	278,938	23%

---

6   See https://www.sketchengine.eu.

7   See https://www.gov.uk/browse/employing-people/contracts.

8   See https://croner.co.uk/resources/employment-law/legislation-list/
and https://www.expatica.com/uk/working/employment-law/employment-law-uk-104502/.

9   See https://www.legislation.gov.uk.

10 Sketch Engine defines a token as the smallest unit that a corpus consists of. A token may refer to: a word form, punctuation, a digit, abbreviations, and anything else between spaces (https://www.sketchengine.eu/my_keywords/token/).

11 Sketch Engine defines a word as a type of token which begins with a letter of the alphabet (https://www.sketchengine.eu/my_keywords/word/).

Legislative document	Tokens	Percentage of the corpus
Equality Act 2010	203,351	17%
Health and Safety at Work etc. Act 1974	160,553	13%
Pensions Act 2008	114,655	9%
National Minimum Wage Act 1998	49429	4%
Working time regulations 1998	40,501	3%
Agency Workers Regulations 2010	22,465	2%
Transfer of Undertakings (Protection of Employment) Regulations 2006	14,796	1%
The Fixed-term Employees (Prevention of Less Favourable Treatment) Regulations 2002	9,772	1%
The Maternity and Parental Leave etc. Regulations 1999	9,175	1%
Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000	6,959	1%
Total	**1,213,676**	**100%**

Table 1 lists legislative documents under analysis dealing with the issue of employment relations in the UK, arranged according to their size. It is worth noting that the two top documents are particularly large, making up nearly half of the corpus (48%). They are followed by three that constitute another 39%. The remaining 13% comprises seven acts that are much shorter, ranging from 4% to 1% of the corpus.

5.1.2. Selection of words for analysis

The study covers the legal term *employee* and its synonyms in legislative documents governing employment relations in the United Kingdom. This term was selected for analysis because it is one of the fundamental terms underlying employment relations in the UK legal system. A simple search using the Sketch Engine Wordlist tool, which excludes non-words, revealed that *employee* ranks 23rd in a list of nouns, positioned behind *employment* (10th) and *employer* (19th). Other nouns that ranked higher than *employee* were mainly words responsible for text organisation, such as *paragraph*, *section*, etc.

A review of the corpus made it possible to single out five synonyms of the term *employee*. The search for synonyms relied on a list of synonyms found in online thesauruses: thesaurus.com[12],

---

12 employee. (n.d.) *thesaurus.com*. `https://www.thesaurus.com/browse/employee`.

Cambridge thesaurus[13] and Collins thesaurus[14]. All nouns and noun phrases listed there as synonyms were searched for in the corpus. Synonyms with the absolute frequency in the whole corpus exceeding 50 were subject to further analysis. Only five words met these criteria: *worker, jobholder, agent, staff* and *workforce*.

The classification proposed by Benson et al. (2009, p. XXXI-XXXIV) has served as an inspiration for classifying the collocations found in the corpus. However, it was modified to meet the specific objectives of this study by including the relation of possession. The discussed classification presents different configurations of the node and collocate; the node is one of the terms under analysis: *employee, worker, jobholder, agent, staff* and *workforce*. The collocate can be a verb, adjective or noun (a premodifier), or a different noun (a possessive relation). The following groups of word combinations are further analysed:
— premodifier + NODE,
— NODE + noun,
— verb + NODE (object),
— NODE (subject) + verb, and
— NODE's+ noun.

Since the selected terms are nominal synonyms and as such may be terminological in nature and not interchangeable, it is assumed that the selected words are not freely used by the legislator, but their choice is dictated by their specialist meaning. Additionally, they are not examples of jurisdictional variation as all the texts in the corpus come from the UK legal system.

5.2. Dictionary and legal definitions of terms under analysis

This part of the paper covers an analysis of dictionary and legal definitions of the words. Legal definitions were looked for in two legal dictionaries: *The Penguin Dictionary of Law* (Webb, 2010) and *A Dictionary of Law* (Martin, 2003). Words which were not found in those dictionaries were looked up in an online general English dictionary and a business English dictionary: *Collins English Dictionary*[15] and *Cambridge Business English Dictionary*[16].

A review of selected dictionaries showed that three out of six words under analysis have an entry in legal dictionaries, which proves their status as well-established legal terms. These are *employee, worker* and *agent*. The remaining three, *jobholder, staff* and *workforce*, seem not to be terminographically-recognised legal terms as they are found in non-legal dictionaries only.

A review of the statutes (the corpus) shows that not only are *employee, worker,* and *agent* legal terms, but so are the other three terms under analysis because they have been defined by

---

13 employee. (n.d.) `dictionary.cambridge.org/thesaurus/`. `https://dictionary.cambridge.org/thesaurus/employee`.

14 employee. (n.d.) *Collinsdictionary.com*. `https://www.collinsdictionary.com/dictionary/english-thesaurus/employee`.

15 Available at `https://www.collinsdictionary.com/`.

16 Available at `https://dictionary.cambridge.org/`.

the legislator, either directly or indirectly. What should be stressed is that such legal definitions are usually applicable only within the act of law in which they appear unless otherwise stated, or in another act when there is a direct reference to that other document. It is common to find phrases such as *in these regulations* appearing next to the definitions, and phrases like *has the meaning assigned by section*, *within the meaning of* or *has the meaning given in*[17] next to terms, which clearly delimits the use of the definitions and clarifies the meaning of the terms.

Employee is a word encountered in both legal dictionaries reviewed. The term is defined as a person working under the direction or control of another for a wage, salary, fee or other payment. The relationship between the employee and the employer is governed by an employment contract. Employees enjoy special rights and are protected by certain laws that no contract of employment may deprive them of (this is the most protected form of employment under employment law). This definition of *employee* seems to be based on the legal definitions found in the statutes (for instance in the Employment Rights Act 1996).

Worker is usually defined in legal dictionaries as a person employed to do work for another under an employment contract or any other contract. On the other hand, *A Dictionary of Law* (Martin, 2003) does not provide a definition of *worker* but equates the word directly with *employee*, thus making *worker* a close synonym of *employee* in legal terms. Legal sources are more specific. For example, the Employment Rights Act 1996 states that the worker performs personally work or service, determines the form of the contract (oral or written) and differentiates a worker's contract from the relationship with a client or customer.

Agent is defined as a person appointed by another (the principal) to act on his/her behalf to perform a service, usually to negotiate a contract between the principal and a third party (Martin, 2003). Such a definition of an *agent* makes him/her a type of a representative. He or she is subordinate to the principal, the same as an employee to the employer, yet his or her liability and rights largely depend on the type of agreement with the principal, which makes an agent different from an employee. *Agent* is defined either directly or indirectly in a few statutes. In two cases, Employment Rights Act 1996 and Equality Act 2010, an agent is a type of a middleman, facilitating the conclusion of an agreement. On the other hand, an agent in the context of trade unions is understood as "a banker or solicitor of, or any person employed as an auditor by, the union or any branch or section of the union" (Section 119 of the Trade Union and Labour Relations (Consolidation) Act 1992). That meaning in fact mentions an employment relationship between an agent and trade unions, in which an agent can be an employee (an employed auditor).

Jobholder is not found in legal dictionaries. In general English dictionaries it is defined as someone who has a regular post of employment in an organisation. *Jobholder* is also a legal term of very narrow applicability (encountered in one statute, that is the Pension Act 2008). There it is precisely defined as a worker who is working or ordinarily works in the UK under the

---

17 The examples have been taken from the corpus under analysis.

worker's contract, who is aged at least 16 and under 75, and to whom qualifying earnings are payable by the employer in the relevant pay reference period.

Similarly *staff* has no entry in legal dictionaries. Its general English and business English meaning is all people working for a particular company or in a particular place. *Staff* is indirectly defined in a few statutes. The National Minimum Wage Act 1998 and Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000 provide us with definitions of *a relevant member of the House of Commons staff* and *a relevant member of the House of Lords staff*. The former means "any person who is employed under a worker's contract with the Corporate Officer of the House of Lords" (National Minimum Wage Act 1998), the latter "any person (a) who was appointed by the House of Commons Commission; or (b) who is a member of the Speaker's personal staff" (National Minimum Wage Act 1998, Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000). Based on the above we can conclude that in the studied corpus a member of staff is a person working in the UK Parliament, either employed or appointed. Therefore, there is a semantic relation between *a member of staff* and *an employee* relying on the subordinate nature of the two and the element of work performed by them.

Workforce is the third word that is not taken as a legal term by those compiling legal dictionaries. Both general and business English reference books define it in two ways: firstly, as the total number of people in a particular area, e.g. a country, who are available for work, and secondly, as the total number of people who are employed by a particular company or organisation or who are engaged in a specific activity. *Workforce* is indirectly defined in the corpus through the term *relevant members of the workforce* understood as "all of the workers employed by a particular employer, excluding any worker whose terms and conditions of employment are provided for, wholly or in part, in a collective agreement" (Working Time Regulations 2010). Further in the documents we find that *representatives of the workforce* "are workers duly elected to represent the relevant members of the workforce". Similarly to *jobholder*, *workforce* is a legal term of narrow applicability in UK employment law.

To sum up, the definitions reviewed indicate that the six terms seem to be synonyms with varying degrees of similarity, but they appear not to be interchangeable in the legal context. *Employee*, *worker*, *jobholder*, *agent*, *staff* and *workforce* share certain characteristics, among others the fact that they refer to individuals who perform work for a hiring person under an employment contract or other type of a contract (they are bound by contractual relations); they are also subordinate to someone. A special case exists for *staff* and *workforce*, which are collective nouns that refer to certain groups of workers or employees.

5.3. The six terms in the corpus

The distribution of the terms under analysis in each legislative document reflects the real use of the terms in the contexts of individual employment law acts (Table 2). *Employee* and *worker*

stand out in terms of frequency in the studied sample, amounting to nearly 2,000[18] per million tokens each, while the remaining four range from 175 in the case of *jobholder* to 47 in the case of *agent*.

Employee is a term preferred over *worker* in five acts, and *worker* is favoured over *employee* in seven acts. *Jobholder* stands out in this group as it is used only in one act from the corpus, the Pension Act 2008. *Workforce* is found in 6 acts out of 12. The words *agent* and *staff* are encountered nearly in the whole corpus, apart from the Maternity and Parental Leave etc. Regulations 1999 (where only *employee* and *workforce* are used).

Table 2. Distribution of *employee, worker, jobholder, agent, staff* and *workforce* and their frequencies in the corpus

Legislative document	employee	worker	jobholder	agent	staff	workforce
1. Trade Union and Labour Relations (Consolidation) Act 1992	371	510	0	7	43	0
2. Employment Rights Act 1996	1,355	785	0	12	32	7
3. Equality Act 2010	53	58	0	12	16	0
4. Health and Safety at Work etc. Act 1974	36	0	0	1	7	0
5. Pensions Act 2008	37	118	213	6	11	0
6. National Minimum Wage Act 1998	35	190	0	10	11	0
7. Working time regulations 1998	17	325	0	5	16	28
8. Agency Workers Regulations 2010	51	286	0	1	9	1
9. Transfer of Undertakings (Protection of Employment) Regulations 2006	169	0	0	1	2	7
10. The Fixed-term Employees (Prevention of Less Favourable Treatment) Regulations 2002	145	5	0	1	9	18
11. The Maternity and Parental Leave etc. Regulations 1999	114	0	0	0	0	21

18 All frequency numbers refer to singular and plural forms taken together.

Legislative document	employee	worker	jobholder	agent	staff	workforce
12. Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000	20	94	0	1	13	0
Absolute frequency	2,403	2,371	213	57	169	82
Average frequency per million tokens	1,980	1,954	175	47	139	68

5.3.1. Free combinations, specialised collocations and multi-word terms

This part of the paper presents the main word combinations with *employee*, *worker*, *jobholder*, *agent*, *staff* and *workforce* in the corpus, including free combinations, collocations and multi-word terms. Candidate collocations in the direct linguistic context are not verified in terms of their strength, and some may be considered to be merely free combinations. Yet, their existence in the corpus adds to the candidate collocational profile of the terms and may shed some light on their meaning. Therefore, they are referred to as collocations in this research. Main collocates are defined as those at the top of the lists in word sketches, sorted by logDice score[19] with an absolute frequency of at least 5 in the case of *employee* and *worker*. No frequency threshold is set for the remaining four nouns due to their relatively low frequency and a limited number of possible collocates.

Table 3 presents the collocates of *employee* and *worker* in five groups, based on the grammatical relationship between the node and collocate. The collocates in each group are listed according to the logDice score, starting with the highest. The selected collocates feature an absolute frequency of at least 5. Words that collocate with both *employee* and *worker* have been bolded.

Table 3. Collocates of employee and worker in the corpus (with an absolute frequency exceeding 5)

NODE	EMPLOYEE	WORKER
premodifier + NODE	affected, fixed-term, **comparable**, permanent, pre-TURERA, deceased, existing, **relevant**	young, part-time, opted-out, full-time, **comparable**, betting, protected, offshore, **relevant**, former, mobile, agency, shop, night, applicant, contract, zero hours, home

---

19 logDice is a statistic measure for identifying co-occurrence. Sketch Engine applies it to identifying collocations as its value indicates the typicality (or strength) of the collocation based on the frequency of the node and the collocate and the frequency of the whole collocation. Theoretically, the higher the score the more typical the collocation. The maximum logDice score value is 14. (https://www.sketchengine.eu/my_keywords/logdice/).

NODE	EMPLOYEE	WORKER
NODE + noun	representative, shareholder, share, information	
verb + NODE[20]	**entitle**, dismiss, select, **employ**, **permit**, **require**, concern, **allow**, **pay**, represent, engage, exclude, involve, qualify, reinstate, suspend, leave, **treat**, notify, re-engage, assign, lay off, affect	**employ**, **entitle**, **require**, **pay**, subject, **permit**, provide, propose, represent, involve, **treat**, **allow**, inform, give, affect
NODE + verb[21]	satisfy, **work**, intend, exercise, propose, die, hold, terminate, **refuse**, elect, sustain, present, return, follow, start, apply	**work**, constitute, fall within, carry out, signify, believe, receive, join, complete, remain, become, **refuse**
NODE'S + noun	**contract** (of employment), **employment**, period (of employment), entitlement, remuneration, right, application, death, notice	**contract**, wage, employer, **employment**, year, agreement

The group of collocations where the collocate modifies the node contains a large number of multi-word legal terms that are either directly or indirectly defined in the statutes and used as independent terms throughout individual documents: *fixed-term employee*, *permanent employee*, *pre-TURERA employee*; or in the case of worker: *agency worker*, *shop worker*, *young worker*, *part-time worker*, *opted-out worker*, *night worker*, *applicant worker*, *contract worker*, *zero-hours worker*, *offshore worker*, *mobile worker*, *home worker*. They are examples of terms nesting other terms; in this case the nested terms are *employee* and *worker*. The expressions *affected employee*, *comparable employee*, *deceased employee*, *existing employee*, and *comparable worker*, *protected worker* and *former worker* seem to be collocations (as legal definitions of these collocations have not been found in the corpus, it has been assumed that they should not be given the status of a term).

Both multi-word terms and specialist collocations are also found in the group where the node modifies a noun. The expression *employee shareholder* is a legal term, while the remaining three collocates form specialised collocations of varying strength.

Verb collocates serve as a source of information about the meaning of the nodes as the use of a given verb attributes certain characteristics to the node or indicates that certain characteristics have been attributed to the node by the legislator. An examination of the verbs with *employee* and *worker* as objects reveals that these terms should be at least to some extent synonymous, as both refer to individuals who are *employed* (which means they are involved

---

20 The verb *be* has been excluded from the list as being too common to form a typical collocation without an additional object that would make it specific.

21 The verbs be, *have, take, give, do* and *make* have been excluded from the list as being too common to form a typical collocation without an additional object that would make them specific.

in a contractual relationship in which they occupy a subordinate position); *entitled to, permitted*, and *allowed* something (which means have some rights); *paid* (which means they receive remuneration for what they provide); *required to* do something (which means that they do not act freely); and *affected* by something (which means that they are influenced by their broadly-meant environment). The information conveyed by these verbs aligns with the definitions of *employee* and *worker* presented earlier in this paper. Interestingly, an *employee* is much more frequently *dismissed* and *engaged* in something (based on the absolute frequency); and only an *employee* is *laid off, suspended, left, reinstated* (all the verbs relate to either employment or the post taken); and *assigned* something. Conversely, only a *worker* is *subjected to* something, *provided with* something, *proposed* something and *informed* about something. Additionally, the verb *to employ* is used twice as often with *worker* as with *employee*.

The group of collocations in which *employee* and *worker* are the subjects illustrates how these two types of individuals can act. Both of them can *work* and *refuse to* do something (for instance to work). However, the other verbs are typically associated with only one of the two. An *employee* has the power to *exercise* rights, *satisfy* needs, *intend, elect, apply, hold*, for instance a position, and *terminate* an agreement (in other words express his/her will). An *employee* needs to *return* something or somewhere and *follow* procedures.[22] A *worker* can *constitute* or *fall within* a group (being treated as an element of a group not as an individual), *join* a group, *signify* and *believe* (generally speaking express his/her opinion) and *remain* or *become* somebody else (for instance a *jobholder*). A review of these particular verbs shows that the term *employee* refers to an individual enjoying a number of rights and at the same time submitted to certain rules imposed on him/her. On the contrary, the term *worker* means a person belonging to a group and treated as its component – verbs indicating such a meaning prevail in terms of frequency and form the strongest collocations. The fact that *workers* are also individuals is (seemingly) secondary as the use of verbs shows. This is additionally stressed by the observation that a *worker* is not said to die, only the *employee* is.

The very last group of collocates with *employee* and *worker's* "possessions" shows that both have a *contract* and *employment*. Yet, only an *employee* has an *entitlement* and *right* to something. Another difference is the type of payment; for example, an *employee* receives *remuneration* and a *worker, wages*.

To sum up, the logDice score and absolute frequency results obtained with the word sketch feature brought interesting results in terms of multi-word terms and specialised collocations of *employee* and *worker*. The noun collocates do not contribute significantly to the meaning of

---

22 Worth comparing is Michta and Mroczyńska's dictionary of legal English collocations based on the UK Supreme Court judgements (Michta & Mroczyńska, 2022), where the term *employee* is a separate entry (one of 100 included therein). Despite the fact that the dictionary is based on US law, the list of collocates of *employee* coincides to a large extent with the one drawn up here, which indicates the typicality and strength of the word combinations.

the nodes. They appear to be a source of multi-word terms nesting the studied terms. However, verb collocates illustrate the meanings of *employee* and *worker* relatively well as they demonstrate what can be done to an employee and worker and what an employee and worker can do. Based on this analysis it seems reasonable to conclude that the shared meaning of *employee* and *worker* is reflected in the use of the same verb collocates, and where the collocates differ, the meaning of the terms may differ in that scope as well.

Table 4 presents the collocates of *jobholder, agent, staff* and *workforce,* which form much less representative groups than those of *employee* and *worker* owing to the low frequency of those words (see Table 2). The frequency of individual collocations formed by those lexical units is also relatively low.

Table 4. Collocates of jobholder, agent, staff and workforce in the corpus

NODE	JOBHOLDER	AGENT	STAFF	WORKFORCE
premodifier + NODE	relevant	estate	House of Commons, House of Lords, Speaker's personal, part-time, requisite, Parliamentary, academic, prison, mobile, full-time, comparable	-
NODE + noun	-	-	cost, overheads, negotiations, references	agreement
verb + NODE[23]	induce, become	authorise	appoint, recruit, exist, exclude, employ	entitle
NODE + verb[24]	become, cease, opt out of, remain, pay, authorise	-	restrict, provide, include	
NODE'S + noun	request membership, employer, right	-	Speaker, officer	-

The contextual relations in which *jobholder* is found are limited. There are no noun collocates apart from those where *jobholder* is followed with the Saxon Genitive. This specific group of collocates illustrates a relation to the terms *employee* and *worker* (the collocate *right* is shared with *employee* and the collocate *employer* with *worker*). Based on the verb collocates one can

---

23 The verb *be* has been excluded from the list as being too common to form a typical collocation without an additional object that would make it specific.

24 The verbs *be, have, give* and *do* have been excluded from the list as being too common to form a typical collocation without an additional object that would make them specific.

deduce that certain aspects of a *jobholder's* existence depend on his/her will (*authorise, opt out of*) and some are imposed on a *jobholder* (*cease, remain*).

As for the term *agent*, the corpus provides even fewer examples. Indeed, there is only one noun collocate *estate* and one verb collocate *authorise*. The former together with *agent* form a legal term defined in the Equality Act 2010.

Similarly, few collocates are listed under *workforce*, but still some similarities with *employee* and *worker* are evident: the fact that there is an *agreement* to which *workforce* is a party, and that *workforce* is *entitled* to something. In fact, a *workforce agreement* is a legal term in the corpus and means an agreement between an employer and his employees or their representatives. The definition implies that *workforce* is a group of employees and shows the semantic relation between this term and *employee*.

Staff, on the other hand, takes on quite a specific meaning based on the context provided in the corpus. A look at the premodifiers of *staff* shows that it primarily relates to the workers of the UK Parliament, including its two houses and the Speaker. Thus, *staff* directly relates to a particular organisation as defined above. *Part-time* and the collocates listed further in that part of the table show low absolute frequency (under 3). Yet the fact that such collocates as *part-time*, *mobile*, *full-time* and *comparable* are used with *staff* indicates that the term shares some characteristics with *employee* and *worker*, as the same collocates are also used with *employee* and *worker*. The legislator's verb selection suggests that *staff* is subordinate to another entity that can *appoint*, *recruit* or *employ* it (in other words bring into existence) and also can *exclude staff* from something.

6. Findings and conclusions

While there have been a number of corpus-based studies on legal terminologies, synonymy in legal language and collocations in legal discourse, this study focuses on the field of law that has not been in the centre of attention in terminological research so far. Employment law varies depending on the legal system and the country in which it applies. The UK legal language of employment law has its special features, including its typical terminology. The objective of this study was to empirically verify whether the terms *employee*, *worker*, *jobholder*, *agent*, *staff* and *workforce* are synonyms based on their dictionary and legal definitions, and whether their distant and immediate context of use can help differentiate between them.

The findings show that the context and co-text provide significant insights into the meaning of the terms under analysis. Although the sample corpus is of moderate size, the word combinations found in it allow us to draw certain conclusions as to the meaning of individual terms and the relations between them.

The study of the distribution of these terms suggests that they are applied in different areas of employment law, and thus the terms are not interchangeable. They are unevenly distributed over the 12 statutes. First of all, the frequency of the terms differs: *employee* and *worker* appear most often, while the remaining four are less frequent: from 175 (*jobholder*) to 47 (*agent*). The

distribution of the terms varies as well: *jobholder* is used only in one act, which shows that its applicability is limited to that area of law; *workforce* is found in six statutes but in three it is much more frequent than in others; *employee* is present in all the legislative documents, but its frequency in each varies and the variability is not proportional to the document size. In some cases, *worker* seems to be preferred over *employee*; *agent* and *staff* are scattered around the whole corpus and do not appear only in one act, where only *employee* and *workforce* are used.

The dictionary definitions of the terms from legal dictionaries, a business dictionary and a general English dictionary as well as the legal definitions found in the 12 statutes indicate that the terms are partial synonyms, spread along the scale of synonymity. The closest synonyms are *employee* and *worker*, but the legislator did not use them interchangeably. The other four, *jobholder*, *agent*, *staff* and *workforce*, are more distant in meaning from *employee* than *worker*, but still they semantically overlap, as demonstrated above. *Agent* appears to be the most distant in meaning, based on its dictionary and legal definitions.

The list of words (nouns, verbs and adjectives) that most frequently collocate with the terms, as identified with corpus analysis tools, has proven useful for understanding the differences between the terms under analysis. In fact, the collocational profiles of the terms seem to reflect at least part of their dictionary meanings. However, not all collocates contribute to that end equally. Most information can be deducted from combinations with verbs, which explicitly indicate what an employee, worker, jobholder, agent, staff and workforce can do and what can be done to them. As a result, it is possible to see elements of the definitions of the terms in their verb collocates. In some cases, the word combinations with the terms even extend the definitions with additional semantic information, thereby contributing to a deeper understanding of the concepts the terms represent. For example, the study of verb combinations with *employee* and *worker* showed that an *employee* is treated more as an individual, whereas a *worker* seems to be perceived as part of a larger whole.

Additionally, the study of word combinations has shown that the terms do not appear to be interchangeable based on their co-text, as the collocates they appear with differ. The extent to which their collocates overlap coincides with their shared meanings, while collocates that seem to be typical to each term suggest differences in meaning. Additionally, word combinations with nouns often form multi-word terms with the terms under analysis, which can serve as examples of nested terms. They usually stand for a specific group of employees, workers, jobholders, etc.

The study has shown that the legislative use of these terms is restricted by context, both immediate and wide, of a given paragraph, section or enactment. This study opens new avenues for exploration. It would be worth investigating how these terms behave in the secondary legal genre and whether they tend to lose at least part of the meaning assigned to them in the statutes when used in the lower-order genre. The focus could also shift towards general language, following the line of research already initiated by other scholars (L'Homme & Azoulay, 2020; Michta, 2022). This could involve comparing the specialised collocations of the six terms with those found in a general language corpus in order to learn more about the typicality of these phrases and their use and meaning in discourse outside the legal domain.

References

Adams, K.A. (2004). *A manual of style for contract drafting*. ABA Publishing, Section of Business Law.

Austin, J. (1962). *How to do things with words*. Oxford University Press.

Benson, M., Benson, E. & Ilson, R. (2009). *The BBI combinatory dictionary of English: your guide to collocations and grammar* (3rd edition). John Benjamins.

Bhatia, V. K. (2006). Legal genres. In K. Brown (Ed.), *Encyclopaedia of language and linguistics* vol. 7 (pp. 1–7). Elsevier.

Bhatia, V.K. (1993). *Analysing genre: language use in professional settings*. Longman.

Biel, Ł. (2012). Areas of similarity and difference in legal phraseology: collocations of key terms in UK and Polish company law. In A. Pamies, J. M. P. Bretaña & L. L. Nadal (Eds.), *Phraseology and discourse: cross-linguistic and corpus-based approaches* (pp. 225–233). Schneider Verlag.

Biel, Ł, & Koźbiał, D. (2020). How do translators handle (near-)synonymous legal terms? A mixed-genre parallel corpus study into the variation of EU English-Polish competition law terminology. *Estudios de Traducción, 10*, 69–90. DOI:10.5209/estr.68054.

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899.

Burkhanov, I. (1998). *Lexicography. A dictionary of basic terminology*. Wyższa Szkoła Pedagogiczna w Rzeszowie.

Buşilă, A. (2016). The issue of quasi-synonymy in the translation of notary acts from English in Romanian. *ANADISS*, *21*, 182–193.

Cao, D. (2007). *Translating law*. Multilingual Matters LTD.

Carvalho, L. (2008). Translating contracts and agreements: a corpus linguistics perspective. *Culturas Jurídicas*, *3*(1), 1–15.

Chroma, M. (2011). Synonymy and polysemy in legal terminology and their applications to bilingual and bijural translation. *Research in Language, 9*(1), 31–50.

Clark, E. V. (1992). Conventionality and contrast: Pragmatic principles with lexical consequences. In A. Lehrer, E.F. Kittay, & R. Lehrer (Eds.), *Frames, fields, and contrasts: new essays in semantic and lexical organization* (pp. 171–188). Routledge.

Cowie, A. P. (1994). Phraseology. In R.E. Asher (Ed.), *The encyclopaedia of language and linguistics* (pp. 3168–3171). Pergamon Press.

Cowie, A. P. (1998). *Phraseology. Theory, analysis, and applications*. Clarendon Press.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.

Crystal, D. (2006). *Stories of English*. Penguin.

Danet, B. (1980). Language in the courtroom. In H. Giles, W. P. Smith & P.M. Robinson (Eds.), *Language: Social and psychological perspectives* (pp. 367–376). Pergamon.

De Groot, G.-R. (1996). *Guidelines for choosing neologisms.* In: *Translation and meaning: Part 4* (pp. 377–380). Rijkshogeschool Maastricht.

Edmonds, P. & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics*, *28*(2), 105–144.

Firth, J. R. 1957/1968. A synopsis of linguistic theory, 1930–1955. In F.R. Palmer (Ed.), *Selected Papers of J.R. Firth 1952–1959* (pp. 168–205). Indiana University Press.

Goźdź-Roszkowski, S. (2012). Legal language. In C.A. Chapelle (Ed.), *Encyclopaedia of applied linguistics*. John Wiley & Sons. `doi/10.1002/9781405198431.wbeal0678`

Goźdź-Roszkowski, S. (2013). Exploring near-synonymous terms in legal language. A corpus-based, phraseological perspective. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, *12*, 94–109.

Halliday, M. (1961). Categories of the theory of grammar. *Word, 17*(2), 241–292.

Halliday, M.A.K. (1999). The notion of "context" in language education. In M. Ghadessy (Ed.), *Text and context in functional linguistics* (pp. 1–24). John Benjamins.

Hartmann, R. & James, G. (2002). *Dictionary of lexicography*. Routledge.

Jopek-Bosiacka, A. (2011). Defining law terms: a cross cultural perspective. In S. Goźdż-Roszkowski & I. Witczak-Plisiecka (Eds.), *Research in language. special issue on legal terminology: approaches and applications* (pp. 9–29). Łódź University Press.

Kaczmarek, K. (2013). Precision and vagueness in the language of the law in Hungarian and Polish legal texts. *Comparative Legilinguistics*, *13*, 51–68.

Klabal, O. (2019). Corpora in legal translation: overcoming terminological and phraseological asymmetries[25] between Czech and English. *CLINA*, *5*(2), 165–186.

Klabal, O. (2022). Synonyms as a challenge in legal translation training. *Białostockie Studia Prawnicze*, *27*(4), 69–82.

Kornacka, M. (2005). Termin. In J. Lukszyn (Ed.), *Języki specjalistyczne. Słownik terminologii przedmiotowej* (p. 131). Uniwersytet Warszawski.

L'Homme, M.-C. & Azoulay, D. (2020). Collecting collocations from general and specialised corpora: A comparative analysis. In G.C. Pastor & J-P. Colson (Eds.), *Computational phraseology* (pp. 151–175). John Benjamins.

Lukszyn, J. (2001). Termin i system terminologiczny w świetle praktyki terminograficznej. In J. Lukszyn (Ed.), *Metajęzyk lingwistyki. Systemowy słownik terminologii lingwistycznej* (pp. 7–25). Wydawnictwo "Aquila".

Lyons, J. (1981). *Language and linguistics. An introduction*. Cambridge University Press.

Maley, Y. (1994). The language of the law. In J. Gibbons (Ed.), *Language and the law* (pp. 11–50). Longman.

Marciniak, M. & Mykowiecka, A. (2014a). NPMI driven recognition of nested terms. In *Proceedings of the 4th International Workshop on Computational Terminology (Computerm)* (pp. 33–41). Association for Computational Linguistics and Dublin City University. `DOI:10.3115/v1/W14-4805`.

---

25 A correction to 'assymetries', the spelling used by the author of the original article.

Marciniak, M. & Mykowiecka, A. (2014b). Terminology extraction from medical texts in Polish. *Journal of Biomedical Semantics, 5*(24). DOI:10.1186/2041–1480–5–24.

Martin, E. A. (2003). *A dictionary of law* (5th edition). Oxford University Press.

Matilla, H. (2006). *Comparative legal linguistics*. Ashgate.

Matulewska, A. (2016). Semantic relations between legal terms. A case study of the intralingual relations of synonymy. *Studies in Logic, Grammar and Rhetoric, 45*(1), 161–174.

Michta, T. & Mroczyńska, K. (2022). *Towards a dictionary of legal English collocations*. Wydawnictwo Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach.

Michta, T. (2022). You shall know a term by the company it keeps: Collocations of the term *evidence* in general and legal corpora. *Beyond Philology: An International Journal of Linguistics, Literary Studies and English Language Teaching*, *19*(1), 65–96.

Murphy, M. L. (2003). *Semantic relations and the lexicon*. Cambridge University Press.

Rycak, M. B. (2008). *Wymiar i rozkład czasu pracy – monografia naukowa*. Wolters Kluwer Polska.

Rzepkowska, A. (2021). Polish-English LSP dictionaries in translation work: Labour-law terminology from the Polish Labour Code in terminographic and translation practice. *Language Culture Politics International Journal*, *1*/2021, 15–46. DOI:10.54515/lcp.2021.1.15–46

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

Sinclair, J. (2005). Corpus and text – basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxbow Books.

Trosborg, A. (1995). Statutes and contracts: An analysis of legal speech acts in the English language of the law. *Journal of Pragmatics*, *23*(1), 31–53.

Ullmann, S. (1972). *Semantics: an introduction to the science of meaning*. Basil Blackwell.

Vázquez y del Árbol, E. I. (2006). La Traducción al español de expresiones binomiales y trinomiales (doublet & triplet expressions) en inglés jurídico: el caso de los testamentos (wills). *Babel Afial*, *15*, 19–27.

Vázquez y del Árbol, E. I. (2014). Binomios, trinomios y tetranomios cuasi sinónimos en los poderes notariales digitales británicos y norteamericanos: análisis y propuesta de traducción. *Revista de llengua i dret*, *61*, 26–46.

Webb, J. E. (2010). *The Penguin dictionary of law*. Penguin.

Wróblewski, B. (1948). *Język prawny i prawniczy* [Language of law and legal language]. Polska Akademia Umiejętności.

Yevchuk, A. (2021). Plesionyms as a vocabulary teaching tool: the case of Estonian EFL learners. *Sustainable Multilingualism*, *19*(1), 203–226.

Legal sources

Agency Workers Regulations 2010

Employment Rights Act 1996

Equality Act 2010

Health and Safety at Work etc. Act 1974

National Minimum Wage Act 1998

Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000

Pensions Act 2008

The Fixed-term Employees (Prevention of Less Favourable Treatment) Regulations 2002

The Maternity and Parental Leave etc. Regulations 1999

Trade Union and Labour Relations (Consolidation) Act 1992

Transfer of Undertakings (Protection of Employment) Regulations 2006

Working time regulations 1998

Websites

`https://dictionary.cambridge.org/` (Date of access 20 April 2023).

`https://www.collinsdictionary.com/` (Date of access 20 April 2023).

`https://www.gov.uk/browse/employing-people/contracts` (Date of access 20 April 2023).

`https://www.legislation.gov.uk` (Date of access 20 April 2023).

`https://dictionary.cambridge.org/thesaurus/` (Date of access 22 April 2023).

`https://www.sketchengine.eu` (Date of access 28 April 2023).

`https://www.thesaurus.com` (Date of access 22 April 2023).

***

Agnieszka Rzepkowska, PhD, is an assistant professor at the Institute of Linguistics and Literary Studies at the University of Siedlce, Poland. She earned a B.A. in Foreign Economic Relations from the Warsaw School of Economics, as well as a B.A. and M.A. in English Studies and a PhD in Applied Linguistics from the University of Warsaw. Her main area of interest is language for special purposes (LSP). She has published a number of papers and a monograph on terminology and terminography, focusing on the structure and applicability of interdisciplinary professional dictionaries. Her recent research focuses on corpus studies in LSP, particularly legal language, specialist collocations, and LSP teaching.