

Investigating “congrams” in the language of contracts and legal agreements

Katja Dobrić Basanež

Abstract Extended units of meaning have been scarcely investigated thus far in legal phraseology. Moreover, dictionaries of law neglect to include extended versions of prototypical binary units. For instance, *nominal consideration* is included in the *English-Croatian Dictionary of Law* (Gačić 2010) but its wider context *transferred at nil or nominal consideration* is not listed. This paper will therefore contrastively investigate extended phraseological units in the language of contracts. The analysis will involve comparing specialized corpora with special purpose (LSP) dictionaries and with reference corpora. The paper proposes that by analysing the extended units of meaning found in contracts and legal agreements one can gain a clearer insight into the natural idiomaticity and formulaicity prevalent in this genre. By comparing the extended units of meaning found in English and Croatian corpora, one can reveal examples of usage of such units with all their associations. The results of such analysis can be applied to translator training, especially where translation is carried out into a second language (L2).

Keywords congrams, extended units of meaning, legal agreements, genre analysis, legal phraseology

1 Theoretical background

Phraseology is usually regarded as a subfield of lexicology “dealing with the study of word combinations rather than single words” (Granger/Meunier 2008: IX). Although phraseology is part of linguistics, it attracts scholars from various disciplines, thereby making it interdisciplinary.

Corpus linguistics, however, has provided more rigorous and reliable methodologies for studying phraseology. Different studies have been conducted since the introduction of specialized computerised resources (cf. Evert 2004 and Nesselhauf 2004) and the notion of the phraseological unit has expanded to denote “the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance” (Gries 2008: 6).

Although general purpose (LGP) phraseology has been researched from many linguistic perspectives in the last twenty years, the same cannot be said of legal phraseology. There are studies proposing a morpho-syntactic (Kordić/Ivković 2010) or lexico-syntactic (Bhatia et al. 2004, Biel 2011) description of “legal phrasemes” (Kjær 2007). Some scholars have focused on the analysis of the so called “doublets” or “triplets” in different legal genres (Bukovčan 2009, Carvalho 2007), while others have studied phraseology for the purpose of compiling legal dictionaries (De Groot 1999). Finally, there are scholars who use large legal corpora for the purpose of analyzing lexical bundles and different types of co-occurrence patterns in order to study intra-register variation in legal discourse (Goźdz-Roszkowski 2011). These surveys represent the few attempts at research on legal phraseology. Kjær (2007) finds one of the reasons for this scarcity of research in the fact that, generally speaking, only the terminological component is taken into account when studying legal phraseology. The other reason pointed

out by Kjær is that this is phraseology of a specialized language, the findings of which cannot be applied universally, thereby restricting LSP phraseology to the periphery of the discipline.

A further reason for the peripheral position of LSP phraseology might be the fact that LSP phraseological units cannot be analyzed without taking into account specific domains and related disciplinary discourses. With legal phraseology this involves taking the legal system and genre into account. Therefore, when studying legal phraseology, one has to keep in mind that the legal genre “establishes constraints on allowable contributions” (Swales 1990: 52) of word combinations. Since this paper focuses on the genre of contract, it is inevitable that phraseological and terminological units found in contracts will also be influenced by the respective legal systems. For instance, the legal concept of *good faith* is very much present in European contract law, but common law does not recognize it. This, of course, does not mean that a common-law system does not impose a standard of fairness on parties entering into a contract; it simply does so through several concepts such as *misrepresentation*, *fraud*, *facility*, *circumvention*.

The approach to interpretation of contracts also differs according to the legal system. This becomes obvious when comparing the length of common-law and civil-law contracts, with the latter being a lot shorter. The length of common-law contracts, however, should come as no surprise, since common law prefers “**particularity**”, i. e. it is “*more worried about not being misunderstood by the specialist community*” (Bhatia 1993: 137), unlike civil law, which is characterized by “**generality**”, i. e. it is “*eager to be widely understood by the ordinary readership*” (Bhatia 1993: 137). Legal translators are thus confronted with many pitfalls when studying phraseology and terminology in common-law and civil-law contracts, especially in cases where a contract written in English is governed by a civil-law system or vice versa (Triebe 2009).

In legal translation it is extremely important to be precise, since translation errors could have serious consequences. This especially applies to the translation of documents from languages used in Continental Europe (civil-law system) into English (common-law system) and vice versa, since the two represent not only different legal systems but also different legal languages resulting in different patterns of phraseological units. Despite these differences, one has to figure out how to find an “approximation between both systems” (Llopis 2007: 17) and overcome the untranslatability of law in contracts.

2 Hypotheses

A model that has had considerable influence on corpus semantics is John Sinclair’s model of extended lexical units (Sinclair 2004: 24), which proposes that focus should be put on large phraseological units rather than on individual words. Sinclair (2004: 30) suggests that phrases “have to be taken as wholes in their contexts for their distinctive meaning to emerge”. Furthermore, if we look at the wider context of lexical phrases, we discover that they are prone to variation. Therefore, Sinclair takes the binary unit *naked eye* as the starting point and by extending it he detects that the unit is dominated by the prepositions *to* and *with*, that it favours verbs of “visibility” and that it involves the semantic prosody of difficulty (e. g. *too faint to be seen with the naked eye*).

Although Sinclair’s model mostly refers to natural language processing, his hypothesis of extending phraseological units can be applied to studying legal phraseology. This paper will therefore provide an insight into the patterns of extended units of meaning in two specialized corpora. Since corpora compiled for the purpose of this research consist of contracts and legal

agreements, the genre of contracts and the legal systems behind it will sometimes have to be taken into account in order to explain the existence or non-existence of phraseological units in contracts. The two language populations stem from different legal systems (English – common law, Croatian – civil law) and we know that “legal language is inextricably intertwined with the legal system” (Kjær 2007: 508). The paper will therefore reveal the reasons behind the different phraseological patterns in legal English and legal Croatian. It will also show that these units consist of constituents that are not easily recognizable as collocating items. That is why they are very problematic when translating both from and into one’s L1.

One would expect that in legal language the structure of extended units of meaning would be more complex, i. e. it would involve phraseological units clustering around terms. The paper will also point out the application of phraseological research to legal lexicography by showing that extended units of meaning extracted from the two corpora of contracts are largely under-represented in legal dictionaries.

3 Methodology

3.1 *Extended units of meaning precisely defined*

“Phraseologists must carefully define the linguistic level(s) at which they observe a potential phraseologism” (Gries 2008: 8). Therefore, the extended unit of meaning will at this point be precisely defined taking into account six parameters established by Gries (2008). Regarding the nature of the elements involved in a phraseologism, the extended unit of meaning will encompass a lexical item extended by other lexical items (e. g. to make good any *damage*; integral part of the *agreement*). Regarding the number of the elements involved in a phraseologism, the extended unit of meaning will include at least three elements (a lexical item extended by at least two other lexical or grammatical items). As regards the number of times an expression must be observed in order to count as a phraseologism, the extended unit of meaning “has to occur a minimum of twice” (Sinclair 2004: 28) in the corpus. Since it was pointed out above that phraseologisms in legal language are frequently discontinuous, the extended unit of meaning will, therefore, include units consisting of both adjacent and non-adjacent elements. As regards the degree of lexical and syntactic flexibility of the elements involved, the extended unit will include flexible patterns, but sometimes only a part of the unit might be flexible. Finally, regarding the role that semantic unity and semantic non-compositionality play in the definition of a phraseologism, the extended unit of meaning has to represent a semantic unity but does not have to be non-compositional. In this paper analysis will focus on extended lexical collocations only, i. e. on combinations consisting of either a noun or a verb as the base extended by at least two other items.

3.2 *Description of the corpora*

As was pointed out above, this paper will focus on extended units of meaning in two corpora of contracts and legal agreements. In addition, *English-Croatian Dictionary of Law* (Gačić 2010), and *Opći pravni rječnik* (Mukić Vidaković 2006) will be consulted to see whether the extracted units are listed there.

It should be pointed out here that the initial intention was to build a corpus consisting of authentic contracts and legal agreements. However, lawyers are not willing to share their legal

documents since they may contain sensitive information, which, if revealed, would violate their clients' privacy.

The corpora, therefore, consist of contract and agreement templates that are used by lawyers on a daily basis. Croatian contracts were extracted from the digital edition of the book *Zbirka ugovora građanskog i trgovačkog prava 4* (Junačko/Rotar 2007), whereas the British English contracts and agreements were downloaded from the online edition of *Encyclopaedia of Forms and Precedents* (Millet/Walker 2014), which is constantly updated. The Croatian source for contracts was chosen after having interviewed lawyers on the most common source they use when drafting and adapting contracts, whereas the English source was chosen after having consulted Dr. Volker Triebel, an English barrister and a German lawyer. Since “corpora intended for LSP can be smaller than those used for LGP studies” (Bowker/Pearson 2002: 48), the Croatian corpus of 158,317 tokens and the English one of 375,223 tokens can be regarded as large enough. As pointed out above, the English corpus is larger than the Croatian one due to the “particularity” and “all-inclusiveness” (Bhatia 1993: 137) of common-law contracts.

In order to show “the deviance of special corpora”¹ and that the extended units of meaning are typical of legal phraseology, in particular of contracts and legal agreements, two reference corpora will be consulted, *hrWaC 2.0* (Ljubešić/Klubička 2014) for Croatian, consisting of 1,404,262,704 tokens, and BNC² for English, consisting of 112,181,015 tokens.

3.3 Procedure

Both specialized corpora of contracts will be searched using *Wordsmith Tools 6.0* (Scott 2011) by means of WSCongram. Scott has adopted the definition from Cheng et al. (2008: 237) who define congrams as “instances of co-occurring words irrespective of whether or not they are contiguous, and irrespective of whether or not they are in the same sequential order”. The search for congrams is fully automated and can find “the associated words even if they occur in different positions relative to one another (i. e. positional variation) and even when one or more words occur in between the associated words (i. e. constituency variation)” (Cheng et al. 2006: 413). This cannot be claimed for n-grams, on the other hand, since “n-gram searches are only helpful in finding instances of collocation that are strictly contiguous in sequence” (Cheng et al. 2006: 412), which results in failing to discover many collocations that occur in non-contiguous sequences. Parameters in the settings will therefore be modified to display clusters which occur at least twice, but will stop at sentence breaks.

The procedure will first involve creating a wordlist and adding the wordlist to an index. This index will be chosen for the procedure, and all items which occur together at least twice (Sinclair 2004: 28) will be saved as potential constituents of each congram. These constituents will appear in frequency order, but one can also sort them alphabetically. A tree view of congrams will then be produced, where “each branch of the tree shows how many sub-items and how many items of its own it has”³. The resulting congrams will be analyzed and compared in their concordances.

It is expected that not all congrams will represent valid phraseological units, which means that the procedure will be semi-automatic, i. e. it will involve both automatic extraction

¹ <<http://www.ilc.cnr.it/EAGLES/corpusstyp/node18.html>>, 10.08.2015.

² <<http://www.natcorp.ox.ac.uk/>>, 28.08.2014.

³ <http://www.lexically.net/downloads/version5/HTML/?viewing_congrams.htm>, 20.01.2015.

of concgrams and linguistic analysis. For instance, ConcGram utility suggests that AGREEMENT-WHICH-THE is a potential phraseological unit. This, however, is common for all such types of software. For instance, SketchEngine lists *which* as one of the collocation candidates for *agreement* (cf. figure 1 for SketchEngine results). But ConcGram utility also produces many valid longer word combinations (cf. figure 2 for concgram TERMINATE-THIS-AGREEMENT-BY-WRITTEN-NOTICE). Therefore, despite the fact that the procedure with concgrams cannot be completely automated, it still remains true that concgram function can save us a lot of time when trying to discover phraseological units.

Collocation	Count	Log Likelihood	Information	Chi-Square
P N not	154	11.375	3.584	10.167
P N :	123	10.258	3.735	10.058
P N any	242	13.514	2.929	10.030
P N or	395	16.798	2.691	10.011
P N ,	389	16.647	2.681	9.998
P N agreement	94	9.139	4.124	9.941
P N Agreement	95	9.059	3.826	9.858
P N for	133	10.239	3.156	9.844
P N parties	71	8.191	5.161	9.829
P N a	152	10.621	2.851	9.752
P N made	73	8.140	4.404	9.736
P N I	174	11.159	2.698	9.723
P N obligations	67	7.907	4.880	9.718
P N are	86	8.552	3.685	9.715
P N 5	69	7.956	4.568	9.702
P N may	86	8.529	3.638	9.700
P N I	171	10.846	2.551	9.615
P N that	110	9.122	2.941	9.601
P N 1	87	8.394	3.321	9.593
P N /	60	7.497	4.957	9.591
P N provisions	64	7.573	4.229	9.549
P N with	94	8.470	2.984	9.509
P N as	111	8.969	2.750	9.504
P N :	78	7.950	3.324	9.497
P N Accepted	50	7.032	7.488	9.495
P N (203	11.248	2.248	9.461
P N 2	74	7.759	3.351	9.458
P N which	72	7.661	3.364	9.437
P N on	96	8.397	2.806	9.436
P N In	51	6.926	5.049	9.395

Fig. 1: Collocations of the term „agreement“ in SketchEngine

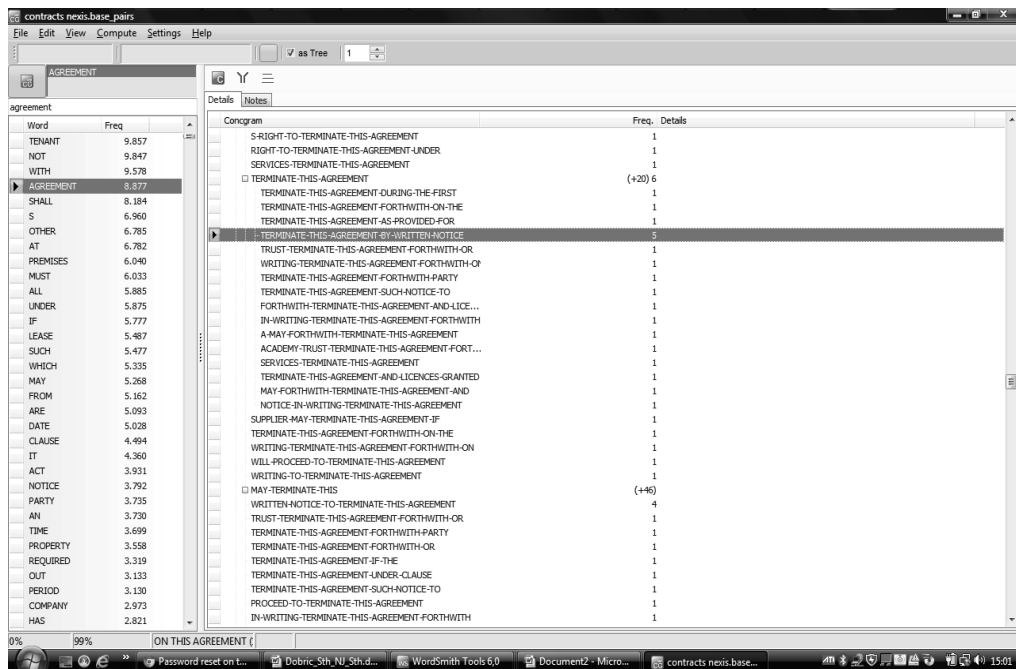


Fig. 2: Concgram TERMINATE-THIS-AGREEMENT-BY-WRITTEN-NOTICE

In order to reveal the frequency of the extracted extended units of meaning in *hrWaC 2.0*. (Ljubešić/Klubička 2014) and BNC⁴, Sketch Engine software (Kilgariff et al. 2014) will be used.

4 Results

The starting point for the selection of elements for comparison was the definition of the term *contract*. One of the definitions provides that contract is “an agreement between two or more persons which creates an obligation to do or not to do a particular thing; its essentials are competent parties, subject matter, a legal consideration, mutuality of agreement, and mutuality of obligation” (Black 2004: 292). In other words, an *agreement* (offer + acceptance) between two or more *parties* must exist and be supported by *consideration* in order to create a valid *contract*. The italicized terms were therefore chosen as the starting point for running congrams in the English corpus and their Croatian equivalents *ugovor* (agreement/contract), *strane* (parties) and *protučinidba* (consideration) as the starting point for running congrams in the Croatian corpus. Since *agreement* is in the English corpus used to denote *contract*, the analysis of congrams in the Croatian corpus will focus on the term *ugovor* only, since the concept in Croatian includes both contracts and agreements.

⁴ <<http://www.natcorp.ox.ac.uk/>>, 28.08.2014.

4.1 Concgrams in the English corpus

4.1.1 Agreement

The term *agreement* results in 8,877 concgrams in the corpus of English contracts and legal agreements. It is far more frequent than the term *contract*, since most of the contracts in the corpus are actually titled agreements and Rossini suggests that “agreement is an acceptable title for any contract” (Rossini 1998: 11).

Extended lexical collocations are among the most frequent combinations with the term *agreement*. Examples from the concgram list can be found in table 1:

Tab. 1: Extended lexical collocations with the term *agreement* and their frequency in the English corpus of contracts and in BNC

Extended lexical collocations	Frequency in the corpus of contracts	Frequency in BNC
1.1 <i>to be in breach of the agreement</i>	11	2
1.2 <i>to enforce any term of this agreement</i>	7	0
1.3 <i>to make an effective agreement</i>	6	0
1.4 <i>to contain the entire agreement between the parties</i>	6	0
1.5 <i>to forthwith terminate the agreement</i>	5	0
1.6 <i>to be expressly incorporated into this Agreement</i>	5	1
1.7 <i>to terminate this agreement by written notice</i>	5	0
1.8 <i>to assign the benefit of this agreement</i>	5	1
1.9 <i>performance of one's duties under the agreement</i>	5	0
1.10 <i>to execute an agreement in counterparts</i>	3	0
1.11 <i>to validly terminate the agreement</i>	3	0
1.12 <i>to be of the essence of this agreement</i>	3	<i>to be of the essence of the contract (10)</i>
1.13 <i>effective period of this agreement</i>	3	0
1.14 <i>integral part of this agreement</i>	3	0
1.15 <i>assignment of the benefit of this agreement</i>	2	0
1.16 <i>to enter into a binding agreement</i>	2	0

Most of the extended lexical collocations listed in table 1 do not occur in BNC and the ones that do occur represent collocations derived from legal agreements. This suggests that these combinations are typical of the genre of contracts.

As far as dictionaries are concerned, *English-Croatian Dictionary of Law* (Gačić 2010) lists *performance of duties*, *integral part* and *essence of the contract*. This supports the claim that dictionaries mostly focus on binary units. They also deprive the translator of the real context in which these units are used, e. g. *to be of the essence of the contract/agreement*.

4.1.2 Parties

The term *parties* results in 1.873 congrams in the English corpus of contracts and legal agreements. The list of extended lexical collocations is here much shorter:

Tab. 2: Extended lexical collocations with the term *parties* and their frequency in the English corpus of contracts and in BNC

Extended lexical collocations	Frequency in the corpus of contracts	Frequency in BNC
2.1 exclusion of third party rights	16	20
2.2 to act as the third party	13	3
2.3 duly authorised representatives of the parties	6	7
2.4 to contain/constitute entire understanding between the parties	3	1
2.5 to embody entire understanding of the parties	2	3

The frequency of extended lexical collocations in BNC is relatively high. But again, all examples found in BNC are from contracts and legal agreements. Of all the above listed units *English-Croatian Dictionary of Law* (Gačić 2010) only lists *duly authorised*, again suggesting that dictionaries favour binary units.

4.1.3 Consideration

In order for an agreement to constitute a contract there must be “something promised or actually done by one party in return for the promise of the other party” (Chartrand et al. 2009: 19). In English contract law this is called *consideration* and without it the contract cannot be legally enforced. In the Croatian legal system, on the other hand, a contract can be legally enforced even if it is not supported by consideration. This difference between the two systems is also reflected in the number of phraseological units clustering around the term *consideration*.

In the English corpus there are 244 congrams for the term *consideration*. There is a binary lexical collocation *to give consideration to something* (six instances), which was not taken here as one of the points for analysis, since the unit displays the meaning of giving “careful thought before making a decision or judgment about something”⁵, therefore making it irrelevant for the phraseology of contracts. Extended lexical collocations include:

Tab. 3: Extended lexical collocations with the term *consideration* and their frequency in the English corpus of contracts and in BNC

Extended lexical collocations	Frequency in the corpus of contracts	Frequency in BNC
3.1 consideration payable under this agreement	4	0
3.2 payment in full of the consideration payable under this agreement	2	0

Extended lexical collocations listed in table 3 do not occur as such in BNC, but some of their constituencies do. For instance, *consideration payable* occurs five times, whereas *payment in*

⁵ <<http://www.macmillandictionary.com/dictionary/british/consideration>>, 05.06.2015.

full occurs ten times. *English-Croatian Dictionary of Law* (Gačić 2010) does not list any of the above listed extended lexical collocations.

4.1.4 Contract

The term *contract* results in 617 concgrams in the English corpus. It is, therefore, expected that the list of extended units of meaning will here be shorter than in the section on extended units of meaning with the term *agreement*. Extended lexical collocations are here less frequent than in the section displaying concgrams of the term *agreement*:

Tab. 4: Extended lexical collocations with the term *contract* and their frequency in the English corpus of contracts and in BNC

Extended lexical collocations	Frequency in the corpus of contracts	Frequency in BNC
4.1 <i>to form part of the contract</i>	10	10
4.2 <i>to reasonably require under the contract</i>	2	0

The same frequency of units in the specialized corpus and in the reference corpus can again be explained by the fact that all instances from BNC contain units derived from contracts and legal agreements. None of the above listed units occurs in *English-Croatian Dictionary of Law* (Gačić 2010).

4.2 Concgrams in the Croatian corpus

Ugovor (*contract; agreement*), *parties* (*strane*) and *protučinidba* (*consideration*) were chosen as candidates for concgrams in the Croatian corpus as equivalents of the above listed English candidates for concgrams.

4.2.1 *Ugovor* (*agreement, contract*)

It has already been pointed out above that the term *agreement*, apart from referring to the first prerequisite leading to the formation of a contract, can also be used interchangeably with the term *contract*. Furthermore, if we look at the list of titles for different types of contracts that Rossini (1998: 11–14) explains, it is clear that the term *agreement* is preferred in terms of phraseology (only six types of contractual undertakings actually use the term *contract* out of a total of thirty-three listed by Rossini). This was noticed in the context of European contract law as well, where “in a large number of texts the word ‘agreement’ is used to refer to a type of ‘contract’” (Fauvarque-Cosson/Mazeaud 2008: 17). In the Croatian corpus, on the other hand, all documents are referred to as *ugovori*.

The term *ugovor* results in 16,632 concgrams in the Croatian corpus. Most of the relevant extended units of meaning include extended lexical collocations:

Tab. 5: Extended lexical collocations with the term *ugovor* and their frequency in the Croatian corpus of contracts and in *hrwac 2.0*

Extended lexical collocations	Frequency in the corpus of contracts	Frequency in <i>hrwac 2.0</i> .
5.1 <i>vlastoručno potpisati ugovor</i> (*to sign the agreement with one's own hand)	257	<i>vlastoručno potpisati</i> (118)
5.2 <i>sastavni dio ovog ugovora</i> (integral part of the agreement)	78	182
5.3 <i>bitan sastojak ovog ugovora</i> (of the essence of the agreement)	13	25
5.4 <i>sklopiti ugovor na vrijeme od</i> (to conclude a contract for the term of)	6	111
5.5 <i>poseban dio ovog ugovora</i> (separate part of this agreement)	5	0
5.6 <i>pravo jednostranog raskida ugovora</i> (the right to unilateral rescission of the agreement)	4	0
5.7 <i>raskinuti ugovor uz otkazni rok od</i> (to rescind the agreement by giving notice)	4	2
5.8 <i>otkazati ugovor uz otkazni rok od</i> (to cancel the agreement by giving notice)	3	2
5.9 <i>jednostrano raskinuti ugovor</i> (to unilaterally rescind the agreement)	2	113

Some of the units listed in table 5 are more frequent in the reference corpus than in the one created by the author. This should not lead to the conclusion, however, that the units listed above are not typical of contracts. Moreover, if one studies the wider context of units found in *hrwac 2.0*, one realizes that they either stem from contracts, newspaper articles reporting on the consequences of certain parties entering into contracts or from certain acts regulating the business of contracts or possible disputes that might arise out of them (e.g. *Consumer Protection Act*; *Civil Obligations Act*).

Opći pravni rječnik (Mukić Vidaković 2006) does not list the above discussed extended units of meaning.

4.2.2 *Strane* (parties)

The term *strane* results in 5,585 congrams in the Croatian corpus. Among the most frequent ones are the prototypical binary units e.g. *ugovorne strane* (contracting parties – 647 instances) and *suglasno utvrditi* (to determine by agreement – 134 instances), the latter representing the most frequent verb phrase that collocates with *ugovorne strane* (contracting parties).

Examples of extended lexical collocations can be found in table 6:

Tab. 6: *Extended lexical collocations with the term strane and their frequency in the Croatian corpus of contracts and in hrwac 2.0*

Extended lexical collocations	Frequency in the corpus of contracts	Frequency in hrwac 2.0.
6.1 <i>predstavljati pravu volju ugovornih strana (to represent the real will of contracting parties)</i>	3	1
6.2 <i>pozivanje ugovornih strana na naknadne izmjene (inviting contracting parties to additional amendments)</i>	3	0

Instances found in the Croatian corpus of contracts are either not present in the reference corpus or display very low frequency. *Opći pravni rječnik* (Mukić Vidaković 2006) does not list the extended units of meaning listed above.

4.2.3 *Protučinidba (consideration)*

The term *protučinidba (consideration)* occurs only four times in the Croatian corpus. It does, however, occur as part of an extended lexical collocation:

Tab. 7: *Extended lexical collocations with the term protučinidba and their frequency in the Croatian corpus of contracts and in hrwac 2.0*

Extended lexical collocations	Frequency in the corpus of contracts	Frequency in hrwac 2.0.
7.1 <i>darovati bez protučinidbe (to donate without consideration)</i>	4	1

As seen above, Croatian contracts use the term *protučinidba* only if it needs to be emphasized that there is no consideration. In the UK common-law system, such documents would not be regarded as contracts or agreements at all. The only exception to the rule that an agreement or contract does not need to be supported by consideration is the so called contract under seal or deed (e.g. *deed of donation*).

The extended unit of meaning listed in table 7 does not occur in *Opći pravni rječnik* (Mukić Vidaković 2006).

5 Discussion

By comparing extended lexical collocations extracted from the two corpora some conclusions can be drawn based on their formation patterns. Figure 3, therefore, suggests that collocations in the English corpus are mostly extended by adjectives (e.g. *integral part of the agreement; to enter into a binding agreement*). Less frequently they are extended by noun phrases and adverbs (e.g. *performance of one's duties under the agreement; payment in full of the consideration payable under this agreement; duly authorised representatives of the parties; to reasonably require under the contract*). Prepositional phrases display the lowest frequency among extensions (e.g. *to terminate the agreement by notice in writing*). Figure 4, on the other hand,

suggests that lexical collocations extracted from the Croatian corpus are mostly extended by prepositional phrases and adjectives (e.g. *darovati bez protučinidbe*; *raskinuti ugovor uz otkazni rok od*; *sastavni dio ugovora*; *bitan sastojak ugovora*). Less frequently they are extended by adverbs and nouns (e.g. *vlastoručno potpisati ugovor*; *pravo jednostranog raskida ugovora*).

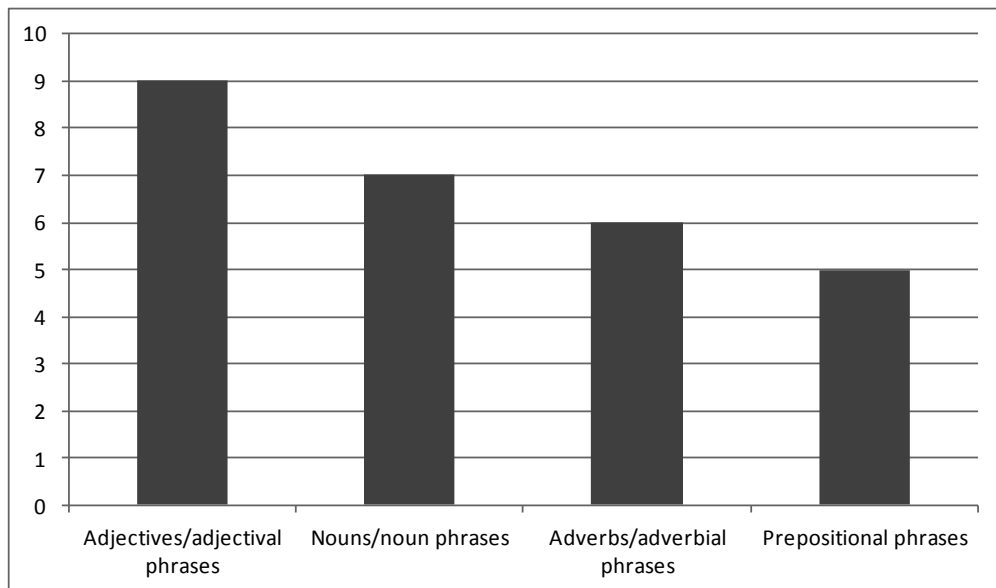


Fig. 3: Types of extensions in the English corpus

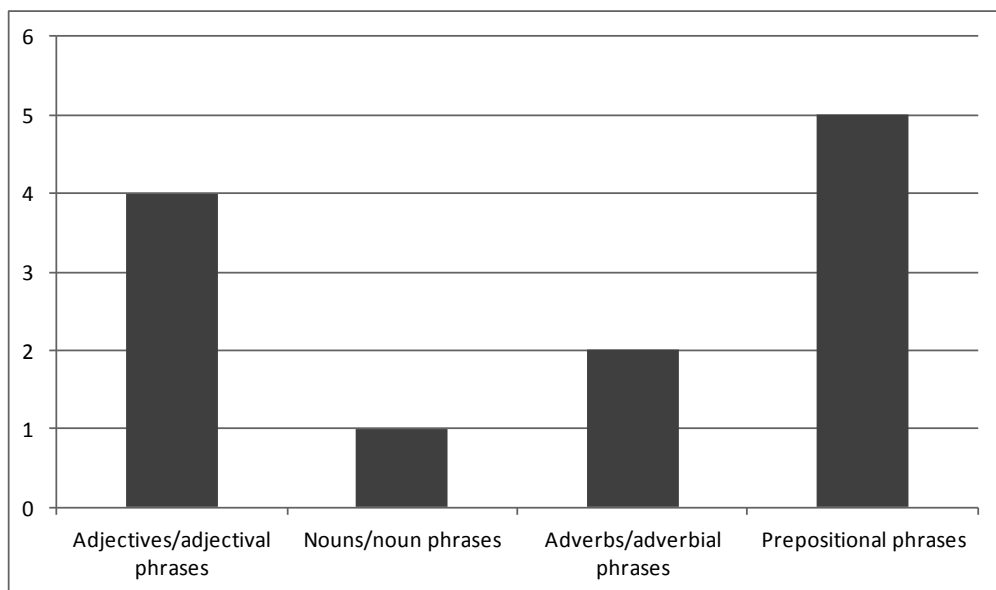


Fig. 4: Types of extensions in the Croatian corpus

There are differences in the way the extracted lexical collocations are rendered in the two languages. For instance, *bitan sastojak ugovora* is not rendered literally in English by means of an Adj+N+N structure (*important element of the agreement*), but with the help of the phrase *to be of the essence of the agreement*. Similarly, the Croatian unit *vlastoručno potpisati ugovor* would not be used in the English legal document in the form which represents the literal translation of the said unit, i. e. *to sign the agreement with one's own hand*; instead it would either use the noun phrase *handwritten signature* or the *routinae formula*: *I have hereunto set my hand*.

As regards frequency, the extended lexical collocations extracted from the two corpora are in most cases less frequent in the reference corpora. Some, however, display higher frequency in the reference corpus. This may be attributed to the size of the reference corpus and the variety of texts it includes, among them newspaper articles reporting on government bodies or companies entering into contracts as well as acts applicable to the business of contracts. Since the context in which these extended lexical collocations occur in the reference corpus is the context of contracts, one may suggest that they are typical of contract language. In dictionaries, on the other hand, extracted extended lexical collocations are highly underrepresented. Some of their constituents do occur, but they often deprive the reader of crucial information on the behaviour of these units in a wider context (e. g. *essence of the contract*).

Extending prototypical binary units sheds light on their possible variations and the relative stability of their constituents. For instance, units 2.4 and 2.5 listed in table 2 suggest that the verbs *contain* and *embody* favour different prepositions (e. g. *to contain entire understanding between the parties* vs *to embody entire understanding of the parties*). Extensions can sometimes lead to detecting translation equivalents (e. g. *sastavni dio ugovora – integral part of the agreement*; *otkazati ugovor uz otkazni rok – to terminate this agreement by written notice*), which are often far from straightforward (e. g. *bitan sastojak ugovora – to be of the essence of the agreement*; *jednostrano raskinuti ugovor – to unilaterally terminate the agreement*).

Some units are extended by vague or “flexible” (Mellinkoff 1963: 301) modifiers (e. g. *to reasonably require under the contract*). The attachment *reasonable* and *reasonably* has been subject to judicial interpretation many times and it has been shown “that a word wild and amorphous can suddenly become tame and purposeful” (Mellinkoff 1963: 302). Therefore, this vague modifier grants very specific meaning to some units (e. g. *to unreasonably withhold or delay a consent*; *reasonable time*). As regards the unit *to reasonably require under the contract*, however, the modifier *reasonably* does not aim at precision. In *Common European Sales Law* this modifier is part of the concept of good faith and is defined as “any reference to what can be expected of or by a person, or in a particular situation”⁶ A common-law system, on the other hand, does not recognize the concept of good faith, which is why courts have been asked to interpret this word on many occasions.

Extensions can also reveal the tendency of drafters of legal agreements to overemphasize a provision in order to avoid possible disputes. For instance, *third party rights* is only once extended by the verb *to grant*, but there are 16 instances of *exclusion of third party rights*. If one takes the genre and the legal system into account, one can easily detect that since 1999 the *Contracts Rights of Third Parties Act* has been in force in England, Wales and Northern Ireland and it grants rights to third parties to enforce the terms of a contract. Although this provision has been underused up until now, drafters of legal agreements still feel they need to

⁶ CESL, cited in <<http://www.trglaw.com/documents/TRGlaw-ActingReasonably.pdf>>, 10.08.2015.

emphasize that nothing in the agreement is intended to confer any benefit upon a person who is not a party to it.

The low frequency of some units in one corpus can help us describe the genre of legal agreements. For instance, the term *consideration*, occurs only four times in the Croatian corpus of contracts and only in the context of a deed of donation, i. e. when it needs to be emphasized that the donor does not expect anything in return from the donee. In the English corpus of contract, on the other hand, the term *consideration* is used even when it is clear that there is mutual exchange between the parties (e. g. *In consideration of the obligations of the Company under this agreement the Guarantor guarantees to the Company the due and punctual performance by the Buyer of the Buyer's obligations under this agreement.*)

Similarly, the unit *predstavljati pravu volju ugovornih strana* is quite frequent in the Croatian corpus, but there is no such thing as *to represent the real will of the parties* in the English corpus. The reason for this lies in the differences between the respective legal systems. Article 249 of the Croatian *Civil Obligations Act* stipulates that the expression of willingness to enter into a contract must be free and serious. In a common-law system, however, everything turns on what the parties declare and not what they want or will. Civil-law jurists, on the other hand, struggle between the two, i. e. whether “a declaration of intent” as an element of contract, should be interpreted on the basis of either the “will theory” or the “declaration theory”.⁷ Therefore, civil-law contracts rely on the real and serious will of the parties, whereas common-law contracts tend to put much more weight on the wording of the contract. This is why common-law contracts tend to be much longer than their civil-law counterparts. Additionally, in English law there are only two acts applicable to transactions between parties. The *Unfair Terms in Consumer Contract Regulations 1999* apply to consumer contracts, whereas the *Unfair Contract Terms Act 1977* applies to business relationships, but not to tenancy or other agreements relating to land. Therefore, we can trace many examples of “**particularity**” (Bhatia 1993: 137) in the corpus of British legal agreements, since in a common-law system “each party presumes, correctly, that the other party can cause it considerable difficulties by advancing a strategic misinterpretation of a contract provision” (Hill/King 2004: 904). All this results in the necessity to allow “little scope for vagueness in judicial interpretation”, which leads to the certainty of expression and wordiness being “the most valued quality in common law drafting” (Bhatia 1993: 139).

6 Concluding remarks

The aim of this paper was to show that the investigation of congrams can reveal the wider contexts of phraseological units, which contributes to many interesting and useful findings for research on legal phraseology. Legal dictionaries, although useful, have the flaw of scarcity of space. They sometimes list one or two examples of extended units of meaning, but provide no examples of usage. Sometimes they even deprive the reader of crucial information on the behaviour of the unit in a wider context. This impairs the work of a legal translator, especially if the translation is carried out into L2. Since legal dictionaries cannot include all possible combinations a lexical item can be part of, developing phraseological resources focusing on different genres (e. g. contracts, wills, judgments etc.) would help legal translators in their work. By

⁷ <<http://www.lexology.com/library/detail.aspx?g=96d2ca88-79be-41dd-8797-f2fae9be6e1b>>, 27.06.2015.

focusing on extended units of meaning we can detect their variations and stability in a wider context. If the study is conducted contrastively, one can also detect functional equivalents and contribute to achieving a translation which sounds natural. The wider context also helps us to reveal which lexical items are more frequently associated with the prototypical binary unit. This again enables us, with the help of genre analysis, to account for the frequency of a unit in the corpus.

Although comparable corpora consisting of non-translated texts cannot produce complete comparability, they represent a useful resource of phraseological patterns used in different legal genres. Most importantly, they provide us with examples of usage, which is something that dictionaries can never fully achieve. Since this paper focuses on extended lexical collocations only, it represents a small insight into patterns of extended units of meaning in contracts. Therefore, future research might profit from analyzing all concgrams from both corpora, and thereby could contribute to a complete analysis of the patterning of these units in the language of contracts. It is expected, however, that such a study would have to involve an in-depth genre analysis in order to account for the differences, both in patterning and frequency, between the two comparable corpora. This especially applies to extended binomial expressions, which are more typical of legal English than legal Croatian.

Studying and analyzing extended units of meaning in contracts can, therefore, contribute to filling yet another gap in the research on legal phraseology. Since the study would provide a contrastive insight into the phraseology of common-law and civil-law contracts, it might also contribute to resolving translation difficulties when translating between these two legal systems.

References

- Agić, Željko/Ljubešić, Nikola/Merkel, Danijela (2013): "Lemmatization and Morphosyntactic Tagging of Croatian and Serbian." *Proceedings of BSNLP 2013*. 20.05.2014 <<http://nlp.ffzg.hr/data/publications/nljubesi/agic13-lemmatization.pdf>>.
- Bhatia, Vijay K. (1993): *Analysing genre*. Harlow: Longman.
- Bhatia, Vijay K./Langton, Nicola M./Lung, Jane (2004): "Legal discourse: opportunities and threats for corpus linguistics." *Discourse in the Professions. Perspectives from Corpus Linguistics*. Eds. Ulla M. Connor/Thomas A. Upton. Amsterdam/Philadelphia: Benjamins. 203–231.
- Biel, Lucja (2011): "Areas of similarity and difference in legal phraseology: collocations of key terms in UK and Polish company law." *Multilingual phraseography: translation and learning applications*, Vol. 2. Eds. Antonio Pamies/Lucia Luque-Nadal/José Manuel Pazos. 20.09.2014. http://www.academia.edu/2602242/Areas_of_similarity_and_difference_in_legal_phraseology_collocations_of_key_terms_in_UK_and_Polish_company_law.
- Black, Henry Campbell (2004): *Black's Law Dictionary*. St. Paul, Minnesota: West.
- Bowker, Lynne/Pearson, Jennifer (2002): *Working with Specialized Language. A practical guide to using corpora*. London/New York: Routledge.
- Bukovčan, Dragica (2009): "Binominal Expressions in the German and English Language of Criminal Law." *Curriculum, Multilingualism and the Law*. Eds. Lelija Sočanac/Christopher Goddard/Ludger Kremer. Zagreb: Nakladni zavod Globus. 61–78.
- Carvalho, Luciana (2007): "Translating contracts and agreements from a corpus linguistics perspective." *Language and the Law: International Outlooks*. Eds. Krzysztof Kredens/Stanislaw Goźdz-Roszkowski. Frankfurt am Main et al.: Lang. 109–121.

- Chartrand, Marcella/Millar, Catherine/Wiltshire, Edward (2009): *English for Contract and Company Law*. London: Reuters.
- Cheng, Winnie/Greaves, Chris/Warren, Martin (2006): "From n-gram to skipgram to congram." *International Journal of Corpus Linguistics* 11.2: 411–433.
- Cheng, Winnie/Greaves, Chris/Sinclair, John/Warren, Martin (2008): "Uncovering the Extent of Phraseological Tendency: Towards a Systematic Analysis of Congrams." *Applied Linguistics* 30.2: 236–252.
- De Groot, Gerard (1999): "Zweisprachige juristische Wörterbücher". *Übersetzen von Rechtstexten. Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache*. Hrsg. Peter Sandrini. Tübingen: Narr. 203–227.
- Evert, Stefan (2004): *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. PhD dissertation: University of Stuttgart.
- Fauvarque-Cosson, Bénédicte/Mazeaud, Dennis (2008): *European Contract Law: Materials for a Common Frame of Reference: Terminology, Guiding Principles, Model Rules*. München: de Gruyter.
- Gačić, Milica (2010): *Englesko-hrvatski rječnik prava i međunarodnih i poslovnih odnosa*. Zagreb: Školska knjiga.
- Goźdz-Roszkowski, Stanislaw (2011): *Patterns of Linguistic Variation in American Legal English*. Frankfurt am Main et al.: Lang.
- Granger, Sylviane/Meunier, Fanny (2008): *Phraseology: An Interdisciplinary Perspective*. Amsterdam/Philadelphia: Benjamins.
- Gries, Stefan (2008): "Phraseology and linguistic theory: A brief survey." *Phraseology: An Interdisciplinary Perspective*. Eds. Sylviane Granger/Fanny Meunier. Amsterdam/Philadelphia: Benjamins. 3–25.
- Hill, Claire/King, Christopher (2004): "How do German Contracts do as much with fewer words?" 15.12.2014. <http://www2.nkfust.edu.tw/~chairman/law/class3/970412/2004_why_german_contract_uses_less_words_than_us_contracts.pdf>.
- Junačko, Zdravko/Rotar, Dragutin (2007): *Zbirka ugovora građanskog i trgovačkog prava 4*. Zagreb: Poslovni zbornik.
- Kilgariff, Adam/Baisa, Vít/Bušta, Jan/Jakubiček, Miloš/Kovář, Vojtěch/Michelfeit, Jan/Richly, Pavek/Suchomel, Vít (2014): "The Sketch Engine: ten years on." *Lexicography*: 1–30. 01.09.2014. <<http://www.sketchengine.co.uk>>.
- Kjær, Anne Lise (2007): "Phrasemes in Legal Texts." *Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*. Hrsg. Harald Burger/Dmitrij Dobrovolskij/Peter Kühn/Niel R. Noerrick. Berlin/New York: de Gruyter. 506–515.
- Kordić, Ljubica, Ivković, Ana (2010): "Frazelogija njemačkoga i hrvatskoga pravnog idioma: komparativni prikaz." *Pravni vjesnik*. Osijek: Pravni fakultet Sveučilišta J.J.Strossmayera. 115–131.
- Llopis, Maria A. (2007): "The Untranslatability of Law? Lexical Differences in Spanish and American Contract Law." *European Journal of English Studies* 11.1: 17–28. 14.02.2014. <www.tandf.co.uk/journals>.
- Ljubešić, Nikola/Klubička Filip (2014): "{bs,hr,sr}WaC – Web corpora of Bosnian, Croatian and Serbian" 01.09.2014. <<http://nlp.ffzg.hr/data/publications/nljubesi/ljubesi14-bs.pdf>>.
- Mellinkoff, David (1963). *The Language of the Law*. Oregon: Wipf and Stock Publishers.
- Millett, Peter/Walker, Robert, eds. (2014): *Encyclopaedia of Forms and Precedents*. 14.01.2014. <<https://www.lexisnexis.com/uk/legal>>.
- Mukić Vidaković, Marta (2006): *Opći pravni rječnik*. Zagreb: Narodne novine.
- Nesselhauf, Nadja (2004): "What are collocations?" *Phraseological Units: Basic Concepts and their Application*. Eds. David J. Allerton/Nadja Nesselhauf/Paul Skandera. Basel: Schwabe. 1–21.
- Rossini, Christine (1998): *English as a Legal Language*. London: Kluwer Law International.
- Scott, Mike (2011): *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software.

Scott, Mike (2015): *WordSmith Tools Help*. 20.01.2015. <http://www.lexically.net/downloads/version5/HTML/?viewing_concgrams.htm>.

Sinclair, John (2004): *Trust the Text: Language, Corpus and Discourse*. London/New York: Routledge.

Stubbs, Michael (2002): *Words and Phrases. Corpus Studies of Lexical Semantics*. Malden/Oxford/Carlton: Blackwell.

Swales, John (1990): *Genre Analysis: English in Academic and Research Settings*. New York: Cambridge University Press.

The British National Corpus, version 3 (BNC XML Edition) (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. 28.08.2014. <<http://www.natcorp.ox.ac.uk/>>.

Triebel, Volker (2009): "Pitfalls of English as a Contract Language." *Translation Issues in Language and Law*. Eds. Frances Olsen/Alexander Lotz/Dieter Stein. Hampshire/New York: Palgrave Macmillan. 147–182.

Additional Links:

Macmillan Dictionary: <www.macmillandictionary.com/dictionary/british/consideration> [05.06.2015].

Reference corpora: <www.ilc.cnr.it/EAGLES/corpus/typ/node18.html> [10.08.2015].

Acting Reasonably: <www.trglaw.com/documents/TRGlaw-ActingReasonably.pdf> [10.08.2015].

The civil law interpretation of declarations of intent by contracting parties: <www.lexology.com/library/detail.aspx?g=96d2ca88-79be-41dd-8797-f2fae9be6e1b> [27.06.2015].

*Katja Dobrić Basanež, prof.
Department of Foreign Languages
Faculty of Law, University of Rijeka
Hahlić 6
51 000 Rijeka, Croatia
kdobric@pravri.hr*