

Volume 2022/52

Comparative Legilinguistics

International Journal
for Legal Communication

Faculty of Modern Languages and Literatures
Adam Mickiewicz University
Poznań, Poland

FACULTY OF MODERN LANGUAGES AND LITERATURES

EDITORIAL BOARD

Editor-in-chief: Aleksandra Matulewska

Co-editor 1: Anne Wagner

Co-editor 2: Joanna Nowak-Michalska

Editorial Secretaries and Assistants: Joanna Nowak-Michalska, Paula Trzaskawka,
Emilia Wojtasik-Dziekan

Editorial Board and Advisory Board available at:
<https://pressto.amu.edu.pl/index.php/cl/about/editorialTeam>

Section editors: Joanna Nowak-Michalska, Paula Trzaskawka, Emilia Wojtasik-Dziekan

Technical editors: Paula Trzaskawka, Emilia Wojtasik-Dziekan

Editorial Office

Faculty of Modern Languages and Literatures
al. Niepodległości 4, pok. 107A
61-874 Poznań, Poland
lingua.legis@gmail.com

The journal has been indexed on ERIH PLUS since 2018 and Scopus since 2021

The electronic version serves referential purposes. Wersja elektroniczna jest wersją referencyjną czasopisma

Copyright by Faculty of Modern Languages and Literatures
Printed in Poland

ISSN 2080-5926
e-ISSN 2391-4491 (<http://pressto.amu.edu.pl/index.php/cl/issue/archive>)

Adam Mickiewicz University, Poznań, Poland

Table of Contents

Patrizia GIAMPERI (ITALY) Translating legal formulae: a corpus-driven approach	293
Panagiotis G. KRIMPAS, Christina VALAVANI (GREECE) Attention mechanism and skip-gram embedded phrases: short and long-distance dependency n-grams for legal corpora	318
Valentina SOPJANI, Vjosa HAMITI (KOSOVO) Challenges in the translation of legal texts: the case in Kosovo	351
Dariusz KOŹBIAŁ (POLAND) Research profile and bibliography of Professor Łucja Biel	381

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.hltlaw.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 McMillan.
- 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

Comparative Legilinguistics
vol. 2022/52
DOI: <http://dx.doi.org/10.14746/cl.52.2022.14>

**Attention mechanism and skip-gram embedded
phrases: short and long-distance dependency n-grams
for legal corpora**

PANAGIOTIS G. KRIMPAS, Tenured Associate Professor

Faculty of Classics and Humanities
Democritus University of Thrace, Greece
P. Tsaldari 1 Str., Komotini 69100
pkrimpas@bscc.duth.gr

ORCID: <https://orcid.org/0000-0001-7271-9653>

CHRISTINA VALAVANI, Phd

Journalism Computational Linguistics Lab (JCL Lab)
National and Kapodistrian University of Athens, Greece
Panepistimiou 30 Str., Athens 10679
cvalavani@hotmail.com

ORCID: <https://orcid.org/0000-0002-2944-0734>

Abstract: This article examines common translation errors that occur in the translation of legal texts. In particular, it focuses on how German texts containing legal terminology are rendered into Modern Greek by the Google translation machine. Our case study is the Google-assisted translation of the original (German) version of the Constitution of the Federal Republic of Germany into Modern Greek. A training method is proposed for phrase extraction based on the occurrence frequency, which goes through the Skip-gram algorithm to be then integrated into the Self Attention Mechanism proposed by Vaswani et al. (2017) in order to minimise human effort and contribute to the development of a robust machine translation system for multi-word legal terms and special phrases. This Neural Machine Translation approach aims at developing vectorised phrases from large corpora and process them for translation. The research direction is to increase the in-domain training data set and enrich the vector dimension with more information for legal concepts (domain specific features).

Keywords: computational linguistics; legal terminology; legal translation; Neural Machine Translation; Self Attention Mechanism; short and long-distance dependency n-grams; skip-gram algorithm.

SELF ATTENTION ΚΑΙ ΦΡΑΣΕΙΣ ΠΟΥ ΕΝΣΩΜΑΤΩΝΟΝΤΑΙ ΣΤΟ SKIP-GRAM: Ν-ΓΡΑΜΜΑΤΑ ΣΕ ΚΟΝΤΙΝΗ ΚΑΙ ΜΑΚΡΙΝΗ ΕΞΑΡΤΗΣΗ ΓΙΑ ΣΩΜΑΤΑ ΝΟΜΙΚΩΝ ΚΕΙΜΕΝΩΝ

Περίληψη: Αυτό το άρθρο εξετάζει συνήθη μεταφραστικά σφάλματα που σημειώνονται κατά τη μετάφραση νομικών κειμένων. Ειδικότερα, εστιάζει στον τρόπο με τον οποίο αποδίδει στα νέα ελληνικά η μηχανή μετάφρασης της Google γερμανικά κείμενα που περιέχουν νομική ορολογία. Η μελέτη περίπτωσης που χρησιμοποιούμε αφορά τη μετάφραση της πρωτότυπης (γερμανικής) εκδοχής του Συντάγματος της Ομοσπονδιακής Δημοκρατίας της Γερμανίας στα νέα ελληνικά, μέσω της μηχανής μετάφρασης της Google. Προτείνεται μια μέθοδος εκπαίδευσης για την εξαγωγή φράσεων βάσει της συχνότητας εμφάνισης τους, η οποία διέρχεται από τον αλγόριθμο Skip-gram για να ενσωματωθεί κατόπιν στον Μηχανισμό Αυτοπροσοχής (Self Attention Mechanism) των Vaswani et al. (2017), προκειμένου να ελαχιστοποιήσει την ανθρώπινη προσπάθεια και να συμβάλει στην ανάπτυξη ενός ισχυρού συστήματος μηχανικής μετάφρασης για πολυλεκτικούς νομικούς όρους και ειδικές φράσεις. Αυτή η προσέγγιση, στο πλαίσιο της Νευρωνικής Μηχανικής Μετάφρασης, αποσκοπεί να αναπτύξει διανυσματοποιημένες φράσεις από μεγάλα σώματα και να τις επεξεργαστεί με στόχο τη μετάφραση. Η έρευνά μας κατευθύνεται προς την αύξηση των συνολικών δεδομένων εκπαίδευσης εντός δεδομένου θεματικού πεδίου και να

εμπλουτίσει τη διανυσματοποιημένη διάσταση με περισσότερες πληροφορίες για νομικές έννοιες (ιδιαίτερα χαρακτηριστικά του θεματικού πεδίου).

Λέξεις-κλειδιά: αλγόριθμος skip-gram; ν-γράμματα σε κοντινή και μακρινή εξάρτηση; νευρωνική μηχανική μετάφραση; νομική μετάφραση; νομική ορολογία; self attention mechanism; υπολογιστική γλωσσολογία.

SELF-ATTENTION-MECHANISMUS UND SKIP-GRAM EINGEBETTETE PHRASEN: N-GRAMME IN NAH- UND FERNABHÄNGIGKEIT FÜR RECHTSKORPORA

Zusammenfassung: Dieser Artikel untersucht häufige Übersetzungsfehler, die bei der Übersetzung von Rechtstexten auftreten. Insbesondere geht es darum, wie deutsche Texte mit juristischer Terminologie von der Google-Übersetzungsmaschine ins Neugriechische übertragen werden. Unsere Fallstudie ist die Google-gestützte Übersetzung der deutschen (originalen) Fassung der Verfassung der Bundesrepublik Deutschland ins Neugriechische. Für die Extraktion von häufigen Phrasen wird eine Trainingsmethode vorgeschlagen, die den Skip-Gram-Algorithmus durchläuft und wird dann in den von Vaswani et al. (2017) vorgestellten Selbstaufmerksamkeitsmechanismus integriert, um den menschlichen Aufwand zu minimieren und zur Entwicklung eines robusten maschinellen Übersetzungssystems für Mehrwortrechtstermini und -phrasen beizutragen. Dieser Ansatz der neuronalen maschinellen Übersetzung zielt darauf ab, vektorisierte Phrasen aus großen Korpora zu entwickeln und sie zur Übersetzung zu verarbeiten. Unsere Forschungsrichtung besteht darin, den domäneninternen Trainingsdatensatz zu erweitern und die Vektordimension mit mehr Informationen um Rechtskonzepte (domänenspezifische Merkmale) anzureichern.

Schlüsselwörter: Computerlinguistik; juristische Terminologie; juristische Übersetzung; n-Gramme von Nah- und Fernabhängigkeiten; neuronale maschinelle Übersetzung; Self-Attention-Mechanismus; Skip-Gramm-Algorithmus.

1. Introduction

This article examines common errors that occur when machine-translating legal texts. In particular, we discuss legalese translated from German into Modern Greek with focus on how the Google

TranslationTM machine renders German legal terms, term elements, appellations and special phrases (hereinafter: legal language units) into Modern Greek at sentence level. Our case study is a Google-translated Modern Greek version of the German (original) text of the Constitution of the Federal Republic of Germany, the so-called *Grundgesetz für die Bundesrepublik Deutschland* (Bundesministerium der Justiz, Bundesamts für Justiz, <http://www.gesetze-im-internet.de/gg/GG.pdf>), hereinafter referred to as the GBD Parallel Text Corpus. We test the accuracy of sentence-level Google translation by comparing, in two adjacent columns, the target text sentence with the source text sentence generated by Google Translate in order to detect specific error types whose study may lead to steps and suggestions for a more sensitive Machine Translation. The errors of the automatic tool were identified within the corpus and were then analysed and classified according to specific criteria (see below in this unit). As explained by Stanisław Goźdz-Roszkowski (2021: 1524):

“The influence of corpus linguistics methodology on how legal phraseology has been investigated extends beyond technological advances in text processing. Rather, corpus linguistics phraseology has paved the way for new and innovative studies which have begun to reveal the potential for investigating various roles and functions performed by different multi-word units in legal discourse.”

Our study falls under what is called corpus-driven approaches (Tognini Bonelli 2001: 84–100), given that we make no prior assumptions and our source of information is the corpus itself (Goźdz-Roszkowski 2021: 1517). The utility of web legal resources as legal corpora has been discussed and supported by various scholars (e.g. Giampieri 2018), while the same is true of the relevance of machine-translation research for legal translation, especially for paedagogical purposes (e.g. Wiesmann 2019).

Since the 1990s there has been a shift from the dominant rule-based methods to statistical approaches. Following this background, deep learning goes further down, and gradually becomes the de facto technique of the mainstream statistical landscape (Liu et al. 2017). Neural Machine Translation (Kalchbrenner and Blunsom 2013) has demonstrated impressive performance in recent years. In this article we propose a training method for multi-word legal language unit extraction that goes through the Skip-gram algorithm (Mikolov et al. 2013) to be then integrated into the Self Attention Mechanism

(Vaswani et al. 2017); this process can minimise human effort and contribute to the development of a robust machine translation system for multi-word units. The aim of this Neural Machine Translation approach is to develop vectorised phrases from large corpora and process them in a way novel for translation.

After a thorough review of the machine translation output, some error types have been recorded, largely based on existing categorisations (Tezcan, Hoste and Mackel 2017; van Brussel, Tezcan, and Mackel 2018: 3800–3803; Krimpas 2017b: 79–96) by distinguishing between fluency and accuracy errors (Tezcan, Hoste and Mackel 2017). The error types recorded in our sample, adapted from Tezcan, Hoste and Mackel (2017), are as follows (our adaptations/additions appear in square brackets; the abbreviation TRM stands for ‘term’):

Accuracy errors

- Mistranslation
 - Multi Word Expressions (MIS-MWE) [MIS-MWE-TRM]
 - POS
 - Sense (MIS-SE) [(MIS-SE-TRM)]
 - Mistranslation of verb tense and voice, number (nouns) (MIS-TVN)
 - Partial (MIS-PA)
 - Semantically unrelated (MIS-SU) [MIS-SU-TRM]
- Do not translate (DNT) (words have been translated unnecessarily e.g. for proper names)
- Untranslated (UT) [German Word/-s (GW)]
- Addition (AD) [AD-TRM]
 - Content Word
 - Function Word
- Omission (OM) [OM-TRM]
 - Content Word
 - Function Word

- Mechanical (non-meaning errors e.g. punctuation) (MECH)

Fluency errors

- Grammar (GR)
 - Word Form (WF)
 - Word Order (WO)
 - Extra Word(s) (EW)
 - Missing Word(s) (MW)
 - Multi Word Syntax (MWS)
- Lexicon (LEX)
 - Nonexistent (LEX-NE)
 - Lexical choice (LEX-CH) [LEX-CH-TRM] [Phrase Lexical Choice Term (PH-LEX-CH-TRM) and Partial Phrase Lexical Choice Term (PPH-LEX-CH-TRM)]
- Orthography (ORTH)
- Multiple errors (MULER)

To alleviate the translation errors documented in this research we propose a better, enriched version of the Skip-Gram and Self Attention mechanism proposed by Vaswani et al. (2018), where we modify the System so as to process legal and, in general, special multi-word units. In general, we make use of the Pointwise Mutual Information (PMI) (Bouma 2009) method (bigram extraction) (see Figures 1, 2, 3) and of the Short and Long-Distance Dependencies Extraction Algorithm (SLDDExAl) before inserting our words into the Skip-gram algorithm (Mikolov et al. 2013); then we insert the output vectors into the Self Attention Mechanism for more meaning, in which case the words become re-embedded. Attention is a concept that has helped improve the performance of NLP applications (Jay Alamar, jalamar.github.io, Visualizing machine learning one concept at a time, article posted June 27, 2018), including Machine Translation.

We also extend the Skip-gram model (Mikolov et al. 2013) by customising it to our needs. This paper largely incorporates a Self

Attention model that takes into account n-grams in relation to the predicted word. We begin with searching for bigrams, trigrams and tetragrams (hereinafter n-grams) to then embed them with Skip gram for better results (Mikolov et al. 2013). Then we train the System in the Self Attention Mechanism in order to develop vectorised n-grams (bigrams, trigrams, tetragrams in short and long-distance dependency).

In short, unit two presents all categories of errors in legal language units with examples extracted from the corpus. Their classification helps identify weak points of the translation tool with special focus on legal language units. Unit three discusses ways of automatic extraction of both bilets and multi-word units in short and long distance dependency. In unit four Skip-gram training takes place in order to obtain embeddings with more meaning. Unit five describes the attention mechanism with the new elements and how exactly the n-grams are integrated for correct translation purposes. Unit six records the proposed mechanism's steps as well as new proposals. Unit seven summarises by presenting advantages and disadvantages of our proposal.

2. Error documentation

The difficult-to-process character of legal language units often results in pronounced discrepancies in both human and machine translation, the DE > EL language pair being no exception.

To reflect the actual will of the legislator it is vital for the legal terminology used in the target language (TL) to cover the same conceptual areas as the source text (ST). In practice, however, the attempt to find legally equivalent terms is not always straightforward due to the asymmetry of legal systems (Duběda 2021: 61, 68, 69; Prieto Ramos 2021: 175–176), even if they belong to the same family of law, as is the case with the Greek and the German ones. At times the asymmetry can be purely terminological-semantic rather than conceptual, but this can be equally problematic for the legal translator (Krimpas 2017a).

Tables 1–27 below show examples of n-grams, some of which are interdependent with other, correctly translated units in the sentence, while others are semantically mistranslated independently of

context. Context is included whenever appropriate. All examples given below are taken from the aforementioned GDB Parallel Text Corpus (see unit 1). This corpus comprises approximately 5,000 pairs of sentences, whose translation into Greek was carried out at the sentence level by Google Translate; approximately 300 sentence pairs out of them were translated by both Google Translate and a human translator; the examples below come from this particular subset. The tables show sentence parts that are essential to illustrate machine translation errors; whole sentences are given only when necessary. Units involved in one or more machine translation errors were manually extracted. Underlined text in the first row of each table (source text) shows translational correspondence with underlined text in the third row of each table (target text), as a way to highlight text involved in the machine translation error. The second row of each table shows the machine-translated target text, accompanied by the error code (see unit 1). In cases of clear correspondence between the first and third row there is no underlining.

Tables 1–11: Examples of unigrams involved in context-dependent errors

1.	Source text:	<u>Grundgesetz</u> für die Bundesrepublik Deutschland
	Google translation:	Βασικός νόμος (MIS-SE-TRM)
	Correct rendering:	Σύνταγμα

2.	Source text:	(weggefallen)
	Google translation:	(εγκαταλείφθηκε) (MIS-SE-TRM)
	Correct rendering:	(καταργήθηκε)

3.	Source text:	Jeder Deutsche hat in jedem <u>Land</u> e die gleichen staatsbürgerlichen Rechte und Pflichten.
	Google translation:	χώρα (MIS-SE-TRM)
	Correct rendering:	ομόσπονδο κρατίδιο

4.	Source text:	...soweit der <u>Bundesrat</u> ihm zustimmt.
	Google translation:	Bundesrat [...]. (GW(s)+ MIS-SE-TRM)
	Correct rendering:	Ομοσπονδιακό Συμβούλιο

5.	Source text:	den <u>Wasserhaushalt</u>
	Google translation:	ισοζύγιο νερού (MIS-SE-TRM)
	Correct rendering:	διαχείριση (των) υδάτινων πόρων

6.	Source text:	[...] zwei Jahren nach der <u>Durchführung</u> der <u>Volksbefragung</u> ein [...]
	Google translation:	<u>πραγματοποίηση</u> του <u>δημοψηφίσματος</u> (LEX-CH-TRM)
	Correct rendering:	διεξαγωγή του δημοψηφίσματος

7.	Source text:	Oberster Gerichtshof für die in Absatz 1 und 2 genannten Gerichte ist der <u>Bundesgerichtshof</u> .
	Google translation:	Ομοσπονδιακό Δικαστήριο (OM-TRM)
	Correct rendering:	Ομοσπονδιακό Ακυρωτικό Δικαστήριο

8.	Source text:	Kunst und Wissenschaft, Forschung und Lehre sind <u>frei</u> .
	Google translation:	δωρεάν (LEX-CH-TRM)
	Correct rendering:	ελεύθερες

9.	Source text:	Ihre <u>Gründung</u> ist frei.
	Google translation:	Η εγκατάσταση σας (LEX-CH-TRM)
	Correct rendering:	Η ίδρυσή της

10.	Source text:	<u>(Vollzitat:)</u>
	Google translation:	(Πλήρες απόσπασμα:) (PPH-LEX-CH-TRM)
	Correct rendering:	(πλήρες παράθεμα:)

11.	Source text:	<u>Bundesrecht</u> bricht <u>Landesrecht</u> . (a context-independent unigrams)
	Google translation:	<u>Ο ομοσπονδιακός νόμος</u> παραβιάζει <u>τον κρατικό νόμο</u> . (MIS-MWE-TRM)
	Correct rendering:	Το ομοσπονδιακό δίκαιο παραβιάζει το πολιτειακό δίκαιο

Tables 12–19: Examples of bigrams involved in short-distance dependency errors

12.	Source text:	daß ein <u>billiger Ausgleich</u> erzielt,
	Google translation:	φθηνή αποζημίωση (MIS-MWE-TRM)
	Correct rendering:	εύλογο συμψηφισμό

13.	Source text:	Die <u>konkurrierende Gesetzgebung</u> erstreckt sich auf folgende Gebiete
	Google translation:	ανταγωνιστική νομοθεσία (MIS-MWE-TRM)
	Correct rendering:	συντρέχουσα νομοθετική αρμοδιότητα

14.	Source text:	Den <u>unehelichen Kindern</u> sind durch die Gesetzgebung die gleichen Bedingungen für ihre leibliche und seelische Entwicklung und ihre Stellung in der Gesellschaft zu schaffen wie den <u>ehelichen Kindern</u> .
	Google translation:	[...] παράνομα παιδιά [...] νόμιμα παιδιά. (MIS-MWE-TRM)
	Correct rendering:	[...] τέκνα εκτός γάμου [...] τέκνα γεννημένα σε γάμο

15.	Source text:	(+++ <u>Textnachweis Geltung</u> ab: 14.12.1976 +++)
	Google translation:	(+++ Η <u>απόδειξη ισχύει</u> από: 14.12.1976 +++) (MIS-MWE-TRM)
	Correct rendering:	Η παρούσα εκδοχή τέθηκε σε ισχύ από: 14.12.1976 +++

16.	Source text:	Zur Wahrung der <u>Einheitlichkeit</u> der <u>Rechtsprechung</u> ist ein <u>Gemeinsamer Senat</u> der in Absatz 1 genannten Gerichte zu bilden.
	Google translation:	[...] <u>ομοιομορφία</u> της <u>νομολογίας</u> , [...] <u>μεικτή σύγκρουση</u> [...] (MIS-MWE-TRM + LEX-CH-TRM)
	Correct rendering:	[...] <u>ενιαίου χαρακτήρα της νομολογίας</u> [...] Μείζων Ολομέλεια

17.	Source text:	Sie soll hierbei ihre <u>Auffassung darlegen</u> .
	Google translation:	[...] <u>εξηγήσει</u> την <u>άποψή</u> της. (MIS-MWE-TRM and LEX-CH-TRM).
	Correct rendering:	[...] καταθέτει τη γνώμη της.

18.	Source text:hat sich das Deutsche Volk <u>kraft seiner verfassungsgebenden Gewalt</u> dieses Grundgesetz gegeben.
	Google translation:	λόγω των συστατικών του δυνάμεων (MIS-MWE-TRM)
	Correct rendering:	δυνάμει της συντακτικής του εξουσίας

19.	Source text:	Zwischen dem Antrage und der Wahl <u>müssen</u> achtundvierzig Stunden <u>liegen</u> .
	Google translation:	να υπάρχουν (MIS-MWE-TRM)
	Correct rendering:	να μεσολαβούν

Tables 20–27: Examples of phrases or sentences with words involved in short- and long-distance dependency errors

20.	Source text:	Der <u>Verlust</u> der <u>Staatsangehörigkeit</u> darf nur auf Grund eines Gesetzes und gegen den Willen des Betroffenen nur dann <u>eintreten</u> [...] (LEX-CH-TRM)
	Google translation:	Η <u>απώλεια</u> της <u>ιθαγένειας</u> <u>μπορεί να συμβεί</u> [...] (NRTDNL)
	Correct rendering:	Η απώλεια της ιθαγένειας επέρχεται [...]

21.	Source text:	Im Falle eines vorsätzlichen Verstoßes <u>kann</u> auf <u>Entlassung</u> <u>erkannt werden</u> .
	Google translation:	<u>η απόλυση</u> <u>μπορεί να αναγνωριστεί</u> . (MIS-MWE-TRM)
	Correct rendering:	μπορεί να τεθεί σε διαθεσιμότητα.

22.	Source text:	Frauen <u>vom vollendeten achtzehnten bis zum vollendeten fünfundfünfzigsten Lebensjahr</u>
	Google translation:	<u>από την ηλικία των δεκαοκτώ έως την ηλικία των πενήντα – Πέμπτο</u> . (MIS-MWE-TRM and EW(s))
	Correct rendering:	με συμπληρωμένη ηλικία από δεκαοκτώ έως πενήντα ετών

23.	Source text:	Ihre <u>hauptamtlichen Richter</u> müssen die <u>Befähigung</u> zum <u>Richteramt</u> haben.
	Google translation:	Οι κριτές πλήρους απασχόλησης πρέπει να είναι κατάλληλοι για να υπηρετούν ως κριτές. (MIS-MWE-TRM)
	Correct rendering:	Οι δικαστές πλήρους απασχόλησης πρέπει να έχουν την ικανότητα ανάληψης του δικαστικού λειτουργήματος.

24.	Source text:	<u>Wehrpflichtige</u> , die nicht zu einem <u>Dienst</u> nach Absatz 1 oder 2 <u>herangezogen sind</u>
	Google translation:	<u>Στρατιωτικοί</u> που δεν <u>εμπλέκονται</u> σε <u>υπηρεσία</u> (MIS-MWE-TRM)
	Correct rendering:	Επίστρατοι/Κληρωτοί που δεν καλούνται να υπηρετήσουν τη θητεία τους

25.	Source text:	Die <u>Verwirkung</u> und ihr <u>Ausmaß werden</u> durch das Bundesverfassungsgericht <u>ausgesprochen</u> .
	Google translation:	Η <u>κατάπτωση</u> και η <u>έκτασή</u> της θα εκφραστεί από το Ομοσπονδιακό Συνταγματικό Δικαστήριο. (MIS-MWE-TRM)
	Correct rendering:	Η έκπτωση δικαιώματος και ο βαθμός της αποφασίζονται από το Ομοσπονδιακό Συνταγματικό Δικαστήριο.

26.	Source text:	<u>Gesetzesvorlagen</u> werden beim <u>Bundestage</u> durch die Bundesregierung, aus der <u>Mitte</u> des <u>Bundestages</u> oder durch den <u>Bundesrat</u> eingebracht
	Google translation:	Οι <u>λογαριασμοί</u> εισάγονται στο <u>Bundestag</u> από την Ομοσπονδιακή Κυβέρνηση, από το <u>κέντρο</u> του <u>Bundestag</u> ή από το <u>Bundesrat</u> . (LEX-CH-TRM and UT-GW(s))
	Correct rendering:	Τα νομοσχέδια εισάγονται στο Ομοσπονδιακό Κοινοβούλιο από την Ομοσπονδιακή Κυβέρνηση, από το βήμα του Ομοσπονδιακού Κοινοβουλίου ή από το Ομοσπονδιακό Συμβούλιο.

27.	Source text:	Gewählt ist, wer die <u>Stimmen</u> der <u>Mehrheit</u> der Mitglieder des <u>Bundestages</u> auf <u>sich vereinigt</u> .
	Google translation:	ποιος <u>έχει</u> την <u>πλειοψηφία</u> των μελών του <u>Bundestag</u> (UT-GW(s) and LEX-CH-TRM)
	Correct rendering:	ποιος συγκεντρώνει την πλειοψηφία των μελών του Ομοσπονδιακού Κοινοβουλίου.

It should be borne in mind that in such cases translation quality depends on the available pool of German legal texts having been translated so far into Modern Greek (in this case not as many) by the machine translator (in this case Google Translate), as well as on the model used by the latter. An additional factor is whether the text to be translated will be machine-translated at the word, sentence (as is here the case) or whole text level.

A closer look at the above data suggests that areas where syntax interference is detected show a high correlation with wrong syntactic interpretation errors, occurring at the stage of decoding or recoding of the syntactic structure of the text (we code them,

respectively, as MWS, WO), while the areas where lexical interference is detected are mainly associated with errors in decoding or recoding of the semantics of the text (LEX-CH-TRM, MIS-SE, MIS-SU).

There are also numerous language register-related errors in the test text (we code them as LEX-CH-TRM); in such cases the texts have been semantically, grammatically and syntactically correctly rendered in Greek, but there is a register discrepancy with respect to Greek legalese. Moreover, Modern Greek LSP is notorious for its heavy learned element, which poses lexical, phonotactic, morphological and syntactic difficulties even for native speakers (Valeontīs and Krimpas 2014: 49–54; cf. Krimpas 2019), a fact reflected also in the translation of LSP texts, which becomes particularly hard especially in institutional thematic areas (Valeontīs and Krimpas 2014: 21) such as law, economics, religion etc. This linguistic landscape favours translation errors due to ‘non-recognition of text-specific deviations from normal language usage’ (Wiesmann 2019: 137) (an error type that we code as MIS-SE, MIS-SU).

Finally, at various places in the test text there is some indication of probable syntax interference of the source language. In such cases, the meaning from the source text is not transferred to the target text, a translation error that we code as MWS, WO.

In the above cases the difficulty of transferring the exact meaning of the source language (German) into the target language (Modern Greek) is obvious, either because the relative concept does not exist in the target language, or because of intersystemic differences, which often lead to wrong term choice or even non-translation; such cases are coded as LEX-CH-TRM.

3. Pre-training of the corpus and n-grams mining

Mining n-grams is the automatic extraction of frequent phrases (Del, Tättar, and Fishel 2018), such as multi-word terms and special phrases, from a corpus. First, we POS tag, parse (syntactic dependencies) (Klemen, Krsnik, and Robnik-Šikonja 2022), lemmatise and tokenise the whole corpus and then extract bigrams, trigrams and tetragrams, hereinafter referred to as n-grams (verbs,

nouns, adverbs, adjectives, participles and prepositions) to subsequently take them as input into Word2Vec, in particular into the Skip-gram algorithm, which generates vectorised words of high dimensionality (Camacho-Collados and Pilehvar 2018) with more meaning (see Figure 9). The threshold for the n-grams will be high, so that high quality legal LSP words (especially with short- and long-distance dependencies), phrases are extracted.

The mechanism for extracting frequent n-grams is as follows: If x and y represent bigrams in the legal corpus, y follows x . Whenever x and y appear together many times, the Pointwise Mutual Information (PMI) (Bouma 2009) will have a high value (see Figures 1, 2, 3), while it will have a value of 0 if x and y are completely independent, i.e. if they appear in different sentences (Moshe Hazoom, Towards Data Science, article posted December 22, 2018). This can be extended to three or four words e.g. a tetragram $[a, b, x, y]$ could collocate in a document by using the Short and Long-Distance Dependency Extraction Algorithm (since the PMI formula is tailored for pairs and combinations of two items).

$$E \in R^{|V| \times d}$$

$|V| = \text{vocabulary size}$

Figure 1: Embedding matrix after Word2Vec training (Moshe Hazoom, Towards Data Science, article posted December 22, 2018).

$$PMI(x; y) = \log \frac{p(x, y)}{p(x)p(y)}$$

Figure 2: Pointwise Mutual Information (PMI). PMI of concrete occurrences of x and y (Moshe Hazoom, Towards Data Science, article posted December 22, 2018).

PMI helps us find bigrams in order to build phrase vectors and embed them (Moshe Hazoom, Towards Data Science, article posted December 22, 2018).

$$NPMI(x; y) = \frac{\log \frac{p(x, y)}{p(x)p(y)}}{-\log p(x, y)}$$

Figure 3: Normalised Pointwise Mutual Information of x and y (Moshe Hazoom, Towards Data Science, article posted December 22, 2018).

Additionally, an n-gram could often co-occur in a sentence but in a long-distance dependency (see Figure 8). With the aim of being able to extract words that are also in a long- distance dependency but syntactically related, we built a simple algorithm, the Short and Long Distance-Dependency Extraction Algorithm (SLDDExAI). First, the corpus is processed and more specifically parsed, lemmatised, all stop words are removed and all tokens per sentence are collected (from within each sentence). Lemmata remain within the sentences they belong to. Co-occurrence counting is only done at parser-defined sentence boundaries. For example, if the word x co-occurs with the word y in sentence s1, then this is registered by the algorithm. Every sentence of the corpus where the word x appears is checked. If, at a later point, the word x appears again in another sentence along with the word y, then the algorithm adds this information to the frequency count list.

In case the word y appears in sentences without the word x, then the word y is not counted since we are only interested in its co-occurrence with the word x. This process is repeated/iterated for all lemmata in the corpus. Thus it is established which words frequently co-appear at any distance within a sentence, while at the same time their syntactical relation is detected by the parser. Some issues may arise with respect to the automatic translation of eventual out-of-domain parts of the corpus.



Figure 4: Examples of frequent n-grams for embedding them with one vector.

In particular, our system builds one vector for n-grams that collocate (with high occurrence frequency) (see Figure 4). In a parsed text, the system is able to know, inter alia, which words are related to each other and focuses on extracting them as they often co-occur in a sentence. We are not looking in the text for words with a specific syntactical relationship between them, e.g. verb-subject, but for frequently co-occurrence words that may have any syntactical relationship with each other. For the aforementioned reasons we don't use the two generalizations for multivariate distributions of Pointwise Mutual Information, presented by Tim Van de Cruys (2011).

4. N-gram embedding with Skip-gram

Word2vec can be applied to a big amount of data and Skip-gram (Mikolov et al. 2013) is one of the unsupervised learning techniques (it can work on any raw text) used to find the most relative words for a given word (Mikolov et al. 2013), especially with infrequent words.

Skip-gram predicts the context words from the target word and -in our proposed approach- it can learn legal concept embeddings from different data sources, including journals and legal narratives. Creating representations for legal concepts by training the System with legal corpora is highly recommended (domain adaption) (Diniz da Costa et al. 2022).

In the output vector of Skip-gram there is semantic information and representation of the relation between words, which is not the case for one-hot representations. Then those n-grams are inserted into Word2Vec (Mikolov et al. 2013) to be trained in the Skip-gram algorithm (Mikolov et al. 2013). The algorithm will take as input one-hot encodings which represent n-grams but will process the collocating n-grams as one vector and the co-occurring ones simultaneously. The architecture of Skip-gram is presented in Figure 5 below:

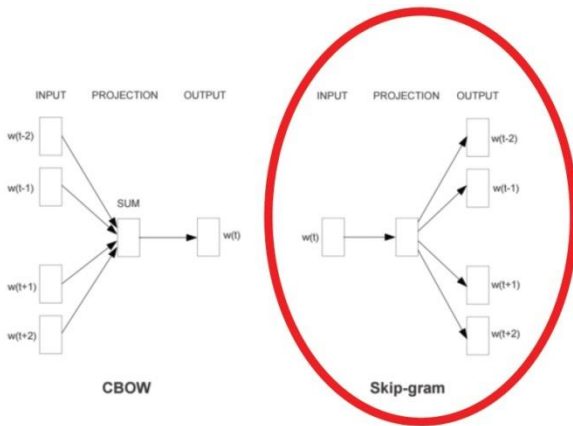


Figure 5: The Skip-gram model architecture (Source: <https://arxiv.org/pdf/1301.3781.pdf> Mikolov et al., 2013)

Skip-gram is a simple Neural Network with only one hidden layer (Mikolov et al. 2013). The input to the network is a one-hot encoded vector representation of a target-word; all of its dimensions are set to zero, apart from the dimension corresponding to the target-word (one-hot representation). The output is the probability distribution over all

words in the vocabulary, which defines the likelihood of a word being selected as the input word’s context (Paula, Cambridge Spark, article posted November 9, 2018). Figure 6 below illustrates the Skip-gram model in more detail.

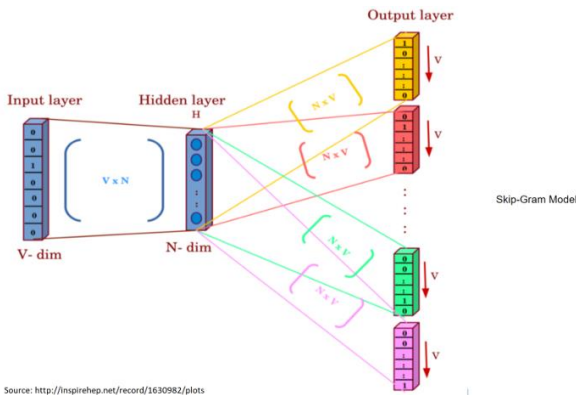


Figure 6: Skip-gram model architecture in detail (Doshi 2019).

With the SLDDExAI we can extract n-grams and use them as one-word (see Figure 4) vector for collocations and as multiple-word vectors for short and long-distance dependencies (depending on the number of words) (see Figure 8). This happens additionally to the one-word vectors vocabulary. When two words appear at a distance then they will be extracted as such and inserted together into the Skip-Gram with a gap, e.g. “Verlust [...] eintreten” will be inserted as a unit into the Skip-Gram. Skip-Gram will train the vectors by simultaneously setting a window of two words left and right of each word, the two words will be trained in the shared context of the words. If four words do not appear between the two, then the system automatically shrinks the window and adapts to those that exist. Moreover, when a gap stands for more words, the system is still trained on the basis of a two-word window, as is usual with Skip-Gram.

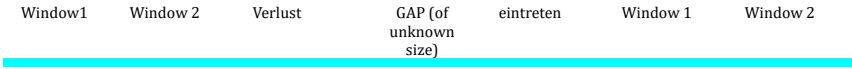


Figure 7: Long-distance dependency legal term inserted in Skip-Gram.

After training in Skip-gram we have the embeddings with more targeted meaning.

5. Self-attention mechanism

Most successful approaches to machine translation (e.g. Wu, Zhao, and Li 2020; Bahdanau, Cho, and Bengio 2016; Vaswani et al. 2017; Gehring et al. 2017) rely on the availability of parallel corpora. Supervised Neural Machine Translation (NMT) (Kalchbrenner and Blunsom 2013) works with the encoder- decoder, where the encoder reads the source sentence and produces its representation, which is then fed to the decoder in order to generate the target sentence word-by-word (Del, Tättar, and Fishel 2018) (see Figure 9). Cross entropy loss is usually used as a training objective and beam search algorithm is used for inference (Del, Tättar, and Fishel 2018). Such neural models rely on vast amounts of parallel data. We employ the Self Attention Mechanism as presented in Vaswani et al. (2017). The closer the vectors are, the bigger the dot product is. By computing the cosine similarity we find the similarity between vectors, and we can also measure the Euclidean distance d for it.

5.1. Why self-attention in legal language

With respect to legal language units long memory might be regarded as not required by a system of Neural Machine Translation, since they can be one-word (simple or complex) terms, multi-word terms or other multi-word (context-conditioned or fixed) special phrases. Legal language, however, is pretty demanding as such and if ones wishes to structure a mechanism that translates correctly while maintaining the

language register and rendering it accordingly to the target language, then the Self Attention Mechanism is quite appropriate. For example, it is worth considering the German sentence: ‘Der Verlust der Staatsangehörigkeit darf [...] eintreten’, which was Google-translated as ‘Η απώλεια της ιθαγένειας μπορεί να συμβεί’. Although the translation is perfectly understandable even with the general-language verbal phrase *μπορεί να συμβεί* ‘can/may happen’, the expected wording in Modern Greek legal language would require the LSP verb *επέρχεται* ‘takes place’. In this case the participle is semantically bound to the legal language. High attention is necessary, especially due to the long-distance dependency between the two words.

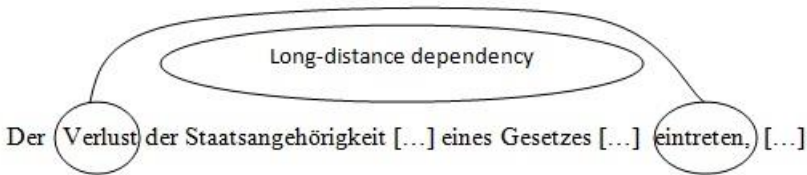


Figure 8: Long-distance dependency unigrams.

Το δικαίωμα [...] εγγυημένο.

Example 1: A legal phrase showing that high attention is required due to the long dependency between the words.

In this case the two unigrams will be simultaneously processed by Skip-Gram (see Figure 7), while the created vectors of the interdependent n-grams will result only from a context where such n-grams co-occur. The numerical representation of such co-occurring units will contain information of their shared context; the same will be done in the target language during training.

5.2. Self-attention with n-grams

Self-attention (Vaswani et al. 2017) is a sequence-to-sequence operation: a sequence of vectors goes in and a sequence of vectors

comes out. The input vectors are e.g. x_1, x_2, \dots, x_n and the corresponding output vectors are y_1, y_2, \dots, y_n with a d dimension (Peter Bloem, peterbloem.nl, article posted August 18, 2019). N-gram embeddings are the mathematical expression of phrases and single units (unigrams) (Jurafsky and Martin 2022).

The Self Attention mechanism (see Figure 9) is applied unaltered, the only difference being that it accepts as input n-grams embedded with the Skip-gram algorithm. The processing of words so as to become vectors before being inserted into the Self Attention mechanism is performed for both the source and the target language. The mechanism enriched with the novel interventions and the exact points of the latter are illustrated below.

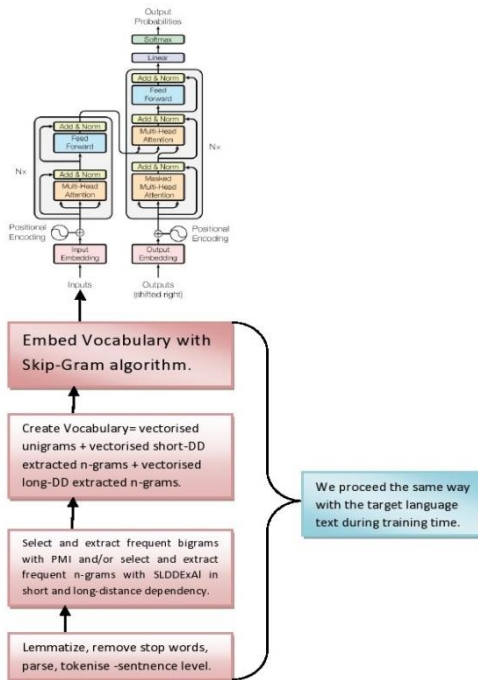


Figure 9: The proposed n-gram-extended Self Attention Mechanism (Vaswani et al. 2017) with Skip-gram, PMI (Bouma 2009) and SLDDExAI.

6. General implementation steps

After stating in detail the main points of the proposed mechanism and the new suggestions, we list below the implementation steps in their actual order, focusing on the proposed steps of the novel system. The main process of the system is the following:

- (1) POS corpus annotation, lemmatisation, tokenisation syntactic parsing and stop word list.
- (2) Creation of an algorithm for extracting short and long-distance dependencies (Short and Long-Distance Dependency Extraction Algorithm).
- (3) Selection of frequent n-grams (bigrams, trigrams and tetragrams) from the corpus by using the Pointwise Mutual Information for bigrams (PMI) (Bouma 2009) method (Moshe Hazoom, Towards Data Science, article posted December 22, 2018) and/or SLDDExAI with a high threshold. The system selects only nouns, verbs, adverbs, adjectives, participles and prepositions.
- (4) Limitation of selection to n-grams with sufficient frequency (Shang et al. 2018).
- (5) Embedment of all the unique words as well as the n-grams (consisting of more words in short and long-distance dependencies) from the vocabulary $|V|$ with one-hot representation, with a dimension of $[1, |V|]$.
- (6) Attribution of a single numerical representation to frequent lexical collocating units.
- (7) Simultaneous insertion of frequent lexical co-occurring n-grams (in short and long-distance dependency) into Skip-gram and training thereof in their shared context.
- (8) Creation of an n-gram vectorised vocabulary $|V|$.
- (9) Integration of the one-hot representations of the n-grams into the Skip-gram algorithm (Mikolov et al. 2013).
- (10) Use of the Skip-gram implementation from the Word2Vec model and of the Gensim library (Mikolov et al. 2013) to train embeddings.

The output of the Skip-gram consists of meaningful vectorised n-grams with only high frequency (Shang et al. 2018). The objective of the Skip-gram is to maximise $P(V_{\text{target}}|V_{\text{source}})$, the probability of V_{target} being predicted as V_{source} context for all training pairs in the corpus. The n-gram vectors are the input vectors for the Self Attention

Mechanism (Vaswani et al. 2017). This preparation is performed for both languages and the words of the source and target language (see Figure 9).

The main features of and requirements for our proposed model to be functional can be summarised as follows:

- (1) Training with legal domain parallel corpora for performance improvement.
- (2) Extension of the embedded words with features of legal content. This minimises ambiguity to the extent possible.
- (3) Reliance on large corpora of legal domain containing hundreds of thousands of documents to help deliver superior performance (Shang et al. 2018).
- (4) Phrase learning from an unsupervised text (Del, Tättar, and Fishel 2018).
- (5) Domain-independence (it can support multiple domains) (Shang et al. 2018).
- (6) Development of a single numeric representation for combining words (e.g. bigrams) (see Figure 4).
- (7) Data-driven approach.
- (8) Reward of frequent phrases, as frequency of the phrase occurrence is important; e.g., if “A B” is frequent, then “A B” is a phrase candidate.
- (9) Choice of high frequency since PMI (Bouma 2009) and SLDDExAl can reflect the frequency counts rather than the quality of the phrases. The assumption is that if it appears in the corpus frequently, then it is a quality multi-word term/appellation/phrase.
- (10) Careful choice of minimum threshold for the selected phrases (in a high rank) in order not to vectorise infrequent n-grams. The set limit for common phrases mainly depends on the size of the parallel Text Corpus and whether it is a domain corpus.
- (11) Support of any language (language agnostic system).

- (12) Improvement of the neural-based translation system capability by modelling both word and phrase (n-grams) (Del, Tättar, and Fishel 2018).
- (13) Learning of phrase embedding by minimisation of the semantic distance between translation equivalents and maximisation of the semantic distance between non-translation pairs (Zhang et al. 2014).
- (14) Introduction of the PMI (Bouma 2009) and SLDDExAl methods to generate phrase level memory in vector form.
- (15) Embedment of both the source and the target phrase with the same vectors, having the same dimension. The idea for phrase embedding has been picked up from LASER (GitHub, Language-Agnostic Sentence Representations, updated July 6, 2022), where sentence embedding takes place.
- (16) Re-embedment of the vectors for more meaning.
- (17) Use of the Self Attention Mechanism as presented in Vaswani et al. (2017).

7. Pros and cons of the proposal

In this article, which attempts to serve as a proposal for improving legal translation at both the lexical and the structural level, we pull existing methods and techniques together in a new way. Admittedly, the main obstacle to the implementation of this novel proposal is the lack of large Modern Greek special text corpora, let alone parallel ones. Anyway, some advantages of our proposal are the following:

- (1) It introduces LSDDExAl.
- (2) Skip-gram processes frequent co-occurring vectorised words simultaneously and trains them in their shared context.
- (3) It combines already existed methods and techniques in a novel way.
- (4) It can be trained for any domain.
- (5) It deals with long-distance dependencies.
- (6) It is language agnostic.
- (7) It focuses in particular on the correct translation of multi-word special phrases.
- (8) It contributes to overcoming the previously known errors in Neural Machine Translation.
- (9) It limits the post-editing errors

Beyond doubt, any model has also disadvantages, and the ones of our approach are probably the following:

- (1) There is no benchmark available.
- (2) The model requires more time and space for training in comparison with the original Self Attention Mechanism.
- (3) In order for the performance to be enhanced, a vast amount of data of legal content is needed, especially bilingual parallel legal corpora for the German-Greek language pair.

8. Future work

In our proposed approach the existing phrase-mining potential is complemented (since units connected meaning-wise but distant syntax-wise are extracted as well) and the Pointwise Mutual Information (PMI) method (Bouma 2009) is presented; this method ranks the extracted phrases by their term frequency. We also extend the Self attention Mechanism (Vaswani et al. 2017) with the Skip-gram algorithm (Mikolov et al. 2013) and the embedded vectors by adding domain specific (legal) features in order to eliminate semantic ambiguities. Our method requires availability of domain parallel corpora. Our baseline system follows principles of the Self Attention Mechanism (Vaswani et al. 2017) where we integrate n-gram vectors. The vectors for n-grams are learned as individual vocabulary entries. Yet, the effect of legal phrase embeddings is still to be investigated. The research direction is to increase the in-domain training data set and enrich the vector dimension with more information for legal concepts (domain specific features).

CONTRIBUTION OF AUTHORS: Both authors have contributed equally to the article.

CONFLICT OF INTEREST: The authors declare that there is no conflict of interest.

Bibliography

- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2016. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015*. arXiv:1409.0473v7 [cs.CL]. DOI: <https://doi.org/10.48550/arXiv.1409.0473>.
- Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5: 135–46. <https://aclanthology.org/Q17-1010.pdf> (accessed December 28, 2022).
- Bouma, Gerlof. 2009. Normalized (Pointwise) Mutual information in collocation extraction. In *From Form to Meaning: Processing Texts Automatically: Proceedings of the Biennial GSCCL Conference 2009*, eds. Christian Chiarcos, Richard Eckart de Castilho and Manfred Stede, 31–40. Tübingen: Gunter Narr.
- Camacho-Collados, José, and Mohammad Taher Pilehvar. 2018. From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research* 63: 743–88. DOI: <https://doi.org/10.1613/jair.1.11259>.
- Diniz da Costa, Alexandre, Mateus Coutinho Marim, Ely Edison da Silva Matos, and Tiago Timponi Torrent. 2022. Domain Adaptation in Neural Machine Translation using a Qualia-Enriched FrameNet. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, eds. Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk and Stelios Piperidis, 1–12. Paris: European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2022/LREC-2022.pdf> (accessed December 28, 2022).
- Duběda, Tomáš. 2021. Direction-asymmetric equivalence in legal translation. *Comparative Legilinguistics* 47: 57–72. DOI: <http://dx.doi.org/10.2478/cl-2021-0012>.

- Giampieri, Patrizia. 2018. The web as corpus and online corpora for legal translations. *Comparative Legilinguistics* 33: 35–55. DOI: <http://dx.doi.org/10.14746/cl.2018.33.2>.
- Goźdz-Roszkowski, Stanisław. 2021. Corpus linguistics in legal discourse. *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique* 34: 1515–1540. DOI: <https://doi.org/10.1007/s11196-021-09860-8>.
- Jurafsky, Daniel, and James H. Martin. 2022. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd edition draft). https://web.stanford.edu/~jurafsky/slp3/ed3book_jan122022.pdf (accessed December 28, 2022).
- Kalchbrenner, Nal, and Phil Blunsom. 2013. Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, USA, 18-21 October 2013 (EMNLP 2013)*, 1700–09. Stroudsburg: Association for Computational Linguistics. <https://aclanthology.org/D13-1176.pdf> (accessed December 28, 2022).
- Klemen, Matej, Luka Krsnik, and Marko Robnik-Šikonja. 2022. Enhancing deep neural networks with morphological information. *Natural Language Engineering* 28(3): 1–26. DOI: <https://doi.org/10.1017/S1351324922000080>.
- Krimpas, Panagiotis G. 2017a. Terminological preciseness or translational and legal effectiveness? Terminology of commodatum in no > el language pair. In *Konteksty súdneho prekladu a tlmočenia VI*, ed. Zuzana Guldánová, 66–84. Bratislava: Univerzita Komenského v Bratislave. https://fphil.uniba.sk/fileadmin/fif/katedry_pracoviska/kgn/transius/na_stiahnutie/Konteksty_sudneho_prekladu_a_tlmocenia_VI_2017.pdf (accessed December 28, 2022).
- Krimpas, Panagiotis G. 2017b. *Eisagōgē stī theōria tīs metafrāsīs* [Introduction to Translation Theory]. Athīna: Grīgorī.
- Krimpas, Panagiotis G. 2019. Pseudologioi typoi kai yperdiorthōsi stī Neoellīnikī Koinī me vasī ta epipeda glōssikīs analysīs [Pseudo-learned forms and hypercorrection in Standard Modern Greek on the basis of linguistic analysis levels]. In *Apo ton oiko sto spiti kai tanapalin... To logio epipedo stī sygchronī nea ellīnikī: Theōria, Istoría, Efarmogē* [From oikos

to *spiti* and vice versa: The learned register in Standard Modern Greek: Theory, History, Practice], eds. Asimakīs Fliatouras and Anna Anastasiadī-Symeōnidī, 57–126. Athīna: Patakī.

- Maksym Del, Andre Tättar, and Mark Fishel. 2018. Phrase-based unsupervised machine translation with compositional phrase embeddings. In *Proceedings of the Third Conference on Machine Translation (WMT), Volume 2: Shared Task Papers, Belgium, Brussels, October 31 - November 1, 2018*, 361–67. Stroudsburg: Association for Computational Linguistics. DOI: <https://doi.org/10.18653/v1/W18-64034>.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2014. Distributed representations of words and phrases and their compositionality. In *27th Annual Conference on Neural Information Processing Systems 2013, December 5-10, 2013 Lake Tahoe, Nevada, USA, Volume 1 of 4*, 3128–36. New York: Curran.
- Prieto Ramos, Fernando. 2014. Parameters for problem-solving in legal translation: Implications for legal lexicography and institutional terminology management. In *The Ashgate Handbook of Legal Translation*, eds. Le Cheng, King Kui Sin, and Anne Wagner, 121–134. Abingdon: Routledge.
- Shang, Jingbo, Jialu Liu, Meng Jiang, Xiang Ren, Clare R. Voss, Jiawei Han. 2018. Automated phrase mining from massive text corpora. *IEEE Transactions on Knowledge and Data Engineering* 30(10): 1825–1837. DOI: <https://doi.org/10.1109/TKDE.2018.2812203>.
- Tezcan, Arda, Véronique Hoste, and Lieve Macken. 2017. SCATE Taxonomy and Corpus of Machine Translation Errors. In *Trends in e-Tools and Resources for Translators and Interpreters*, eds. Gloria Corpas Pastor and Isabel Durán Muñoz, 219–248. Leiden: Brill. DOI: https://doi.org/10.1163/9789004351790_012.
- Tognini Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam and Philadelphia: John Benjamins.
- Valeontīs, Kōnstantinos E., and Panagiōtīs G. Krimpas. 2014. *Nomikī Glōssa, Nomikī Orologia: Theōria kai Praxī* [Legal Language, Legal Terminology: Theory and Practice]. Athīna: Nomikī Vivliothikī/Ellīnikī Etaireia Orogias.

- van Brussel, Laura, Arda Tezcan, and Lieve Macken. 2018. A Fine-grained Error Analysis of NMT, PBMT and RBMT Output for English-to-Dutch. In *Proceedings of the 11th Conference on Language Resources and Evaluation (LREC 2018)*, eds. Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asunción Moreno, Jan Odiijk, Stelios Piperidis and Takenobu Tokunaga, 3799–3804. Paris: European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2018/index.html> (accessed December 28, 2022).
- Van de Cruys, Tim. 2011. Two Multivariate Generalizations of Pointwise Mutual Information. In *Proceedings of the Workshop on Distributional Semantics and Compositionality (DiSCo'2011)*, 16–20. Stroudsburg: Association for Computational Linguistics.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. *Attention Is All You Need*. DOI: <https://doi.org/10.48550/arXiv.1706.03762>.
- Vaswani, Ashish, Samy Bengio, Eugene Brevdo, Francois Chollet, Aidan N. Gomez, Stephan Gouws, Llion Jones, Łukasz Kaiser, Nal Kalchbrenner, Niki Parmar, Ryan Sepassi, Noam Shazeer, and Jakob Uszkoreit. 2018. *Tensor2Tensor for neural machine translation*. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: MT Researchers' Track)*, eds. Colin Cherry and Graham Neubig, 193–99. Association for Machine Translation in the Americas. <https://aclanthology.org/W18-1819.pdf> (accessed December 28, 2022).
- Wiesmann, Ewa. 2019. Machine translation in the field of law: A study of the translation of Italian legal texts into German. *Comparative Legilinguistics* 37: 117–153. DOI: <http://dx.doi.org/10.14746/cl.2019.37.4>.
- Zhang, Jiajun, Shujie Liu, Mu Li, Ming Zhou, and Chengqing Zong. 2014. Bilingually-constrained Phrase Embeddings for Machine Translation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: Volume 1: Long Papers, ACL 2014, June 22–27, Baltimore,*

111–21. Stroudsburg: Association for Computational Linguistics. DOI: <https://doi.org/10.3115/v1/P14-1011>.

Websites and blogs

Bundesministerium der Justiz, Bundesamts für Justiz – www.gesetze-im-internet.de (accessed October 12, 2021).

Cambridge Spark. Tutorial: Build your own skip-gram embeddings and use them in a neural network. <https://blog.cambridgespark.com/tutorial-build-your-own-embedding-and-use-it-in-a-neural-network-e9cde4a81296> (accessed March 25, 2022).

Jay Alammar, [jalamar.github.io](https://github.com/jalammar). Visualizing machine learning one concept at a time. The Illustrated Transformer. <https://jalamar.github.io/illustrated-transformer/> (accessed March 25, 2022).

peterbloem.nl. Transformers from scratch. <https://peterbloem.nl/blog/transformers> (accessed March 25, 2022) (accessed March 25, 2022).

LASER Language-Agnostic SEntence Representations. GitHub - facebookresearch/LASER: Language-Agnostic SEntence Representations (accessed July 22, 2022).

Towards Data Science. Word2Vec for phrases learning embeddings for more than one word. <https://towardsdatascience.com/word2vec-for-phrases-learning-embeddings-for-more-than-one-word-727b6cf723cf> (accessed March 25, 2022).

Comparative Legilinguistics
vol. 2022/52
DOI: <http://dx.doi.org/10.14746/cl.52.2022.15>

Challenges in the translation of legal texts: the case in Kosovo

VALENTINA SOPJANI, Ph.D. cand.
University of Prishtina “Hasan Prishtina”
Prishtina/Kosovo 10 000
valentina.sopjani1@student.uni-pr.edu

ORCID: <https://orcid.org/0000-0002-0031-1648>

&

VJOSA HAMITI, Ph.D.
Associate Professor at the University of Prishtina “Hasan
Prishtina”
Prishtina/Kosovo 10 000
vjosa.hamiti@uni-pr.edu

ORCID: <https://orcid.org/0000-0003-2916-516X>

Abstract: Our paper conducts a contrastive analysis between Albanian and English lexical units in the language of laws using corpora analysis. It fills a literature gap related to corpus linguistics in order to better comprehend patterns of legal lexicon. We use KWIC Concordance to extract the top frequently used words and concordance lines in order to analyse differences and similarities between/among lexical units in Albanian (Source Language) and English (Target Language) laws, compared to those identified in respective corpora. Additionally, we identify types of errors in translation, difficulties in translating legal texts, and factors that influence translators' errors in translating certain laws of the Republic of Kosovo from Albanian into English. According to our analysis, both languages have in common the use of the same grammatical patterns such as: conjunctions, prepositions, and the use of the common noun forms. The rest have significant differences in their usage, especially with regards to particles and determiners. Our contrastive approach demonstrates that some of these laws were not originally written in Albanian and then translated into English language, as we assumed, considering that many articles within these laws lack the appropriate word structure and word order and in some cases are semantically ill-formed in the Source Language.

Key words: analysis; corpus; challenges; language; legal; translation; KWIC Concordance.

SFIDAT NË PËRKTHIMIN E TEKSTEVE LIGJORE: RASTI NË KOSOVË

Abstrakt: Punimi ynë ka për qëllim ta bëjë një analizë kontrastive midis njësive leksikore të shqipes dhe anglishtes në gjuhën e ligjeve duke përdorur analizën kontrastive. Gjithashtu, plotëson boshllëkun në literaturë në lidhje me linguistikën e korpusit për të kuptuar më mirë modelet e leksikut ligjor. Ne do të përdorim programin KWIC Concordance për të nxjerrë fjalët e përdorura më së shpeshti në mënyrë që të analizohen dallimet dhe ngjashmëritë midis njësive leksikore në gjuhën e ligjeve të shqipes (Gjuha burimore) dhe përkthimin në gjuhën angleze (Gjuha e synuar), krahasuar me ato të identifikuar në korpuset përkatëse. Gjithashtu, ne do identifikojmë llojet e gabimeve në përkthim, vështirësitë në përkthimin e teksteve ligjore dhe faktorët që ndikojnë në gabimet e përkthyesve gjatë përkthimit të ligjeve të caktuara të Republikës së Kosovës nga shqipja në anglisht. Sipas analizës sonë, të dyja gjuhët kanë të përbashkët përdorimin e të njëjtave forma gramatikore siç janë: lidhëzat, parafjalët dhe përdorimin e emrave të njëjtë të përdorur më së shpeshti. Pjesa tjetër kanë dallime të konsiderueshme në përdorimin e tyre, veçanërisht në

lidhje me nyjet dhe pjesëzat. Qasja jonë kontrastive tregon se disa nga këto ligje nuk janë shkruar fillimisht në gjuhën shqipe dhe më pas janë përkthyer në gjuhën angleze, siç kemi supozuar, pasi shumë nene brenda këtyre ligjeve nuk e kanë strukturën e duhur të fjalëve ose renditjen e fjalëve në fjali, si dhe në disa raste janë të shkruara pa kuptim në gjuhën burimore.

Fjalët kyçe: analizë; korpus; sfidat; gjuhë; juridike; përkthim; KWIC Concordance.

1. Introduction

The act of translation involves not only stating what has been reproduced from one language into another, but it also enables the communicative transfer of what has been produced in a certain language and culture. (Okyayuz 2019) Moreover, according to Kuleli (2019) the act of translation can on no condition be rationalised to find the target language equivalents of words and expressions in a source text. According to him the source text lends itself to various interpretations provided by different translators, therefore it is in the translation process that the cultural differences between two languages manifest themselves and require the translator to take on the role of the 'mediator' between two distinctive cultures. (Kuleli 2019: 1106–1107)

Translation has existed since the ancient times when there was a need to communicate with people who did not speak each other's mother tongue and there was also the need to exchange the goods among people of different countries (Stein 2018). There are various definitions about the term of translation, however some of them will be highlighted in this research. According to Venuti (1995) translation is "a process by which the chain of signifiers that constitutes the source-language text is replaced by a chain of signifiers in the target language which the translator provides on the strength of an interpretation". (Venuti 1995: 17)

Meanwhile, Nida and Taber (1982: 12) see translating as a process of reproducing in the receptor language the closest natural equivalent of the source language message, firstly in terms of meaning and secondly in terms of style. Moreover, Larson states that "translation

is transferring the meaning of the source language into the receptor language". (Larson 1984: 3)

However, legal translation is the translation of the language used in legal environments and for legal purposes. Nevertheless, translation of laws and bylaws must be accurate by preserving all the elements of the original message and using the syntactic and semantic forms of the translated language. This allows the parties to the proceedings to appear satisfied and unharmed, and to convey the message to the recipient without causing confusion and erroneous decision-making against the accused as a result of misinterpretation or bad translation. Moreover, taking into consideration the progress reports of the European Union and the general public in Kosovo we are aware of the quality of the translated documents into three languages. Even though English is not an official language in Kosovo, still every single material has always been translated into English language, considering the existence of a vital international factor in our country. Moreover, according to the Law on the Use of Languages in Kosovo¹:

“Every person has the right to freedom of expression. This right shall include freedom to receive, seek and impart information and ideas in the language of one’s choice without interference”.

People have had high expectations for positive interactions between different languages since the advent of translation and the need for translation of various documents (Pham, Nguyen and Pham 2022). They further argue that translation has always been manifesting itself through each stage of civilized development. Nevertheless, there are still certain obstacles to approach this subject.

Therefore, the objective of this research paper is to identify types of error in translation, difficulties in translating legal texts, and factors that influence translators’ errors in translating legal documents. Furthermore, the main purpose of this research is to investigate the top most common words frequency lists, concordance tables, by using electronic files and comparing them in both English and Albanian languages. The KWIC concordance tool will be used to analyse the most frequently used words of Kosovo laws, namely the Law on Courts,

¹ For more information, please see <https://gzk.rks-gov.net/ActDetail.aspx?ActID=2440> (accessed July 15, 2020).

the Law on the Kosovo Judicial Council, the Law on the Use of Official Languages, the Law on Protection from Discrimination, the Law on Protection of the Rights of the Child, the Law on the Kosovo Prosecutorial Council, the Law on Access to Public Documents, the Law on Civil Service, the Law on Personal Data Protection, and the Law on Organization and Functioning of State Administration and Independent Agencies.

Hence, first, a quantitative description of the Source Language and Target Language texts is provided, followed by most frequency lists and keyness. The tools of the KWIC Concordance computer program for linguistic data processing have been used to establish the source of the study, on the basis of which a corpus is formed. Consequently, such data can be used in various scientific studies (Perkhach and Shyika 2018). The numerical values of the Albanian and English corpora summarized in tables below will convey keywords and top high frequency lists, as shown on each table separately.

1.1. Literature review

The few studies that have looked at cross-cultural differences in community interpreting clearly show that there is no consistency in the way interpreters approach potential cross-cultural misunderstandings. In consequence, the results point to a need for greater guidance and clearer protocols for interpreters working in the legal system (Hale 2014).

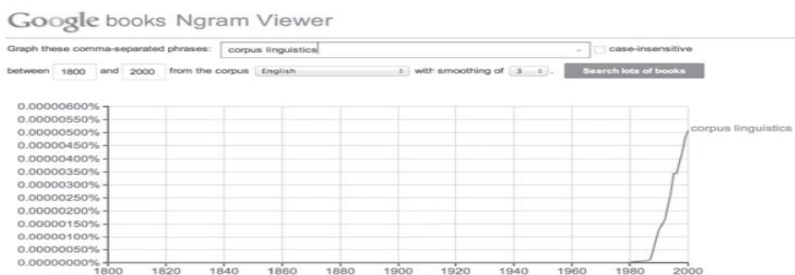
With the advance of corpus linguistics, its corpus is commonly used in language teaching and research, while its function of concordance and key word in context can promote the acquisition of lexical units (Chun-Guang 2014).

According to Hunston (2002), computational text analysis has become an exciting research field with many applications in communication² research. It can be a difficult method to apply, however, because it requires knowledge of various techniques, and the

² For more about the communication issue, see Koljada, Kowalskaia, and Melkonyan (2019).

software required to perform most of these techniques is not readily available in common statistical software packages (Hunston 2002). He further states that “The corpus investigation package Wordsmith Tools includes a program which automatically compares two corpora – usually a smaller, more specialised, one and a larger, more general, one – and lists the keywords for the more specialised corpus” (Hunston 2002: 68).

Below we will see how the term “corpus linguistics” has been published in literature over time.



Source: Generated using Google Ngram viewer: <https://books.google.com/ngrams> (cited from Vaughan, Elaine and O’Keeffe, Anne, 2015: 2)

However, teaching methodologies used nowadays in translator training programmes are mainly out of date, focusing on translation theories and trends only, without using current professional practices (Bolaños García-Escribano, Díaz Cintas and Massidda, 2018).

Byrne (2007) states that at the very heart of translation studies is the issue of translation quality. He argues that there are numerous methods for assessing the quality of translations, however there is no evidence nor research in our country about the consequences of producing a poor quality translation. He states that translation error, as a whole, can have significant consequences for both translator and client. This is analyzed by examining a number of case studies gathered from official reports and communications, court records, newspaper articles and books that illustrate the diversity of situations, which can arise as a result of translation errors. Consequently, the issue of liability and negligence can be used to illustrate the legal means by which translators can be held accountable for the quality of their work. He

further argues that by understanding how liability for faulty translations arises, it will be possible to see the implications of laws and instructions governing technical translations (Byrne 2007).

Additionally, different versions of translated texts may have a different effect on personality when taking their shape (Koljada, Kowalskaia and Melkonyan 2019). Similarly, Scott and John consider that sometimes online translation tools may generate different translations of the same original text as individual expressions or wording can be issued incoherently³.

On the other hand, Ralarala (2014) argues that Law and language are inextricably linked to each other. Consequently, sworn statements, taken from members of the public, initiate court processes, and their role culminates in court, as evidence for proceedings.

However, according to Koljada, Kowalskaia and Melkonyan (2019)⁴, “*The translation continues to be a powerful means of communication that influences the development of both the nation and the individual personality*” (Koljada, Kowalskaia and Melkonyan 2019: 77).

In addition, recent technological advances in the production of specialist audiovisual translation software and web-based applications have opened avenues for further changes and improvement in the ways the translation services are offered (Bolaños-García-Escribano, Díaz-Cintas and Massidda 2021).

1.2. The objectives of the study

This study aims at investigating the similarities and differences in corpora of the language of laws using a computerized text analysis by providing a detailed concordance lines, frequency lists and high frequency words.

This study seeks answers to the following questions:

³ For more about translation, see Scott and John (2022).

⁴ More about translation issue see Baker and Saldanha (2001).

1. What are the major challenges of translating legal texts in Kosovo?
2. What are the most common frequency used words in legislative texts' corpora?

The following hypotheses have been expressed:

1. Albanian to English translators face difficulties in translating laws from Albanian into English.
2. Translation of laws in Target Language (English) is better than the original text in the Source Language (Albanian) according to our selected corpus.

2. The KWIC concordance

The KWIC Concordance is a corpus systematic instrument for doing comparative word frequency lists, concordances and collocation tables by using electronic files, and is a central tool for conducting text analysis⁵. At the time when KWIC was developed, computer screens had a limited number of characters that could be produced on each line and using them was basically an experience of looking at text on screen in a form which was very different from other reading materials. This situation has dramatically changed today considering the development of technologies and the internet in general including modern apps and applications (Jeaco 2017: 25).

According to Hunston (2002), a concordance is a list of target words extracted from a given text or set of texts, often presented in such a way as to indicate the context in which the word is being used. He states that this format of presenting information is called 'KWIC': Key Word in Context arguing that concordance software can usually extract and present other types of information too, e.g. identifying the words that most commonly appear near a target word.

As noted by Yunus and Awab (2012), computer technology has revolutionised English Language Teaching and Learning with the advent of Corpus Linguistics. Yet, the application of Corpus Linguistics in classrooms, in the so-called Corpus (Data)-Driven Learning (DDL)

⁵ More about KWIC Concordance see O'Donnell, Brook Matthew (2008).

approach, makes use of ‘real-life’ concordance data and the concordance tools train learners to apply concordancing. He further states that it also trains learners to originate linguistic rules and meanings based on observations of repetitive words or collocation patterns as the KWIC in the concordance lines (Yunus and Awab 2012).

On the other hand, Jeaco (2017) argues that corpus tools provide several different ways to display relationships between words within texts and across texts, while the main format for viewing concordance data is the Key Word in Context (KWIC). He states that in Computer Aided Language Learning, concordance lines in the KWIC format may be accessed inside a concordancer or within other software in the concordance line horizontally across the screen with the number of words or letters to the right and left (Jeaco 2017: 22–23).

3. Method

This research presents a corpus-based approach to discourse analysis that starts with a detailed analysis of ten selected laws in a corpus that can be generalized across all texts of a corpus, providing a list of 30 top frequently used words in four columns in both Albanian and English languages, including keywords and frequency lists.

Corpus is taken from the website of the Official Gazette of the Republic of Kosovo in both languages, that is in the Source language (Albanian) and the Target language (English). Therefore, ten laws are selected to complete this research, as mentioned before, through KWIC Concordance tool. Data collection is analysed in detail by comparing their differences and similarities as well as errors in translation.

4. Results

The results from the KWIC Concordance tool is presented on the tables listed below. The results are presented for the Law on Courts, the Law on the Kosovo Judicial Council, the Law on the Use of Official

Languages, the Law on Protection from Discrimination, the Law on Child Protection, the Law on the Kosovo Prosecutorial Council, the Law on Access to Public Documents, the Law on Civil Service, the Law on Personal Data Protection, and the Law on Organization and Functioning of State Administration and Independent Agencies, including key words the most frequently used words.

Table 1 shows an example of a word list created using KWIC Concordance tools which is the most common format for concordance lines.

Table 1. Law on the Use of Official Languages in Kosovo

Key word	Freq.	Key word	Freq.
the	342	të	422
of	221	në	236
and	164	e	226
in	163	dhe	181
official	108	zyrtare	102
language	99	do	63
to	98	i	62
languages	79	për	62
shall	62	nga	56
by	48	gjuhëve	46
article	47	me	46
their	44	që	43
a	41	tyre	41
or	39	neni	40
use	39	gjuhën	37
be	38	nuk	33
institutions	38	si	32
as	36	është	31
Kosovo	36	apo	30
an	35	institucionet	27
is	35	amtare	26
any	33	Kosovës	26
law	33	së	26
with	33	një	23
public	32	publike	23
right	31	ligji	22
not	30	ka	21

have	27	këtij	21
this	27	drejtë	20
proceedings	26	person	20

As we can see from table 2, most of the single-word items are functional, or grammatical operators like determiners (the: 945 occurrences), prepositions (of: 676 occurrences, to: 400 occurrences etc.):

Table 2. Law on the Civil Service of the Republic of Kosovo

Key word	Freq.	Key word	Freq.
the	945	të	1288
of	676	e	659
to	400	në	538
and	384	dhe	368
civil	381	civil	366
in	245	për	359
for	212	me	231
servants	204	i	226
shall	199	nga	162
be	191	nëpunësit	154
a	138	së	121
or	136	neni	112
service	121	ose	112
article	116	një	106
by	111	punës	101
with	97	do	86
law	94	që	84
on	88	nuk	69
are	87	janë	67
is	83	shërbimin	67
public	83	tyre	67
their	83	nëpunësve	63
this	80	mund	61
from	79	publike	60
procedures	72	procedurat	59
not	71	eshtë	52
disciplinary	69	kosovës	47
that	69	disiplinore	43

servant	65	ligj	43
positions	63	këtij	42
administrative	61	kanë	41

Table 3 presents the top 30 items in the corpus of the Law on Protection from Discrimination.

Table 3. Law on Protection from Discrimination

Key word	Freq.	Key word	Freq.
the	312	të	452
of	301	e	259
to	141	në	198
and	134	për	133
in	111	dhe	132
or	93	i	89
on	92	me	80
discrimination	86	ligji	64
law	83	një	64
a	65	apo	52
this	61	këtij	51
article	56	nga	49
for	55	që	48
be	43	së	46
is	39	nr	42
with	38	ose	42
shall	38	neni	32
no	37	mund	29
from	31	diskriminimit	26
by	30	kosovës	26
which	27	është	25
person	26	çdo	24
republic	26	ka	24
treatment	26	republikës	24
may	25	barabartë	22
any	24	mbi	21
equal	24	mbrojtjen	21
grounds	23	nenin	20
who	22	sipas	20
court	21	bazat	17

A look at the frequency list in table 4 shows us the ranking order for almost the same top frequently used words between both languages (English and Albanian). One method for doing this is to normalize the frequency figures and that is achieved by using the following calculation:

- $nf = (\text{number of examples of the word in the whole corpus} \div \text{size of the corpus})$
- $\times (\text{base of normalization}) (1)$ (Vaughan, Elaine and O’Keeffe, Anne, 2015: 7)

Table 4. Law on Courts in Kosovo

Key word	Freq.	Key word	Freq.
the	701	të	472
of	537	e	320
court	192	në	249
and	144	për	186
shall	132	dhe	146
in	124	me	127
for	112	gjykata	125
to	97	i	116
Kosovo	90	së	107
law	89	kosovës	79
a	80	themelore	77
basic	79	neni	51
by	72	republikës	48
judges	59	nga	46
be	58	nr	44
article	57	ligji	40
courts	57	siç	38
on	57	çdo	36
republic	56	që	30
no	52	gjyqtarët	29
president	50	një	29
department	48	komunë	27
judge	47	prishtinë	27
with	47	dega	26
as	45	ka	25
branch	42	kryetari	25
cases	42	supreme	25
council	38	apelit	24

have	33	departamenti	24
or	31	gjyqësor	24

Table 5 shows which words are occurring most frequently in the language of laws and it results that both languages have in common the use of the same grammatical forms such as: determiner: the (1978 occurrences), prepositions such as: of (559 occurrences in English and 2867 occurrences in Albanian), to (806 occurrences in English and 778 in Albanian), conjunction such as: and (591 occurrences in English and 590 in Albanian language).

Table 5. Law on Protection of Personal Data

Key word	Freq.	Key word	Freq.
the	1978	të	2867
of	559	e	1265
to	806	në	778
data	743	për	702
and	591	dhe	590
or	544	dhënavë	581
in	500	i	540
a	404	me	502
this	367	ose	419
personal	364	personale	363
article	357	këtitj	312
shall	330	një	278
for	321	nga	268
law	315	që	254
be	258	neni	246
on	242	ligji	232
processing	242	së	198
controller	205	nëse	181
paragraph	185	mbrojtjen	140
by	181	është	127
protection	179	kontrolluesi	125
with	163	nuk	124
subject	160	dhënat	122
is	159	nr	112
agency	131	mund	108
an	131	sipas	105
no	127	paragrafi	96

from	124	përpunimit	92
that	123	çdo	89
as	113	kontrolluesit	86

Table 6 comparing with the corpus in Albanian language, we see that many of the noun forms in this table are common with English language, for example: *article* (alb. *neni*), *rights* (alb. *të drejtat*), *law* (alb. *ligji*), *protection* (alb. *mbrojtjen*), *child* (alb. *fëmijës*), *relevant* (alb. *përkatëse*), *social* (alb. *sociale*). However, in the number of generated keywords, we see high frequency of the preposition: *of* (with 559 occurrences in English and 2867 occurrences in Albanian language) *to* (with 806 occurrences in English and 778 in Albanian) and the article *the* (with 1978 occurrences) and the conjunction: *and* (838 occurrences in English language and 836 in Albanian).

According to Baker, a concordance view entails the analyst to have firstly identified, more probably through frequency lists in terms of which item they would like to investigate, and therefore the frequency list is most of the time the first entry point into a data set (Baker 2006).

Table 6. Law on Child Protection

Key word	Freq.	Key word	Freq.
the	1620	të	1604
of	1001	e	1137
and	838	dhe	836
child	605	për	681
to	468	në	613
in	401	fëmijës	517
or	321	me	316
for	277	i	276
protection	233	që	219
a	202	së	197
with	176	ose	187
shall	171	nga	167
by	151	mbrojtjen	161
on	151	si	115
is	150	apo	103
be	139	përkatëse	100
law	118	duhet	92

rights	108	është	91
as	104	nr	83
that	103	ligji	81
relevant	102	drejtat	80
other	100	një	80
children	99	neni	74
this	91	çdo	70
no	86	kosovës	65
article	84	fuqi	60
any	74	gjitha	60
are	73	sociale	60
social	73	nuk	58
all	70	republikës	57

Table 7 shows that there are some slight differences in both numbers of the keywords generated where the determiner: *the* is the most frequently used word, followed by preposition *of* and conjunction *and*.

Table 7. Law on Kosovo Judicial Council

Key word	Freq.	Key word	Freq.
the	888	të	584
of	485	e	326
council	236	dhe	202
and	199	për	201
for	126	në	196
to	119	i	154
a	117	me	125
in	115	këshilli	120
shall	100	kosovës	77
kosovo	85	nga	75
on	85	së	74
by	75	nr	59
judges	72	një	56
law	69	neni	55
members	68	gjyqësor	50
be	66	gjyqtarëve	49
court	62	që	49
or	62	apo	44
no	61	si	39
article	60	gjykatave	36

with	55	ligji	35
judicial	54	do	34
president	50	republikës	33
as	44	nuk	30
republic	37	mund	29
chair	35	zyrtare	29
this	34	anëtar	28
assembly	29	gazeta	26
official	29	është	25
gazette	26	anëtarët	24

Table 8 presents the top 30 items in the corpus of the Law on Kosovo Prosecutorial Council.

Table 8. Law on Kosovo Prosecutorial Council

Key word	Freq.	Key word	Freq.
the	794	të	575
of	443	e	342
council	214	dhe	222
and	209	për	198
in	111	në	175
for	109	i	134
to	100	me	129
a	98	këshillit	115
shall	93	nga	85
on	77	së	84
prosecutors	77	kosovës	65
law	73	një	62
by	71	që	53
Kosovo	70	prokurorëve	51
prosecution	60	neni	47
members	57	nr	47
article	54	si	35
be	54	apo	34
office	49	shtetit	34
no	48	ligji	33
with	47	prokurorial	31
or	44	republikës	31
as	41	anëtar	29
chair	40	do	27

state	37	mund	25
prosecutorial	36	nuk	25
performance	34	performances	25
republic	33	zyrtare	23
this	33	gazeta	21
member	32	ështëë	19

Table 9. Law on Organization and Functioning of State Administration and Independent Agencies

Key word	Freq.	Key word	Freq.
of	576	të	600
the	453	e	472
and	277	dhe	285
for	141	për	219
agency	140	i	200
in	123	në	176
to	116	me	96
law	105	së	93
by	88	nga	84
agencies	87	neni	80
article	87	një	74
is	85	kosovës	68
on	85	apo	65
or	82	agjencie	64
administration	81	shtetërore	60
kosovo	74	agjencisë	59
performance	71	performances	59
a	69	agjencia	48
independent	66	nr	48
state	62	që	48
responsible	61	ligji	46
organization	59	pavarura	46
shall	59	administrates	44
public	58	ekzekutive	39
with	55	organizimin	38
this	54	ështëë	38
minister	51	këtij	37
executive	49	rregullatore	35
no	49	përgjegjës	34
be	48	republikës	34

The results in tables 9 and 10 have in common several nouns used in both languages such as: article (*alb. neni*), agency (*alb. agjencia*), public (*alb. publike*), independent (*alb. pavarura*) and the use of the same grammatical forms such as determiner: *the*, prepositions: *of, to, in, for* and the conjunction: *and*.

Table 10. Law on Access to Public Documents

Key word	Freq.	Key word	Freq.
the	586	të	580
of	338	e	326
public	331	në	301
to	268	për	266
and	187	publike	202
documents	154	dhe	193
shall	135	publik	166
for	130	me	107
in	124	i	102
access	121	dokumente	99
a	96	qasje	92
or	94	së	69
law	89	nga	68
on	84	që	63
by	73	apo	61
institution	73	ligji	51
this	70	institucionet	46
be	68	një	45
article	56	neni	44
no	42	këtij	41
agency	40	nr	39
with	38	kosovës	35
data	37	mund	32
request	37	ose	32
right	36	ka	30
applicant	35	nuk	30
official	35	republikës	30
kosovo	34	ështëë	29
which	34	agjencia	27
as	32	si	26

5. Discussion

This research paper examined concordancing in a group of laws in a bid to investigate the most frequently used words. The study introduced concordance-based contrastive analysis⁶ of legal lexicon of original texts in Albanian language and its translation in the English language, examining common items and patterns. All these findings point to important gains and represent a preliminary step in the development of corpus-based contrastive analysis.

Regarding the first research question about challenges that Albanian to English translators face when translating laws from Albanian into English language, it is worth mentioning that the key challenge is that some of these laws were not originally written in Albanian and then translated into English, as many articles within these laws lack the appropriate word structure or word order and in some cases are semantically ill-formed. Another key challenge has been lack of legal dictionaries from English to Albanian and vice versa and lack of translator's training in legal terminology in Kosovo. Moreover, lack of professional translators resulted in translation errors and poor translation of legal texts. It was also emphasized on ASI Newsletter⁷ that the majority of qualified English to Albanian translators and interpreters were working for International Organizations at the time of drafting the legislation in Kosovo, and subsequently, many draft laws were either being written in English or were strongly influenced by English speakers, such as international experts who were on the draft law working groups, and they needed to be translated into Albanian and Serbian which consequently resulted in translation errors. This can be confirmed by the fact that certain parts of these laws in Albanian have used inappropriate or meaningless words as follows: "*Linguistic Freedom has been translated*" as "*Liria linguistike*"; "*Transitional and Final Provisions*" as "*Dispozitat kalimtare dhe të fundit*" (*Law on the Use of Official Languages*)⁸; "*Other justified treatments*" as "*Trajtimi i ndryshëm i përligjur*" (*Law on Protection from*

⁶ More about text analysis see Welbers, Van Atteveldt and Benoit 2017.

⁷ For more information, please consult <https://www.osce.org/files/f/documents/1/d/15261.pdf> (accessed June 24, 2022)

⁸ <https://gzk.rks-gov.net/ActDetail.aspx?ActID=2440> (accessed June 30, 2022)

Discrimination); “*Submission of an appeal to the...*” as “*Paraqitja e ankimit tek...*”, “*Correction of legal violations*” as “*Ndreqja e shkeljeve ligjore*”, “*The detailed competencies*” as “*Kompetencat detaje*”, “*The Provisional Institutions of Self-Government are obliged to adopt detailed regulations...*” as “*Institucionet e Përkohshme të Vetëqeverisjes janë të obliguara të miratojnë rregullore të hollësishme*”, “*From the day the damaged party becomes aware of the violation*”, as “*Nga dita që i dëmtuari merr dijeni për këtë shkelje...*”.

Meanwhile, regarding the second research question about the most common frequency used words from Albanian into English language, it is concluded that both languages had in common the use of the same parts of speech such as conjunctions and prepositions, as well as the use of the same top frequent nouns, as indicated above. The rest had significant differences in their usage, especially with regard to particles and articles.

Translation of words is analyzed in terms of the most frequent word lists in both languages and they were interpreted by comparing the likelihood value if there was a significant difference in their usage or not. The results illustrated that the frequency lists per 30 words of the most prominent occurrences in the ten selected laws of the Republic of Kosovo has resulted in the usage of the following grammatical forms (Law no. 06/L-084 on Child Protection) in English laws: prepositions “*to, of, in, on, with, by, from, with*”, conjunctions “*and, or, for*”, articles (both definite and indefinite), demonstrative pronouns “*this, that*”, determiners “*any, all*”, nouns “*article, Kosovo, official, person, Republic, law, person, Council, judge, member*”, modal verbs “*shall, may*”, abbreviation “*no*”, particle “*not*” and the verb “*to be*” (Law no. 03/L-224 on the Kosovo Prosecutorial Council).

Whereas, in Albanian the most frequently used words were: particles, articles, short form of possessive pronouns, prepositions: “*në*” (in), “*me*” (with), “*nga*” (from), “*mbi*” (on), “*për*” (for); nouns: “*shteti*” (state), “*Kosova*” (Kosovo), “*gjyqtar*” (judge), “*dega*” (branch), “*kryetar*” (President), “*Këshilli*” (Council), “*gjykata*” (court), “*neni*”(article), “*ligji*” (law), “*republika*” (Republic), “*anëtar*” (member), “*zyrtar*” (officer), “*publike*” (public), “*institucion*” (Institution), “*agjencia*” (Agency); auxiliary verb “*ka*”

(have), modal verb “*mund*”⁹, abbreviation “*nr*”(no.), indefinite pronoun “*çdo*” (every) and conjunctions: “*dhe*” (and), “*edhe*” (also), “*ose*” (or), “*siç*” (as), “*apo* (or) (Law no. 06/L-055 on Kosovo Judicial Council).

Therefore, depending on the context of the sentences, sometimes the same word appeared differently. For instance, we found the word “*të*” (of) as the most frequent used word in Albanian (Law no.05/L-021 On the Protection from Discrimination) which sometimes appeared as a genitive case article when used before nouns for instance: “*të shtetit*” (of state), “*të prokurorit*” (of prosecutor), when it appeared before the adjectives or as a petrified form of the verb. The same occurred with the other most frequent preposed articles¹⁰ “*i*” and “*e*”, when they were used before adjectives, for instance “*e barabartë....*” (equal to...), “*...e përcaktuar*” (defined by...), and when used before verbs as short form of possessive pronouns for instance: “*i propozon...*” (proposes), “*i siguron...*” (ensures), “*i plotëson ...*” (fulfils), and when used before nouns as a genitive case article for example: “*e prokurorëve....*” (of prosecutors), “*e politikave....*” (of policies), “*i ligjit...*” (of law), “*e prokurorive....*” (of prosecution offices), “*e rregullave....*” (of rules) etc.

Finally, the conclusion is that the main difficulties¹¹ were lack of vocabulary, grammatical issues, word order, then translating word for word and most translators lack a robust background on the content of the text.

⁹ The modal verb in Albanian “*mund*” is equivalent with “*can, could, may, might, be able to*” in English language.

¹⁰ Newmark et al. (1982: 179):” Adjectival articles are proclitics that indicate that the following adjective, noun, number, or pronoun is attributed to a noun. These preposed articles are connected historically with postposed articles which lost their independence and turned into definite case endings. Note the similarity of form of the proclitics and the definite case endings for the same word in the following examples: **I mir-I** ‘the good one (masc.)’, **TË mirë-T** ‘the good ones (masc.)’, **SË mirë-S** ‘to the good one (fern.)’ **TË miri-T** ‘to the good one (masc.)’. In the nominative case adjectival articles have three forms: *i* for the masculine singular, *e* for the feminine singular, **të** or **e** for the neuter singular as well as for the plural of all three genders.”

¹¹ For more about difficulties on translation, see Pham, T. A., Nguyen, L. T. D., Pham, V. T. T. (2022). English language students’ perspectives on the difficulties in translation courses: Implication for language education. *Journal of Language and Linguistic Studies*.

6. Research limitations

This present research paper demonstrates only a preliminary step in investigating major translation problems in translating legislative texts in our country, Kosovo. This could be explained, on the one hand, by the scarcity of research in this field and, on the other hand, by lack of a standardized legal dictionary, difficulties resulting from inadequate technical facilities, educational background of the translators, the subjective approach of the translators for correct interpretation, as well as lack of a unique software that would include the legal terms and expressions and be comprehensive, enabling the translators to acquire the legal terminology and apply it in practice on a daily basis.

7. Practical implications

Translation as a process is a very important field of study, thus a good and faithful translator is a decent asset. One of the possible limitations of this research paper may be the authors' inability to review and consult all documents, laws, and other legal acts, which would provide a more accurate overview of the findings of translations in legal terminology. Another obstacle would be that not all acts are translated into more than two languages. Therefore, it would be worthy to conduct a research in the future that would scientifically deal with the terminological diversity of the implementation of the translation of laws and by-laws of all minority communities' languages in Kosovo (Serbian, Turkish, Bosnian, Roma and Ashkali).

As Newmark (1981) states:

“Legal documents also require a special type of translation, basically because the translator is more restricted than in any other form. Every word has to be rendered, differences in terminology and function noted and as much attention paid to the content as to the intention and all possible interpretations and misinterpretations of the text – all legal texts are definitions.” (Newmark 1981: 47)

8. Originality

Taking into account that Kosovo is a multi-ethnic and multicultural state and considering that according to the Constitution of the Republic of Kosovo, more specifically the Law on the Use of Languages in Kosovo, everyone has the right to express themselves in their native language, the translation process is even more complicated but at the same time responsible for a unified legal terminology, where everyone would be able to be served equally and comprehensively in their native language.

Legal translation is considered as one of the most complex types of translation (Newmark 1981) considering its importance to a fair trial and rendering the right decision towards all parties in the proceedings. Therefore, the right to an equal access to justice can only be realized by overcoming and improving the linguistic barriers to all citizens.

9. Recommendations

To sum up, some of the recommendations of this research would include:

1. Drafting a contemporary legal dictionary from Albanian into English and vice versa;
2. Standardization of legal terminology among all justice institutions;
3. Establishment of a unique software to all translators using the latest technological translation tools;
4. Proofreading of all legislative texts by a professional legal translators' working group, including lawyers, on the basis of the standardized dictionary, correct legal terminology, an applicable and useful translation software, that would contribute to the complete harmonization of legal terminology from one language to another.

10. Conclusions

Translators face difficulties in translating laws from Albanian into English. This is in keeping with our hypothesis that linguistic barriers, lack of a standardized legal dictionary, difficulties resulting from inadequate technical facilities to handle translation, inadequate educational background of the translators, work experience, as well as the subjective approach of the translators for correct interpretation, all these factors have led to the legal texts in the English language being more comprehensive and understandable than those in the Albanian language. In addition, lack of a standardized legal terms dictionary and a software that would include the terms and be comprehensible to all enabling language experts to acquire the legal terminology and apply it in practice have also resulted in difficulties in translating laws from Albanian into English. A joint software that translates everything in an intelligent device that makes a good translation, whether from Albanian to English language or vice versa, is lacking.

It is worth mentioning that the English language has some legal expressions that are found and comprehensive in comparison to Albanian, which does not consist of the standardized legal terminology and as a consequence this might have affected the improper translation of legal texts.

The second hypothesis, that Translation of laws in Target Language (English) is better than the original text in the Source Language (Albanian) according to our selected corpus, has been proven, namely that most of the laws mentioned above were much better written in the Target (English Language) than in the Source language (Albanian Language). Moreover, there is little or almost no research in our country about types of errors in the translation of laws, difficulties in translating legal terminology and factors contributing to errors in translation.

Therefore, given that the laws in force are already being implemented and amended in their current form and wording, when the legislation of the Government provides that any law, which contains substantial errors, be subjected to the amendment procedure, a comprehensive review of the laws should take place, so that they are compatible with expressions in the English language. Experts based on

a contemporary legal dictionary, which would be unique in the regions where the Albanian language is spoken, should do this review. This would make it possible to avoid the use of different expressions in the areas where the Albanian language is the Source Language (Kosovo, Albania, North Macedonia, South Serbia and Montenegro).

CONTRIBUTION OF AUTHORS: The authors worked collaboratively on conceiving and designing the analyses, collecting data, selecting the analysis tools, and writing and proofreading the paper.

CONFLICT OF INTEREST: The authors declare that there is no conflict of interest.

Bibliography

- Baker, Mona, and Gabriela Saldanha. 2001. *Routledge Encyclopedia of Translation Studies*. London: Routledge.
- Baker, Paul. 2006. *Using corpora in discourse analysis*. London: Continuum.
- Bolaños-García-Escribano, Alejandro, Jorge Díaz-Cintas, and Serenella Massidda. 2021. Latest advancements in audiovisual translation education. *The Interpreter and Translator Trainer* 15(1): 1–12. DOI: <https://doi.org/10.1080/1750399X.2021.1880308>.
- Bolaños-García-Escribano, Alejandro, Jorge Díaz-Cintas, and Serenella Massidda. 2018. Recent trends in translator training – audiovisual translation and applied technologies at stake. *The Interpreter and Translator Trainer* 12(2): 254–255. DOI: <https://doi.org/10.1080/1750399X.2018.1478476>.
- Byrne, Jody. 2007. Caveat translator: Understanding the legal consequences of errors in professional translation. *The Journal of Specialised Translation* 7: 2–23.
- Chun-Guang, Tian. 2014. An Empirical Research on the Corpus-Driven Lexical Chunks Instruction. *International Journal of English Language Teaching* 2(2): 1–36.

- Hale, Sandra. 2014. Interpreting culture. Dealing with cross-cultural issues in court interpreting. *Perspectives Studies in Translatology* 22(3): 321–331. DOI: <https://doi.org/10.1080/0907676X.2013.827226>.
- Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: University of Birmingham.
- Jeaco, Stephen. 2017. Helping Language Learners Put Concordance Data in Context. *International Journal of Computer-Assisted Language Learning and Teaching* 7(2): 22–39.
- Koljada, Natalija, Svetlana Kowalskaia, and Ashhen Melkonyan. 2019. Translation as a means of interpersonal and interethnic communication. *XLinguae* 12(2): 72–79.
- Kuleli, Mesut. 2019. Identification of translation procedures for culture specific items in a short story. *Journal of Language and Linguistic Studies* 15(3): 1105–1121. DOI: <https://doi.org/10.17263/jlls.631551>.
- Larson, Mildred L. 1984. *Meaning-based translation: A guide to cross-language equivalence*. New York, Oxford: University Press of America.
- Newmark, Leonard, Philip Hubbard, and Peter Prifti. 1982. *Standard Albanian*. Stanford, California: Stanford University Print.
- Newmark, Peter. 1981. *Approaches to Translation*. London: Pergamon.
- Nida, Eugene A., and Charles Taber. 1974. *The Theory and Practice of Translation*. Leiden: E.J. Brill.
- Okayayuz, Ayşe Şirin. 2019. Translation and Language and Cultural Policies: The Importance of Political Cognizance in Audiovisual Translator Training. *Journal of Language and Linguistic Studies* 15(3): 937–952.
- O'Donnell, Brook Matthew. 2008. KWICgrouper – Designing a tool for corpus-driven concordance analysis. *International Journal of English language* 8 (1): 107–121.
- Perkhach, Roksolana-Yustyna, and Yuliia Shyika. 2018. The Methodology of Frequency Dictionaries to the Instructions to Medical Products. In *IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*: 26–29. DOI: <https://doi.org/10.1109/STC-CSIT.2018.8526600>.

- Pham, Anh Tuan, Linh Thi Diem Nguyen, and Vi Thi Tuong Pham. 2022. English language students' perspectives on the difficulties in translation courses: Implication for language education. *Journal of Language and Linguistic Studies* 18(1): 180-189.
- Ralarala, Monwabisi K. 2014. Transpreters' translations of complainants' narratives as evidence: Whose version goes to court? *The Translator* 20(3): 377–395. DOI: <https://doi.org/10.1080/13556509.2014.934002>.
- Scott, Juliette, and John O'Shea. 2022. Translation in Libel cases: Reputations at Stake! *Comparative Legilinguistics* 50: 123–179. DOI: <https://doi.org/10.14746/cl.50.2022.7>.
- Stein, Daniel. 2018. Machine translation. Past, present and future. In *Language technologies for a multilingual Europe: TC3 III*, eds. Georg Rehm, Felix Sasaki, Daniel Stein and Andreas Witt, 5–17. Berlin: Language Science Press.
- Yunus, Kamariah, and Su'ad Awab. 2012. The Effects of the Use of Module-Based Concordance Materials and Data-Driven Learning (DDL) Approach in Enhancing the Knowledge of Collocations of Prepositions among Malaysian Undergraduate Law Students. *The International Journal of Learning: Annual Review* 18(9): 181–198. DOI: <https://doi.org/10.18848/1447-9494/CGP/v18i09/47758>.
- Vaughan, Elaine and Anne O'Keeffe. 2015. Corpus Analysis. In *The International Encyclopedia of Language and Social Interaction*, eds. Karen Tracy, Cornelia Ilie and Todd Sandel, 252–268. Denver, CO: John Wiley and Sons, Inc.
- Venuti, Lawrence. 1995. *The Translator's Invisibility. A History of Translation*. London: Routledge.
- Welbers, Kasper, Wouter Van Atteveldt, and Kenneth Benoit. 2017. Text Analysis in R. *Communication Methods and Measures* 11(4): 245–265. DOI: <https://doi.org/10.1080/19312458.2017.1387238>.

Internet Resources

- ASI (Assembly Support Initiative Newsletter). OSCE Mission in Kosovo. February 2005, No. 15, Special Edition. <https://www.osce.org/files/f/documents/1/d/15261.pdf> (accessed June 27, 2022).
- Law no. 05/L-021 on Protection from Discrimination. Official Gazette of the Republic of Kosova. <https://gzk.rks-gov.net/ActDetail.aspx?ActID=10924> (accessed July 14, 2022).
- Law no. 02/L-037 on the Use of Official Languages. Official Gazette of the Republic of Kosova. <https://gzk.rks-gov.net/ActDetail.aspx?ActID=2440> (accessed July 15, 2020).
- Law no. 06/L-084 on Child Protection. Official Gazette of the Republic of Kosova. <https://gzk.rks-gov.net/ActDocumentDetail.aspx?ActID=20844> (accessed June 25, 2022).
- Law no. 03/L-224 on the Kosovo Prosecutorial Council. Official Gazette of the Republic of Kosova. <https://gzk.rks-gov.net/ActDetail.aspx?ActID=2709> (accessed July 14, 2022).
- Law no. 06/L-055 on Kosovo Judicial Council. Official Gazette of the Republic of Kosova. <https://gzk.rks-gov.net/ActDetail.aspx?ActID=18335> (accessed June 28, 2022).
- Law no. 05/L-021 on the Protection from Discrimination. Official Gazette of the Republic of Kosova. <https://gzk.rks-gov.net/ActDetail.aspx?ActID=10924> (accessed August 12, 2022).
- Law no. 06/L-081 on Access to Public Documents. Official Gazette of the Republic of Kosova, <https://gzk.rks-gov.net/ActDetail.aspx?ActID=20505> (accessed August 28, 2022).
- Law no. 03/L-149 on Civil Service. Official Gazette of the Republic of Kosova. <https://gzk.rks-gov.net/ActDetail.aspx?ActID=2679> (accessed August 29, 2022).
- Law no. 06/L-082 on Personal Data Protection. Official Gazette of the Republic of Kosova. <https://gzk.rks-gov.net/ActDetail.aspx?ActID=20844>

gov.net/ActDetail.aspx?ActID=18616 (accessed August 30, 2022).

Law no. 06/L-113 on Organization and Functioning of State Administration and Independent Agencies. Official Gazette of the Republic of Kosova. <https://gzk.rks-gov.net/ActDetail.aspx?ActID=18684> (accessed August 30, 2022).

Comparative Legilinguistics
vol. 2022/52
DOI: <http://dx.doi.org/10.14746/cl.52.2022.16>

Research profile and bibliography of Professor Łucja Biel

DARIUSZ KOŻBIAŁ

Institute of Applied Linguistics
University of Warsaw, Poland
d.kozbial@uw.edu.pl

<https://orcid.org/0000-0003-4395-3138>

Abstract: This article consists of two sections. The first outlines the research profile of Dr hab. Łucja Biel, Prof. ucz., a Polish linguist recognised nationwide and internationally who specializes in the analysis of legal varieties of Polish and English in the context of legal translation studies, corpus linguistics and translator training. The second part contains a detailed list of publications (co-)authored or co-edited by Łucja Biel and published between 2004 and 2022.

Keywords: research profile; Łucja Biel; bibliography; applied linguistics; corpus linguistics; translation studies; legal translation.

**PROFIL NAUKOWY I BIBLIOGRAFIA DR HAB. ŁUCJI BIEL,
PROF. UCZ.**

Abstract: Niniejszy artykuł składa się z dwóch części. Pierwsza z nich przedstawia w zarysie profil naukowy dr hab. Łucji Biel, prof. ucz., językoznawczyni o dorobku uznanym zarówno w Polsce, jak i za granicą, specjalizującej się w analizie prawnych odmian języka polskiego i angielskiego w kontekście badań nad przekładem prawnym i prawniczym, językoznawstwa korpusowego oraz kształcenia tłumaczy. Druga część artykułu zawiera szczegółowy spis publikacji, których prof. Biel jest (współ)autorką lub współredaktorką, wydanych między 2004 r. a 2022 r.

Słowa kluczowe: profil naukowy; Łucja Biel; bibliografia; językoznawstwo stosowane; językoznawstwo korpusowe; przekładoznawstwo; przekład prawny i prawniczy.

Research profile of Professor Łucja Biel

Łucja Biel, an acclaimed legal translation scholar, corpus linguist and translator trainer, was born in 1974 in Nowy Sącz, Poland, where she completed I Liceum Ogólnokształcące im. Jana Długosza (Jan Długosz Secondary School) in a class with extended curriculum in mathematics and physics. In 1998, she completed her MA studies in English philology (specialization in translation studies) at the Jagiellonian University of Kraków, where she was introduced to the field of cognitive linguistics through Professor Elżbieta Tabakowska's seminars. Łucja Biel's MA thesis, written under the supervision of Professor Tabakowska, is entitled: "A Cognitive Analysis of Distance in Kazuo Ishiguro's *The Remains of the Day* and its Polish Translation by Jan Rybicki". In 1998, in addition to obtaining her MA diploma, she completed a postgraduate Management and Business Course at the same university.

Łucja Biel further pursued her interest in cognitive linguistics throughout her doctoral studies at the University of Gdańsk under the supervision of Professor Roman Kalisz. In 2004, she defended with honours her PhD dissertation in Linguistics entitled "Distance in English and Polish". The dissertation involves a contrastive analysis of the linguistic ways of expressing social and psychological distance in

both English and Polish based on the methodology of cognitive linguistics. In 2005, she received a postgraduate diploma in American law from the Chicago-Kent College of Law and the University of Gdańsk, whereas, in 2007, she attained a Diploma in an Introduction to English Law and the Law of the European Union from the University of Cambridge. The postgraduate programmes enabled her to broaden the scope of her research interests beyond cognitive linguistics.

After the defence of her PhD dissertation, that is in the years 2004–2013, she worked as an Assistant Professor at the Department of Translation at the Institute of English and American Studies of the University of Gdańsk, where she, *inter alia*, co-founded the translation programmes, taught classes in specialised translation, and acted as the Deputy Director for Student Affairs (2005–2008). In the period from 2009 to 2014, she was also a Visiting Lecturer at the City, University of London in the MA programme in Legal Translation, where she gave workshops on legal translation as well as supervised MA students in the process of dissertation writing.



Photo 1. Łucja Biel during the Quality Aspects in Institutional Translation conference, 25 November 2016, Prague¹

From 2013 on, she has been working at the Institute of Applied Linguistics of the University of Warsaw (ILS UW; Instytut Lingwistyki Stosowanej, Uniwersytet Warszawski), where she continues to teach specialised translation. In the years 2014–2016, she acted as the Deputy Director for International Cooperation and Organization, and, in 2016, the Acting Head of the ILS UW. In 2015, she obtained her *habilitacja* postdoctoral degree in Linguistics at the University of Warsaw as a result of the work performed within the framework of a *habilitacja*-

¹ Origin of the photograph: official website of the Quality Aspects in Institutional Translation conference. Institute of Translation Studies, Charles University, Prague. <https://qait.ff.cuni.cz/photo-gallery/> (accessed October 24, 2022).

degree (postdoctoral) research grant awarded by the Polish Ministry of Science and Higher Education², in which she conducted a corpus-based study of legal collocations and EU Polish, that is a hybrid variant of legal Polish emerging via translation. The outcomes of that research project were published in the monograph *Lost in the Eurofog: The Textual Fit of Translated Law* (see publication no. 1³) as well as in several articles (see publications nos. 20, 64) and chapters (see publications nos. 52, 53)⁴. In 2020, she was promoted to the position of a university professor. She is the Head of EUMultiLingua Research Group⁵ which was formed in 2020 on the basis of the Polish Eurolect research team created by her for the purposes of a research grant funded by the Polish National Science Centre (NCN) (2015–2020)⁶. The overall objective of the research group is to carry out interdisciplinary quantitative and qualitative research into multilingual specialised communication in the European Union, with a focus on legal translation.

To date, Łucja Biel has supervised nearly 80 BA and MA students in the process of writing their dissertations on specialised translation at the University of Warsaw and the University of Gdańsk. So far, she has supervised two PhD dissertations, both of which were defended with honours at the Faculty of Applied Linguistics, University of Warsaw, namely: *Translation of Judgments: A Corpus Study of the Textual Fit of EU to Polish Judgments* by Dariusz Koźbiał (2020)⁷, and *Administrative Reports: A Corpus Study of the Genre in the EU and Polish national settings* by Katarzyna Wasilewska (2021)⁸.

² “Corpus-based study of legal collocations”, 2010–2012, Polish Ministry of Science and Higher Education, grant agreement no. 2251/B/H03/2010/38.

³ Bibliographic entries of publications referred to in this article can be found in the enumerated list in Section 2.

⁴ More information about the concept of textual fit and its application in Biel’s research on EU legal translation can be found below.

⁵ More information about the EUMultiLingua Research Group can be found at: <https://www.ils.uw.edu.pl/en/institute/departments/zespol-badawczy-eumultilingua/> (accessed October 24, 2022).

⁶ More information about the project can be found below.

⁷ The thesis has been published by Peter Lang: Koźbiał, Dariusz. 2020. *The Language of EU and Polish Judges: Investigating Textual Fit through Corpus Methods*. Berlin: Peter Lang.

⁸ The thesis has been published by Peter Lang: Wasilewska, Katarzyna. 2022. *Administrative Reports: A corpus Study of the Genre in the EU and Polish National Settings*. Berlin: Peter Lang.

As regards participation in research projects, Biel participated in a number of internationally and nationally funded research projects related primarily to legal and institutional translation. The most extensive research project in which she acted as the Principal Investigator is “The Eurolect: An EU variant of Polish and its impact on administrative Polish”; it was financed with a Sonata BIS research grant from the NCN and ran in the period of 2015–2020⁹. The project investigated the Polish Eurolect, which is a hybrid, translator-mediated variant of Polish used in the EU context, across four genres: legislation, judgments, reports and websites for citizens, being the first large scale empirical study of its kind. The project also analysed how the Eurolect has evolved since the late 1990s and how the inflow of hybrid legal and semi-legal texts has affected administrative Polish. The methodology applied in the project relied mostly on the methods of corpus linguistics and discourse analysis. Other noteworthy research projects in which Biel took part include: (1) “Understanding Justice” (DG Justice, European Commission, action grant coordinated by Middlesex University, function: local coordinator, 2014–2016), (2) “Eurolect Observatory” (Università degli Studi Internazionali di Roma, Italy, function: investigator, 2013–2018), and (3) “Training action for legal practitioners: Linguistic skills and translation in EU” (DG Competition, European Commission, action grant coordinated by Università degli Studi dell’Insubria, Italy, 2016–2018¹⁰). At present, she is involved in an international project coordinated by the University of Amsterdam the goal of which is to develop and implement across various European healthcare settings a multilingual, culturally-sensitive, sustainable digital information and communication platform enhancing access to mental healthcare for third-country nationals with low language proficiency in the host country’s language, namely “MENTAL HEALTH 4 ALL: Development and implementation of a digital platform for the promotion of access to mental healthcare for low

⁹ More information about the Polish Eurolect research project can be found at the project website: <http://eurolekt.ils.uw.edu.pl> (accessed October 24, 2022). The website contains an extensive EU translation bibliography with ca. 360 entries.

¹⁰ Training of national judges and judicial cooperation in the field of EU competition law: https://competition-policy.ec.europa.eu/single-market-programme-smp/calls-proposals-grants/training-national-judges_en (accessed October 24, 2022).

language proficient third-country nationals in Europe” (Asylum, Migration and Integration Fund, European Commission, 2022–2024¹¹).

In addition to research-related activities, Biel undertakes numerous activities which are closely related to her academic work. In the years 2013–2022, she has been a Board Member of the European Society for Translation Studies (EST) and Secretary of the Society for the two terms of office. Currently, she is the Editor-in-Chief of the *Journal of Specialised Translation* (JoSTrans)¹², which is indexed in, *inter alia*, SCOPUS and Web of Science, a secretary of the Council of Editors of Translation & Interpreting Studies for Open Science¹³ which promotes diamond open access and other open science practices in Translation and Interpreting Studies, as well as a member of the Editorial and Advisory Boards for, among others, *Comparative Law and Language, Terminology, Across Languages and Cultures, Comparative Legilinguistics* and *Lingua Legis*. She is also a peer reviewer for numerous national and international journals. Furthermore, she is a member of various academic and professional associations, most notably the Polish Society of Sworn and Specialised Translators (TEPIS), and a member of various research groups based outside the University of Warsaw, such as the International Research Group CAL² (Computer-Assisted Legal Linguistics)¹⁴. It is worth noting that in the years 2012–2017, she was an Expert in the Polish Normalisation Committee (TC 37 – Terminology and other language and content resources, Sub-Committee SC 5 – Translation, interpreting and associated technologies) in which she worked on ISO standards on translation and interpreting.

In 2006, following the acquisition of her PhD degree in Linguistics, Biel’s scholarly interests started to revolve around the relationship between law and language, which is owed to her professional practice as a specialised translator since 1997 and postgraduate diplomas in American, English and EU law (see above). Thanks to the hands-on knowledge gained as a freelance translator and, later on, a sworn translator, she is apt in tackling topics and solving issues which are of importance to both the academia and professional

¹¹ For more information about the project see: <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/projects-details/31077795/101038491/AMIF> (accessed October 24, 2022).

¹² Journal’s website: <https://www.jostrans.org/> (accessed October 24, 2022).

¹³ Council of Editors of Translation & Interpreting Studies for Open Science: <https://tisopencouncil.eu/> (accessed October 24, 2022).

¹⁴ CAL² Research Group: <https://cal2.eu/index.php/team> (accessed October 24, 2022).

translators who work individually as well as for translation agencies, companies and (supra)national institutions.

Overall, the main research trends clearly discernible in Łucja Biel's publications are related to:

- 1) legal translation;
- 2) corpus linguistics;
- 3) terminology and phraseology;
- 4) translator education¹⁵.

The following description of the major research trends aims to provide a perfunctory overview of Biel's research output in order to familiarise the reader with the range of topics tackled in her works¹⁶.

In her publications related to **legal translation**, Biel focuses on theoretical and methodological issues in legal translation studies. According to Biel, the predominant methodology in contemporary legal translation studies is corpus-based (see publication no. 21), as a result of which there takes place a shift from qualitative to quantitative methods. She fits in that trend owing to her extensive use of corpus methods (see below). Biel advocates a multi-perspective and mixed-method approach to research on legal translation within legal translation studies. The research framework proposed by her embraces not only the end product of the translation process, as it is typically the case in legal translation studies, but also four other key dimensions of translation, namely: the context of production, the process, the participants and text reception (see publications nos. 15, 29). Understandably, this type of approach is very demanding, as it requires researchers to engage in even greater interdisciplinarity beyond the domains of language and law. Biel is also interested in the concept of genre which is discussed and applied by her in numerous publications, both theoretical and empirical ones (see publication no. 41). In her opinion, genres lend themselves to qualitatively- and quantitatively-oriented research on legal translation and translator education, both of which should pay attention to distinctive features (conventions) of genres, their structure as well as social, communicative, cultural, cognitive and ideological factors behind their practical use. Issues discussed by Biel in her theoretically-

¹⁵ An example of a publication authored recently by Biel which does not exactly fit the main research trends indicated above includes a paper on practical issues related to machine translation (MT), such as types of MT and its applications, factors influencing output generated by MT, MT-related ethical issues, and competences which are required from professional post-editors (see publication no. 58).

¹⁶ Section 2 contains a comprehensive bibliography of Biel's publications which appeared between 2004 and the end of 2022.

oriented research on legal translation set the ground for similarly oriented research on **EU legal translation**, which is described by her as a sub-genre (i.e. a special type) within the field of legal translation (see publication no. 57). In general, Biel deems EU legal language to be particularly well suited for the study of law and language, since the EU legal culture is shaped through the processes of mediation, filtering of law through the EU's official languages and translation, and is thus presented by her as a hybrid supranational construct mutually dependent on national legal cultures (see publication no. 14). Additionally, the translation of EU law is viewed by her as complex to operationalise for research due to its hybrid nature and the challenges posed to the central concepts of translation studies, such as: source text and target text (versus language versions), translation process (translation as part of a multistage and multilingual drafting process), equivalence and translation quality (uniform application and interpretation of law) (see publication no. 60). The key theoretical and methodological challenges in EU legal translation research are thus said to rely in, among others, complex multistage drafting processes which are combined with translation at all the stages, with the constant switching of languages, rewriting and changes to draft documents. Biel also holds that that EU translation should be studied through five interrelated axes, namely: the institutional, political, supranational, legal and multilingual axis, and that EU translation should be operationalised from the perspective of fundamental intertextual relations, that is through the dimensions of: concordance, continuity and **fit** (see publication no. 37). This approach emerges as conducive to conducting research on legal translation in the EU context owing to its all-inclusivity and, most importantly, the involvement of the concept of **textual fit**, which is a constantly recurring theme in Biel's research on EU translation as well as translator training (see below).

Since 2010, multifaceted analyses of the Polish Eurolect make up the foundation of Biel's research (see *inter alia* publications nos. 1, 10–14, 19, 20, 33–36, 39–40, 42, 44–45, 48, 51). Initially, her research on EU legal translation revolved strictly around the genre of legal acts, and, subsequently, several other related genres — court judgments, administrative reports and websites for citizens. Biel's novel contribution to the development of legal translation studies relies in adapting the concept of **textual fit** to the empirical study of EU legal translation with the simultaneous employment of the methodology of corpus linguistics (see below) which was still underutilised at the outset

of her research into the nature of the Polish Eurolect. With the help of corpus methods, she has been able to conduct systematic, reliable and objective analyses of large sets of legal texts. Originally, the concept of **textual fit** (also referred to as naturalness owing to translation recipients' expectations concerning the overall readability and quality of translations) has been coined by Andrew Chesterman (see 2004, 2010¹⁷). It relates to the hypothesis of translation universals (levelling out, normalisation/conventionalisation, untypical collocations and colligations, underrepresentation of unique TL-specific items), or, more specifically, to the textual relation between translated and non-translated texts in a target language. Prior to Biel taking up research on the **textual fit** of translated EU legal texts and comparable non-translated Polish legal texts under the assumption that translated EU legislation differs from non-translated Polish legislation, the relation of **textual fit**, as opposed to the intertextual relation of equivalence (which concerns rather parallel translated and non-translated texts), had been rarely studied as an object of legal-linguistic research. Biel operationalised this concept by going beyond the context of translation universals and defined it as the linguistic distance between translations and target language non-translated texts of a comparable genre at the level of both the macrostructure (text-structuring and grammatical patterns) and the microstructure (term-embedding, term-forming and lexical collocations)¹⁸. Her initial research into the Polish Eurolect was focused solely on legislation (regulations and directives) which she had chosen owing to its status as a primary legal genre which is characterised by a high degree of standardization and formulaicity. By virtue of these features, legal acts are an ideal research object for studying repetitive lexical and grammatical patterns using corpus methods. Biel's monograph *Lost in the Eurofog: The Textual Fit of Translated Law* (see publication no. 1) is one of the few in-depth and systematic investigations into the nature of EU legal translation and its impact on the national legal language of an EU Member State (Poland).

¹⁷ See, among others: (1) Chesterman, Andrew. 2004. "Hypotheses about translation universals" In *Claims, Changes and Challenges in Translation Studies*, eds. Hansen, Gyde, Kirsten Malmkjær, Daniel Gile, 1–13. John Benjamins: Amsterdam, and (2) Chesterman, Andrew. 2010. "Why study translation universals?" In *Kiasm. Acta Translatologica Helsingiensia (ATH) 1*, eds. Ritva Hartama-Heinonen and Pirjo Kukkonen, 38–48. Helsingfors: Helsingfors universitet, Nordica, svensk översättning.

¹⁸ At first, Biel studied solely EU and national legislation. After the completion of the Eurofog project, she extended her research onto other genres, such as court judgments, administrative reports and websites for citizens.

As it has already been mentioned, Biel takes advantage of corpus methodology. She expanded her methodological repertoire with the methods of **corpus linguistics** in 2009. Thanks to these methods, she efficiently conducts quantitative and qualitative research on extensive empirical material made up of translated and non-translated EU and national legal texts, primarily within the research trend related to terminology and phraseology (see below). Nevertheless, several of her publications describe corpus methodology and its potential applications from the theoretical perspective. In the chapter *Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential* (see publication no. 54), Biel demonstrates the potential of corpus-based studies as a methodology for, among others, researching legal translation, as a tool in translator training and translation practice. According to her, corpus-linguistic methods open up new perspectives on legal language and legal translation, e.g. the equivalence relation between STs and TTs as well as the **textual fit** of translations to comparable non-translations, all of which may help to improve the naturalness of translations by lessening the effects of translation universals and source language interference. She also stresses that the pedagogical potential of corpora relies in fostering inductive, data-driven learning of translation. Moreover, Biel's article written in Polish which is entitled *Translatoryka korpusowa* [Corpus Translation Studies] (see publication no. 61) discusses the assumptions of corpus translation studies as a new empirical, inductive, descriptive and data-driven methodology for translation research and positions it against the background of the latest trends in translation research. It also examines its theoretical foundations which are derived from corpus linguistics and descriptive research on translation and lays out the theoretical and practical applications of corpus research, its development prospects and limitations. In that article, Biel stresses that corpora are well-suited for identifying and comparing frequent and typical lexico-grammatical patterns. The current limitations, in her view, and rightly so, include technological and corpus design constraints as well as the need for stronger triangulation of data and methodologies. In the co-authored chapter *Metody korpusowe w analizie gatunków specjalistycznych – założenia, perspektywy i ograniczenia* [Corpus methods in the analysis of specialised genres – assumptions, perspectives and limitations] (see publication no. 78), Biel analyses the advantages of corpus methods and their limitations in genre analysis, that is identification and description

of various genres' features (conventions). Corpus methods listed as conducive to genre analysis include: keyword analysis, analysis of lexical bundles (multi-word units), and multi-dimensional register analysis (which covers, among others, analysis of lexical distribution, grammatical features such as nominalisation and passive voice, syntactic and lexico-grammatical structures). The latter lends itself particularly well to measuring the degree of **fit** between genres. Therefore, corpus methods should be universally viewed as indispensable; however, they are also viewed as insufficient when it comes to addressing pragmatic, rhetorical, (inter)discursive, contextual (social), and cognitive issues, all of which call for qualitative approaches and triangulation of methods. Lastly, In the chapter *Corpora in Institutional Legal Translation: Small Steps and the Big Picture* (see publication no. 43), Biel stresses the usefulness of corpora as an analytical tool in research on institutional translation within legal translation studies oriented at the final outcome of the translation process, as well as the need to obtain more quantitative data on major and lesser-used languages, including various branches of law and translational settings, and to triangulate them with qualitative data obtained through mixed methods in research. Furthermore, she states that corpora, as a research tool, provide valuable empirical data thanks to which it is possible to generalise about institutional translation. They are also viewed by her as a valuable resource for translators. All in all, Biel's theoretically-oriented publications related to the research trend of corpus linguistics clearly demonstrate that she is a committed proponent of corpus methods in research on EU legal translation. Evidence of this can be gathered from her other works, including terminology- and phraseology-related publications, and publications on translator training (see below).

The **terminological and phraseological research trend** in Łucja Biel's publications concerns practical issues related to decision-making in the process of choosing equivalents of legal terms, the linguistic **fit** of translated EU texts to non-translated national texts, as well as other associated aspects. At the outset of her academic career, Biel's terminology- and phraseology-related research interests revolved around EU law (see publication no. 71), company law (see publications nos. 70, 69, 72, 68), human rights and international law (see publication no. 65), and labour law (see publication no. 67). Later works published by Biel (and her co-authors) within the same research trend concern the area of EU competition law, touching upon issues such as key terms in

EU competition law in English and the examination of their collocational environment via corpus methods, or hybridity of EU competition law concepts which is owed to their travelling via translation (see publications nos. 44, 39, respectively). In a similar vein, Biel has also studied how source-language synonymy (terminological variation, i.e. synonymy at the term level) is handled in the process of translation across institutional genres using a parallel-comparable corpus of EU English-Polish legislation, soft law and judgments in the area of competition law (see publication no. 12) and how EU supranational terms related to consumer protection transfer into national legal systems of three English-language jurisdictions (the UK, Ireland and Malta) during the transposition of EU directives (see publication no. 11). Biel's studies related to the phraseological trend concern primarily the analysis of the Polish (and English) Eurolect using corpus methods (see above). These analyses, as it has been indicated above, center around the crucial concept of **textual fit**. Biel has taken advantage of the intertextual relation of textual fit to, among others, explore phraseology in comparable monolingual corpora of EU legal acts (regulations and directives) translated into Polish against non-translated Polish legal acts (see publication no. 51), to measure the **fit** of translated EU law to non-translated Polish law at the level of deontic modality patterns (see publication no. 20), and to analyse the distribution and discourse functions of complex prepositions in multilingual EU legislation (regulations and directives) and national legislation (see publications nos. 19, 79). Biel has also examined formulaicity using the concept of genre and lexical bundles, for instance, in EU translations into Polish across several institutional genres (legislation, judgments, reports, websites) with reference to the corresponding English-language texts in order to understand how formulaicity is affected by the variable of genre (see publications nos. 13, 73).

Biel's comprehensive research into the Polish Eurolect, including research within the framework of the project "The Eurolect: An EU variant of Polish and its impact on administrative Polish" (see above) which has been widely popularised by her at numerous international conferences, has awakened wider interest in the legislative varieties of EU official languages other than Polish. One example of this is the volume *Observing Eurolects. Corpus analysis of linguistic variation in EU Law* (see publication no. 40) which provides insights into eleven EU official and working languages (Dutch, English,

Finnish, French, German, Greek, Italian, Latvian, Maltese, Polish, Spanish) which have been examined using corpus methods and the same protocol. Biel's chapter *Observing Eurolects. The Case of Polish*, which is a part of that volume, investigates the Polish Eurolect and its features induced through contact with EU English, and shows that the Polish Eurolect has developed a set of features that diverge from the generic conventions of Polish law (e.g. high frequency of semantic Europeisms, a low overlap of lexical bundles, interference from EU English in the form of borrowings and calques from EU English). Owing to those features it may be perceived as a distinct variety of legal Polish expressing multilingual EU law. As such, the Polish Eurolect is subject to constant evolution, and has been found by Biel to have developed significantly since its formative pre-accession stage to the mature post-accession variant (see publication no. 10). Biel has found thereby that the Polish Eurolect is dynamic and to some extent subject to controlled changes, which is a finding that warrants further analyses. Therefore, Biel acknowledges the legitimacy of research into other official EU languages and extends her Eurolects-related research beyond the enclave of the Polish Eurolect. In 2018, she expanded her research with the analysis of the English Eurolect (see *inter alia* publications nos. 13, 33, 42). Her recent chapter entitled *Eurolects and EU Legal Translation* (see publication no. 34) overviews contemporary studies into Eurolects set in the same methodological current, namely corpus linguistics, and which, at the same time, focus on the development of the complex concept of Eurolects and their **textual fit** to domestic non-translated varieties of legal languages. Thanks to Biel's research into EU legal language and translation, studies into the nature of different varieties of Eurolects are starting to become more prominent in the field of legal translation studies *sensu largo*.

The research trend related to **translator education** focuses on the acquisition of competences related to translation and service provision by aspiring (sworn) translators. Biel has developed her own integrative functional approach for teaching legal translation which she has described in two articles — *Professional realism in the legal translation classroom: translation competence and translator competence* and *Integracyjne podejście funkcjonalne w dydaktyce przekładu prawniczego [Integrative functional approach in legal translation teaching]* (see publications nos. 23, 63, respectively), as well as a chapter entitled *Przekład prawny i prawniczy [Legal translation]* (see publication no. 73). Her approach consists in allowing

translation trainees to develop a responsible and reflective attitude towards legal translation, at the heart of which lies the awareness of differences between legal systems (and hence legal concepts) and sensitivity towards differences between genres. Possible implementations of this approach have been presented in an article (see publication no. 56) and a chapter in a monograph (see publication no. 82). Biel has also published on other topics related to translator education. In *EU translation as an institutional translation: an advanced course with focus on information mining competence* (see publication no. 52), Biel presents her advanced course on EU legal translation with focus on information mining competence dedicated for student translators. According to her, such a course is a mandatory component of specialised translator training within the EU context. In the article *The textual fit of legal translations: focus on collocations in translator training*, she focuses on the phraseological competence of an aspiring legal translator, stating that the training process should pay more attention to equivalence beyond word level, i.e. collocations and colligations (prefabricated phraseological units) which have the potential for increasing the **textual fit** of trainee translations to (comparable) TL texts (see publication no. 55). The article *Enhancing the communicative dimension of legal translation: comparable corpora in the research-informed classroom* demonstrates a two-fold manner of implementing comparable corpora in translation classes at MA level with the aim of increasing the communicative dimension of legal translations from the perspective of the English-Polish language pair, namely using corpora as a decision-making aid during the translation process to increase the **textual fit** of trainee translations, and using them to reflect on the translation process *ex post facto* (see publication no. 16). In another article (see publication no. 59), Biel provides a critical analysis of the written part of the sworn translator examination in light of the European Master's in Translation (EMT)¹⁹ competences framework and the reality of the translator's profession. Biel argues, among others, that the written part of the examination should be adjusted to a greater degree to skills and competences required of sworn translators in real life and that the written part should be separated from

¹⁹ The Institute of Applied Linguistics of the University of Warsaw is an EMT member institution which offers MA level training for translators in line with the EMT standards. More information can be found at: https://ec.europa.eu/info/resources-partners/european-masters-translation-emt/list-emt-members-2019-2024_en#poland (accessed October 24, 2022).

the interpreting part, as a result of which aspiring sworn translators could potentially choose whether they would like to become sworn translators or sworn interpreters. Her remarks strike one as particularly on point, considering the fact that today's translators and interpreters need to become highly specialised in order to adequately satisfy the demands of the market. Her most recent publications (nos. 2, 31) focus on institutional translator training, continuing professional education and institutional outreach to universities.

Łucja Biel's research- and teaching-related accomplishments demonstrate that she has contributed greatly to the legal-linguistic research and translator education in Poland. Having actively attended numerous national and international conferences and workshops closely tied with the fields of translation studies and legal linguistics, she has been able to share her knowledge and experience with, among others, like-minded scholars, practitioners dealing with the translation of specialised texts and university students. Numerous prizes and distinctions awarded to Biel are a proof of the impact of her works. Most notably, in 2017, she was awarded the Bronze Cross of Merit²⁰ in recognition of her contributions to the training of sworn and specialised translators in Poland, whereas, in 2022, she was distinguished with *Laur Thumacza* by the Polish Society of Sworn and Specialised Translators (TEPIS)²¹ in recognition of the work performed for the community of specialised translators. Apart from those distinctions she has been honoured by rectors of several universities for *inter alia* teaching and research achievements (University of Gdańsk – 2011, University of Warsaw – 2017, 2019, 2020).

More detailed information about Łucja Biel's academic career as well as her publications and achievements can be found at the following websites:

- ORCID profile: <https://orcid.org/0000-0002-3671-3112> (accessed October 24, 2022).
- Google Scholar profile: <https://scholar.google.com/citations?user=aP3-J7YAAA&hl=en&oi=ao> (accessed October 24, 2022).

²⁰ In Polish: *Brązowy Krzyż Zasługi*. The Cross of Merit is a state decoration awarded to citizens who make great contributions to the state and society. More information can be found at: <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WMP20170001105> (accessed October 24, 2022).

²¹ Website of the Polish Society of Sworn and Specialized Translators (TEPIS): <https://tepis.org.pl/the-polish-society-of-sworn-and-specialized-translators/> (accessed October 24, 2022).

- Profile on the website of the Institute of Applied Linguistics, University of Warsaw: <https://www.ils.uw.edu.pl/en/institute/staff/lucja-biel/> (accessed October 24, 2022).
- Profile in the Polish Science database: https://nauka-polska.pl/#/profile/scientist?id=128443&_k=4a9obb (accessed October 24, 2022).
- ResearchGate profile: <https://www.researchgate.net/profile/Lucja-Biel> (accessed October 24, 2022).

Thus far, Łucja Biel has published one monograph and over seventy chapters and articles in both English and Polish, and co-edited five volumes. Section 2 presents a comprehensive bibliography of her publications, all of which are deemed to have positively influenced the development of legal translation studies in both Poland and abroad. The bibliography is divided into eight sections: monographs, co-edited volumes, co-edited special issues of journals, journal articles (in English), chapters in monographs (in English), journal articles (in Polish), chapters in monographs (in Polish), and other publications. Each section is chronologically ordered and starts with the latest publications.

Bibliography of Professor Biel's publications (2004–2022)

A. Monographs

1. Biel, Łucja. 2014. *Lost in the Eurofog: The Textual Fit of Translated Law*. Frankfurt am Mein: Peter Lang; 2nd ed. in 2017. DOI: 10.3726/978-3-653-03986-3.

B. Co-edited volumes

2. Svoboda, Tomáš, Łucja Biel, and Vilelmini Sosoni, eds. 2022. *Institutional Translator Training*. New York: Routledge. DOI: 10.4324/9781003225249.
3. Biel, Łucja, and Hendrik Kockaert, eds. Forthcoming. *The Handbook of Legal Terminology*. John Benjamins.

4. Biel, Łucja, Jan Engberg, M. Rosario Martín Ruano, and Vilelmini Sosoni, eds. 2019. *Research Methods in Legal Translation and Interpreting. Crossing Methodological Boundaries*. London: Routledge.
5. Marino, Silvia, Łucja Biel, Martina Bajčić, and Vilelmini Sosoni, eds. 2018. *Language and Law: The Role of Language and Translation in EU Competition Law*. Cham: Springer. DOI: 10.1007/978-3-319-90905-9.
6. Svoboda, Tomáš, Łucja Biel, and Krzysztof Łoboda, eds. 2017. *Quality Aspects in Institutional Translation*. Berlin: Language Science Press. <http://langsci-press.org/catalog/book/181>. DOI: 10.5281/zenodo.1048173.

C. Co-edited special issues of journals

7. Sosoni, Vilelmini, and Łucja Biel, eds. 2018. EU Legal Culture and Translation. *International Journal of Language & Law* 7.
8. Biel, Łucja, and Vilelmini Sosoni, eds. 2017. The Translation of Economics and the Economics of Translation. *Perspectives. Studies in Translation Theory and Practice* 25(3). DOI: 10.1080/0907676X.2017.1313281.
9. Biel, Łucja, and Jan Engberg. 2013. Research models and methods in legal translation. *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 12.

D. Journal articles (in English)

10. Biel, Łucja. 2022. From National to Supranational Institutionalisation: A Microdiachronic Study of the Post-Accession Evolution of the Polish Eurolect. *Perspectives* (February 3). DOI: 10.1080/0907676X.2022.2025870.
11. Biel, Łucja, and Agnieszka Doczekalska. 2020. How do supranational terms transfer into national legal systems?: A corpus-informed study of EU English terminology in consumer protection directives and UK, Irish and Maltese transposing acts. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 26(2): 184–212. DOI: 10.1075/term.00050.bie.

12. Biel, Łucja, and Dariusz Koźbiał. 2020. How do translators handle (near-)synonymous legal terms? A mixed-genre parallel corpus study into the variation of EU English-Polish competition law terminology. *Estudios de Traducción* 10: 69–90. DOI: 10.5209/estr.68054.
13. Biel, Łucja, Katarzyna Wasilewska, and Dariusz Koźbiał. 2019. The formulaicity of translations across EU institutional genres. A corpus-driven analysis of lexical bundles in translated and non-translated language. *Translation Spaces* 8(1): 67–92. DOI: 10.1075/ts.00013.bie.
14. Sosoni, Vilemini, and Łucja Biel. 2018. EU Legal Culture and Translation. *International Journal of Language & Law* 7: 1–7. DOI: 10.14762/jll.2018.001.
15. Biel, Łucja. 2017. Researching Legal Translation: A multi-perspective and mixed-method framework for legal translation. *Revista de Llengua i Dret, Journal of Language and Law* 68: 76–88. DOI: 10.2436/rld.i68.2017.2967.
16. Biel, Łucja. 2017. Enhancing the communicative dimension of legal translation: comparable corpora in the research-informed classroom. *The Interpreter and Translator Trainer* 1(4): 316–336. DOI: 10.1080/1750399X.2017.1359761.
17. Biel, Łucja, and Vilemini Sosoni. 2017. The translation of economics and the economics of translation. *Perspectives. Studies in Translation Theory and Practice* 25(3): 351–361. DOI: 10.1080/1750399X.2017.1359761.
18. Vogel, Friedemann, Hanjo Hamann, Dieter Stein, Andreas Abegg, Łucja Biel, and Lawrence M. Solan. 2017. “Begin at the beginning”: Lawyers and Linguists Together in Wonderland. *International Journal of Language & Law* 6: 90–100. DOI: 10.15200/winn.148184.43176.
19. Biel, Łucja. 2015. Phraseological profiles of legislative genres: complex prepositions as a special case of legal phrasemes in EU law and national law. *FACHSPRACHE — International Journal of Specialized Communication* 37(3-4): 139–160. DOI: 10.24989/fs.v37i3-4.1286.
20. Biel, Łucja. 2014. The textual fit of translated EU law: a corpus-based study of deontic modality. *The Translator* 20(2): 332–355. DOI: 10.1080/13556509.2014.909675.

21. Biel, Łucja, and Jan Engberg. 2013. Research models and methods in legal translation. *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 12: 1–11, <https://lanstts.uantwerpen.be/index.php/LANS-TTS/article/view/316/225> (accessed October 24, 2022). DOI: 10.52034/lanstts.v0i12.316.
22. Biel, Łucja. 2011. Training translators or translation service providers? EN 15038:2006 standard of translation services and its training implications. *The Journal of Specialised Translation* 16: 61–76, https://jostrans.org/issue16/art_biel.pdf (accessed October 24, 2022).
23. Biel, Łucja. 2011. Professional realism in the legal translation classroom: translation competence and translator competence. *Meta. Translators' Journal* 56(1): 162–178. DOI: 10.7202/1003515ar.
24. Biel, Łucja. 2009. Organization of background knowledge structures in legal language and related translation problems. *Comparative Legilinguistics. International Journal for Legal Communication* 1: 176–189. DOI: 10.14746/cl.2009.01.13.
25. Biel, Łucja. 2008. Legal terminology in translation practice: dictionaries, googling or discussion forums? *SKASE Journal of Translation and Interpretation* 3(1): 22–38.
26. Biel, Łucja. 2008. Working for Translation Agencies as a Freelancer: A Guide for Novice Translators. *Translation Journal* 12(2), <http://translationjournal.net/journal/44freelancer.htm> (accessed October 24, 2022).
27. Biel, Łucja. 2006. Incongruity of Company Law Terms: Categorization of Polish Business Entities and their English Equivalents. *Translation Journal* 10(4), <http://translationjournal.net/journal/38legal.htm> (accessed October 24, 2022).
28. Biel, Łucja. 2004. New English-Polish and Polish-English Dictionaries: Some Problems Related to Legal, Financial and Insurance Terminology. *Translation Journal* 8(3), <http://translationjournal.net/journal/29poldic.htm> (accessed October 24, 2022).

E. Chapters in monographs (in English)

29. Biel, Łucja. 2022. Translating Legal Texts. In *The Cambridge Handbook of Translation*, ed. Kirsten Malmkjær, 379–400. Cambridge: Cambridge University Press. DOI: 10.1017/9781108616119.020.
30. Doczekalska, Agnieszka, and Łucja Biel. 2022. Interlingual, Intralingual and Intersemiotic Translation in Law. In *Translation Beyond Translation Studies*, ed. Kobus Marais, 99–118. London: Bloomsbury. DOI: 10.5040/9781350192140.ch-6.
31. Biel, Łucja, and María Rosario Martín Ruano. 2022. Institutions' outreach to and involvement with universities: How international organizations collaborate with universities in training translators. In *Institutional Translator Training*, ed. Tomáš Svoboda, Łucja Biel and Vilemini Sосoni, 151–169. London: Routledge. DOI: 10.4324/9781003225249.
32. Biel, Łucja. 2022. Research into Legal Translation: An Overview of the 2010s Trends from the Perspective of Translation Studies. In *Comparative Perspectives on Law and Language*, ed. by Agustín Parise and Olivier Moréteau, Maastricht Law Series, 59–89. The Hague: Eleven Publishing.
33. Biel, Łucja, Dariusz Koźbiał, and Dariusz Müller. 2022. The judicial English Eurolect: A genre profiling of CJEU judgments. In *Law, Language and the Courtroom: Legal Linguistics and the Discourse of Judges*, eds. Stanisław Goźdź-Roszkowski and Gianluca Pontrandolfo, 3–25. London: Routledge. DOI: 10.4324/9781003153771.
34. Biel, Łucja. 2021. Eurolects and EU Legal Translation. In *The Oxford Handbook of Translation and Social Practices*, eds. Meng Ji and Sara Laviosa, 479–500. New York: Oxford University Press.
35. Biel, Łucja. 2021. EU institutional websites: Targeting citizens, building asymmetries. In *Translating Asymmetry – Rewriting Power*, eds. Ovidio Carbonell i Cortés and Esther Monzó-Nebot, 227–252. Amsterdam: John Benjamins. DOI: 10.1093/oxfordhb/9780190067205.013.15.
36. Biel, Łucja, and Izabela Pytel. 2020. Corrigenda of EU Legislative Acts as an Indicator of Quality Assurance Failures: A Micro-Diachronic Analysis of Errors Rectified in the Polish Corrigenda. In *Institutional Translation and Interpreting: Assessing Practices and Managing for Quality*, ed. Fernando Prieto Ramos, 150–173. New York: Routledge. DOI: 10.4324/9780429264894.

37. Biel, Łucja. 2019. Theoretical and methodological challenges in researching EU legal translation. In *Legal Translation. Current Issues and Challenges in Research, Methods and Applications*, ed. Ingrid Simonnæs and Marita Kristiansen, 25–39. Berlin: Frank & Timme.
38. Biel, Łucja, Jan Engberg, Rosario Martín Ruano, and Vilelmini Sosoni. 2019. Introduction to research methods in legal translation and interpreting: Crossing methodological boundaries. In *Research Methods in Legal Translation and Interpreting. Crossing Methodological Boundaries*, eds. Łucja Biel, Jan Engberg, Rosario Martín Ruano, and Vilelmini Sosoni, 1–12. London: Routledge.
39. Biel, Łucja, and Vilelmini Sosoni. 2019. EU Legal Culture and Translation in the Era of Globalisation. The Hybridisation of EU Terminology on the Example of Competition Law. In *Legal Linguistics beyond the borders: Language and Law in a World of Media, Globalisation and Social Conflicts. Relaunching the International Language and Law Association*, ed. Friedemann Vogel, 208–229. Berlin: Duncker and Humblot. DOI: 10.3790/978-3-428-55423-2.
40. Biel, Łucja. 2018. Observing Eurolects. The Case of Polish. In *Observing Eurolects. Corpus analysis of linguistic variation in EU Law*, ed. Laura Mori, 295–327. Amsterdam: John Benjamins. DOI: 10.1075/scl.86.12bie.
41. Biel, Łucja. 2018. Genre analysis and translation. In *The Routledge Handbook of Translation Studies and Linguistics*, ed. Kirsten Malmkjær, 151–164. London: Routledge. DOI: 10.4324/9781315692845-11.
42. Biel, Łucja. 2018. Lexical bundles in EU law: the impact of translation process on the patterning of legal language. In *Phraseology in Legal and Institutional Settings: A Corpus-based Interdisciplinary Perspective*, eds. Stanisław Goźdź-Roszkowski and Gianluca Pontrandolfo, 11–26. Abingdon: Routledge. DOI: 10.4324/9781315445724.
43. 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.

44. Biel, Łucja, Agnieszka Biernacka, and Anna Jopek-Bosiacka. 2018. Collocations of Terms in EU Competition Law: A Corpus Analysis of EU English Collocations. In *Language and Law: The Role of Language and Translation in EU Competition Law*, eds. Silvia Marino, Łucja Biel, Martina Bajčić and Vilelmini Sosoni, 249–274. Cham: Springer. DOI: 10.1007/978-3-319-90905-9_14.
45. Biel, Łucja, Agnieszka Biernacka, and Anna Jopek-Bosiacka. 2018. The Glossary of EU English Competition Collocations and Terms. In *Language and Law: The Role of Language and Translation in EU Competition Law*, eds. Silvia Marino, Łucja Biel, Martina Bajčić, and Vilelmini Sosoni, 275–324. Cham: Springer. DOI: 10.1007/978-3-319-90905-9_15.
46. Biel, Łucja. 2017b. Quality in institutional EU translation: Parameters, policies and practices. In *Quality Aspects in Institutional Translation*, eds. Tomáš Svoboda, Łucja Biel, and Krzysztof Łoboda, 31–57. Berlin: Language Science Press. DOI: 10.5281/zenodo.1048183.
47. Svoboda, Tomáš, Łucja Biel, and Krzysztof Łoboda. 2017. Quality Aspects in Institutional Translation: Introduction. In *Quality Aspects in Institutional Translation*, eds. Tomáš Svoboda, Łucja Biel and Krzysztof Łoboda, 1–13. Berlin: Language Science Press. DOI: 10.5281/zenodo.1048175.
48. Biel, Łucja. 2016. Mixed corpus design for researching the Eurolect: a genre-based comparable-parallel corpus in the PL EUROLECT project. In *Polskojęzyczne korpusy równoległe. Polish-language Parallel Corpora*, eds. Ewa Gruszczyńska and Agnieszka Leńko-Szymańska, 197–208. Warszawa: Instytut Lingwistyki Stosowanej.
49. Biel, Łucja, Anna Jopek-Bosiacka. 2016. Country Report: Poland. In *Understanding justice: An enquiry into interpreting in civil justice and mediation*, ed. Brooke Townsley, 149–159. London: Middlesex University, Short Run Press.
50. Biel, Łucja, Stanisław Goźdź-Roszkowski. 2015. Legal Translation. In *Ways to Translation*, eds. Łukasz Bogucki, Stanisław Goźdź-Roszkowski and Piotr Stalmaszczyk, 249–275. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
51. Biel, Łucja. 2014. Phraseology in Legal Translation: A Corpus-based Analysis of Textual Mapping in EU Law. In *The Ashgate*

- Handbook of Legal Translation*, eds. Le Cheng, King Kui Sin and Anne Wagner, 177–192. Farnham: Ashgate/Routledge.
52. Biel, Łucja. 2012. EU translation as an institutional translation: an advanced course with focus on information mining competence. In *Teaching Translation and Interpreting. Advances and Perspectives*, eds. Łukasz Bogucki, Mikołaj Deckert, 23–37. Newcastle: Cambridge Scholars Press.
53. Biel, Łucja. 2012. Areas of Similarity and Difference in Legal Phraseology: Collocations of Key Terms in UK and Polish Company Law. In *Phraseology and Discourse: Cross-Linguistic and Corpus-based Approaches*, eds. Antonio Pamies, José Manuel Pazos Bretaña and Lucía Luque Nadal, 225–233. Baltmannsweiler: Schneider Verlag.
54. Biel, Łucja. 2010. Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential. In *Reconceptualizing LSP. Online proceedings of the XVII European LSP Symposium 2009*, eds. Carmen Heine and Jan Engberg. Aarhus: University of Aarhus. <https://asb.dk/fileadmin/www.asb.dk/isek/biel.pdf> (accessed October 24, 2022).
55. Biel, Łucja. 2010. The textual fit of legal translations: focus on collocations in translator training. In *Teaching Translation and Interpreting: Challenges and Practices*, ed. Łukasz Bogucki, 25–39. Newcastle upon Tyne: Cambridge Scholars Publishing.
56. Biel, Łucja. 2010. Collaborative projects with external stakeholders in the translation classroom. In *Quality and Qualifications in Translation and Interpreting*, ed. Liudmila Sereda, 47–55. Grodno: Y. Kupala State University of Grodno.
57. Biel, Łucja. 2007. Translation of Multilingual EU Law as a Sub-genre of Legal Translation. In *Court Interpreting and Legal Translation in the Enlarged Europe*, ed. Danuta Kierzkowska, 144–163. Warsaw: Translegis.

F. Journal articles (in Polish)

58. Biel, Łucja. 2021. Postędycja tłumaczeń maszynowych [Post-editing of machine translation]. *Lingua Legis* 29: 11–34.
59. Biel, Łucja. 2017. Egzamin na tłumacza przysięgłego a realia wykonywania zawodu: głos krytyczny w sprawie weryfikowania

- kompetencji kandydatów na tłumacza przysięgłego [The sworn translator examination and the realities of practicing the profession: Critical remarks on how sworn translator candidates' competences are verified]. *Rocznik Przekładoznawczy* 12: 33–46. DOI: <https://doi.org/10.12775/RP.2017.002>.
60. Biel, Łucja. 2016. Tłumaczenie prawa unijnego a centralne pojęcia przekładoznawstwa [Translation of EU law versus the central concepts of Translation Studies]. *Między Oryginałem a Przekładem* 22(1): 107–119. DOI: <https://doi.org/10.12797/MOaP.22.2016.31.06>.
61. Biel, Łucja. 2015. Translatoryka korpusowa [Corpus Translation Studies]. *Rocznik Przekładoznawczy* 10: 15–40. DOI: [10.12775/RP.2015.001](https://doi.org/10.12775/RP.2015.001).
62. Biel, Łucja, Anna Jopek-Bosiacka, and Katarzyna Wasilewska. 2015. Mediacje w sprawach cywilnych w Polsce – analiza terminologii polskiej i angielskiej [Mediation in civil cases in Poland: An analysis of Polish and English terminology]. *Lingua Legis* 23: 35–52.
63. Biel, Łucja. 2013. Integracyjne podejście funkcjonalne w dydaktyce przekładu prawniczego [Integrative functional approach in legal translation teaching]. *Między Oryginałem a Przekładem* 19(3): 11–27. DOI: [10.12797/MOaP.19.2013.20.02](https://doi.org/10.12797/MOaP.19.2013.20.02).
64. Biel, Łucja. 2013. Tekst prawny jako hipertekst – o intertekstualności tłumaczeń prawnych na przykładzie międzynarodowych instrumentów z zakresu handlu ludźmi [Legal text as a hypertext: Intertextuality of translated international law related to trafficking in human beings]. *Comparative Legilinguistics. International Journal for Legal Communication* 13: 127–143. DOI: [10.13140/2.1.2027.0727](https://doi.org/10.13140/2.1.2027.0727).
65. Biel, Łucja. 2011. *Trafficking in human beings* – anglojęzyczna terminologia dotycząca handlu ludźmi i jej polskie ekwiwalenty [English terminology related to trafficking in human beings and its Polish equivalents]. *Lingua Legis* 20: 46–59.
66. Biel, Łucja. 2011. Jakość przekładu prawnego i prawniczego w świetle normy europejskiej PN-EN 15038 oraz hipotezy uniwersaliów translatorycznych [Quality of legal translation in light of PN-EN 15038 standard of translation services and hypotheses of translation universals]. *Rocznik Przekładoznawczy* 6: 13–28.

67. Biel, Łucja. 2011. Terminologia, frazeologia i stylistyka *umowy o pracę* oraz *contract of employment* i *statement of employment particulars* [Terminology, phraseology and stylistics of a contract of employment and statement of employment particulars]. *Lingua Legis* 19: 47–62.
68. Biel, Łucja. 2010. Terminologia i frazeologia reprezentacji spółki (*pełnomocnik*, *prokurent* i *przedstawiciel*) oraz ich angielskie ekwiwalenty [Terminology and phraseology connected with company representation in Poland (*pełnomocnik*, *prokurent* and *przedstawiciel*) and their English equivalents]. *Lingua Legis* 18: 26–42.
69. Biel, Łucja. 2009. Terminologia polskiego, angielskiego i amerykańskiego prawa spółek: organy spółki [Terminology of Polish, English and US company law: Company governing bodies]. *Lingua Legis* 17: 46–57.
70. Biel, Łucja. 2008. *Memorandum of Association*: Terminologia związana z zawiązaniem i rejestracją spółki *company limited by shares* [Memorandum of Association: Terminology related to the formation and registration of a company limited by shares]. *Lingua Legis* 16: 10–23.
71. Biel, Łucja. 2007. Współpraca unijna w sprawach karnych. Podstawowe terminy w praktyce translacyjnej [Union cooperation in criminal matters. Basic terms in translational practice]. *Lingua Legis* 15: 3–9.
72. Biel, Łucja. 2007. Analiza porównawcza nazw spółek w prawie polskim, angielskim i amerykańskim w wybranych słownikach [Comparative analysis of company names in Polish, English and US law in selected dictionaries]. *Lingua Legis* 15: 65–74.

G. Chapters in monographs (in Polish)

73. Biel, Łucja. 2021. Przekład prawny i prawniczy [Legal translation]. In *Perspektywy na przekład*, ed. Maria Piotrowska, 121–140. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
74. Giczela-Pastwa, Justyna, and Łucja Biel. 2021. Kooperatywne uczenie się przekładu – 15 lat działalności Studenckiego Koła Naukowego Przekładu Specjalistycznego LINGUANA [Cooperative learning of translation – 15 years of activity of the LINGUANA Students’ Specialized Translation Research Club]. In

- Dydaktyka przekładu: nauczanie przez projekty*, eds. Olga Kubińska, Wojciech Kubiński and Maciej Kur, 79–98. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.
75. Biel, Łucja, Dariusz Koźbiał, and Katarzyna Wasilewska. 2018. Zbitki wielowyrządowe w polskim eurolektie – analiza gatunkowa formułczości dyskursu administracyjnego. Badanie pilotażowe [Lexical bundles in the Polish eurolect: Genre analysis of the formulaicity of administrative discourse. Pilot study]. In *Wybrane zagadnienia lingwistyki tekstu, analizy dyskursu i komunikacji międzykulturowej – In memoriam Profesor Anny Duszak (1950–2015)*, eds. Urszula Okulska-Łukawska, Urszula Topczewska and Anna Maria Jopek-Bosiacka, 259–300. Warszawa: ILS UW.
76. Biel, Łucja. 2017. Wyrażenie *tak zwany (tzw.)* jako leksykalny wykładnik modalności epistemicznej w języku prawniczym: analiza korpusowa orzecznictwa Sądu Najwyższego [The expression *tak zwany (the so called)* as a lexical marker of epistemic modality in legal language: A corpus analysis of the Polish Supreme Court’s jurisprudence]. In *Język prawny i prawniczy. I Kongres ogólnopolski*, eds. Dorota Kondratczyk-Przybylska, Adam Niewiadomski and Ewa Walewska, 19–39. Warszawa: Uniwersytet Warszawski.
77. Biel, Łucja, Karolina Broś, and Anna Jopek-Bosiacka. 2017. Polityka zagraniczna i integracja europejska w uchwałach Sejmu w latach 1950-2014. Swoi – Inni – Obcy [Foreign policy and European integration in the resolutions of the Sejm in 1950-2014. Us – Them – Strangers]. In *Polskie sprawy 1945-2015. Warsztaty analizy dyskursu*, eds. Marek Czyżewski, Anna Horolets, Krzysztof Podemski and Dorota Rancew-Sikora, 225–259. Warszawa: Sedno.
78. Biel, Łucja, and Justyna Giczela-Pastwa. 2016. Metody korpusowe w analizie gatunków specjalistycznych – założenia, perspektywy i ograniczenia [Corpus methods in the analysis of specialised genres – assumptions, perspectives and limitations]. In *Pod pretekstem słów. Księga jubileuszowa dla Profesora Wojciecha Kubińskiego*, eds. Paula Gorszczyńska and Wioleta Karwacka, 140–168. Gdańsk: Części Proste.
79. Biel, Łucja. 2016. Wyrażenia przyimkowe w krajowych i unijnych aktach prawnych – analiza korpusowa funkcji przyimków w polskim języku prawnym [Prepositional phrases in national and EU legal acts: Corpus analysis of the functions of prepositions in Polish

- legal language]. In *Język polskiego prawa: nowe wyzwania*, eds. Dorota Kondratczyk-Przybylska, Adam Niewiadomski and Ewa Walewska, 9–35. Warszawa: Uniwersytet Warszawski. DOI: 10.14746/cl.2018.35.4.
80. Biel, Łucja. 2016. Kształcenie tłumaczy prawniczych w Polsce i w Europie: trendy i dobre praktyki [Education of legal translators in Poland and Europe: Trends and good practices]. In *Przyszłość zawodu tłumacza przysięgłego i specjalistycznego – współczesne wyzwania*, eds. Marta Czyżewska and Aleksandra Matulewska, 9–25. Warszawa: Polskie Towarzystwo Tłumaczy Przysięgłych i Specjalistycznych TEPIS.
81. Biel, Łucja. 2015. Dosłowność jako strategia w przekładzie unijnych aktów prawnych [Literalness as a strategy in the translation of EU legal acts]. In *(Nie)dosłowność w przekładzie. Od literatury dziecięcej po teksty specjalistyczne* [Język a komunikacja 36], ed. Joanna Dybiec-Gajer, 137–147. Kraków: Tertium.
82. Biel, Łucja. 2012. Kompetencja w zakresie świadczenia usług tłumaczeniowych a uczenie się przez projekt [Translation service provision competence and project-based learning]. In *Kompetencje tłumacza. Tom poświęcony Profesor dr hab. Elżbiecie Muskat-Tabakowskiej*, eds. Maria Piotrowska, Artur Czesak, Aleksander Gomola and Sergiy Tyupa, 97–105. Kraków: Tertium.

H. Other publications (incl. interviews and lexicographic works)

83. Biel, Łucja. 2017. Interview with Jan Engberg on Legal translation and knowledge mediation (Part 1 and 2). *The Journal of Specialised Translation* 27, <https://www.jostrans.org/issue27/interviews.php> (accessed October 24, 2022).
84. Biel, Łucja. 2015. Interview with Karine McLaren, Director of the Centre de traduction et de terminologie juridiques (CTTJ), Université de Moncton, Canada. *The Journal of Specialised Translation* 23: 12–17, https://jostrans.org/issue23/art_mclaren.php (accessed October 24, 2022).

85. Biel, Łucja. 2009. Słowniczek kolokacji polskiego prawa spółek [Glossary of collocations related to Polish company law] http://www.academia.edu/1746040/Słowniczek_kolokacji_polskiego_prawa_spolek_Glossary_of_Polish_company_law_collocations (accessed October 24, 2022).

List of reviewers in 2022

Hanka Błaszowska, Adam Mickiewicz University, Poznań, Poland
Onorina Botezat, Dimitrie Cantemir Christian University, Bucharest, Romania
Tomáš Duběda, Charles University, Prague, Czech Republic
Marcus Galdia, International University of Monaco, Monaco
Ksenia Gałuska, University of Silesia in Katowice, Katowice, Poland
Karolina Gortych-Michalak, Adam Mickiewicz University, Poznań, Poland
Jan Gościński, Pedagogical University in Kraków, Kraków, Poland
Stanisław Gózdź-Roszkowski, University of Lodz, Łódź, Poland
Joanna Grzybek, Jagiellonian University in Kraków, Kraków, Poland
Liu Guanghua, Lanzhou University, Lanzhou, China
Sonia Halimi, University of Geneva, Geneva, Switzerland
Clara Ho-yan Chan, University of Hong Kong, Hong Kong, China
Karolina Kaczmarek, Adam Mickiewicz University, Poznań, Poland
Alex Kasonde, Africa University, Mutare, Zimbabwe
Ljubica Kordić, University of Osijek, Osijek, Croatia
Stavros Kozobolis, Ionian University, Corfu, Greece
Agata de Laforcade, Paris Panthéon-Assas University, Paris, France
Nicholas Léger-Riopel, University of Moncton, Moncton, Canada
Pietro Manzella, University of Milan, Milan, Italy
Aleksandra Matulewska, Adam Mickiewicz University, Poznań, Poland
Federica Monti, University of Macerata, Macerata, Italy
Lavinia Nadrage, Ovidius University of Constanța, Constanța, Romania
Dan Ophir, Ariel University, Ariel, Israel
Karolina Paluszak, Jan Długosz University in Częstochowa, Częstochowa, Poland
Ivo Petrů, University of South Bohemia in České Budějovice, České Budějovice, Czech Republic
Piotr Pieprzyc, Pedagogical University in Kraków, Kraków, Poland
Theodorus du Plessis, University of the Free State, Bloemfontein, South Africa
Hammouda Salhi, University of Tunis, Tunis, Tunisia
Elsa Skenderi, University of Tirana, Tirana, Albania
Ekaterina Strati Gjergo, Aleksander Moisiu University, Durres, Albania
Virginia Vecchiato, University of Turin, Turin, Italy
Edyta Więclawska, University of Rzeszów, Rzeszów, Poland
Emilia Wojtasik-Dziekan, Adam Mickiewicz University, Poznań