Exploring near-synonymous terms in legal language. A corpus-based, phraseological perspective

Stanisław Goźdź-Roszkowski

University of Łódź, Poland
gozdz.roszkowski@gmail.com

This paper aims to determine the extent to which a corpus-based, phraseological approach can be effectively applied to discriminate among near-synonymous, semantically-related terms which often prove troublesome when translating legal texts. Based on a substantial multi-genre corpus of American legal texts, this study examines the collocational patterns of four legal terms 'breach', 'contravention', 'infringement' and 'violation', first in the genre of contracts and then in the multi-genre context of the entire corpus. The findings highlight the area of overlap as well as specificity in the usage of these terms. While collocational constraints can be argued to play an important disambiguating role in the semantic and functional analysis of both source and target text items carried out by translators prior to the interlingual translation, this study emphasizes the applicability of the phraseological approach to English source texts.

1. Introduction

In his well-known book on legal language, professor Peter Tiersma (1999), both a lawyer and a linguist, notes, with brutal candour, that "the legal profession has a very schizophrenic attitude toward synonyms" (p. 113). On the one hand, law professionals are encouraged to follow the fundamental rule of legal writing, that is, same meaning same form, whereby the same term should be used consistently in reference to a given concept. For example, if a legal drafter chooses to adopt the term *residence* to refer to a particular concept in a contract, then this term, rather than *domicile,* should be used throughout the entire document. On the other hand, legal language is notorious for employing strings of semantically-related words, the so-called binomial or trinomial expressions (Gustafsson, 1984), such as *null and void, last will and testament* or *make, constitute and appoint.* The importance of the concept of near-synonymy to legal translation is well-recognised. Translators dealing with legal language inevitably face a bewildering range of synonymous or near-synonymous terms or words appearing in virtually all legal texts. For example, the semantic field 'cancel' contains lexical

items such as: 'annul', 'revoke', 'dismiss', 'overrule', 'quash', 'strike out', 'recall', or 'reverse', to name just a few (see Alcaraz & Hughes, 2002). The presence of near-synonymous lexis in English legal texts is likely to be confusing for various types of readers including law students and law professionals (esp. those who are not familiar with the Anglo-American common law system) as well as translators (Chromà, 2011). Yet, relatively little work has been done on this area in English legal language (see, however, Goźdź-Roszkowski, 2009). Existing specialized lexicographical resources are usually of limited use as they do not effectively clarify the nuances of meaning and usage involved in such terms. Magris (2004) is one of few notable contributions which recognizes the fact that properly constructed and reliable language resources (such as terminological databases) should cater to the specific terminological needs of legal translators.

This problem seems to be particularly acute in the context of translation. Traditionally, synonymy in LSP (language for specific purposes) texts has been approached from the onomasiological and terminological perspective, according to which synonyms are perceived as two or more terms from the same language representing the same concept (Felber, 1984, p. 98).Consequently, equivalents could be defined as two or more words representing the same concept and derived from different languages. In the semasiological approach, synonymy and equivalence are treated as relations between lexemes with the same denotative meaning. As Rogers (1997) demonstrates, both approaches are not without certain limitations. The concept-based approach is useful in determining whether the fact of synonymy occurs but it does not enable one to propose any criteria for selecting particular synonyms (equivalents) for a particular concept. Importantly, the semasiological approach assumes substitutability in all contexts, a feature reserved for the rather elusive notion of absolute synonymy. This paper aims to pursue the aspect of syntagmatic relations, i.e. collocability, by proposing a corpus-based phraseological perspective on discriminating among semantically similar legal terms. The importance of syntagmatic relations in constraining synonymy is well recognized in general language. In fact, the role of collocational patterns in discriminating among semantically similar items has been explored extensively since the advent of corpus linguistics and the accessibility of its analytical tools (e.g., Partington, 1998). However, legal communication, just like any specialized communication, operates under different constraints resulting from the intention to eliminate or reduce ambiguity. Tiersma (1999, p. 182) insists that for lawyers, even if words are similar, they are apparently never identical. Instead, they exhibit subtle differences in their connotational value. The choice of such items may be strategic (see, for example, Danet, 1980 for a study of lexical variation used in reference to the concept of abortion) or it may depend on the knowledge of extra-linguistic entities, processes, generic conventions, etc. The arbitrary or

non-arbitrary nature of synonym variation appears to be of central importance for translators when resolving the question of interpreting the meaning of semantically-related source language items (intralingual translation) and deciding upon interlingual equivalence.

In the remainder of this paper I first turn to the concept of synonymy and briefly discuss the type of synonymous relations usually encountered in legal texts (Section 2). I then move on to present a case study illustrating the strengths and weaknesses of the phraseological approach in discriminating among four semantically-related terms: *breach, violation, infringement* and *contravention.* Section 3 specifies the methodology and data used, while Section 4 provides the results and discussion. Finally, Section 5 brings conclusions and directions envisaged for future research.

2. Synonymy in legal language

The phenomenon of synonymy has been the object of linguistic enquiry for a long time. Semanticists, in particular, have devoted much attention to formulating basic relations between words. Classic studies in this area include, for example, Nida (1975), Lyons (1977), Leech (1981) and Cruse (1991). It turns out that it is extremely rare for words to be totally interchangeable (see also Ullmann, 1962). Instead, words may share a number of identical features but still they may differ considerably in their actual use. Such differences may be related to, for example, literary and non-literary usage, neutrality versus marked evaluation, formal versus informal usage, etc. Linguistic research into synonymy has resulted in the proposal of several different definitions and typologies of synonymy. For example, Lyons (1981, pp. 148–149) distinguishes between *complete* and *absolute* synonymy. The former occurs "if and only if [two items] have the same descriptive, expressive and social meaning (in the range of contexts in question)" (Lyons, 1981, pp. 148–149).Conversely, two items are "absolutely synonymous if and only if they have the same distribution and are completely synonymous in all their meanings and in all their contexts of occurrence". Cruse (1991, pp. 265–295) employs the term *cognitive* synonymy to refer to two utterances which fulfil the same truth conditions, even if a part of the first utterance has been substituted by something else in the second. Finally, the concept of *plesionym* characterising the quasi-hyponymy of items such as the above-mentioned terms of 'annul', 'revoke', 'dismiss', etc. under the superordinate 'cancel' seems particularly relevant to capture the relations between semantically-related terms in legal English. In addition, a certain degree of synonymy can exist when two utterances are found to perform the same function, i.e. the same speech act (Austin, 1962). In this case, context of use is clearly of fundamental importance. This very brief overview demonstrates two essential points. First, synonymy is a matter of degree, that is, certain

items may come closer to being absolute synonyms of each other than others. Second, determining synonymy involves examining contextual relations of two or more items.

One would not expect to find many instances of full (e.g., complete or absolute) synonymy in legal language. Chromà (2011, p. 45) provides the examples of *causal link*, *causal nexus* or *causal connection* as those rare cases when legal terms can be used interchangeably in all legal and linguistic contexts. Propositional synonymy, also referred to as paraphrase (e.g., Murphy, 2008, p. 144), describes the relations between syntactic units. For example, the meaning of phrases often found in contractual instruments such as *unless the contract provides otherwise, in the absence of a provision to the contrary, except when otherwise provided by the contract* is determined in terms of their overall function of indicating that "whatever has been expressly agreed upon, or has been implied by contracting parties, should apply as long as there is no explicit statement (in the law) overriding it" (Chromà, 2011, p. 40). As a result, these phrases can be translated by one phrase in the target language that best reflects the message contained in the source language phrase in the target law.

Synonymy in legal English can also be found in the age-old practice of employing binomial or trinomial expressions. In one of the earliest definitions of binomials, Malkiel (1959) explains that these are "[…] the sequence of two words pertaining to the same form-class, placed on an identical level of syntactic hierarchy, and ordinarily connected by some kind of lexical link" (p. 113). This definition should be expanded to cover semantic and functional unity characteristic of this type of multi-word expressions. Strictly speaking, fixed and non-compositional binomials or trinomials should be distinguished from long lists of near-synonymous lexical items (also referred to as *synonymical chains* in Chromà, 2011) created ad hoc by legal drafters. They tend to be repetitive and thus more or less fixed. Whether they are also non-compositional can sometimes be problematic. They are generally employed for technical accuracy and for the sake of precision and unambiguity, but there are cases where doubling-up serves no specific purpose (Gustafsson, 1984, p. 123). In other words, it is very often up to the translator to decide whether there is a meaningful distinction between semantically-related words in such phrases as, for example, *give, devise and bequeath, due and owing, ordered, adjudged and decreed,* or *power and authority,* etc. He or she must determine the degree of fixedness and idiomaticity of synonymous chains and, ultimately, determine whether such lexical items serve any function. Importantly, the specialized meaning of such expressions does not often reside in individual words. For example, *full faith and credit* is a term in American law and it cannot be tampered with by modifying it as *full faith* or *full credit* (Tiersma, 1999, p. 113).

3. Data and methodology

The notion of substitutability brings up an important semantic aspect with respect to syntagmatic relations, or collocability. It is only through a large amount of texts that one can obtain evidence of collocability. It is well known that syntagmatic relations play an important role in constraining synonymy in general language. In the case of legal terms, one needs to take into account both denotational factors and their linguistic contexts. Indeed, apart from the scope of similarity or resemblance, permissible or acceptable differences and the distributional potential of alleged synonyms, it is the linguistic contexts (or, more precisely, co-texts)that have a decisive impact on determining the meaning of a lexical unit or units (Murphy, 2008, p. 145). Partington (1998) convincingly demonstrates how corpus-based collocational analysis can discriminate among semantically similar items, such as *sheer, pure, complete* and *absolute*. While corpus linguistics has become a much used methodology in mainstream linguistics for the study of language for general purposes (LGP) in its various aspects, the application of corpus linguistics tools in exploring legal English in the context of legal translation seems to be somehow lagging behind (see Biel, 2010). In an attempt to at least partially fill the void, the present study seeks to find out whether, by analogy, differences between near-synonymous legal terms can be, to some extent accounted for by building their collocational profiles, identifying larger phraseological patterns (Gries, 2008, pp. 6–7) and adding extralinguistic parameters such as subject-specific domain (e.g., contract law, intellectual property law), and genre (e.g., statute, judgment or contract). This paper attempts to explore to what extent combining both corpus methodology and the notion of phraseology could provide a solution to the rather daunting task of distinguishing between near-synonymous or semantically related terms in the domain of law by exploring their collocational and contextual restrictions. It should be noted that the 'textual' and 'phraseological' approach is meant to complement rather than supersede any other, especially the conceptual, ontological approaches to this problem.

The study is based on a new multi-genre corpus of legal texts. The collection of texts (hereinafter called the *American Law Corpus* or the ALC) contains over 5,500,000 words and represents seven major genres which are part of the American legal culture and education. Table 1 below shows the overall composition of the ALC by genre category. These range from primary genres such as federal legislation and the Supreme Court opinions, through operative documents (briefs, contracts, powers of attorney) to academic genres (journal articles and textbooks).

Table1: Composition of the American Law corpus

Genre	# of texts	# of words
Expository/descriptive		
Academic journals	71	552,487
Textbooks	104	519,116
Operative/prescriptive		
Contracts	177	1,178,616
Legislation	60	1,178,516
Legal representation/descriptive		
Briefs	64	763,222
Advice	100	201,404
Descriptive/prescriptive		
Judgements	114	1,182,246
Total	687	5,578,393

The textual (genre) categories selected for the American Law Corpus were randomly sampled from a range of written activities associated with American legal culture. The present corpus resources are being revised and expanded. In some cases, more recent texts are added while some older ones are removed. At the same time, new genres will be added in order to ensure the corpus representativeness. A more detailed description of the ALC can be found in Goźdź-Roszkowski (2011).

The corpus was analyzed using the popular *WordSmith Tools* (version 5.0). From a number of possible synonymous sets, it was decided to focus on four common legal terms *breach, infringement, violation* and *contravention.* This choice was motivated by the importance of these terms (they all signal the fundamental concept of non-compliance with law) and their relatively high frequency (two terms, *breach* and *violation* are found in the ALC 299 and 246 times per million words respectively).Through the analysis of the co-texts of these nouns – that is, through the computer-assisted tool Concordance that made it possible to retrieve concordance lines of these key nouns (methodology known in corpus linguistics literature as KWIC, key words in context) – it was possible to isolate interesting collocational patterns in the corpus. 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, that is, the term under investigation. In addition, 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 (2000), the author of the *WordSmith Tools*, refers to such constructs as *clusters.*

4. Results and discussion

We start by considering four typical dictionary definitions of the terms in question provided by one of the most authoritative legal dictionaries, that is, *the Black's Law Dictionary.* In doing so, I would like to demonstrate that dictionary definitions are of limited use to LSP users without a legal background.

Violation is defined as injury, <u>infringement</u>; <u>breach</u> of right, duty or law; ravishment; seduction. The act of breaking, infringing, or transgressing the law. (Black's Law Dictionary)

Breach the <u>breaking</u> or <u>violating</u> of a law, right, obligation, engagement , or duty, either by commission or omission. Exists where one party to contract fails to carry out term, promise, or condition of the contract.

Contravention in French law, an act which <u>violates</u> the law, a treaty, or an agreement which the party has made.

Infringement A <u>breaking</u> into; a trespass or encroachment upon; a <u>violation</u> of a law, regulation, contract, or right. Used especially of invasions of the rights secured by patents, copyrights, and trademarks

---

Figure 1: Dictionary definitions of 'violation', 'breach', 'contravention' and 'infringement' in Black's Law Dictionary (1990)

The four words found in this group share the sense of breaking the law, failure to comply with a legal rule. The information provided by the dictionary suggests that 'violation' seems to be the most general term denoting deliberate breaking of a law. It also seems that it is the most wide-ranging term which could be used with reference to various kinds of wrongdoing, even including rape. The term *contravention* is marked as having a civil law origin. As such, the scope of the term is fairly broad ranging from international law (treaty) to private law (agreement). With regard to the latter, it appears to overlap with *breach.* The terms *breach* and *infringement* seem to denote more specific concepts. The term *breach* is associated with civil law contexts related to contractual instruments, while 'infringement' appears in the legal area which deals with

intellectual property rights. A dictionary user without a legal training may be confounded by the recycling of the apparently synonymous terms in the definitions. For example, *infringement* is defined as a "breaking into; a trespass or encroachment upon; a violation", while *violation* is defined, inter alia*,* as "infringement". As a result, a pertinent question arises as to the interchangeability of the terms. For example, is it possible to use both *contravention* and *breach* in reference to agreements? Are *violation* and *infringement* used interchangeably?

     We turn to look at the overall frequency counts of the four related words in the entire corpus material.

Table 2: Absolute frequency distributions of 'breach', 'breaches', 'contravention', 'infringement', 'infringements', 'violation' and 'violations'

Term	Number of words
breach	1666
breaches	81
contravention	35
contraventions	0
infringement	593
infringements	26
violation	1310
violations	437

Table 2 shows that singular forms of these terms checked against the entire corpus tend to be employed far more frequently than the plural variants. The analysis will therefore focus on the singular forms not only because of their much larger frequencies but also because of the growing evidence that phraseological patterns tend to be attached to word forms rather than lemmas (see Baker, 2006) and such patterns should turn out to be more effective in discriminating between these items. The large frequency of *breach* is not surprising given the inclusion of the category of contracts in the corpus. The definitions included in Figure 1 above suggest that this term could be treated as domain- and, indeed, genre-specific. It would, however, be interesting to find out whether *breach* is used exclusively in this particular legal domain. Thus, the next stage is to identify frequency counts for these four terms only in the textual category of contracts.

     The quantitative findings provided in Table 3 show that all these terms occur in contracts albeit with varying frequency. Overall, the

frequencies for the single domain reflect the frequencies already provided for the multi-genre data in Table 2. Breach is the most frequent term but only one-third of its occurrences belong to the contracts category. Rather surprisingly, the second-most frequent term, violation is almost half as frequent as breach in contracts. However, its distribution tends to be spread across the other genres. These figures raise the question of genre specificity, suggesting that breach is not the only term employed to denote non-compliance in the context of contractual provisions. The term infringement is similar to breach in that it is associated with a particular domain (i.e., intellectual property); unlike breach, it is not likely to be genre-specific.

Table 3: Frequency counts of breach, violation, infringement and contravention in the genre of contractual instruments

breach	510
violation	218
infringement	72
contravention	15

After these general observations, it is time to examine the collocational behaviour of the four items as they occur in the textual category of contracts. It is expected that the comparison of their collocates should throw some light on the question of genre-specificity and interchangeability of terms within the same genre category.

agreement (103); material (51); default (49); result (44); warranty (43); party (38); shall (38); constitute (26); representation (26); terms (26); waiver (26); contract (22); obligations (22); threatened (22); event (21); covenant 21); duty (19); failure (18); company (17); violation (16); section (15); damages (14); provided (14); deemed (13); provision (13); termination (13); remedy (13); arising (13); cure (12); conditions (12); provisions (11); claims (10)

Figure 2: Most frequent collocates (32) of breach (with at least ten occurrences found within 5 words left and right of the node word)

Figure 2 shows 32 most frequent collocates identified by the collocates function of the *WordSmith Tools* for the term *breach*. Their presence corroborates our earlier, rather obvious observation that *breach* is firmly

associated with the domain of contract law. The examination of the relevant co-texts by means of the *Concordance* and *Clusters* tools leads to the identification of the following list of terminological phrases or clusters: *breach of this Agreement, breach of the terms, breach of any provision, a material breach of, breach of any representation, breach of any covenant, a waiver of any breach, breach of its obligations, breach of duty,* etc. The collocates are thus a useful starting point to uncover repetitive phraseologies of the term *breach,* and evoke concepts necessary to understand this legal term.

If we repeat the analytical procedure and analyze clusters with *breach* (of a size between 3 and 5 words with the minimum frequency of 10) in the entire corpus material, the phraseological picture does not substantially differ from what has been already established.

Table 4: Eight most frequent clusters with 'breach' in left search term position

Rank	Frequency	Cluster
1.	201	breach of contract
2.	33	breach of this Agreement
3.	27	breach of warranty
4.	25	breach of the covenant
5.	13	breach of trust
6.	13	breach of the terms
7.	11	breach of condition
8.	11	breach or default

Clusters with *breach* in right search term position provide more evidence corroborating that breach is a highly technical term, the occurrence of which is largely confined to the domain of contract law. The technicality of this term is underlined by the collocating adjectives as in *repudiatory breach* and *a fundamental breach.* The results provided in Table 4 show that *breach* has a tendency to appear in attributive phrases with "of". *Breach* seldom appears as a single-word term. Instead, it tends to be used as a compound term in which it is a head.

Table 5: Eight most frequent clusters with 'breach' on right search term position

Rank	Frequency	Cluster
1.	52	damages for breach
2.	23	a repudiatory breach
3.	22	a material breach
4.	20	action for breach
5.	16	is in breach
6.	13	waive the breach
7.	12	a fundamental breach
8.	12	the party in breach

The term *breach* appears in the contexts related to possible legal sanctions (note the most frequent cluster in Table 5, *damages for breach,* and *action for breach*) which could be imposed as a result of being in the state of violating the law, e.g. *is in breach, the party in breach.* In other words, indicating that a legal or natural person (a party to some contractual relations) violates specific contractual provisions constitutes a significant proportion of the clusters. For example, the cluster *is in breach* is invariably followed by a specification of the type of breach. The results obtained so far indicate that the *breach* collocates seem to be marked by relative systemic homogeneity. The findings presented so far in relation to the term *breach* evoke the following scenario: If one party to a contract fails to perform his obligations (*breach of the terms*) or indicates his intention not to do so (*repudiatory breach*),then the other party is entitled to sue for *damages*. The innocent party may choose to affirm the contract (*waive the breach*) unless it is a *breach of condition* or a *fundamental breach*. These findings should, however, be seen in light of collocate analysis carried out for the term *violation.*

Interesting findings emerge from the examination of the most frequent words co-occurring with the term *violation.* As Figure 3 shows, eight out of 14 collocates of *violation* are shared by the term *breach.* The shared words are marked in bold in Figures 2 and 3. This suggests that there is considerable overlap in the way both terms are used in some contexts. At the same time, it should be pointed out that the proportion of shared words is naturally much lower in the case of *breach*.

law (24); alleged (24); <u>agreement</u> (23); <u>material</u> (17); <u>result</u> (16); breach (16); <u>constitute</u> (14); environmental (16); <u>section</u> (13); <u>shall</u> (13); <u>default</u> (12); <u>party</u> (11); rights (10); related (10)

Figure 3: Most frequent collocates (14) of *violation* (with at least ten occurrences found within 5 words left and right of the node word)

The presence of the shared collocates can be at least partially accounted for by the statistical fact that both terms are also mutual collocates. There are 16 attested cases where both *breach* and *violation* are found in the same co-texts. Three corpus examples are provided below:

(1) Tenant shall reimburse Landlord for all expenses, damages or fines incurred or suffered by Landlord, by reason of any breach, violation or nonperformance by Tenant,

(2) (…) result in, or give rise to, a violation or breach of or a default under any of the terms of any material contract, (…)

(3) The execution, delivery and performance of this Agreement by the Seller will not (i) constitute a breach or a violation of the Corporation's Certificate of Incorporation, By-Laws, or of any law, agreement, indenture, deed of trust, mortgage, loan agreement or other instrument to which it is a party, or by which it is bound; (ii) constitute a violation of any order, judgment or decree to which it is a party or by which its assets or properties is bound or affected; or (iii) result in the creation of any lien, charge or encumbrance upon its assets or properties, except as stated herein.(emphasis added)

The examples show that the textual environment in which these two terms are found is highly repetitive and formulaic. It contains strings of related words so characteristic of contractual texts. Many of them could be regarded as examples of the bi- or trinomial expressions discussed in Section 2.Example 3 provides further evidence that *breach* and *violation* are found in contractual provisions replete with numerous lists of semantically-related terms and bi- and trinomials, such as *any order, judgment or decree, any lien, charge or encumbrance.* This suggests that the interpretation of the two terms' co-occurrence should be considered in light of the generic and strategic practice of ensuring all-inclusiveness in contractual provisions. Worth noting is that in the examples above, *breach* and *violation* are followed by other related nouns, for example, *non-performance* (example 1) and *default* (example 2). The data provided above can also be used to explain the occurrence of other shared collocates. For example, Figure 1 shows that word *constitute* can

collocate with both nouns. In those examples, *material* precedes *breach* and the co-occurrence of *material breach* and *violation* results in *material* being identified by the computer tool as a collocate of violation within the span of 5 words to the left of the node word. Still, there are five attested examples of *material* forming a separate phrase with *violation,* as for example in *alleged material violation.*

Extending the analysis of the most frequent clusters with *violation* to cover the other genres represented in the corpus, reveals that it is most frequently used in the phrase *in violation of* (342 occurrences). Other very frequent clusters (*violation of the law* (46), *violation of Section* (24), *a constitutional violation* (21) *violation of federal law* (12), a *Sixth Amendment violation* (11), *violation of due process* (10)) are used in reference to a wide range of legal objects, some of them general, such as *rights*, *the law(s)* or *constitution* or more specific, i.e. *section*, etc. Apart from indicating a range and types of legal rules that can be violated, there are clusters that suggest that violation is often described as intentional. There are 36 instances of the phrase *intentional or willful violation.* In addition, there are several verbal clusters such as *to be in violation, is in violation, is a violation, constitute/constituted a violation, result in a violation,* etc. used to establish whether a particular action undertaken by a legal or natural person amounts to an act of breaking the law.

We now turn to consider the usage of the two remaining terms in the textual category of contracts. Figure 4 lists the most frequent words co-occurring with *infringement*. Given the overall low frequency of this term (only 72 instances) the minimum frequency cut-off point was lowered to 5. This yielded 15 collocates which corroborate the restricted and highly technical use of this term.

---

misappropriation (16); rights (12); present (10), future (10); past (10); violation (9); patents (8); claims (7); intellectual (6); suit (6); respect (6); immunity (6); property (5), action (5); patent (5)

Figure 4: Most frequent collocates (15) of *infringement*

The frequencies of some collocates signal their occurrence in formulaic and repetitive multi-word sequences. For example, the words *present, future* and *past* all appear ten times because of the formula *the right to sue for past, present or future infringement, misappropriation or violation of rights* routinely inserted in contractual clauses dealing with trademark infringement. Licence agreements also contain a separate section which deals with copyright infringement. There are several instances of this formulaic multi-word string and their scrutiny helps to account for the presence of other collocates such as *violation, misappropriation* and

rights. The co-occurrence of *infringement* and *violation* in such long and repetitive sequences is similar to what we observed in the case of *breach* and *violation*. Extending the analysis to cover other genres does not change the overall phraseological picture radically. *Infringement* most frequently co-occurs with the word *act* in the phrase *act of infringement* (68 hits), and in some domain-specific terms, as in c*opyright infringement* (36 hits) and *patent infringement* (28), thereby confirming the close association between the term *infringement* and the domain of intellectual property rights. The clusters tend to occur in highly prescriptive legislative provisions (308 out of 593 instances are found in legislation) where the focus is on specifying what constitutes an act of infringement and what legal consequences may follow. Apart from the extremely frequent sequence of *act of infringement, infringement* co-occurs with *liability.* There are 41 instances of the phrase *liability for infringement.* For reasons of space, it is not possible to provide other co-occurring patterns. In one other example, examining the extended co-texts of the three most frequent clusters, leads to the identification of a formulaic sequence of *actionable as an act of infringement under section* with its characteristic legislative flavour. Other collocates belonging to the semantic field of 'litigation' include *suit, action* and *remedies.*

Finally, contravention with just 15 attested occurrences in contracts is usually found in the phrase *in contravention of* as shown below:

(4)   conflict with or result in any <u>breach or contravention</u> of, or the creation of any lien, security interest, or charge under, any material agreement, contract, indenture, document, or instrument to which the Borrower (from *Deed of Partnership*);

(5)   In the event the Founder sells any Co-Sale Securities of the Company in contravention of the participation rights of the Investor under this Agreement. (from *Co-Sale Agreement*)

As such, it seems that it is interchangeable with the phrase *in violation of.* Note also the presence of *breach* in Example 4 above. However, the extremely low frequency of *contravention* precludes any more exhaustive analysis.

5. Conclusions

In this paper, I have considered the collocational and phraseological behaviour of four terms in specialist legal communication based on the textual (generic) category of contracts and other major legal genres (e.g. legislation and judgments). The analysis has highlighted a number of constraints which are of relevance to translators when they read, analyze and interpret ST meanings. The data which has been analysed suggests

that the concept of near-synonymy needs to take into account a number of factors at the linguistic level which are not usually considered in terminology work or indeed in lexicographical work. These factors include the role of a specific domain and genre. The concept of synonymy, seen as a relation between word forms rather than as a relation between decontextualised lexemes, is highly constrained, revealing many complex relations of overlap and exclusion. The collocational information can be treated as a clue or a prompt to evoke a generic scenario in which a particular legal concept functions. Such is the case of *breach,* which reflects a unity of domain and genre with a well-defined and homogenous class of objects this term refers to. Similarly, the use of *infringement* is marked by domain-specificity. This tendency for certain legal terms to co-occur with other terms or phrases marked by semantic resemblance could also be accounted for by referring to the concept of semantic preference (Stubbs, 2001). In contrast, *violation* cuts across legal domains and genres and it is the most 'inclusive' of all the terms. Finally, *contravention* illustrates a heavy phraseological restriction to virtually one form of (a) phrase. The apparent interchangeability of some of these terms (e.g. *breach* and *violation*) can be accounted for in terms of their occurrence in bi- or tri-nomial expressions or even longer highly formulaic multi-word sequences. Corpus tools turn out to be particularly useful in identifying such textual patterns.

In their book *Legal Translation Explained*, the authors bemoan the lack of linguistic tools that would deal with "the troublesome area of synonymy" (Alcaraz & Hughes, 2002, p. 38). They go on to stress that "we would like instant clarification of the nuances of meaning and usage (…). Unfortunately, nothing of the kind is available at present, or likely to be forthcoming" (Alcaraz & Hughes, 2002, p. 38). While these words are, by and large, still true today, the present study is intended as a small step towards filling this gap.

References

Alcaraz, E., & Hughes, B. (2002). *Legal translation explained.* Manchester: St. Jerome.

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

Baker, P. (2006). *Using corpora in discourse analysis.* London: Continuum.

Biel, Ł. (2010). Corpus-based studies of legal language for translation purposes: Methodological and practical potential. In C. Heine & J. Engberg (Eds.), *Reconceptualising LSP. On-line Proceedings of the XVII European LSP Symposium 2009*. Aarhus 2010.

Chromà, M. (2011). Synonymy and polysemy in legal terminology and their applications to bilingual and bijural translation. *Legal terminology: Approaches and Applications.* A special issue of *Research in Language, 9*(1), 31–50.

Cruse, J. (1991). *Lexical semantics.* Cambridge: Cambridge University Press.

Danet, B. (1980). 'Baby' or 'Fetus'?: Language and the construction of reality in a manslaughter trial. *Semiotica 32,* 187–219.

Felber, H., (1984). *Terminology manual.* Paris: UNESCO/INFOTERM.

Goźdź-Roszkowski, S. (2011). *Patterns of linguistic variation in American legal English: A corpus-based study.* Frankfurt: Peter Lang.

Goźdź-Roszkowski, S. (2009). Using co-text to cope with vagueness in legal lexical units: A case study in the opinions of the United States Supreme Court. In L. Sočanac, C. Goddard, & L. Kremer (Eds.), *Curriculum, multilingualism and the law* (pp. 193–206). Zagreb: Nakladni zavod Globus.

Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 3–25)*.* Amsterdam: John Benjamins.

Gustafsson. M. (1984).The syntactic features of binomial expressions in legal English. *Text, 4*(1–3), 123–141.

Leech, G. N. (1981). *Semantics: The study of meaning* (2nd ed). Harmondsworth: Penguin.

Lyons, J.(1977). *Semantics*. Cambridge: Cambridge University Press.

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

Magris, M. (2004). Verso una terminografia per il traduttore giuridico. In R. Temmerman & U. Knops (Eds.), *Linguistica Antverpiensia New Series – Themes in Translation Studies3: The translation of domain-specific languages and multilingual terminology management* (pp. 53–65). Antwerp: Antwerp University Press.

Malkiel, Y. (1959). Studies in irreversible binomials. *Lingua, 8*, 113–160.

Murphy, M. L. (2008). *Semantic relations and the lexicon.* Cambridge: Cambridge University Press.

Nida, E.(1975). *Componential analysis of meaning.* The Hague: Mouton.

Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching.* Amsterdam: John Benjamins.

Rogers, M. (1997). Synonymy and equivalence in special-language texts: A case study in German and English texts on genetic engineering. In A. Trosbor (Ed.), *Text Typology and Translation* (pp. 217–245). Amsterdam: John Benjamins.

Scott, M. R. (2000). Focusing on the text and its keywords. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 103–122)*.* Frankfurt: Peter Lang.

Stubbs, M. (2001). *Words and phrases.* Oxford: Blackwell.

Tiersma, P. (1999). *Legal language.* Chicago, IL: The University of Chicago Press.

Ullmann, S. (1962). *Semantics.* Oxford: Basil Blackwell.