

Comparative Legilinguistics

vol. 2022/52

DOI: <http://dx.doi.org/10.14746/cl.52.2022.13>

Translating legal formulae: a corpus-driven approach

PATRIZIA GIAMPIERI, Adjunct Professor

School of Law

University of Camerino, Italy

patrizia.giampieri@unicam.it

ORCID: <http://orcid.org/0000-0003-4443-1593>

Abstract: Fixed lexical or syntactical expressions and formulae hallmark legal language. They serve both linguistic and legal purposes, and should be rendered accordingly in a target language and legal system. Most of the times, however, formulaic expressions are translated by resorting to calques, false cognates, or phrases that are uncommon in the target legal language (and legal system). This paper is aimed at exploring how and if corpus analysis can dispel doubts and help find acceptable translation candidates. As there are currently no publicly available legal corpora addressing corporate documents such as contracts and agreements, this paper wishes to bridge this gap by building and relying on an *ad hoc* corpus of authentic agreements written in English as a first language according to the laws of England and Wales. In this way, corpus evidence can help find equivalents and, possibly, address recurrent mistranslations from Italian into English. During the corpus analysis process, the paper shows and discusses search queries and how equivalents can be obtained. At the same

time, it questions dictionary entries. The paper findings highlight that the consultation of the *ad hoc* corpus allows to find acceptable translations of Italian legal formulae and address recurrent mistranslations. English formulaic expressions, in fact, can be rendered satisfactorily thanks to the possibility of noticing word usages in context, keywords in contexts and collocations. Further research can encompass a wider variety of formulae and/or legal documents so that scholars and translators can be equipped with useful reference tools.

Keywords: corpus-based translation; legal translation; legal linguistics; corpus analysis; legal language.

TRADURRE FORMULE GIURIDICHE ATTRAVERSO I CORPORA

Riassunto: Espressioni e formule lessicali o sintattiche predefinite caratterizzano il linguaggio giuridico e sono utilizzate sia per finalità linguistiche che legali, quindi devono essere necessariamente adattate alla lingua ed al sistema giuridico di arrivo. Tuttavia, molto spesso espressioni e formule sono tradotte ricorrendo a calchi, falsi affini o frasi non frequenti nella lingua giuridica (e nel sistema giuridico) di arrivo. Il presente articolo ha lo scopo di verificare se la consultazione di un corpus di contratti possa aiutare a dissipare dubbi linguistico-giuridici e a trovare traduzioni accettabili. Poiché al momento non esistono corpora giuridici pubblicamente disponibili contenenti documenti aziendali quali contratti, questo articolo si pone l'obiettivo di tentare di colmare questa lacuna creando e consultando un corpus *ad hoc* costituito da contratti autentici redatti in lingua inglese secondo la legge dell'Inghilterra e Galles. In questo modo, il corpus può aiutare a trovare equivalenti e, possibilmente, correggere ricorrenti traduzioni errate dall'italiano all'inglese. Durante il processo di analisi del corpus, si mostra come è possibile ottenere equivalenti. I risultati dell'articolo evidenziano che la consultazione del corpus consente di trovare traduzioni accettabili di formule giuridiche italiane e di correggere frequenti errori di traduzione. Le formule inglesi, infatti, possono essere rese in modo soddisfacente grazie alla possibilità di notare gli usi delle parole nel contesto, le parole chiave ricorrenti e le collocazioni. Ulteriori ricerche possono riguardare una più ampia varietà di formule e/o documenti legali in modo che studiosi e traduttori possano avvalersi di utili strumenti di riferimento.

Parole chiave: traduzione mediante corpora; traduzione giuridica; linguistica giuridica; analisi dei corpora; linguaggio giuridico.

1. Theoretical background

Coulthard and Johnson (2007: 37) claim that understanding legal style implies developing interpretative skills that are necessary “in order to make sense of legal texts”. Legal documents are, in fact, characterised by “apparently meaningless repetitions and archaisms” (Bhatia 1993: 101), as well as formulaic expressions and fixed phrases (Tiersma 1999; Bhatia et al 2004; Kast-Aigner 2009; Bhatia 2010; Coulthard and Johnson 2007: 37). Formulae are argued to be necessary in order to ensure the validity of the actions (Gotti 2012: 52). Nonetheless, legal phraseology is particularly challenging (Garzone 2007: 218–219; Prieto Ramos 2014b: 16) and translators must search for and become acquainted with prefabricated patterns in both the source and target language (Garzone 2007: 218). Moreover, not only do they need to source and find corresponding fixed expressions in the target language, but they also have to make necessary comparisons between the source and target legal system (Šarcevic, 1997: 18–19, 68). As legal systems change over time and are subject to influences from others (Stein 2009), translators need to find a way to mediate terms and meanings. Before engaging in legal translation, in fact, they have to become acquainted with the features of the different legal systems in question (Engberg 2020). They need to build a bridge for the target text readers by using target words that most relevantly represent the concepts (and institutions, or law codes) of the source text (Engberg 2021).

In order to become acquainted with sector-specific language and address specialised texts, scholars, amongst others, suggest consulting corpora. Corpora are collections of texts available in electronic format (Bowker and Pearson 2002: 9). They are considered useful as they show samples of language (Farr and O' Keeffe 2019) as well as collocations and colligations (Lehecka 2015). Collocations and colligations refer to the co-occurrences of lexical and grammatical categories, respectively (Lehecka 2015: 1). For these reasons, corpora are considered particularly useful when addressing sector-based documents, as they help notice word usages in context and patterns of language.

As far as legal corpora are concerned, there are manifold research papers and corpus-based studies dedicated to legislation (Biel 2018; Prieto Ramos and Guzmán 2021) and court judgements or similar interlocutory acts (see, for example, the BoLC corpus, Rossini Favretti

et al. 2007; Vogel et al. 2017; Nikitina 2018). However, there is at present no corpus dedicated to contracts and agreements (also referred to as “private legal texts”, Prieto Ramos 2014a: 263). The main reason is due to the private nature of such documents (Biel 2018; Dani 2019: 26–27).

In order to try and bridge this gap, Giampieri (2018) explored how some frequent fixed expressions and phrases used in legal documents and contracts could be tackled by consulting free online parallel and comparable legal corpora, such as the EuroParl v7 (Koehn 2005; Tiedemann 2012) and the BoLC (Rossini Favretti et al. 2007). To this aim, the following syntactic and lexical items were investigated: “*come sancito*” (which can be rendered “as set out” or “as laid down”), and “*in deroga a*” (whose best translation option is “by way of derogation from”) (Giampieri 2018: 240–244). This initial trial study, however, revealed some challenges, which were mainly due to the fact that the corpora consulted were not composed of private documents such as contracts, agreements, memoranda of understanding, and the like. Therefore, the number of legal lexical and grammatical phrases focussed on was not exhaustive.

Given the unavailability of legal corpora composed of private documents, one might argue that legal formulae may be sourced from multi-language databases, which are available online for free. In this regard, the literature has often warned against the pitfalls of such language resources (Durán Muñoz 2012: 78; Giampieri and Milani 2021: 56–57). Genette (2016), for example, claims that terminological databases should only be consulted by expert users of the language and of the field in question in order to disambiguate the terms proposed. Other researchers posit that the multi-language resources available at EU level present language shortcomings due to both “translationese” (Seracini 2020) and interferences from the source language (Prieto Ramos 2021: 179). These issues are claimed to arise mainly from the EU adaptation process, whereby the English language used at EU-level is a *lingua franca* and, hence, deprived of a legal system of reference and a well-rooted legal tradition (Jacometti and Pozzo 2018).

In light of the above, it can be asserted that there is a gap in the literature regarding a corpus-driven study of the language of contracts and, in particular, of the legal formulae used in contracts and agreements.

1.1 Research question

The aim of this paper is to bridge this gap and propose a set of acceptable English translations of some Italian legal formulae that are commonly used in contracts and agreements, but are often mistranslated. To this aim, an *ad hoc* DIY (do-it-yourself) corpus is composed of various types of English contracts and agreements and it is analysed. In this way, by applying corpus advanced search techniques, collocations and word usages in context are explored, so that recurrent language patterns can come to the fore and acceptable equivalents in English are found.

The research questions that this paper aims to answer are, hence, the following: can an *ad hoc* corpus composed of English contracts and agreements be consulted in order to find equivalents of Italian recurrent formulaic expressions? Can the corpus address the recurrent mistranslations of such Italian legal formulae? To what extent can the corpus be considered reliable and the equivalents acceptable?

In order to answer these questions, an *ad hoc* corpus of contracts and agreements is built. The contracts are drawn up according to the laws of England and Wales (i.e. “English law”).

2. Methodology

This paragraph describes the way the DIY corpus is built and the software tools are used.

2.1. The software solution

The BootCaT freeware software solution (Baroni and Bernardini 2004) allows to build a corpus in a matter of few minutes, either automatically or semi-automatically, depending on the Internet connection speed. For the purpose of this paper, the fully automated mode is ruled out for a variety of reasons. With such corpus building function, it is necessary to input at least 5 unique words, or sets of words, which are then searched

online automatically. In this case, the key terms to search for revolve around the words “contract” and “agreement”. Any document of this type suffices, provided that it is written in English as a first language, and according to English law. Therefore, it is considered unnecessary to search for as many contract types as possible (such as “land lease agreement”, “employment agreement”, “supply agreement”, “agency agreement” etc.). As a matter of fact, it would be arduous to list and comprise them all. Furthermore, if this option is chosen, the BootCaT software solution would combine the sets of keywords together and, hence, look for “land lease agreement” together with “employment agreement” and/or “supply agreement”, and so on. This would obviously cause unnecessary inconsistencies in the search process and in the results.

For these reasons, the BootCaT semi-automatic mode is preferred. With this option, it is possible to search for a few sets of words (or phrases) on the Internet and then save the Google's results pages in a folder. In this way, the software downloads as many documents as possible by retrieving them from the html results pages saved.

2.2. Building the corpus

In order to build the *ad hoc* corpus, the following two phrases are queried on Google:

"the law/laws of England" "agreement/contract" site:.onecle.com and
"in accordance with English law" "agreement/contract" site:.onecle.com.

As regards the first query, the phrase “the law|laws of England” is written in order to obtain documents which contain either the word “law” or “laws”. The OR Boolean operator, in fact, is triggered by the straight line (“|”) and enables to search either for “law” or “laws”. The aim of this search string is to retrieve contracts governed by the laws of England and Wales, or contracts signed by companies incorporated under the laws of England and Wales. The same can be said of the second query, where the expression “in accordance with English law” allows to obtain contracts or documents governed and interpreted according to English law. It is self-evident that these phrases are aimed

at obtaining legal documents issued and drafted (or, at least, read, understood and signed) by native speakers of English. Hence, the purpose is to source authentic documents drawn up in English as a first language, and build a corpus of contracts written in authentic legal language, according to the laws of England.

As regards the phrase “agreement|contract”, the Boolean OR operator serves the purpose of sourcing either agreements or contracts. These two terms are, in fact, considered synonyms in the majority of native legal contexts and texts¹, despite the fact that at common law, “a contract is an agreement between two or more parties for the doing or not doing of some specified thing” (Campbell Black 2008: 261). On the basis of this definition, an “agreement” should have a wider meaning than “contract”. Nonetheless, in practice, they are considered equivalent.

Finally, the command “site:.onecle.com” helps retrieve documents only from the Onecle.com domain, which is a website containing a vast *repertoire* of legal documents drafted by native and non-native speakers of English.

As anticipated, the two phrases are written in the Google search field (one query at a time) and the results pages are saved in a folder. As regards the first query (i.e. “*the law/laws of England*” “agreement|contract” site:.onecle.com), the first 10 Google results pages are saved in a separate folder, whereas as concerns the second (i.e. “*in accordance with English law*” “agreement|contract” site:.onecle.com), the first 8 results pages are considered. This process is necessary in order to launch the BootCaT semi-automatic corpus building mode.

The final corpus is composed of 161 documents, 37,795 word types and 2,337,255 tokens. The contract types are many and varied, such as “Loan and Security Agreement”; “Licensed Publisher Agreement”; “Employment Agreement”; “Service Agreement”;

¹ For example, see the definition of a “Tenancy Agreement” provided in the Stafford Borough's website: “A tenancy agreement is a legally binding contract” (<https://www.staffordbc.gov.uk/tenancy-agreements-and-inventories>); see also the Terms and Conditions of Contract for Goods of the Crown Civil Nuclear Constabulary, which quotes “Agreement” means the contract between (i) the Customer acting as part of the Crown and (ii) the Supplier constituted by the Supplier’s countersignature of the Award Letter (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1030891/Shortform_Terms_and_Conditions_Goods_FINAL.pdf).

“Executive Service Contract”; “Consultancy Contract”, just to mention a few.

2.3. Corpus consultation tool

In order to consult the corpus, the AntConc (Anthony 2022) freeware software solution is used. AntConc is an offline concordancer which allows to analyse a corpus offline. Amongst other functions, the software tool generates concordances, collocates, keywords in context (KWIC) and lemmas. In particular, collocations are listed in the “collocate” window tab by writing the word (or phrase) one wishes to obtain collocates of. For example, it is possible to list collocations within a span of 5 words to the left and to the right of the node word. The word span is modifiable. Also, collocates are listed by frequency of occurrences (i.e. from the most frequent to the least) or in alphabetical order. In this way, searching for and obtaining collocations is quite straightforward. The KWIC function is incorporated in the software by default; therefore, words in the proximity of the node word are always marked with different colours, allowing to notice collocations and/or colligations, as well as specific word usages in context. Lemmatised searches are triggered by the asterisk, which functions as a wildcard character (Zanettin 2012: 124, 190). The asterisk can also be used between words, in order to search for specific words or grammar words. For example, by searching for “contract* * the part*”, the system retrieves phrases such as “contract between the parties”, “sub-contractors of the parties”, or “contracts. However, the parties”. Hence, the asterisk replaces alphanumerical and non-alphanumerical characters. Finally, searches are case-insensitive, but a case-sensitive search can be optioned. This is particularly useful when looking for specific phrases as clause titles or words at the beginning of phrases or sentences.

2.4. The Italian formulae

For the purposes of this research paper, the following Italian legal formulae are addressed: “*premesse che*” (back-translation: “considering

that”, or “given that”); “*premessa*” (back-translation: “premises” or “preamble”); “*tutto ciò premesso, le parti convengono quanto segue*” (back-translation: “considered all that, the parties agrees as follows/the following”); “*società fondata/costituita*” (back-translation: “founded/created company”); “*sede legale*” (back-translation: “legal seat” or “headquarters”); “*domiciliato a*” and “*domiciliato all'indirizzo*” (back-translations: “domiciled in” and “domiciled at the address”); “*trattamento dati personali*” (back-translation: “treatment of personal data”); “*di cui sopra*” (back-translation: “as of above”); “*di cui in oggetto*” (back-translation: “as of in the subject”), and “*in fede di che*” (back-translation: “in faith”) and the acronym “*LCS*” meaning “*letto, confermato e sottoscritto*” (back-translation: “read, understood and signed”).

Some formulae are written at the beginning of documents (e.g. “*premesso che*” and “*premessa*”, or “*tutto ciò premesso, le parti convengono quanto segue*”). Other formulaic expressions appear at the end, before the parties' signature, such as “*in fede di che*” or “*LCS*”. Some lexical and syntactical phrases can be difficult to find equivalents of in view of the words they collocate with. For example, the verbs “*fondata*” or “*costituita*”, collocating with “*società*”, can be challenging as bilingual dictionaries may not propose the right translation candidates used in legal contexts. The same can be said of “*sede legale*”, which refers to the legal “seat” of a company.

The paragraphs which follow address some controversial aspects by shedding light on the best translation options which can be sourced from the corpus.

3. Analysis

This paragraph describes and comments on the corpus analysis carried out in the search for English equivalents of the above formulae.

3.1. Corpus analysis and equivalent formulae

As regards the expression “*premessso che*”, most of the general bilingual dictionaries suggest “given that”². This translation option, however, may be wrongly used by those who do not disambiguate terms. In practice, wrong translations may be proposed when translators do not verify whether the terms suggested belong to the legal field. In the case in point, “given that” is not the right rendering as it is not used in legal documents. As a matter of fact, corpus evidence only lists 12 hits with “given that”, and the concordances retrieved only show unrelated contexts, such as “transferred into the escrow account given that the value of eBay Shares may fall”. As can be noticed, the expression “given that” does not serve the same purposes of “*premessso che*”; also, it is not written at the beginning of a contract in order to set the “premises”. Some other dictionaries and online forums may suggest “whereas” as a legal or formal translation³. If “whereas” is queried in the corpus, 224 hits are retrieved and the examples obtained are consistent with the Italian “*premessso che*”. Sample phrases are “WHEREAS, Company and Executive agree that the contract shall be amended as follows”; “Recitals. Whereas, the parties entered into (...); and Whereas, the Parties desire to (...)”, or “Recitals. Whereas [parties' names] have entered into a Licence Agreement (...)”.

Although it may appear that the next word, “*premesssa*”, is related to the expression “*premessso che*”, it is helpful to know that it is rendered differently in English. Some bilingual dictionaries, in fact, propose “premises” as the legal translation of “*premesssa*”⁴, whereas others suggest “preamble”⁵. As regards the first option, the corpus shows 1,125 hits; however, the meaning in context of this word is “site” or “building”. The following phrases are self-explanatory: “to enter onto the Premises”, and “have access to the Premises after the lease is granted”. If the word “preamble” is queried, 55 occurrences are

² See, for example, the Collins dictionary: <https://www.collinsdictionary.com/it/dizionario/italiano-inglese/premettere>.

³ See, for example, the Proz online forum: <https://www.proz.com/kudoz/italian-to-english/law-contracts/922454-premessso-che.html>.

⁴ See, for example, the Hoepli dictionary: <https://dizionari.repubblica.it/Italiano-Inglese/P/premesssa.html>.

⁵ See the Sansoni dictionary: https://dizionari.corriere.it/dizionario_inglese/Italiano/P/premesssa.shtml.

retrieved, but only once is this term used at the beginning of a document, as in the following phrase: “Preamble. Whereas, [parties' names] entered into a stock purchase, reorganization and joint venture agreement”. On the basis of the results obtained previously, in the search for “*premessso che*” (e.g. “Recitals. Whereas, the parties entered into...”), it can be assumed that an acceptable translation of “*premesssa*” can be “recitals”. If this word is searched in the corpus, in fact, 120 hits are found and the related concordances are consistent with the usages and meaning of “*premesssa*”. A clarifying example is the following phrase (companies' names are replaced by “X” and “Y”): “Recitals: A) X is a pharmaceutical company, with activities in the area of (...); B) Y is a pharmaceutical company, with activities in the marketing and sales of pharmaceutical products (...); C) X wishes Y to manufacture the Product on its behalf (...) THE PARTIES AGREE AS FOLLOWS”.

The formula “*tutto ciò premessso, le parti concordano quanto segue*” is generally written after the “premises” (or, better, the “recitals”) to set forth the details of the parties' mutual agreement. Several Italian documents translated into English contain literal renderings⁶, such as “whereas, the parties stipulate and agree as follows”; “having stated the above, the parties stipulate and agree as follows”; “in light of the above, the parties convene and agree as follows”, and “given the above, the parties stipulate and agree as follow” [sic.]. On the basis of the phrase obtained in the search for “*premesssa*”, however, it can be assumed that a possible translation of the formula in question is “the parties agree as follows”. In order to corroborate this assumption, the expressions “as follows” and “the parties agree” are queried in the corpus (one query at a time). In both cases, the following formulaic expressions are retrieved: “NOW, THEREFORE, the parties agree as follows:”, and “NOW, THEREFORE, in consideration of the mutual covenants contained herein and intending to be legally bound hereby, the parties hereto agree as follows:”.

⁶ See, for example, the documents retrieved from the following weblinks: https://en.unibs.it/sites/sten/files/ricerca/allegati/2016ModelloconvOK_0.doc; https://www.mixsrl.it/sites/default/files/2019-09/0970000EN_191.pdf; <https://ilcairo.aics.gov.it/wp-content/uploads/2020/06/Hend-Contract.pdf>; http://www.securpolgroupamministrazionestraordinaria.it/downloads/sg_bando/190412_Securpol_Confidentiality_Agreement_ENG_Final.pdf.

The verbs “*fondata*” and “*costituited*” collocating with “*società*” are used to mention the year of foundation of a company or the laws according to which it was formed. The phrase “*società costituita/fondata*” is often mistranslated due to wrong verbs. By consulting the Hoeppli dictionary, for example, the following translations of “*fondare*” (back-translation: “to found”) and “*constituire*” (back-translation: “to constitute”) are found: to found, to establish, to set up, to start (up), to constitute, and to incorporate. These verbs are now analysed in the corpus. If the lemma “found*” is searched in the corpus, unrelated results come to the fore, such as “is found by any court” or “provisions so found to be void”, as well as “founder” and “foundations”. Also, the past participle “founded” never collocates with “company”, “firm”, or “partnership”. Therefore, it bears different meanings than the Italian verb. The following phrases are some examples: “whether founded in contract or tort”, and “that the defects specified in the Defects Notice are well founded”. The past participle “established” does not apparently have the same meaning of the source words “*fondare*” or “*constituire*”. The following phrases are self-explanatory: “a Debtor has established adequate reserves”; “a separate found shall be established”, and “whether legally binding or established by custom”. However, “established” collocates with “company” and the following phrase resembles the Italian expression: “the Company was established to act as such joint venture company”. Nonetheless, there is only one occurrence with this word combination. It might be assumed that another verb can better serve the purpose. The verb phrase “set up” shows three collocations with the word “company”. However, they are poorly related to the original meaning, as the following phrases show: “any company or business organisation set up in connection with their services or rights”, and “a new company (“Newco”) to be set up as a new member of the Vendor’s Group”. As regards “started up”, the corpus shows no hits. Also, if the lemma “start* up” is searched for, unrelated concordances come to the fore, such as “start-up activities”, or “starting up a system restaurant”. The same occurs to the lemma “constitute*”, as it is mainly used in phrases such as “the security constituted by this deed”. Finally, if the past participle “incorporated” is queried, there are many interesting phrases mirroring the Italian meaning of “*costituited*” or “*fondata*”, such as “a company incorporated in England and Wales”; “a company incorporated under the laws of England and Wales”, and “a company incorporated on 27 January 2000

in Bermuda”. Therefore, the right translation candidate of the two Italian past participles appears to be “incorporated”.

Another interesting and intriguing translation revolves around the noun phrase “*sede legale*”, which refers to the main offices, or “headquarters”, of a company. This expression is generally followed by the company's physical address. Italian legal documents tend to use a wrong equivalent, such as “legal seat”⁷. By consulting some bilingual dictionaries, the word “headquarters” is, instead, found. By searching for “headquarter*” in the corpus, however, only 9 hits are retrieved and they seem unrelated to the Italian usage and meaning. The following phrases are self-explanatory (the company's name is anonymised by “XX”): “to inspect at XX's headquarters”; “the Executive shall be based at the Group's headquarters”. The Hoepli dictionary suggests “corporate domicile” and “registered office”. The former is not present in the corpus, whereas the latter shows 313 hits, with consistent results, such as “XX Limited, whose registered office is at [address]”. Also, given that the company's “registered office” is generally followed by an address, the words “is at” can be queried in the corpus. In this way, other translation options can be sourced, such as “principal office” and “principal place of business”. The following phrases are insightful: “a company incorporated under the laws of England and Wales whose principal office is at [address]”, and “(incorporated in Delaware) whose principal place of business is at [address]”. In particular, the expression “principal office” shows 20 hits, whereas “principal place of business” 73. There is also one occurrence of “principal business office”.

If a company has a “*sede legale*”, a physical, or natural, person has, instead, a “domicile”. According to art. 43 of the Italian Civil Code, a “*domicilio*” is a “*luogo in cui essa ha stabilito la sede principale dei suoi affari e interessi*” (back-translation: “where a person has established the principal place of his/her affairs and interests”). In Italian contracts and agreements, the word “*domicilio*” and the deriving verb phrase “*domiciliato a*” (back-translation: “domiciled at”) are very

⁷ See, for example, the following company's websites:

<https://winehunter.it/cookie-policy/>;

<https://www.friulcamion.it/en/legal-notice/>;

<https://www.hlrlaw.it/en/privacy/>

and the following legal documents released by Italian public authorities:
https://www.enac.gov.it/ContentManagement/information/N1162120931/FO_CERT_00037_003.doc;

<https://www.salute.gov.it/portale/allegatoModulo?idMat=CSM&idAmb=CLV&idSrv=M1&idFlag=P&idModulo=7>.

frequent. Both terms are generally followed by a physical address. The same cannot be said of English agreements, where possible equivalents of “*domicilio*” or “*domiciliato*” are hardly ever mentioned. If the Italian section of the BoLC corpus is consulted, for example, the words “*domicilio*” and “*domiciliato*” show 2,973 and 10,801 occurrences, respectively, whereas the English equivalents “domicile” and “domiciled” only 299 and 457, respectively. Although it might be argued that the BoLC is not composed of corporate documents such as contracts and agreements, in does, nonetheless, show a word usage pattern in both the Italian and English legal language. Therefore, it may be speculated that legal English resorts to other ways of expressing a “domicile” as a place (or address) of one’s interests or affairs. In this respect, dictionaries are not particularly helpful as they generally suggest the calque “domicile”. However, it might be resourceful to search for collocations of “address” in the corpus. In this case, the words “notice” and “notices” would come to the fore. The corresponding concordances are revealing, as they show interesting results such as the following: “Address for Notice: [address]”; “Address for Notices: [address]”; “our address for notices is [address]”, and “the address at which they are to receive notices”. The expression “address for notice(s)” produces 91 hits in the corpus. An alternative of “address for notice” is “address for service”, which shows 18 occurrences. Interestingly, if “address for notice(s)” and “address for service(s)” are searched for in the English section of the BoLC, the former has no occurrences, whereas the latter shows 256 hits. In light of the above, it might be speculated that the English equivalents of “*domicilio*” can be “address for notice(s)” or “address for service”, and of “*domiciliato*” may be “whose address for notice(s) is”. Further research would, however, be called for in order to either corroborate or confute these findings. For example, larger legal corpora on corporate documents (or on court’s decisions) might be consulted.

The expression “*trattamento dati personali*” is generally used in privacy policy notices and it concerns the management of personal data. In many Italian documents translated into English, the word “*trattamento*” tends to be rendered with a false cognate (i.e. “treatment”)⁸. This is a mistranslation which can be addressed quite

⁸ See, for example, the privacy policies of companies and public institutions at the following websites:

<https://stage-air.com/circulars/>;

straightforward. It suffices to look for “personal data” in the corpus and notice the words to the left. In this way, many phrases with “processing (of) personal data” and “process personal data” come to the fore. Also, by exploring the collocates of “personal data”, it is possible to read “treat”, but this term only occurs once in the following phrase: “Employee agrees to treat any personal data (...) in accordance with the Data Privacy Policy”. As can be seen, the word “treat” has a different meaning from “process”; in the phrase above, in fact, it can be considered a synonym of “deal with”.

The prepositional phrases “*di cui sopra*” and “*di cui in oggetto*” might be challenging due to false equivalences or false cognates. In this respect, it could be useful to know the function they serve in contracts. Both phrases are deictic (Galdia 2009: 40, 74; Mooney 2014: 30), as they refer to something already mentioned in the text. In particular, the first relates to what was cited previously, as in the Italian phrases “*le finalità di cui sopra*”, or “*ai fini di cui sopra*” (back-translation of both: “the above purposes”). The second expression, instead, relates to something that is quoted or stated in the document subject or that is the subject-matter of the contract. Italian sample phrases are “*i servizi di cui in oggetto*” (back-translation: “the services in the subject”) and “*le attività di cui in oggetto*” (back-translation: “the activities in the subject”). As regards frequent mistranslations, the expression “*di cui sopra*” tends to be wrongly rendered “as of above”⁹. If the adverb “above” is queried in the corpus, the following concordances are found: “set out above”; “as provided above”; “above-mentioned”; “above mentioned”, and “above-named”. In particular, “set out above” shows 42 hits; “as provided above” 17 hits; “above-named” 14 hits, and “above-mentioned” or “above mentioned” 10 hits. As regards the prepositional phrase “*di cui in oggetto*”, the Hoepli dictionary and some online forums suggest “the matter in hand” or “the matter at hand” as

<https://openinnovability.enel.com/personal-data-treatment/>;
<https://whalesanddolphins.tethys.org/personal-data-treatment-policy/>;
<https://www.ilpolodelcaffè.it/en/treatment-of-personal-data/>;
https://www.interno.gov.it/sites/default/files/bando_protezione_internazionale_2018_eng_crui.pdf.

⁹ See, for example, the following governmental document: <https://www.sicurezza nazionale.gov.it/sisr.nsf/wp-content/uploads/2018/07/accordo-sicurezza-Israele.pdf>.

translation candidates¹⁰. These options may not be the legal equivalents of the Italian expression. The corpus, in fact, shows no hits of “matter * hand”. Therefore, instead of focussing on the whole prepositional phrase, it would be helpful to search for a translation of “*oggetto*”. The Hoepli dictionary suggests the following options: “subject”, “subject-matter”, “object”, and “theme”. If “subject” is queried, interesting phrases are noticed, such as “with respect to the subject matter contained herein”; “in respect of the subject matter contained herein”; “concerning the subject-matter hereof” and “relating to the subject matter hereof”. Unrelated phrases are, instead, retrieved with “object”, such as “right to object to the Purchaser's calculation”, or “object code”. Only one occurrence is finally retrieved with “theme”; i.e. “the intended styling theme of the XX vehicle”. This term is obviously not a suitable translation option. Therefore, acceptable translations of “*di cui in oggetto*” can be the following expressions: “concerning the subject-matter hereof”, “relating the the subject matter hereof”, or “with respect to the subject matter contained herein”.

The last Italian formulae are “*in fede di che*” or “LCS” (“*letto, confermato e sottoscritto*”), which are reported at the end of legal documents, before the parties' signatures. These phrases, in particular “*letto, confermato e sottoscritto*”, are generally translated literally, such as “read, understood and signed”, or “read, confirmed and signed”¹¹. A good strategy to search for suitable translation options would be to notice the corresponding formulae reported at the end of English documents, before the signatures. Alternatively, it is possible to consult a dictionary. The Hoepli suggests “in witness (whereof)” as a translation of “*in fede (di che)*”. If “in witness” is searched for in the corpus, relevant formulaic expressions emerge, such as “In witness whereof, the parties hereto have caused this Agreement to be executed and delivered as of the date first set forth above”, and “IN WITNESS WHEREOF, this Agreement has been executed the day and year first above written”.

¹⁰ See, for example, the WordReference forum: <https://forum.wordreference.com/threads/la-pratica-in-oggetto.409748/>.

¹¹ See, for example, the following consent forms released by Italian universities: https://en.unito.it/sites/sten/files/informativa_liberatoria_v20_inglese.pdf; <https://www.unimib.it/sites/default/files/Sistemi%20informativi/LiberatorieVideoconferenze/liberatoriaennew%20%281%29.doc>.

See also some translators' forums or dictionary entries: <https://ita.proz.com/kudoz/italian-to-english/law%3A-contracts/2687044-lcs-or-lcs.html>; <https://www.wordreference.com/iten/letto,%20confermato,%20sottoscritto>.

These phrases can be considered perfect equivalents, as they express the willingness of the parties to be bound by a contractual relationship and the acknowledgement to have read and understood the related contract.

4. Discussion

On the basis of the analysis carried out above, Appendix 1 summarises the Italian formulaic expressions discussed, their recurrent mistranslations and the English correct equivalents according to corpus evidence.

It is interesting to highlight that corpus consultation has helped find acceptable equivalents and address recurrent mistranslations.

From an analysis of Appendix 1, it is evident that the formulae analysed in this paper range from long phrases such as “*tutto ciò premesso, le parti convengono quanto segue*”, to short prepositional or verb phrases, such as “*di cui sopra*”, “*di cui in oggetto*”, “*premessi che*”, or “*domiciliato a*”, as well as acronyms (“LCS”).

Appendix 1 also highlights recurrent mistranslations, mostly due to wrong collocates or false cognates, as in “*trattamento dati personali*” (where “*trattamento*” is rendered literally) and “*fondata / costituita*” referring to a “*società*” (where the various translation options of “*fondata / costituita*” proposed by dictionaries are not particularly satisfactory). As discussed in the analysis, wrong renderings can be due to imprecise or out-of-context dictionary suggestions, as in the case of “*sede legale*”, whose translation is generally “headquarters”. Although this term is correct in a business context, it is not in a legal document.

As mentioned, the majority of the shortcomings noticed above are due to literal translations, or calques, of original terms and phrases. Unfortunately, legal advisers, lawyers and translators do not always explore or understand legal formulae fully. The literature reports, in fact, that contracts are increasingly drafted in (legal) English as a *lingua franca* (Anesa 2019: 16); hence, they are written in a language that is not grounded in a particular legal system. At the same time, many international law firms tend to draw up legal documents following an Anglo-American drafting style, without bothering too much about the differences in language conventions, styles and/or in the legal systems (Jacometti and Pozzo 2018: 198). Generally, in fact, they apply clagues

so that the sense, or communicative intent of a phrase or formula of the source text is conveyed. As claimed by Tiersma (1999), lawyers have developed linguistic “quirks” with little communicative function (Tiersma 1999: 51). Legal prose, in fact, is argued to be “the largest body of poorly written literature ever created” (Coulthard and Johnson 2010: 46). This means that legal texts do not (only) have a communicative intent, but they also aim at conveying form and compliance with (language) norms (Tiersma 2015: 29ff).

Therefore, the acknowledgement of and conformity to formulae in a first and second language are pivotal when both drafting and translating legal documents.

5. Conclusions

This paper was aimed at exploring if and how recurrent Italian formulae used in contracts (and often mistranslated) can be satisfactorily rendered into English by consulting an *ad hoc* corpus. To this aim, a DIY corpus was composed in order to find acceptable equivalents and tackle specific translation issues.

Sometimes corpus analysis was quite straightforward thanks to dictionary entries suggesting the right translation candidate(s). An examples of this type was the word “whereas” translating “*premessochè*”. Other times, instead, the translation options proposed by bilingual dictionaries were misleading, as in the case of “*fondata*” and “*costituita*” (referring to a “*società*”) whose suggested terms were too many *vis-à-vis* the correct word to use (i.e. “incorporated”). Finally, at times dictionaries provided incorrect solutions, as with “*domiciliato*”, given that the English equivalent is not the calque “domiciled” but “whose address for service/notice is”, or similar expressions.

Despite the difficulties posed by the various, and sometimes mixed search strategies, corpus consultation yielded insightful results and helped dispel doubts as regards the best translation candidates. Thanks to word frequencies, analyses of word usages in context and of collocates, in fact, the corpus allowed to disambiguate terms and find English equivalents. It is self-evident that searches were not unproblematic, and sometimes finding or confirming translation candidates was arduous. Nonetheless, in light of the results obtained,

corpus analysis can be considered a successful tool to source authentic formulae and expressions, provided that the corpus is reliable; i.e. representative of its genre. Exploring and commenting on the corpus reliability, or better “representativeness” (McEnery et al. 2010), would go beyond the scope of this paper. However, translators (and corpus builders) should consider this aspect when composing an *ad hoc* corpus as a reference tool.

On the basis of the paper findings, it can be claimed that corpus consultation helped shed light not only on the best translation options, but also on the reasons why the recurrent translations into English are wrong. In practice, it showed how mistranslations can be tackled. From this perspective, it can be argued that corpus consultation is useful not only when translating legal texts, but also when reviewing past (or others') translation choices.

The limits of this paper lie in the reduced number of phrases taken into consideration. A larger number of formulaic expressions would have, for example, helped shed light on more frequent mistranslations. Also, further research could encompass other documents, such as notary's deeds, powers of attorney, testaments, or other private legal texts.

CONFLICT OF INTEREST: The author declares that there is no conflict of interest.

Bibliography

- Anthony, Lawrence. 2022. AntConc (Version 4.0.3) [Computer Software]. Tokyo, Japan: Waseda University. <https://www.laurenceanthony.net/software> (accessed September 20, 2022).
- Anesa, Patrizia. 2019. Towards a Conceptualization of Legal English as a Lingua Franca? *International Journal Of English Linguistics* 9(6): 14–21. DOI: <https://10.5539/ijel.v9n6p14>.
- Baroni, Marco and Bernardini, Silvia. 2004. BootCaT: Bootstrapping corpora and terms from the web. *Proceedings of LREC 2004*. sslmit.unibo.it/~baroni/publications/lrec2004/bootcat_lrec_2004.pdf (accessed September 20, 2022).

- Bhatia, Vijay K. 1993. *Analysing Genre: Language Use in Professional Settings*. London: Longman.
- Bhatia, Vijay K. 2010. *Textbook on legal language and legal writing*. New Delhi: Universal law publishing Co. Pvt. Ltd.
- Bhatia, Vijay K., Nicola M. Langton, and Jane Lung. 2004. Legal Discourse: Opportunities And Threats For Corpus Linguistics. In *Discourse In The Professions. Perspectives From Corpus Linguistics*, eds. Ulla Connor and Thomas A. Upton, 203–231. Amsterdam/Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/scl.16.09bha>.
- Biel, Łucja 2018. Corpora in Institutional Legal Translation: Small Steps and the Big Picture. In *Institutional Translation for International Governance: Enhancing Quality in Multilingual Legal Communication*, ed. Fernando Prieto Ramos, 25–36. London: Bloomsbury. DOI: <https://10.5040/9781474292320.0010>.
- Bowker, Lynne and Jennifer Pearson. 2002. *Working with specialized language: A practical guide to using Corpora*. London/New York: Routledge.
- Campbell Black, Henry. 2008. *A Dictionary of Law*. Second Edition. Clark: The Lawbook Exchange.
- Coulthard, Malcolm and Alison Johnson. 2007. *An Introduction to Forensic Linguistics. Language in Evidence*. London/New York: Routledge.
- Coulthard, Malcolm, and Alison Johnson. 2010. *The Routledge Handbook of Forensic Linguistics*. Abingdon: Routledge.
- Dani, Giuliana. 2019. English and Italian land contracts: a corpus-linguistics analysis. In *Corpus-based Research on Variation in English Legal Discourse*, eds. Teresa Fanego and Paula Rodríguez-Puente, 25–50. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/scl.91.02dia>.
- Durán Muñoz, Isabel. 2012. Meeting translators' needs: translation-oriented terminological management and applications. *JoSTrans Special issue on Terminology, Phraseology and Translation*: 77–92.
- Engberg, Jan. 2020. Comparative Law for Legal Translation: Through Multiple Perspectives to Multidimensional Knowledge. *International Journal for the Semiotics of Law – Revue*

- Internationale de Sémiotique Juridique* 33(2): 263–282. DOI: <https://10.1007/s11196-020-09706-9>.
- Engberg, Jan. 2021. Legal translation as communication of knowledge: On the creation of bridges. *Parallèles* 33(1): 6–17. DOI: <https://doi.org/10.17462/para.2021.01.02>.
- Farr, Fiona, and Anne O’Keeffe. 2019. Using corpora to analyse language. In *Routledge Handbook of English Language Teacher Education*, eds. Steve Walsh and Steve Mann, 268–282. London: Routledge.
- Galdia, Marcus. 2009. *Legal Linguistics*. Frankfurt Am Main: Peter Lang.
- Garzone, Giuliana. 2007. Osservazioni sulla didattica della traduzione giuridica. In *Tradurre le microlingue scientifico professionali. Riflessioni teoriche e proposte didattiche*, eds. Patrizia Mazzotta and Laura Salmon, 194–238. Turin: UTET.
- Genette, Marie. 2016. *How reliable are online bilingual concordancers? An investigation of Linguee, TradooIT, WeBiText and ReversoContext and their reliability through a contrastive analysis of complex prepositions from French to English* (Master's Thesis). https://www.duo.uio.no/bitstream/handle/10852/51577/MAThesis_Marie_GENETTE.pdf?sequence=1 (accessed September 20, 2022).
- Giampieri, Patrizia. 2018. Online parallel and comparable corpora for legal translations. *Altre Modernità* 20: 237–252. DOI: <https://doi.org/10.13130/2035-7680/10842>.
- Giampieri, Patrizia, and Giorgia Milani. 2021. Legal corpora: A trial lesson with translators and lawyers. *International Journal of Language Studies* 15(1): 47–66.
- Gotti, Maurizio. 2012. Text and Genre. In *The Oxford Handbook of Language and Law*, eds. Peter Tiersma and Lawrence M. Solan, 52–66. New York: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199572120.013.0005>.
- Jacometti, Valentina, and Barbara Pozzo. 2018. *Traduttologia e linguaggio giuridico*. Milan: Wolters Kluwer.
- Kast-Aigner, Judith. 2009. Terms in context: a corpus-based analysis of the terminology of the European Union’s development cooperation policy with the African, Caribbean and Pacific group of states. PhD thesis. Abstract available:

- <https://journal.equinoxpub.com/IJSL/article/view/6005>
(accessed September 20, 2022).
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž, and David Tugwell. 2004. Itri-04-08 the sketch engine. *Information Technology*. Available: <http://www.sketchengine.eu> (accessed September 20, 2022).
- Koehn, Philipp. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of Machine Translation Summit X: Papers*, 79–86. Phuket, Thailand. <https://aclanthology.org/2005.mtsummit-papers.11.pdf> (accessed September 20, 2022).
- Lehecka, Tomas. 2015. Collocations and Colligations. *Handbook of Pragmatics* 19: 1–20. DOI: <https://doi.org/10.1075/hop.19.col2>.
- McEnery, Tony, Richard Xiao, and Yukio Tono. 2010. *Corpus-based language studies: an advanced resource book*. London: Routledge.
- Mooney, Annabelle. 2014. *Language and Law*. New York: Palgrave MacMillan.
- Nikitina, Jekaterina. 2018. *Legal style markers and their translation in written pleadings before the european court of human rights*. Chisinau: Edizioni Accademiche Italiane.
- Prieto Ramos, Fernando. 2014a. Legal Translation Studies as Interdiscipline: Scope and Evolution. *Meta* 59(2): 260–277. DOI: <https://10.7202/1027475ar>.
- Prieto Ramos, Fernando. 2014b. Quality Assurance in Legal Translation: Evaluating Process, Competence and Product in the Pursuit of Adequacy. *International Journal for the Semiotics of Law* 28: 11–30. DOI: <https://doi.org/10.1007/s11196-014-9390-9>.
- Prieto Ramos, Fernando. 2021. Translating legal terminology and phraseology: between inter-systemic incongruity and multilingual harmonization. *Perspectives* 29(2): 175–183. DOI: <https://doi.org/10.1080/0907676X.2021.1849940>.
- Prieto Ramos, Fernando, and Diego Guzmán. 2021. Examining institutional translation through a legal lens: A comparative analysis of multilingual text production at international organizations. *Target* 33(2): 254–281. DOI: <https://doi.org/10.1075/target.21003.pri>.

- Rossini Favretti, Rema, Fabio Tamburini, and Enrico Martelli. 2007. Words from Bononia Legal Corpus. In *Text Corpora and Multilingual Lexicography*, ed. Wolfgang Tuebert, 11–30. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/ijcl.6.si.03ros>.
- Šarcevic, Susan. 1997. *New Approach to Legal Translation*. The Hague/London/Boston: Kluwer Law International.
- Seracini, Francesca. 2020. *The Translation of European Union legislation. A corpus-based study of norms*. Milan: LED Edizioni Universitarie.
- Stein, Dieter. 2009. Introduction: Forensic Translation — Practical Issues and their Theoretical Underpinnings. In *Translation Issues in Language and Law*, eds. Frances Olsen, Alexander Lorz and Dieter Stein, 1–4. Palgrave Macmillan, London. https://doi.org/10.1057/9780230233744_1.
- Tiedemann Jörg. 2012. Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'2012)*, ed. Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis, 2214–2218. Istanbul: European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf (accessed September 20, 2022).
- Tiersma, Peter. 2015. Some Myths about Legal Language. In *Speaking of Language and Law: Conversations on the Work of Peter Tiersma*, eds. Lawrence Solan, Janet Ainsworth and Roger W. Shuy, 27–34. Oxford: Oxford University Press.
- Tiersma, Peter. 1999. *Legal Language*. Chicago and London: The University of Chicago Press.
- Tiersma, Peter, and Lawrence Solan. 2012. *The Oxford Handbook Of Language And Law*. New York: Oxford University Press.
- Vogel, Friedemann, Hanjo Hamann, and Isabelle Gauer. 2017. Computer-assisted legal linguistics: Corpus analysis as a new tool for legal studies. *Law & Social Inquiry* 42: 1–24. DOI: <https://doi.org/10.1111/lsi.12305>.
- Zanettin, Federico. 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. London: Routledge.

Appendix 1

Italian legal formulae, their recurrent mistranslation and the corpus-sourced English equivalents

Italian legal formula	Recurrent mistranslation	English correct corresponding formula
<i>Premesso che</i>	Given that	Whereas,
<i>Premessa</i>	Premises	Recitals [<i>more frequent</i>]; Preamble
<i>Tutto ciò premesso, le parti convengono quanto segue</i>	Whereas, the parties stipulate and agree as follows; Having stated the above, the parties stipulate and agree as follows; In light of the above, the parties convene and agree as follows; Given the above, the parties stipulate and agree as follow. [<i>sic.</i>]	NOW, THEREFORE, the parties agree as follows: NOW, THEREFORE, in consideration of the mutual covenants contained herein and intending to be legally bound hereby, the parties hereto agree as follows:
<i>(Società) Fondata / Costituita</i>	Founded; established, set up, started (up), constituted	Incorporated
<i>Sede legale</i>	Headquarter(s); legal seat	Registered office; Principal office; Principal place of business
<i>Domicilio; Domiciliato (a)</i>	Domicile; Domiciled (at/in)	Address for Notice(s); Address for Service; Whose address for Notice(s) is
<i>Trattamento dati personali</i>	Treatment of personal data	Processing of personal data; Personal data processing
<i>Di cui sopra</i>	As of above	Above-mentioned; above mentioned;

Comparative Legilinguistics 2022/52

		above-stated; as provided above; set out above
<i>Di cui in oggetto</i>	The matter in/at hand	With respect to the subject matter contained herein; In respect of the subject matter contained herein; Concerning the subject-matter hereof; Relating to the subject matter hereof
<i>In fede di che; LCS (Letto Confermato Sottoscritto)</i>	Read, confirmed and signed; Read, understood and signed	In witness whereof, the parties hereto have caused this Agreement to be executed and delivered as of the date first set forth above; IN WITNESS WHEREOF, this Agreement has been executed the day and year first above written