

*Katja Dobrić Basanež*  
*Pravni fakultet Sveučilišta u Rijeci*  
*kdobric@pravri.hr*

## **Interpreting Phraseological Units in Contracts: The Case of Extended Term–Embedding Collocation**

Extended units of meaning (Sinclair 2004) have been scarcely investigated thus far in legal phraseology with the exception of research into lexical bundles (Goźdz–Roszkowski 2006, 2011; Kopaczyk 2013; Breeze 2013; Tománkova 2016; Biel 2017). This paper is therefore an attempt to show that the Sinclairian wider–context–perspective may prove to be especially useful for the study of phraseological units in legal genre since it helps us to reveal their collocational framework, allowing both grammatical and phraseological patterns to emerge. The paper focuses on extended ‘term–embedding collocations’ (Biel 2014b) extracted from the English and Croatian comparable corpora of contracts by means of *WordsmithTools 6.0* (Scott 2012). It highlights some of the most striking examples supporting the above hypothesis and it accounts for their interpretation by means of analysing the extralinguistic context of phraseological units in contracts. It may be suggested that this study represents an attempt to fill a gap in research on legal phraseology due to the fact that private legal documents tend to be largely underrepresented in this specialized phraseology. It may also be suggested that by focusing on extended units of meaning in legal Croatian the paper attempts to fill yet another gap in corpus–based studies of legal language, which tend to be largely Anglocentric. Finally, the paper may, apart from revealing the stability of legal phrasemes, also represent a useful resource for translator training since it offers the wider context of a term or an expression in contract language.

---

### **1. Introduction**

While it is true that phraseology of general–purpose language (LGP) has been researched from many linguistic perspectives in the last twenty years, this does not apply to phraseology of special–purpose (LSP) language. As suggested by Pontrandolfo and Goźdz–Roszkowski (2015), this trend may be attributed to a variety of reasons, such as, the absence of rigorous methodologies to identify phraseological units in a specialized discourse, the focus on the ter–

minological aspect of these units, the domain-specificity of a given disciplinary discourse, etc.

Corpus linguistics, however, has marked a significant shift in the studies on LSP phraseology in general and legal phraseology in particular. As proposed by Pontrandolfo and Goźdz-Roszkowski (2015), corpus-based research of legal phraseology can be divided into five main groups:

- studies that analyse lexico-syntactic combinations in legal language, with a preference for specialised collocation (Biel 2011);
- studies that deal with the formulaic nature of legal language, although it needs to be pointed out that this group of studies has recently been expanded by research into lexical bundles carried out from different angles: synchronic (Goźdz-Roszkowski 2006), standardisation of early legal discourse (Kopaczyk 2013), variation in legal discourse (Goźdz-Roszkowski 2011)
- lexicographic investigations aimed at compiling legal dictionaries or glossaries (De Groot 1999)
- corpus-based analysis of phraseology applied to *contrastive analysis* of legal language (Pontrandolfo 2013, Tabares Plasencia 2014) and/or *legal translation* (Biel 2014a)
- studies that focus on the way legal patterns weave an intricate web of semantic meanings by resorting to a wider notion of phraseology (Mazzi 2010; Goźdz-Roszkowski and Pontrandolfo 2013, 2014)

The proposed classification can by no means be regarded as “a ‘sealed off’ box with fully defined borders” (Goźdz-Roszkowski and Pontrandolfo 2015: 134). As a matter of fact, some of the recent studies (Biel 2015; Monzó 2015; Dobrić Basanež 2015) suggest a constant interaction between the five groups. This paper is no exception in that regard.

It is hypothesised in this paper that Sinclair’s model of extended lexical units (Sinclair 2004), which proposes that focus in both corpus and phraseological research should be put on large phraseological units rather than on individual words, may yield useful results for the study of these units since it helps us to reveal the typical behaviour of a unit in terms of its usage in context. Since phraseologists must carefully define the object of their study (Gries 2008), let us at this point recall the classifications of phraseological units in legal language. Although these have been rarely undertaken, there are “few remarkable exceptions” (Goźdz-Roszkowski and Pontrandolfo 2015). According to Kjær (1990), most frequently found in legal discourse are:

- (1) multi-word terms,
- (2) specialized collocations and
- (3) formulaic expressions and standard phrases.

More recently, there have been two classifications of lexical bundles in legal discourse, one focusing on cross-genre classification (Goźdz-Roszkowski

2011) and the other on the early legal discourse of Scottish burghs (Kopaczyk 2013).

Finally, combining Kjær's (1900, 2007) and Goźdz-Roszkowski's (2011) approaches to typologies of word combinations in legal language, Biel (2014b) suggests her own typology based on corpus-based analyses of textual mapping in EU law.

Her typology includes:

- (1) text-organizing patterns (e.g., amending and closing formulas);
- (2) grammatical patterns (e.g., patterns which express deontic modality such as *shall, must, should, may*);
- (3) term-forming patterns (multi-word terms);
- (4) term-embedding collocations, i.e., collocates of terms which embed terms in cognitive scripts, evidencing their combinatory properties (e.g., *to hold shares*);
- (5) lexical collocations (e.g., *subject to this Regulation*).

Although term-embedding collocations would in phraseological research traditionally be regarded as lexical collocations, Biel distinguishes between non-terminological collocations and specialised phrasemes “clustering around terms” (Pontrandolfo 2015: 148). She thus defines lexical collocations as “routine formulae at the microstructural level which are not built around terms” (2014b: 181). Her category of lexical collocations, includes, among others, non-terminological lexical bundles (e.g. *within the meaning of; in accordance with the*), whose identification is, unlike the one of term-embedding collocations and multi-word terms, based on recurrence. Additionally, as pointed out by Biel, “lexical bundles do not fit the existing categorizations of legal phraseology” (Biel 2017: 12); hence, they rather “cut across all these categories” (ibid.). Thus, for instance, although Breeze (2013) views units such as *the Articles of Association* and *request for confidential treatment* as content word bundles, we may also view them as multi-word terms. This also applies to some of the content-abstract concepts (e.g. *the Treaty of Lisbon*) discussed by Tománková (2016). Similarly, some of the verb bundles listed by Breeze (2013) (e.g. *contemplated by this agreement*) may be regarded as term-embedding collocations, while some examples of content-abstract concepts in Tománková's study (2016), (e.g. *the ordinary legislative procedure*) can also be categorized as nominal term-embedding collocations. Although lexical bundles can also be viewed as ‘extended collocations’ (Biber et al. 1999: 989), in this paper the extended term-embedding collocation is viewed as a “structurally complete sequence” (Gabrovšek 2014: 10) consisting of the “prototypical, i.e. binary collocation” (ibid.) and at least one additional grammatical or lexical element. Extensions may thus range from conjunctions (*u razumnom roku; within reasonable time*), modifiers (e.g. *automatically terminate this agreement; vlastoručno potpisati ugovor*), lexical bundles (e.g. *at the time of signing the contract; sklopiti ugovor na vrijeme od*), or even non-terminological collocations (e.g. *imati pravo raskinuti ugovor*). We may thus claim that the extended term-

embedding collocations discussed in this paper represent phraseological units in the broad sense. Nevertheless, “they illustrate – however selectively – typical actual use” (ibid.: 18).

The paper attempts to answer the following research questions:

- 1) What does the wider context of term–embedding collocations reveal?
- 2) To what extent can comparable corpora contribute to producing close or equivalent extended term–embedding collocations in two languages and legal systems?
- 3) To what extent can an even wider context and the extralinguistic context reveal the equivalent extended term–embedding collocation in the target text if it turns out downright for the corpus data to do so?

## 2. Data

In order to answer the research questions, we rely upon a comparable bilingual corpus of English and Croatian contracts. It must be pointed out that the initial intention was to create a corpus consisting of authentic contracts, but this proved to be impracticable due to the confidentiality of information included in private legal documents. As a result, legislation and documents that can be easily accessed from the Web (e.g., *Acquis Communautaire*) prevail in legal corpora, whereas other text types are largely under-represented.

This study is an attempt to fill this gap by focusing on corpora consisting of contract and agreement samples that are used by lawyers on a daily basis, with Croatian contracts extracted from the digital edition of the book *Zbirka ugovora građanskog i trgovačkog prava 4* (Junačko and Rotar 2007) and the English ones mainly downloaded from the online edition of *Encyclopaedia of Forms and Precedents* (Millet and Walker 2014). Since “corpora intended for LSP can be smaller than those used for LGP studies” (Bowker and Pearson 2002: 48), the Croatian corpus (CroCon) of 105,583 and the English one (EnCon) of 434, 118 tokens can be regarded as large enough. Since corpora consist of contracts from two different legal systems (EnCon – common law; CroCon – civil law), the discrepancy between the corpora in terms of their size can on one hand be attributed to the “**particularity**” of a common-law system, i.e., its concern with respect to “*not being misunderstood by the specialist community*” (Bhatia 1993: 137) and on the other, to the complete freedom to contract given to parties in a common-law system. Additionally, the judge in a common law system is expected “to read the contract exclusively on the basis of the provisions that are written therein” (Moss 2007: 5). As a result, common-law contracts tend to be significantly longer than their civil-law counterparts given the fact that the latter are usually regulated by statute law.

It can be claimed that corpora are comparable linguistically, since contracts tend to display a high degree of repetitiveness (Goźdz-Roszkowski 2000 cited in Goźdz-Roszkowski 2006) and low variation in word choice (Goźdz-Roszkowski 2011).

In order to show “the deviance of special corpora”<sup>1</sup> and that the extended units of meaning are typical of legal phraseology, in particular of contracts and legal agreements, two reference corpora are consulted, hrWaC 2.0. (Ljubešić and Klubička 2014) for Croatian, consisting of 1,397,757,548 and the BNC<sup>2</sup> for English, consisting of 112,181,015 tokens.

### 3. Methodology

The extraction of extended term–embedding collocations is based on computing collocates of the chosen nodes and analysing their patterns by means of Wordlist and Concordance in *Wordsmith Tools 6.0*. (Scott 2012). In order for the extension to be classified as a phrase in its own right, it needs to occur at least twice (Sinclair 2004: 28) in the whole corpus. Although phraseological units both in general and special–purpose language need to occur together at least 5 times, this criterion cannot be applied to all types of phraseological units, and especially not to the extended ones given the fact that “each extended sequence is typically (but not unavoidably) less frequent and phraseologically “looser” than the corresponding basic form” (Gabrovšek 2014: 10). Additionally, as asserted by Biel (2014a), frequency is not the main criterion for all types of phraseological units (e.g., term–embedding collocations and multi–word terms).

The extraction of extended term–embedding collocations was at first intended to start from the wordlist of each corpus ordered in terms of frequency. When the wordlists were studied, however, it was revealed that some of the most frequent words tend to represent different parties of contractual undertakings (e.g., landlord, tenant, buyer, seller, etc.), usually performing the function of a subject. In order to produce a nourishing ground for the analysis of the wider–context–perspective and in order to reveal equivalent or close phraseological patterns it was thus decided that the extraction of extended term–embedding collocations will be based on Pontrandolfo’s methodology for the extraction of lexical collocations (2015). He chooses an innovative method for the extraction of nodes by following Schank and Abelson’s notion of *script*, i.e., “a structure that describes appropriate sequences of events in a particular context” (Schank and Abelson 1977: 141 cited in Pontrandolfo 2015: 144).

It may thus be suggested that the following represents a typical sequence of events in a contract:

One *party* makes an *offer* to another party, which suggests that they make a *contract* on certain *terms*. If the offer finds *acceptance* with the other party, it is deemed that both parties have the same understanding of the terms of the *agreement*. In English law, however, the contract needs to be supported by *consideration* in order to exist. This usually refers to the key *obligations*, which

---

1 <http://www.ilc.cnr.it/EAGLES/corpus/typ/node18.html> (Accessed 10 August 2015)

2 <https://www.sketchengine.co.uk/british-national-corpus/> (Accessed 23 September 2016)

must be completed within a certain period of *time*. Once these obligations have been completed, the parties have reached the moment of essential *termination*. Other reasons to terminate a contract include: impossibility of *performance*, *breach* of contract, prior agreement, and *rescission*.

Following the proposed depiction of a contract script, the italicized 12 terms were then chosen to represent the nodes of extended term–embedding collocations. Their frequency in the two corpora is listed in Table 1 below.

<b>Nodes in EnCon and their frequency</b>	<b>Nodes in CroCon and their frequency</b>
<i>agreement</i> (2,731)	
<i>contract</i> (538)	<i>ugovor</i> (2,654)
<i>party</i> (1,628)	<i>strana</i> (1,294); <i>stranka</i> (49)
<i>offer</i> (105)	<i>ponuda</i> (39)
<i>acceptance</i> (44)	<i>prihvat</i> (194)
<i>consideration</i> (105)	<i>protučinidba</i> (4)
<i>term</i> (1,301)	<i>uvjet</i> (114)
<i>time</i> (1,479)	<i>vrijeme</i> (206); <i>rok</i> (366)
<i>obligation</i> (775)	<i>obveza</i> (392); <i>obaveza</i> (8)
<i>termination</i> (251)	<i>otkaz</i> (34) / <i>otkazivanje</i> (6)
<i>performance</i> (171)	<i>ispunjenje</i> (37) / <i>ispunjavanje</i> (4)
<i>breach</i> (269)	<i>kršenje</i> (4)
<i>rescission</i> (3)	<i>raskid</i> (33)

Table 1. Nodes chosen to represent a contract script and their frequency in EnCon and CroCon

As witnessed by Table 1 the terms *agreement* and *contract* are in EnCon used interchangeably. As a matter of fact, most contracts are usually called agreements (Dobrić Basanež 2015). This is also supported by the frequency of the terms *agreement* and *contract* in EnCon, with the latter being less frequent. Croatian language, on the other hand, refers to both contracts and agreements as *ugovori*. The term *agreement* in EnCon also denotes the meaning of “the meeting of the minds”. Croatian language, on the other hand, refers to both contracts and agreements as *ugovori*. As far as the term *party* is concerned, however, the Croatian language disposes of two variants, hence, *strana* and *stranka*, although the latter variant occurs less frequently. In addition, the term *time* is usually rendered as *vrijeme* in Croatian, but in some cases it can also be rendered as *rok* (*deadline*).

Finally, as far as the analysis of reference corpora is concerned, Sketch Engine software (Kilgariff 2014) is used so as to reveal the frequency of the extracted extended term–embedding collocations in hrWaC 2.0. (Ljubešić and Klubička 2014) and the BNC<sup>3</sup>.

3 <https://www.sketchengine.co.uk/british-national-corpus/> (Accessed 23 September 2016)

## 4. Extended term–embedding collocations

### 4.1. Extended term–embedding collocations in EnCon

The wider context of terms chosen to represent a contract script has yielded useful results in terms of classifying extended term–embedding collocations as phraseological units in their own right. Due to the scarcity of space, this paper analyses only the most striking examples clustering around some of the nodes both in EnCon and in CroCon (see Tables 2 and 3).

Extended term–embedding collocations in EnCon	Frequency in EnCon (normalized)	Frequency in the BNC (normalized)
<b>Verbal collocations</b>		
<i>to vary the agreement by agreement between X and Y</i>	27	1
<i>to terminate this agreement by written notice</i>	8	0
<i>to unreasonably withhold agreement</i>	6	44 (to unreasonably withhold consent)
<i>to be entitled to terminate this agreement</i>	5	7
<i>to execute an agreement in counterparts</i>	5	0
<i>to terminate this agreement with immediate effect</i>	4	2 (to terminate the contract with immediate effect)
<i>to automatically terminate this agreement (used in passive: This Agreement shall automatically terminate/terminate automatically)</i>	3	0
<i>to form part of the contract</i>	10	10
<i>not to unreasonably withhold acceptance (whose acceptance may not be unreasonably withheld)</i>	2	2
<i>to impose an obligation to use reasonable endeavours</i>	4	1 (obligation to use reasonable endeavours)
<i>to give notice of termination</i>	3	2
<b>Nominal collocations</b>		
<i>at the time of signing the contract</i>	16	4
<i>upon acceptance of the offer</i>	6	0

<i>after the date of the offer</i>	4	0
<i>integral part of this agreement</i>	2	0
<i>as at the date of the offer</i>	2	14 ( <i>as at the date of</i> )
<i>payment in full of the consideration payable under this agreement</i>	2	5 ( <i>payment in full</i> )
<i>during the term of this agreement</i>	2	1

Table 2. Extended term–embedding collocations in EnCon

Most of the extended term–embedding collocations listed in Table 2 do not occur in the 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. Sometimes, however, only parts of these extended units occur in the reference corpus, suggesting that extensions are not typical of the phraseology of contracts, but are used in general language as well (e.g., *payment in full*).

As witnessed by Table 2, some of the collocations extracted from EnCon occur more frequently in the BNC. This can on one hand be attributed to different corpus sizes, with the reference corpus being significantly larger than EnCon. On a different note, if we study the wider context of units found in the BNC, we realize that they all stem from legal sources.

#### 4.2. Extended term–embedding collocations in CroCon

The analysis of nodes chosen to represent the contract script in CroCon has also confirmed the usefulness of Sinclairian wider context perspective. The Croatian dataset, although not as extensive as the English one, lists numerous extended term–embedding collocations typical of the phraseology of contracts (see Table 3 below).

Extended term–embedding collocations	Frequency in CroCon (normalized)	Frequency in hrvac 2.0. (normalized)
<b>Verbal collocations</b>		
<i>vlastoručno potpisati ugovor</i>	257	118 ( <i>vlastoručno potpisati</i> )
<i>sklopiti ugovor na vrijeme od</i>	10	111
<i>zaključiti ugovor na vrijeme od</i>	6	0
<i>automatski produžiti ugovor</i> (used exclusively in passive: <i>Ovaj se ugovor automatski produžuje</i> )	6	0
<i>zaključiti ugovor za razdoblje od</i>	4	0
<i>raskinuti ugovor uz otkazni rok od</i>	4	2

<i>sklopiti ugovor na neodređeno vrijeme</i>	4	14
<i>sklopiti ugovor na vrijeme od</i>	4	1
<i>otkazati ugovor uz otkazni rok od</i>	3	2
<i>imati pravo raskinuti ugovor</i>	3	75
<i>sklopiti/sastaviti ugovor u X istovjetna primjerka</i>	2	12
<i>raskinuti ugovor s trenutnim učinkom</i>	2	1
<i>izmijeniti odredbe uz suglasnost svih ugovornih strana</i>	3	0
<i>otkloniti materijalne nedostatke u primjerenom roku</i>	3	0
<i>imati pravo otkaza ugovora</i>	3	1
<b>Nominal collocations</b>		
<i>sastavni dio ovog ugovora</i>	78	182
<i>bitan sastojak ovog ugovora</i>	10	25
<i>prilikom sklapanja/potpisivanja/zaključivanja ovog ugovora</i>	12	247 (prilikom potpisivanja ugovora); 286 (prilikom sklapanja ugovora); 17 (prilikom zaključivanja ugovora)
<i>po potpisu ovog ugovora</i>	8	91
<i>s danom potpisa ovog ugovora</i>	7	0
<i>pri potpisu ovog ugovora</i>	6	22
<i>za vrijeme trajanja ovog ugovora</i>	5	109
<i>u vrijeme sklapanja ovog ugovora</i>	5	107
<i>tijekom trajanja ovog ugovora</i>	4	52
<i>u razumnom roku</i>	4	1,063

Table 3. Extended term–embedding collocations in CroCon

Some of the units listed in Table 3 are more frequent in the reference corpus than CroCon. This can again on one hand be attributed to different corpus sizes and, on the other, to the text types in which the units occur, these being 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*).

## 5. Analysis

The analysis of extensions of the chosen nodes tends to reveal their grammatical, phraseological and genre-specific features. As far as grammatical structures are concerned, for instance, it is interesting to note that some extensions reveal the usage of a double negative. (e.g., *such agreement not to be unreasonably withheld or which agreement shall not be unreasonably withheld*). Other co-texts suggest that the unit *to unreasonably withhold* is also used as a collocate of *consent*, *approval*, and *acceptance*, all of which are exclusively used in negative contexts, displaying the same structure as above (e.g., *such consent not to be unreasonably withheld*).

Some extensions seem to allow more variation than others (e.g., *to terminate this agreement by written notice may also be rendered as to terminate this agreement by giving written notice/by notice in writing or to give written notice of termination*). The equivalent extended unit in CroCon, however, does not allow variation of their constituencies (e.g., *otkazati ugovor uz otkazni rok*). CroCon does in turn allow variation for the unit *zaključiti ugovor na vrijeme od*, which may also be rendered as *zaključiti ugovor za razdoblje od*. Similarly, the unit *at the time of signing the contract* does not allow variation, unlike its counterpart in CroCon which may be rendered as *s danom potpisa ovog ugovora, prilikom potpisivanja ovog ugovora or pri potpisu ovog ugovora*. In a similar vein, when we wish to express that something occurs during the time period for which the agreement or contract has been concluded, this is in EnCon rendered as *during the term of this agreement*, whereas CroCon witnesses two variants, *za vrijeme trajanja ovog ugovora* and *tijekom trajanja ovog ugovora*.

Some units are extended by vague or “flexible” (Mellinkoff 1963: 301) modifiers, but nevertheless tend to express a purposeful meaning. This is the case with the Croatian unit *u primjerenom roku* (*within appropriate time*) given the fact that the agent specifying appropriate time is defined by statute law, hence, the *Civil Obligations Act* (hereinafter: COA). Thus, in the case of the unit *otkloniti uočene nedostatke u primjerenom roku* or *otkloniti materijalne nedostatke u primjerenom roku* Art. 608 of the COA states that “a contractee who has duly informed a contractor that there are some defects in the performed works may request removal of such defects and specify the time appropriate for this removal.”<sup>4</sup> Similarly, in the case of the defect in the lease, the lessee is the one who determines appropriate time (Art. 559). In the case of the purchase on trial, the purchaser determines appropriate time (if such time is not defined by the wording of the contract) within which the buyer must inform him/her whether he/she consents to purchase (Art. 456 para. 1).

Another collocation extended by a vague modifier and deserving attention is the unit *to impose an obligation to use reasonable endeavours*, which “coupled with a clear objective [...] is capable of constituting an enforceable obligation that may not always be easy to satisfy”<sup>5</sup>. Needless to say, the meaning of

4 My translation.

5 <http://www.linklaters.com/Insights/Publication1403Newsletter/PublicationIssue20070605/Pages/PublicationIssueItem2385.aspx> (Accessed 12 November 2016)

this unit has very often been subject to judicial interpretation. Most recently judges contrasted the meaning of this unit with the meaning of the unit *an obligation to use best endeavours*, suggesting that

“there may be a number of reasonable courses which could be taken in a given situation to achieve a particular aim. An obligation to use reasonable endeavours to achieve the aim probably only requires a party to take one reasonable course, not all of them, whereas an obligation to use best endeavours probably requires a party to take all the reasonable courses he can.”<sup>6</sup>

The unit *an obligation to use best endeavours* may further be contrasted with the unit *an obligation to use all reasonable endeavours*, the latter implying an even stricter meaning.<sup>7</sup>

Some phraseological units occur in one corpus, but are non-existent in the other. Reasons may range from differences in phraseological conventions to differences between the principles of contract law. The unit *vlastoručno potpisati ugovor* has no direct equivalent in EnCon (*\*to sign the agreement with one’s own hand*), which suggests that English contracts tend to express the fact that the parties have signed the agreement using different patterns, hence, either by ending a contract with the collocation *handwritten signature* or by using the “closing formula” (Biel 2014b) *I have hereunto set my hand*. A similar trend is also depicted by the units in EnCon which refer to the time before or after the making of an offer (e.g., *as at the date of the offer*, *upon acceptance of the offer*, *after the date of the offer*). No direct equivalents of the units occur in CroCon. Furthermore, units clustering around the term *ponuda* in CroCon only refer to the content of an offer (e.g., *prema ponudi*, *pod uvjetima istovjetnim ponudi*). This can be attributed to the fact that any time periods with respect to the acceptance of an offer are specified by the COA. Therefore, Art. 263 of the Act stipulates that

“[a]n offer made to a person that is present shall be deemed to be rejected if it is not accepted without delay, unless it may be inferred from the circumstances that the offeree is entitled to a certain period of time for consideration.”<sup>8</sup>

The co-text of the unit *after the date of the offer* suggests that the acceptance of offer is in English law not regulated by statute, but is subject to the negotiation of the parties, which in our case amounts to 5 business days, e.g.

6 Statement 33 of *Rhodia International Holdings Ltd v Huntsman International LLC* (2007), [http://www.glovers.co.uk/news\\_article232.html](http://www.glovers.co.uk/news_article232.html) (Accessed 12 November 2016)

7 See, for instance, Statement 123 of the case *Yewbelle Ltd v London Green Developments Ltd and Knightsbridge Green Ltd* (2006), [http://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2006/3166.html&query=\(title:\(+Yewbelle+\)\)+AND+\(title:\(+v+\)\)+AND+\(title:\(+London+\)\)+AND+\(title:\(+Green+\)\)+AND+\(title:\(+Developments+\)\)](http://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2006/3166.html&query=(title:(+Yewbelle+))+AND+(title:(+v+))+AND+(title:(+London+))+AND+(title:(+Green+))+AND+(title:(+Developments+))) (Accessed 12 November 2016)

8 <http://www.zakon.hr/z/75/Zakon-o-obveznim-odnosima> (Accessed 19 November 2016)

(1) *Each Offer made under clause 3.1 above may be accepted by the Buyer within a period of 5 Business Days after the date of the Offer.*

It thus follows that, although in this case the formation of a pattern in the target text is straightforward (e.g., *upon acceptance of the offer – po prihvatu ponude*), the usage of this unit in the target text may be regarded as untypical due to the principles of both contract law and the functioning of the respective legal system in general. In such cases it would be best to advocate the application of international contract principles and preference of non-state law (Moss 2007).

The term *rescission*, on the other hand, occurs only 3 times in EnCon, while the term *raskid* has 33 entries in CroCon. This may on one hand be explained by the tendency of the English language towards rendering the concept of rescission by means of the verb *to rescind*, e.g.

(2) *If the Buyer fails to perform the obligations on its part contained in clause 2 within 20 Working Days after receipt of a notice from the Seller specifying the particular breach complained of and stating that it is being served pursuant to the provisions of this clause; ... then and in any such case the Seller may **rescind**<sup>9</sup> this Agreement by notice to the Buyer.*

CroCon, on the other hand, witnesses instances in which the term *rescission* is used interchangeably with the term *termination*, thus referring to the general termination of contract regardless of the circumstances, e.g.

(3) *Svaka strana može raskinuti ugovor s otkaznim rokom od 15 dana.*

This trend was recently reinforced by the European Union in its efforts to harmonize private contract law in order to deal with the problem of the abundance of different terminology referring to the concept of termination, especially with respect to common-law and civil-law legal systems (Keglević 2013: 680).<sup>10</sup> In light of this effort, there is a tendency to widen the meaning of some terms referring to the mentioned concept, which also seems to apply to the concept of *raskid* in CroCon. The Croatian legal system also tends to allow the remedy of *raskid* in situations which the English law treats by means of different doctrines. For instance, if events occur after the formation of contract which make its performance impossible, English law allows for the discharge of contract under the doctrine of frustration (Treitel 1995: 778). Croatian law, on the other hand, calls for the rescission or variation of a contract in the case of changed circumstances (Petrić 1991).

---

9 My emphasis.

10 It remains to be seen how lawyers will deal with tensions created by translations between civil law and common law contracts now that Great Britain has decided to exit the EU and that the application of non-state law such as European PECL (European Principles of Contract Law) will no longer be an option.

Finally, extensions can sometimes lead to detecting translation equivalents (e.g., *sastavni dio ugovora* – *integral part of the agreement* or the more frequent variant *to form part of the contract*; *otkazati ugovor uz otkazni rok* – *to terminate this agreement by written notice*). Some, on the other hand, although on its face revealing the same patterns, convey different meanings given the fact that they use semantically-related collocates. For instance, the unit *to execute an agreement in counterparts* and *sklopiti ugovor u X istovjetna primjerka*, although both implying the meaning of producing an agreement in several copies, differ significantly, since the first unit refers to the act of signing all these copies and the latter to concluding the agreement in several copies. Similarly, when *terminate* and *raskinuti* become part of larger extended units of meaning, hence, *to be entitled to terminate this agreement* and *imati pravo raskinuti ovaj ugovor*, it may remain unnoticed that they produce different consequences, although they both refer to the ending of the contract, the difference lying in the fact that rescission treats contract as though it had never existed. Nevertheless, this tendency towards “free lexical co-occurrence” (L’Homme and Bertrand 2000) may serve as a model for the formation of the target text pattern since the only constituent that needs to be replaced is the collocate.

Usually, however, detecting translation equivalents is far from straightforward or the rendering might not be the first option a translator would choose.<sup>11</sup> For instance, if we would want to change the existing contract, the provision in Croatian would read *Odredbe ovog ugovora mogu se izmijeniti samo uz suglasnost svih ugovornih strana*. Since “to change, correct, revise” (Black 2004: 74) a document is in English rendered by means of the verb *to amend*, a translator not familiar with typical phraseological combinations in contracts might choose to render the above extended unit of meaning as *Provisions of this agreement may be amended by consent between parties*, where the natural rendering, as supported by the data from EnCon would be *Provisions of this agreement may be varied by agreement between the parties*.

Discovering the equivalent extended unit of meaning in the other corpus might sometimes be more complex and require research both on an even wider and the extralinguistic context. The unit *payment in full of the consideration payable under this agreement*, for example, has no direct counterparts in CroCon. Yet, this does not mean that there are no contracts in CroCon which require the exchange of consideration; they merely render the same meaning by means of different units. A clue for the search of the most appropriate unit of meaning in CroCon can be found when analysing the extension of the above-mentioned unit from EnCon, which suggests that what is exchanged under this type of agreement is money. One of the concordance lines of the search unit *consideration payable* in EnCon reveals that the consideration is payable by the Buyer, suggesting that it is used in the context of a Sale and

11 This assumption may be supported through the fact that nowadays most EU Member States witness a general lack of university programs on legal translation and interpretation (Bajčić 2015), which results in various (or sometimes none) certification schemes by EU Member States (Bajčić and Dobrić Basaneže 2016).

Purchase Agreement. Therefore, *payment in full of the consideration* in this context refers to the *payment in full of the purchase price*. If we examine the context of *kupoprodajna cijena (purchase price)* in CroCon, it is revealed that in the case of a Sale and Purchase Agreement the English unit may in Croatian be rendered as *isplata kupoprodajne cijene iz ovog ugovora u cijelosti*. Corpus data thus offer a more natural-sounding translation than the one that would be produced if we were to adhere to the English pattern only (e.g., *isplata u cijelosti protučinidbe u novcu iz ovog ugovora*).

## 6. Conclusion

It has been shown in this paper that the wider context of term-embedding collocations not only reveals their grammatical structures and phraseological patterns but also their genre-specific features. The wider context also suggests that some units allow more variation than others (e.g., *to terminate this agreement by written notice/by giving written notice*), which results in “relative stability of legal word combinations that apply not only to varying degrees of stability of *different* word combinations, but also to varying degrees of stability of *one and the same* word combination, depending on the situation and use” (Kjær 2007: 514). Although some units are extended by vague modifiers, (e.g., *reasonable/unreasonable; razuman/primjeren*) and would thus make us refrain from consulting the extralinguistic context, it seems that in legal phraseology such words never denote non-specific meaning. This especially applies to legal English, where it is of utmost importance to consult case citations in order to reveal lengthy discussions over the meaning of words (e.g. *to impose an obligation to use reasonable endeavours*). In the case of legal Croatian, however, such words are often given more precise meaning by the wording of statute law (e.g. *u primjerenom roku*). Finally, it seems that building lists of units based on the nodes chosen to represent close equivalents in the two corpora may produce equivalent patterns in two unrelated legal systems. If the list fails to produce equivalent patterns, consulting an even wider context in the two corpora might give useful hints for the creation of the target text pattern (e.g., *payment in full of the consideration payable under this contract – isplata kupoprodajne cijene iz ovog ugovora u cijelosti*). Although we might argue that comparable corpora cannot produce complete congruence, it has been shown in this study that in the case of legal phraseology this very notion of discrepancy between corpora reveals pitfalls that are to be taken into consideration to successfully deal with the formation of phraseological patterns in the case of non-equivalence (e.g., *upon acceptance of the offer – po prihvatu ponude*). This in turn supports Kjær’s claim that we cannot study legal phraseology without taking into account the science or profession (Kjær 2007), or in this case, the principles of both English and Croatian contract law as well as the functioning of the two legal systems underlying these phraseological units.

In conclusion, it may be suggested that this study aims to fill a gap in the research on legal phraseology due to the fact that private legal documents

are largely underrepresented in it. Although the corpora built for the purpose of this study are not based on authentic contracts, it has been shown that building corpora based on contract precedents can yield useful results for legal phraseology. Needless to say, future research might significantly benefit from focusing on authentic contracts instead and possibly reveal additional phraseological units created by individual contractual wordings which result from the freedom of the parties to contract. Results of this study, although reporting on phraseological units in English and Croatian contracts, can also be applied to studies focusing on contract phraseology in other languages, of which one is based on common law and the other on civil law. By focusing on legal Croatian, however, the study tends to fill yet another gap in corpus-based studies of legal language, which tend to be overwhelmingly Anglocentric. Finally, shifting the focus from the binary collocation to the Sinclairian wider-context-perspective highlighted in this study can, apart from helping us determine the stability and variations of phraseological units in legal language, also offer a useful resource for translator training, since it reveals the typical lexical environment of a term in contract language.

## References

- Bajčić, Martina (2015). The Way Forward for Court Interpreting in Europe. Šarčević, Susan, ed. *Language and Culture in EU Law: Multidisciplinary Perspectives*. Great Britain: Ashgate, 219–239
- Bajčić, Martina and Katja Dobrić Basaneže (2016). Towards the Professionalization of Legal Translators and Court Interpreters in the EU: Introduction and Overview. Bajčić, Martina, and Katja Dobrić Basaneže, eds. *Towards the Professionalization of Legal Translators and Court Interpreters in the EU*. Newcastle upon Tyne: Cambridge Scholars Publishing, 1–11
- Biber, Douglas, Stieg Johansson, Geoffrey N. Leech, Susan Conrad, and Edward Finnegan (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson
- Biel, Łucja (2011). Areas of similarity and difference in legal phraseology: collocations of key terms in UK and Polish company law. Pamies, Antonio, Lucia Luque-Nadal and José Manuel Pazos, eds. *Multilingual phraseography: translation and learning applications 2*. [http://www.academia.edu/2602242/Areas\\_of\\_similarity\\_and\\_difference\\_in\\_legal\\_phraseology\\_collocations\\_of\\_key\\_terms\\_in\\_UK\\_and\\_Polish\\_company\\_law](http://www.academia.edu/2602242/Areas_of_similarity_and_difference_in_legal_phraseology_collocations_of_key_terms_in_UK_and_Polish_company_law) (Accessed 11 December 2014)
- Biel, Łucja (2014a). *Lost in the Eurofog: The Textual Fit of EU Translated Law*. Frankfurt am Main et al.: Peter Lang
- Biel, Łucja (2014b). Phraseology in Legal Translation: A Corpus-Based Analysis of Textual mapping in EU Law. Cheng, Le, King Kui Sin and Anne Wagner eds. *The Ashgate Handbook of Legal Translation*. Surrey/Burlington: Ashgate, 177–193
- Biel, Łucja (2015). Phraseological profiles of legislative genres: complex prepositions as a special case of legal phrasemes in EU law and national law. *Fachsprache* XXXVII (3/4): 139–161
- Biel, Łucja (2017). Lexical bundles in EU law: the impact of translation process on the patterning of legal language. Goźdz-Roszkowski, Stanisław and Gianluca Pontrandolfo, eds. *Phraseology in Legal and Institutional Settings. A Corpus-Based Interdisciplinary Perspective*. UK: Routledge, 11–26
- Black, Henry C. (2004). *Black's Law Dictionary*. St. Paul, Minnesota: West Publishing Company
- Bowker, Lynne and Jennifer Pearson (2002). *Working with Specialized Language. A practical guide to using corpora*. London/New York: Routledge
- Breeze, Ruth (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics* 18: 229–253 <http://doi.org/10.1075/ijcl.18.2.03bre>

- De Groot, Gerard (1999). Zweisprachige juristische Wörterbücher. Sandrini, Petar, Hrsg. *Übersetzen von Rechtstexten. Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache*. Tübingen: Narr, 203–227
- Dobrić Basaneže, Katja (2015). Investigating “congrams” in the language of contracts and legal agreements. *Fachsprache* XXXVII (3/4): 176–193
- Gabrovšek, Dušan (2014). Extending Binary Collocations: (Lexicographical) Implications of Going beyond the Prototypical a – b. *ELOPE* 11: 8–20 <http://doi.org/10.4312/elope.11.2.7-20>
- Goźdź-Roszkowski, Stanislaw (2000). *Corpus-based Analysis of English- and Polish-language Contracts*. Phd dissertation, University of Łódź
- Goźdź-Roszkowski, Stanislaw (2006). Frequent phraseology in contractual instruments. A corpus-based study. Gotti, Maurizio and Davide Giannoni, eds. *New Trends in Specialized Discourse Analysis (Linguistics Insights. Studies in Language and Communication)*. Frankfurt: Peter Lang, 147–161
- Goźdź-Roszkowski, Stanislaw (2011). *Patterns of Linguistic Variation in American Legal English*. Frankfurt: Peter Lang
- Goźdź-Roszkowski, Stanislaw and Gianluca Pontrandolfo (2013). Evaluative patterns in judicial discourse: a corpus-based phraseological perspective on American and Italian criminal judgments. *International Journal of Law, Language and Discourse* 3(2): 9–69
- Goźdź-Roszkowski, Stanislaw and Gianluca Pontrandolfo (2015). Legal phraseology today: corpus-based applications across legal languages and genres. *Fachsprache* XXXVII (3–4): 130–139
- Goźdź-Roszkowski, Stanislaw and Gianluca Pontrandolfo (2017). Introduction: cross-linguistic approaches and applications to phraseology in legal and institutional discourse. Goźdź-Roszkowski, Stanislaw and Gianluca Pontrandolfo, eds. *Phraseology in Legal and Institutional Settings. A Corpus-Based Interdisciplinary Perspective*. UK: Routledge, 1–8
- Junačko, Zdravko and Dragutin Rotar (2007). *Zbirka ugovora građanskog i trgovačkog prava 4*. Zagreb: Poslovni zbornik
- Keglević, Ana (2013). Neispunjenje ugovora u Europskom privatnom pravu – perspektive harmonizacije i razvoja zajedničkog pravnog okvira. *Zbornik Pravnog fakulteta Sveučilišta u Rijeci* 34 (2), 671–700
- Kilgariff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel (2014). The Sketch Engine: ten years on. *Lexicography*, 1–30 <https://doi.org/10.1007/s40607-014-0009-9>
- Kjær, Anne L. (1990). *Normbetingede ordforbindelser i tysk juridisk sprog*. Phd dissertation, University of Copenhagen
- Kjær, Anne L. (2007). Phrasemes in Legal Texts. Burger, Harald, Dmitrij Dobrovol'skij, Pieter Kühn, and Niel R. Noerrick, eds. *Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*. Berlin/New York: de Gruyter, 506–515
- Kopaczyk, Joanna (2013). *The Legal Language of Scottish Burghs. Standardization and Lexical Bundles 1380–1560*. Oxford: Oxford University Press
- L’Homme, Marie-Claude and Claudine Bertrand (2000). Specialized Lexical Combinations: Should they be Described as Collocations or in Terms of Selectional Restrictions?. <http://olst.ling.umontreal.ca/pdf/berlhom.pdf> (Accessed 11 October 2015) <https://doi.org/10.13140/2.1.1126.2726>
- Ljubešić, Nikola and Filip Klubička (2014). {bs,hr,sr}WaC – Web corpora of Bosnian, Croatian and Serbia. <http://nlp.ffzg.hr/data/publications/nljubesi/ljubestic14-bs.pdf> (Accessed 1 September 2014) <https://doi.org/10.3115/v1/W14-0405>
- Mazzi, Davide (2010). “This Argument Fails for Two Reasons...”: A Linguistic Analysis of Judicial Evaluation Strategies in US Supreme Court Judgments. *International Journal for the Semiotics of Law* 23 (4), 373–385 <https://doi.org/10.1007/s11196-010-9162-0>
- Mellinkoff, David (1963). *The Language of the Law*. Oregon: Wipf and Stock Publishers
- Millett, Peter and Robert Walker (eds.) (2014). *Encyclopaedia of Forms and Precedents*. UK: Lexis-Nexis Butterworths

- Monzó, Esther (2015). (Re)producing habits in international negotiations: a study on the translation of collocations. *Fachsprache* XXXVII (3/4): 193–210
- Moss, Giuditta Cordero (2007). International Contracts between Common Law and Civil Law: Is Non-state Law to Be Preferred? The Difficulty of Interpreting Legal Standards Such as Good Faith. *Global Jurist* 7(1): Article 3. Accessed 11 October 2015
- Petrić, Silvija (1991). Izmjena ili raskid ugovora zbog promijenjenih okolnosti prema novom Zakonu o obveznim odnosima. *Zbornik Pravnog fakulteta Sveučilišta u Rijeci* 28(1): 107–155
- Pontrandolfo, Gianluca (2013). *La fraseologia en las sentencias penales: un estudio contrastive español-italiano-inglé basado en corpus*. Phd dissertation, University of Trieste
- Pontrandolfo, Gianluca (2015). Investigating judicial phraseology with COSPE: a contrastive corpus-based study. Fantinuoli, Claudio and Federico Zanettin, eds. *New directions in corpus-based translation studies*. Berlin: Language Science Press, 137–160
- Reference corpora: <http://www.ilc.cnr.it/EAGLES/corpusstyp/node18.html> (Accessed 10 August 2015)
- Schank, Roger C. and Robert P. Abelson (1977). *Scripts, plans, goals and understanding: an inquiry into human knowledge structure*. Hillsdale, N.J: Lawrence Erlbaum Associates
- Scott, Mike (2012). *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software
- Sinclair, John M. (2004). *Trust the Text: Language, Corpus and Discourse*. London/New York: Routledge
- Statement 33 of Rhodia International Holdings Ltd v Huntsman International LLC (2007): [http://www.glovers.co.uk/news\\_article232.html](http://www.glovers.co.uk/news_article232.html) (Accessed 12 November 2016)
- Statement 123 of the case Yewbelle Ltd v London Green Developments Ltd and Knightsbridge Green Ltd (2006): [http://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2006/3166.html&query=\(title:\(+Yewbelle+\)\)+AND+\(title:\(+v+\)\)+AND+\(title:\(+London+\)\)+AND+\(title:\(+Green+\)\)+AND+\(title:\(+Developments+\)\)](http://www.bailii.org/cgi-bin/format.cgi?doc=/ew/cases/EWHC/Ch/2006/3166.html&query=(title:(+Yewbelle+))+AND+(title:(+v+))+AND+(title:(+London+))+AND+(title:(+Green+))+AND+(title:(+Developments+))) (Accessed 12 November 2016)
- Tabares Plasencia, Encarnación (ed.) (2014). *Fraseologia juridical contrastive español-alemán/ Kontrastive Fachphraseologie der spanischen und deutschen Rechtsprache*. Berlin: Frank & Timme
- The British National Corpus, version 3 (the BNC XML Edition), 2007. Distributed by Oxford University Computing Services on behalf of the the BNC Consortium. <https://www.sketchengine.co.uk/british-national-corpus/> (Accessed 28 August 2014)
- Tománková, Veronika (2016). Lexical bundles in legal texts corpora – Selection, classification and pedagogical implications. *Discourse and Interaction* 9: 75–94 <https://doi.org/10.5817/DI2016-2-75>
- Treitel, Günter Heinz (1995). *The Law of Contract*. London: Sweet and Maxwell
- Triebel, Volker (2009). Pitfalls of English as a Contract Language. Olsen, Frances, Alexander Lotz, and Dieter Stein, eds. *Translation Issues in Language and Law*. Hampshire/New York: Palgrave Macmillan, 147–182
- UK: Should you endeavour to be reasonable?. <http://www.linklaters.com/Insights/Publication1403Newsletter/PublicationIssue20070605/Pages/PublicationIssueItem2385.aspx> (Accessed 12 November 2016)
- Zakon o obveznim odnosima (NN 35/05, 41/08, 125/11, 78/15)

## *Tumačenje frazeoloških jedinica u ugovorima: analiza proširenih terminoloških kolokacija*

Ako bi se izuzela istraživanja leksičkih isječaka (Goźdz–Roszkowski 2006, 2011; Kopaczyk 2013; Breeze 2013; Tománková 2016; Biel 2017), moglo bi se tvrditi da su proširene jedinice značenja (Sinclair 2004) slabo istražene u pravnoj frazeologiji. Stoga je cilj ovog rada ukazati na korisnost Sinclairove perspektive šireg konteksta u istraživanju frazeoloških jedinica u pravnom žanru s obzirom na to da širi kontekst nudi njihov kolokacijski okvir te njihove gramatičke i frazeološke obrasce. Rad se temelji na analizi proširenih ‘terminoloških kolokacija’ (engl. *term-embedding collocations*) (Biel 2014b) izvučenih iz usporedivog korpusa engleskih i hrvatskih ugovora s pomoću alata *WordsmithTools 6.0* (Scott 2012). U radu se ističu primjeri koji najbolje podupiru gore navedenu hipotezu, a tumači ih se uz pomoć analize izvanjezičnog konteksta frazeoloških jedinica u ugovorima. Budući da su privatnopravni dokumenti rijetko predmet istraživanja pravne frazeologije, ovo istraživanje predstavlja pokušaj njezina dopunjavanja. S obzirom na to da se dio istraživanja temelji na frazeološkim jedinicama u hrvatskom pravnom jeziku, ovaj rad nastoji nadopuniti korpusno utemeljena istraživanja koja su uglavnom usredotočena na engleski pravni diskurs. Naposljetku, ovo istraživanje, osim što pruža uvid u stabilnost frazeoloških jedinica u području prava, također može predstavljati koristan izvor u obuci pravnih prevoditelja i sudskih tumača jer raspravlja o širem kontekstu nazivlja u jeziku ugovora.

**Keywords:** language corpora, extralinguistic context, extended term-embedding collocation, collocations, legal contracts

**Ključne riječi:** jezični korpusi, izvanjezični kontekst, proširena terminološka kolokacija, kolokacije, pravni ugovori