CORPUS STUDY CARRIED OUT ON THREE “LEGAL” VERBS TO DEMONSTRATE THEIR SIMILAR AND DIFFERENT USAGE FOR THE PURPOSES OF LEGAL TRANSLATORS AND LAWYER-LINGUISTS

Louise Frances Denyer
Birmingham University

Abstract:
The objective of this research using corpora in translation studies and teaching was to take a small number of words or phrases and carry out a corpus study to show how they are used in similar and different ways. I chose the three “legal” verbs “breach, violate + infringe”, due to the problems that they have caused me working as a Legal translator (Spanish-English) and whilst teaching Legal English to Spanish and French lawyer-linguists. I have concentrated on their legalistic meanings and analysed their similar and different use generally, and specifically regarding rights and laws by examining the concordances, collocations and statistical measurements found in the “Bank of English” (BOE), the “Spanish Real Academia Corpus” (CREA), the “Corpus del Español” (CE) and the “Concordancier Corpus FranÇaise” (CCF) compared to these verbs’ “legal” definitions found in various dictionary types, monolingual, legal and bilingual plus comments from a French Commercial Lawyer-linguist and a practising Spanish Legal Interpreter/Translator, and considered their implications for both legal translators and lawyer-linguists.

Keywords:
Legal verbs “breach, violate + infringe”; problematic for Legal Translators + lawyer-linguists; Spanish/French to English; comparing Concordances information; comments from practitioners; future implications

I. INTRODUCTION

The objective of this research using corpora in translation studies and teaching was to take a small number of words or phrases and carry out a corpus study to show how they are used in similar and different ways. I chose to carry out a corpus study on the three “legal” verbs “breach, violate + infringe”, due to the problems that they have

1 ldenyer@wanadoo.es
caused me working as a Legal translator (Spanish/French-English) and whilst teaching Legal English to Spanish and French lawyer-linguists.

I have concentrated on their “legal” meanings and analysed their similar and different use generally, and specifically regarding rights and laws by examining the concordances, collocations and statistical measurements found in the following corpora.

First, the “Bank of English” (BOE), a corpus of 320 million words in English which is jointly held by COBUILD and Birmingham University. Secondly, the “Spanish Real Academia Corpus” (CREA), a corpus of 200 million Spanish words taken from a wide range of subjects and Spanish speaking countries, however I have restricted my search to mainland Spain and it is found under www.corpus.rae.es. Thirdly, the “Corpus del Español” (CE), a 100 million word corpus of Spanish texts funded by NEH created by Professor Mark Davies of Brigham University and consists of literature, oral texts, newspapers and encyclopedias from various centuries and is found under www.corpusdelespañol.org. Lastly, the “Concordancier Corpus FranÇaise” (CCF), a French corpus of 1,110,392 words, taken from the national French newspaper, “Le Monde” and is found under www.lextutor.ca/concordancers/concord-f.html, and consequently considered their implications for both legal translators and lawyer-linguists.

The above corpora will be referred to throughout this discussion as BOE, CREA, CE and CCF respectively.

References are made to fifteen Appendices which contain concordance lines, statistics, and translations where necessary, however due to the present purposes of this discussion, they cannot be included herein but would be available for future consultation if required.

Section 1 explains Corpus Linguistics’ background, available Corpora types, reasons why corpus studies are useful to translators and their advantages for translating and legal-language teaching.

Section 2, provides the “legal” definitions of the examined verbs from a General English Dictionary, a Monolingual Law Dictionary and both Spanish and French Bilingual Legal Dictionaries, plus linguistic comments from one Legal English student, who is a French Commercial Lawyer-linguist and a practising Spanish Legal Interpreter/Translator, upon which I propose to compare their legal-linguistic experience and knowledge with information obtained from a corpus study on the same verbs. Finally, Section 3 demonstrates my findings arising from analysing each verb in turn from the four mentioned corpora, and discusses how they could assist legal translators and lawyer-linguists.

---

II. CORPUS LINGUISTICS

2.1 A corpus is a word derived from the Latin meaning “body” and refers to any text in written or spoken form taken from large collections that represent a sample of particular varieties or uses of language, which nowadays are presented in machine-readable form. Such computer-readable corpora can consist of raw text i.e. plain text with no additional information, or include annotations. Although, “corpus-based research is often assumed to have begun in the early 1960s with the availability of electronic, machine-readable corpora.” (Kennedy 1998: 13), the use of corpus and collocation of text in language study dates back to the Middle Ages with Bible concordancing. However, improved computation capacity and speed have increased the development of computer-stored based corpora and the use of corpora to study natural language.

Thus, “the essential vision underlying Corpus Linguistics is that computer-assisted analysis of language gives access to data….. previously unobservable … can now profoundly change our understanding of language.” (Stubbs 1996: 46).

Sinclair states,

Analysis of extended naturally-occurring texts, spoken and written, and in particular, computer processing of texts have revealed quite unsuspected patterns of language… The big difference has been the availability of data….The major novelty was the recording of completely new evidence about how the language is used….The contrast exposed between the impressions of language detail noted by people, and the evidence compiled objectively from texts is huge and systematic….The language looks rather different when you look at a lot of it at once (ibid: 46).

However, corpus studies also “depend on the interpretation of frequency and distributional data” (ibid: 47).

2.2. Modern Corpus Linguists such as Sinclair challenge the traditional linguists like Chomsky concerning native-speaker intuition that it provided the key to language, that the “legitimacy of example sentences ..have to pass the empirical test of the judgement of native speakers as to whether they are possible within the language” (Kennedy: 1998: 270), and although it is clearly important while translating into a Target Language to ask a TL native speaker to check translations for accuracy and idiomatic expressions, intuition can be unreliable in areas concerning collocations, frequency, lexis and grammar. However, modern Corpus Linguists can provide empirical information and contribute to language pedagogy, “in contrast to the possible arbitrariness and unreliability of intuitive judgements” (Aijmer 1991: 110).

Furthermore, corpora have had great impact on the classical distinction between grammar and lexis and corpus linguists like Sinclair have through corpora argued that grammar and lexis operate dependently, and have merged together the traditional belief that grammar concerned the rules for language and lexis the detail namely the actual words, “that grammatical generalisations do not rest on a rigid foundation, but are the
accumulation of the patterns of hundreds of individual words and phrases. The language looks rather different when you look at a lot of it at once”. (Sinclair 1991: 100). Therefore, words are effectively dependent on their contexts.

2.3. Corpuses have grown in both size and areas of speciality and are used by a varying range of specialists from different disciplines including translators and language teachers.

Some examples of commonly-used corpora are:

i. Specialized: aims to be representative of a given type of text i.e. academic articles, textbooks e.g. Cancode Corpus of Spoken English.

ii. General: “...is made up of texts assumed to be representative of everyday, non-specialized language” (Laviosa 1997: 292), including as a wide spread of texts as possible if not representative and large e.g. BOE.

iii. Comparable: where 2 or more corpora in different languages e.g. Spanish and English designed to compare language across languages.

iv. Parallel: 2 or more corpora in different languages containing texts translated from one language to another. Both iii. and iv. are particularly interesting to translators and can be used for finding possible translation equivalents in each language. Other corpora used in translation include bilingual, multilingual or monolingual comparable and both bilingual and multilingual corpora may be parallel or comparable.

2.4. There are two mainstream corpus linguistics methodologies:

a. Word-form based: where only minimal tagging is used and is carried out automatically by the computer solely, thus the emphasis is on the behaviour of individual words and phrases rather than categories, and the linguist uses “raw data” for example collocation and phraselogy are all by definition word based.

b. Category based: where methods annotate or “tag” the corpus so that particular categories can be counted and compared and the corpus data is already classified.

Hunston and Laviosa favour the word-form approach that “raw, unannotated corpus is the best starting point in any investigation” (Hunston and Laviosa 2003: 103) as it removes pre-conceived suppositions and “challenges our existing views about language” (ibid: 104). However, she concedes that perhaps both the two different approaches to corpus design and research should be seen more as complimentary and unconflicting “perhaps the fairest thing to say about (both approaches) is that each one can be used to answer particular questions (ibid: 103).

2.5. Corpora especially bilingual corpora are useful to translators to understand SL texts to identify translation equivalents, improve fluency and accuracy and check idiomatic and fixed expressions especially for legal translators needing to understand complex legal concepts and choose correct terms and expressions. Furthermore, corpus based research is very relevant to translators who can draw on the insights gained in
different and similar languages, become aware of strategies adopted by expert and experienced translators, recognize features of translational language that are independent from the SL and gain knowledge of the most frequently used translation equivalents of a given word or phrase. Native speakers cannot always rely on their intuitions, and corpora can demonstrate to a translator how best to find appropriate translation equivalents and be in keeping with specialized, technical jargon such as legal translational styles.

2.6. The advantages of corpora are that they use authentic texts, are valuable in identifying translation equivalents, their frequencies and quick language retrieval. Obviously, a computer unlike humans can provide relative frequency of lexical and grammatical items, collocations, semantic prosody, pragmatic meaning and phraselogy details.

However, corpora do have shortcomings, computers make mistakes, written language tends to be vastly overrepresented and spoken language underrepresented in corpora. Also, concordance lines for translators are often too short to offer enough context or background especially essential for specialized areas such as legal translation.

Baker argues, that by using Corpus Linguistics techniques and tools to study translation “..allows translators to understand what actually happens in the process of translation “ (Baker 1996: 177), but stresses that, “translation scholars are not interested in the words or syntactic structures…. (but) in abstract global notions such as explicitation and simplification “ but concedes that “some applied extensions such as using concordances to establish equivalents are useful to translators” (ibid: 185) and the translator should not use corpora for solely identifying language patterning but to understand “ why a particular translational language exhibits certain features” (Hunston + Laviosa: 177).

Furthermore, corpora can demonstrate how language in use transmits culture in “cultural keywords” that reveal dominant ideology in the culture that “evidence from corpus, concordance and collocation can help to explain class, codes and control” (Stubbs 1996: 195).

III. DICTIONARY VERB DEFINITIONS AND LEGAL COMMENTS

3.1. I chose to carry out a corpus study on these three legal verbs as some of my legal-linguists students asked how they should be using them correctly generally and regarding laws and rights. Also, translating I discovered that one verb was more suitable in specific contexts in one language, but not in the same context in the second language. It could be questioned whether such differences are that important, and apart from consulting the standard monolingual or bilingual legal dictionaries which tend to record “..only collocations representing well-established legal concepts (Hunston + Laviosa 2003: 167), legal translators/legal-linguists seeking answers for less established legal
concepts need more information which perhaps standard legal dictionary explanations, intuition or professional legal experience and knowledge cannot fully provide therefore they should consult their corpora tool.

All three verbs could be classified as partial synonyms as, “lists of near synonyms are a common general characteristic of legal English” (Stubbs 1996: 109) as, “both synonyms and polysemy are abundant in the (English) legal language” (Alcaraz 1994: 84) and quotes the English “legal verb to “annul” which has also many partial synonyms such as “abolish, override, set aside, quash” (ibid: 84).

Legal language is “ambiguous and complex” (Stubbs 1996: 101), and I frequently explain to legal students and translation clients that legal English needs to be learnt and understood like foreign languages before its linguistic implications can be accurately rendered into TT, as finding translation equivalents are often complicated especially when they do not exist between two radically, different legal systems such as the English Legal System (common-law based) and the Spanish and French Legal Systems (civil-law based).

TT has distinctive, universal features such as “distinctive legal vocabulary” (Aijmer 1991: 225) and these features of translated text, “include simplification, (the idea that translators subconsciously simplify the language or message of both), the explicitation (tendency to spell things out…) and normalization or conservatism (the tendency to conform to patterns and practices which are typical of the TL, even to the point of exaggerating them)” (Baker 1996: 176) and such features regularly appear in legal translation.

3.2. Tables showing Dictionary Definitions:
3.2.1. “Breach” is defined as:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I.4.1. Collins Concise Dictionary</td>
<td>I.5. “breaking, infringing or violation of a promise or obligation” e.g. the contract was breached.</td>
<td>I.6. This clearly reflects the merging of such partial synonyms. I.7. It is not clearly defined for linguist-lawyers and remains ambiguous and confusing.</td>
</tr>
<tr>
<td>I.8.2. Peter Collins Legal Dictionary</td>
<td>I.9. “i. failure to carry out the terms of an agreement ii. failure to obey the law”</td>
<td>I.11. More specifically defined.</td>
</tr>
<tr>
<td>I.12.3. Bilingual Spanish Legal Dictionary</td>
<td>I.13. The following translation equivalents are listed; incumplir, contravenir, violar,</td>
<td>I.15. Certain collocates of breach are given and some fixed legal expressions, e.g. breach of contract,</td>
</tr>
</tbody>
</table>
vulnerar”, breach of confidence, breach of the peace, breach of trust, followed by their Spanish translations: Breach of contract = incumplimiento de contrato.

I.17. Breach of the peace = delito de alteración del orden público.
I.19.
I.20.

I.21.4. Bilingual French Legal Dictionary

I.22. The following translation equivalents are listed;
I.23. “enfreindre, rompre, ne pas respecter, manquer.”

As above, certain collocates of breach are given and some fixed legal expressions with French translations, e.g. Breach of contract = manquement à une obligation contractuelle/ rupture de contrat.

I.27. Breach of trust = manquement aux obligations de fidéicommis.

Table 1: Definition of Breach

2.2.2. “Violate” is defined as:

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Definition</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collins Concise Dictionary</td>
<td>“to break, disregard or infringe on (a law or agreement)”</td>
<td>Mention is made to rape and sexual assault.</td>
</tr>
</tbody>
</table>
I.34.2. Peter Collins Legal Dictionary

I.35.“To break a rule or a law” e.g. the Council has violated the planning regulations.

I.36.

I.37.3. Bilingual Spanish Legal Dictionary

I.38.Provides the following Spanish translation equivalents, “violar, infringir, vulnerar”.

I.39.All again partial synonyms and it is explained that “violation” is used under English law in expressions such as parking offences, minor and regulatory offences.

I.40.4. Bilingual French Legal Dictionary

I.41.Provides the following French translation equivalents, “violer, contrevenir, enfreindre”.

I.42. It gives examples of, I.43. to violate a law = violer un loi, and I.44. It is used in French law for major criminal offences like rape.

I.45. Dictionary I.46. Definition I.47. Comments

I.48.1. Collins Concise Dictionary I.49.“To violate or break (law, agreement).” I.50.Again the meaning is blurred.

I.51.2. Peter Collins Legal Dictionary I.52.“To break a law or right” and gives the typical examples of to infringe a copyright, a patent.

I.53.More specific.

I.54.3. Bilingual Spanish Legal Dictionary I.55.Provides the following translation equivalents, “infringir, violar, vulnerar” and offers the explanation that infringe is used mainly regarding patent (derechos de patentes) and trademarks (marcas)

I.56.More helpful.

Table 2: Definition of Violate

2.2.3. “Infringe” is defined as:
comerciales) and copyrights (derechos de autor) e.g. it was a clear infringement of copyright.

I.57.4. Bilingual French Legal Dictionary
I.58. Provides the following translation equivalents, “passer outre, contrefaire, ne pas respecter, infraction”.
I.59. “Contrefaire” is normally used for Intellectual Property (IP) matters for patents (brevets) and trade marks (marques).

Table 3: Definition of Infringe

These definitions taken from conventional reference material concur with the comments made below by the French legal English student and the Spanish Legal Translator.

2.2.4. Legal Comments:

|------------------|------------------|--------------|
| I.63.a. Spanish Legal Translator/Interpreter | I.64. “…these verbs can be used in any of the forms, they can cause difficulties of course, but there are also main differences e.g. infringe is used principally with patents, trademarks, copyrights, violate with parking offences, petty offence with intention, whereas breach is more general and can be used instead of infringe or breach. Breach has direct translation equivalents such as “breach of contract = incumplimiento del contrato, breach of confidence = abuso de confianza, breach of authority = abuso de poder”.
| | | L.65. Her comments are based on her use of legal dictionaries and Internet legal consultations and clearly re-echo the bilingual legal dictionary definitions. |
| I.66.b. French lawyer-linguist student | I.67. “Breach is more neutral more black and white, regardless |
| | | L.69. Similar perspective. |
of intention and is used in relation to contracts (infraction, rupture or ne pas respecter).

I.68. Whereas, violate (violer) is applied to more general law and has greater negative and wilful intention. Infringe (enfreindre/ne pas respecter) is closer to “breach” in meaning but has less intention and is often used for Intellectual Property and Patent Law.

I.70. I.71. I.72. Perhaps both of them would benefit from using corpora?

Table 4: Legal Comments

Their opinions are obviously based on traditional materials however a corpus search could throw light on such issues as frequency, collocations and challenge us to move into the exciting corpora world.

It appears that these verbs are indiscriminately employed and there is no clear evidence of any linguistic or syntactic patterning, which means do legal translators and students have to rope-learn these well-established translation equivalents? Or can the linguist’s corpora tool help to develop reasoning for such equivalents and provide translators with clearer examples and explanations?

IV. CORPORATE STUDY ON THE THREE LEGAL VERBS UNDER EXAMINATION.

IV.1. Corpus study of “breach”.

First, I consulted the BOE corpus for the search string, “breach@+1,5NOUN”, 10,203 matching lines were produced from which I selected 100 random lines and deleted irrelevant lines (Appendix 1). OZ News (Australian) has the highest total number of occurrences 1,514 with 43.3 million (average number per million words)³, followed by the British Times 2,050/ 39.5 million, whereas all the US corpora rank near the bottom e.g. US News 195/ 19.5 million or US Acad 44/ 6.9 million.

It would appear that breach is more commonly used in Australian and British English corpora and particularly in UK broadsheet newspapers, which is important as translators

³ The BOE sets out the corpus statistics in three columns, the first lists the corpus source, the second the total number of occurrences and the third the average number million per words.
should always check their client’s nationality and “gear” TT to their linguistic culture and background thereby upholding accuracy, fluency and idiomatic expressions. Accuracy is essential for legal translation as negligent inaccuracies could incur personal liability for translators and/or their clients.

There are two principal, grammatical patterns, the fixed structures of “in breach of (the) +noun” (10%) e.g. “in breach of the agreement” (L8) and “a breach of (the) +noun” (14%), e.g. “a breach of the peace” (L60) that appear almost more frequently than the conjugated verb. Translators should be aware of these patterns and accurately transfer them to TT.

Surprisingly, adjectives indicating the extent of the breach are revealed such as “serious, almighty, shocking, significant, flagrant, clear” (L4, 35, 75, 67, 72, 96). The conventional materials maintain that “breach” has a general, less wilful intention but these express the opposite.

Furthermore, “breach” forms part of various fixed expressions, it frequently collocates with the following words, trust 2%, fiduciary 3%, confidentiality 4%, security 5%, contract 6% and peace 8% that demonstrate how these idiomatic expressions are often used and translators/lawyer-linguists should recognise them and know their relevant translation equivalents.

I.73. station and charged with a breach of the peace. I was just.
I.74. a logjam. Charges of assault, breach of the peace and vandalism.
I.75. using conduct likely to cause a breach of the peace by squirting.
I.76. expansion is a technical breach of the peace accord with the.
I.77. <p> A: Any offence, sir, is a breach of the peace. Blocking a.
I.78. altogether model citizen, his breaches of the peace and penal code.
I.79. this is comparable to a breach of the peace or a road traffic.
I.80. they should all be charged with breach of the peace and criminal.
I.81. she was being sued with fraud, breach of contract and ‘conversion”,
I.82. against the energy company in a breach-of-contract dispute with a.
I.83. to against the energy company in a breach-of-contract dispute with a.
I.84. to sue the League for breach of contract. As Pearson was a.
I.85. claiming unfair dismissal and breach of contract. <hl> I’m free.
I.86. gossipy tales for fear of breaching confidentiality agreements.
I.87. This was considered a breach of confidentiality and.
I.88. the report, Mr Kiley would be breaching confidentiality agreements.

Table 5: Concordance Lines for Breach

Some of these fixed, legal concepts are translated using Bakers translational techniques of explication and simplification such as “breach of trust” and “breach of fiduciary duty”, because these English legal concepts do not easily translate into Spanish or French as they do not strictly mean the same or exist under Spanish or French law. The former is given five lines of explanation and examples in the bilingual Legal dictionary therefore translators must decide which concept or expression more accurately renders the concept in the TL, whereas the latter is simplified or normalized and given a literal translation that may be insufficient for TL clients.

Also, only 4% of the lines refer to laws (including Acts) e.g. European and Take-Over, and 3% to rights.

Its T-score\(^4\) highest lexical collocates include “peace, security, contract, confidentiality, rules, rights, fiduciary” that coincide with the above findings, and its MI

\(^4\) “T-score is a statistical significance measure based on the difference between observed and expected results” (Barnbrook, 1996: 170).
score\textsuperscript{5} highest lexical collocates include, “fiduciary, confidentiality, treaty, peace contract, laws, security, rights”.

From Picture\textsuperscript{6}, we discover three words occurring to the right of the node word, “breach”, that “peace, the contract, in confidentiality rights, fiduciary, laws” all rank high, (Appendix 2), therefore indicating how these fixed expressions are recurrent and their idiomatic translation equivalents should be used.

IV.1.1. I then typed in the search string, \texttt{“breach@+1,5right”}, only 65 matching lines appeared (Appendix 3). The UK Times Corpus had the highest total number of occurrences 35 with 0.7 million average number per million words, followed by the Canadian and UK Guardian confirming again that “breach” is more frequently used in UK/British corpora than US.

Breach is used for a general and wide range of rights such as basic human rights, free speech, constitutional, privacy, fair trial (L31,33,46,48,49+58), right to sleep, silence (L39), education and are “individual, personal” rights of children, motorists, etc. Curiously, 15% of these matching lines include various modal verbs and adverbs of probability e.g. may amount, probably, possible (L21,31,25+23), and the subjects concerned are varied, the Home Secretary, government, dog owners and operations (L35,14,13+50).

IV.1.2. However, under the string \texttt{“breach@+1,5law”}, 396 matching lines were generated, and from 100 random lines (Appendix 4), the degree of intention is not neutral but wilful as words like “clear, flagrant, unjustified, substantive, very serious, wicked, gross, grave” appear (L74,53,53,61,71,98,81+19). Also, the verb types found also express criminal connotations, “guilty, claimed, alleged, to try, charged with, punish, (L26+71,17,47+65+85,48,56+60)”. Breach mainly refers to general and modern “social” types of laws e.g. education, criminal, environmental, health and safety, planning, company, lottery law, financial + take-over.

However International and European Law rank the highest with 9% and 13% respectively and the (general) law with 17%. Thus breach is commonly collocated and employed with International and European law and translators should remember this when translating International and European legislation.

IV.1.3. I also consulted the CREA\textsuperscript{7} and looked at the verb “incumplir” the Spanish equivalent to “breach”, 155 matching lines were generated and 6% refer to laws and 7% to contractual issues, which shows that it is employed in Spanish contractual matters, (Appendix 5).

However, where the English uses “breach the ceasefire”, the Spanish verb “violar” is employed “violar el alto fuego” (Appendix 10, L100).

\textsuperscript{5} “Mutual Information (MI) score is a statistical significance measure based on the ratio between observed and expected results” (ibid: 169).

\textsuperscript{6} The CAS feature “Picture” shows the collocates of a node word arranged according to where they occur before or after the node word.

\textsuperscript{7} Technically, it has been impossible to copy and reproduce the concordance lines from the CREA Spanish Corpus, therefore full access references have been provided and relevant examples are set out in Appendix 6.
IV.1.4. In the CE⁸, 14 matching lines appeared for “incumplir” and 4/14 refer to acuerdos and 3/14 to compromisos. This corpus also lists its synonyms which include “contravenir, eludir” which conform to its general, contractual nature.

IV.1.5. Furthermore, when I consulted the French CCF, I discovered that only one matching line appeared for the verb “enfreindre”, which was used in relation to “eating habits” and that its actual meaning had changed, as it was best translated into English by the verb “to upset”, “s’interroge, peut-on, sans enfreindre les rites alimentaires, consommer des…” (it is argued whether one can, without upsetting eating habits, consume some.” (Appendix 15, L1).

However, for the verb “rompre”, 30 matching lines appeared which were used for legal and political contexts such as “the truce, with the central wing of the UDF coalition, this alliance, the strike, the isolation of the Trade Unions” (Appendix 15, L2), and more importantly included contractual issues such as, “the local council should breach the contract with the Lyonnaise Bank” (Appendix 15, L3) and other references to contracts are made.

Similarly, for the verb “ne pas respecter”, 30 matching lines appeared but this seemed to be used for more general economic and business matters like, “tariff regulations, international undertakings, corporate aids, salary negotiations and freedom of the press”, (Appendix 15, L4-8). Although, the French lawyer-linguist student had maintained that this verb was used with regard to contract law.

Nevertheless, when I carried out a search under the French noun “droits” (rights), 239 matching lines appeared, and this verb principally collocates with human and personal rights, as does the verb “violier” especially with reference to basic individual rights as discussed later in 3.2.5.

Finally, for the verb “manquer” there were only 4 matching lines, but these, when translated into English, were more in line with the legal sense of “breach”, plus they matched the standard bilingual dictionary definitions. In the examples given, this verb was used “to sanction (countries) for breaching their undertakings” and “reinforcing sanctions for breaching professional codes of practice” (Appendix 15, L9+10).

However, it could be argued by both lawyer-linguists and legal translators that it would be more accurate to translate this sentence by using the English verbal structure “failing to comply with”. Accordingly, this point highlights the fact that legal language is extremely ambiguous and only enhances the difficulties that legal translators encounter when translating between different languages and legal systems.

IV.2. Corpus study of “violate”:

I then consulted the string, “violate@+1,5NOUN”, 6,199 matching lines were generated, significantly less than “breach”, 100 random lines were selected and certain irrelevancies deleted (Appendix 6).

There is a striking contrast with the source corpora as it is US dominated and all US corpora rank the highest, e.g. US News 594/59.4 million, NPR 1207/54.3 million, US Acad 302/47.6, whereas British corpora rank the lowest, unlike “breach”, e.g. Econ 184/11.7 million.

---

⁸ The contents of Footnote no. 11 also apply to the CE Spanish Corpus.
Translators should remember this and verify whether they are translating for a North-American readership or legislation in order to comply with standard translation requirements and produce TT that meets with the cultural requirements and expectations of TL readers.

There is no fixed grammatical patterning unlike “breach”, it is quite variable and includes the “verb + the Noun” e.g. violate the charter (L54), or “verb + ( indefinite article + Noun” e.g. violates any law (L4). However, violate does appear in over 40% of the lines in V-ed form e.g. “had violated election law, has violated any provision, the company had violated” (L8,56+12). There is no indication of wilful intention as only positive adverbs are found “consciously, certainly, actually” (L26,29+54). Also, over 10% of the lines come from US contexts as indicated by their lexical contents, e.g. “had violated congressional conditions, he’s filing a lawsuit, executives violated the federal utility holding” (L19,75+93).

Violate effectively collocates 7% with law and 5% with rights, similar to “breach” and collocates with wide-ranging legal nouns such as “rules, codes, covenants, terms, prohibitions, court injunction” (L7,25,31,58,41+40).

IV.2.1. In order to check my findings I ran another search with the string, “violator@+1,5right” and 132 matching lines were generated, more than double compared to breach (Appendix 7). Again, the majority from US corpora, e.g. US News 15/1.5 million, NPR 33/1.5 million and is clearly indicated from 20% of the lines’ lexical context, (L2,13,20,62,128+129). These “cultural keywords” highlight these examples’ cultural context and this small sample provides interesting examples like the controversial US guns and arms laws (L13) plus the unique US religious reference to the Amish people (L62). A mass of different types of rights are listed including, freedom 4%, fair trial/hearing 4%, privacy 9%, free speech 11% and almost 20% dedicated to constitutional rights.

Therefore, translators when translating texts concerning US constitutional rights should employ this verb.

Another interesting cultural aspect is that the parties violating the right concerned are often official or national organizations, e.g. the police (L44,45,79+121), the government, Congress, the administration and officials (L33,16,78+111) which indicate that “violate” is applied to national matters in the US, contrary to what the traditional, British-based legal materials indicate namely minor offences which the US corpora fail to mention.

IV.2.2. Then, I did the search string for “violator@+1,5law” and 478 matching lines were produced, more than the search with right and the same type of search with the verb “breach”. From 100 random lines (Appendix 8), the corpora source is decidedly US dominated e.g. US News 58/5.8 million, NPR 128/5.8 million and the lines’ lexical content demonstrate this, as 10%+ refer to US federal/state law and 12% refer to International law.

However, there are two mentions to “misdemeanours, (L92+76) which are effectively minor offences, but the violating parties are individuals like, “clients, Baptist
Church, anti-abortion protesters (L95,38+58), another cultural reference to the US controversial topic of anti-abortion.

The t-score top lexical collocates of, “rights, law, rules, privacy, constitution, international” and the MI score top collocates, “privacy, constitution, code, guilty, laws, rights, rules, international” are similar and uphold the above corpus results.

“Picture” clarifies again what was discovered in the corpus search namely that violate as the node word collocates most highly “with the law of …this rights….an European to …not privacy…international” (Appendix 9).

IV.2.3. From the CREA the Spanish equivalent verb “violar” refers to rape cases and from 100 random lines, 26% refer to rape cases, inexperienced translators should be aware of this as modern legal English uses the criminal term of “rape” and not “violate”, 13% to laws and only 7% to rights and 10%+ of the link verb used is “acusar” (accuse). It is also used with a number of different nouns, “law, privacy, undertakings, secret, ceasefire, airspace”.

Interestingly, another Spanish verb “vulnerar” is applied similarly to law 17% (ley) and rights 14% (derechos) in the matching lines. “Vulnerar” is often used for privacy rights (derecho a la intimidad) and EC free competition rights (la libre competencia), perhaps translators should use this verb when referring to such rights. (Appendix 10).

IV.2.4. However, from the CE collocations, 33/57 refer to laws (leyes) and only 7/57 to rights (derechos), e.g. los derechos del ciudadano, (citizens rights), el derecho de las gentes de la hospitalidad (peoples’ right to hospitality). Its highest synonym is “forzar” to “force” that implies wilful intention in Spanish, unsurprisingly considering the numerous rape examples.

IV.2.5. From the French CCF, 3 matching lines appeared for the French equivalent verb “violer”, the first example referred to “airspace” (as mentioned above in 3.2), and the second like the Spanish example also concerned the subject of rape, “…even if at the end, on the 13th of September 1993, while raping and murdering the little girl Karine…” (Appendix 15, L13). However, the third example was curiously in relation to “violating controlling export regulations” (Appendix 15, L12).

Also, I noticed that this verb is often used in a criminal context and many criminal related verbs such as “to condemn, damage, harm, investigate, abuse” and nouns like “corruption, aggression, libel, slander, suspicious grounds and danger to others” are just several listed in the matching lines of the CCF. Furthermore, this confirms what the French lawyer-linguist student commented that “violer” has greater negative and wilful intention and therefore it is unsurprising that this verb collocates with other criminal-based words.

I decided to explore this French verb further and typed in the French word “droits” (rights) and from 239 matching lines, I found that the verb “violer” was employed for a vast selection of rights, in particular 30% were used for human and individual rights (human, personal, children, prisoners), 6% for civil rights such as voting and constitutional rights and surprisingly, 3% were employed for Intellectual Property rights like copyright (droits d’auteur), in contrast to the definitions given in the traditional bilingual dictionary.

Thus, it would seem that although the French verb “violer” is used for specific legal areas such as rape crimes and airspace, at the same time it is a very flexible verb and can be used almost indiscriminately for a much wider range of rights.
Consequently, this only further complicates the work of legal translators and means that thorough linguistic understanding and legal knowledge is required from them, in order to be able to correctly deal with such legal concepts in their everyday translation tasks.

IV. 3. Corpus study of “infringe”.

I consulted the string, “infringe@+1,5NOUN” and 937 matching lines were given, considerably less than both “breach” and “violate” and a random 100 lines were selected (Appendix 11).

The corpora source is completely mixed although the highest average number per million words comes from the UK New Scientist 51, 6.5 million, US News 57, 5.7 million, bbc 66, 3.5 million, the Times 176, 3.4 million and Strathy (Canadian) 53, 3.3 million, thus it is used primarily in British plus to a lesser extent US/other corpora.

It has two main grammatical patterns, “Verb + upon + the Noun” e.g. infringe upon the privacy” (L68) and “Verb+on+ the Noun” e.g. infringe on the rights” (L79), translators should use these structures when translating from their SL into English.

Additionally, there are several adverbs of wilful intention e.g. “systematically, grossly, directly, inadvertently, undoubtedly, wilfully” (L9,27,26,38,45+89) that indicate that infringe implies wrongful intention contrary to the traditional materials definitions but similar to my corpus study findings for breach, e.g. “rights were systematically infringed”(L9).

Infringe expectedly collocates with Intellectual Property (IP) rights in over 18% of the selected matching lines and significantly 24% with any rights unlike breach and violate that have much lower matching lines 3% and 5% respectively, and only 7% with law.

IP rights dominate and there are various examples including patents, copyrights, (L17,18,64,63,19+30).

The rights listed are again very “personal” and include “fundamental, basic (human) rights, workers, MP, British-born Asians”, plus “social-activity” rights, “to hunt and fish, privacy, fair trial,” and “business type” rights such as “commerical, set up a business”.

IV.3.1. I typed in the search string, “infringe@+1,5right” and only 57 lines were matched (Appendix 13).

The corpora source is predominantly British and Canadian and UK broadsheet newspapers feature high like the Times, Independent, e.g. The Times 15 total number of occurrences and 0.3 average number per million words whereas the US corpora rank the lowest.

The grammatical patterning is the same as above although more personal possessives are used 14% e.g. his right, its right, their right etc. and over 42% of the matching lines use the V-ed patterning, “this infringed its right”, “provisions infringed upon a womans’ right” (L49+56).

The rights are, “personal” and cover the varied rights of the accused/defendant, prisoner, dancer, womens to travel, to medical treatment, to have a family, to undisturbed sleep, to privacy, fair trial, (L4/32,41,7,36,40,44,45,47,53) and only 4% refer to IP. Translators should note when translating into British texts that “infringe” should be used for rights, unlike US texts or legislation where “violate” is idiomatically appropriate.
IV.3.2. Finally, I looked up the string, “infringe@+1,5law” and only 26 lines appeared (Appendix 13).

The source corpora is also mixed but seems to lean towards UK corpora as the Economist has the highest score followed by the US News and the Guardian, 4/2/8 and 0.3/0.2/0.2 million respectively.

From this tiny sample, the grammatical patterning is based on “Verb + (determiner) Adjective + Noun” e.g. infringed the Islamic law (L11).

There are 4 lines regarding European law and two provocative US examples, “infringed a Michigan obscene-speech law, infringing the Gun and Sword Control Law (L 16, 5). Plus, examples of severe consequences for infringement, “penalties, found guilty, huge fines, arrested for” (L7,15,15+5).

The top t-score lexical collocates of infringe, “rights, law, copyright, patents, sovereignty and the top MI score collocates “patents, copyright, sovereignty, privacy, rights “all demonstrate how highly it collocates with rights especially IP rights as shown by its raw frequency score of 18.

Furthermore, “Picture” confirms this as two words to the left of the node, rights figure high and likewise to the right of the node, “we have, their rights, by rights, copyright” all showing how important rights are to this verb, (Appendix 14).

IV.3.3. However, from the Spanish CREA, only 5/52 lines refer to rights and three regarding IP and 7/52 refer to laws. It is difficult to establish any conclusive evidence from such short concordance lines, however it is interesting to identify and discover translation equivalents and indications of when to use them.

I.89. Line 29: “from data bases with identical contents without infringing the intellectual property rights”.
I.90. Line 31: “finished invention may be put into practice without infringing the rights derived from a patent”.

Table 6: Spanish Concordance Lines for Infringe

IV.3.4. The same applies from the CE concordance lines for “infringir”, there are 65 lines in total and 20/65 (i.e. 32%) refer to laws, whereas only 1 refers to rights, which does not indicate that “infringir” is mostly used with rights and IP, however Spanish legal practice does use “infringir” for IP rights.

IV.3.5. Finally, from the French CCF, the French equivalent verb for “infringe” such as “passer outre”, appears to take on a completely different meaning from that explained in the standard French bilingual dictionary, plus it is used in a political rather than a legal context.

I have found that it is better to translate the 5 listed matching lines that appear in the CCF, (Appendix 15, L15-19), into English by using the verb “to disregard”, which renders the true meaning of the examples given such as:

I.91. “the elected members of the five departments were ready to disregard the EDF and would throw in the towel (give up) in 1988”.
I.92. “how the psychological impact of the debates from next Monday or Tuesday would be

---

9 Translated from the Spanish Concordance lines, for the verb “infringir” screen 2: no. 29 +31, (Concordancias. Pantalla 2: No. 29+31)
29. “..de bases de datos con idénticos contenidos sin infringir los derechos de propiedad intelectual.”
31. “..terminada invención puede ponerse en práctica sin infringir los derechos de una patente.”
disregarded”.
I.93. “he decided to disregard the advice of his friends and maintain his candidature”.
I.94. “it was politically impossible for Chancellor Kohl to disregard the opinion of the Central Bank”.
I.95. “finished by accepting Italy into the Euro and disregarded the expected reservations”.
Table 7: French Concordance Lines for Infringe

There was no mention of the French verb “contrefaire”, which the traditional dictionary materials uphold are used for Intellectual Property matters but I did discover that the French verb “violer” was often employed for copyright matters (see 3.2.5.).

Finally, I run a corpus search for the French word “lois” (laws) and found from 492 matching lines that the noun “infraction”, (noun equivalent for infringement or breach), not to be confused with the French criminal term that indicates general offences, was employed in 4% of the lines regarding the “infringing of laws”. Interestingly, the French lawyer-linguist student also mentioned that this noun was often used, but was normally translated into English by the equivalent verb of “breach”. Also, 6% of the lines contained the verb “violer” i.e. “violating laws”. Again, this demonstrates the blurry divisions that exist between these verbs and highlights the arduous undertaking facing legal translators when deciding how such verbs should be best translated into English.

Furthermore, this illustrates how under French law, the verb “violer” is extremely versatile and is often employed for wide-ranging legal areas and can be translated into English, depending on the legal context, by any of the three legal verbs currently under examination. Whereas, the standard legal verbs mentioned both by the French lawyer-linguist student and the relevant bilingual dictionary entries are not even listed in the examples thrown up by the CCF.

V. CONCLUSION.

I have discovered from this corpus study regarding the similar and different use of these verbs firstly from the BOE that “breach” is employed significantly more in all legal senses than “violate” and “infringe”. However, the search for these “verbs + rights”, reveals that “violate” is used considerably more than “breach” and “infringe” and the search for these “verbs + law”, also demonstrates that “violate” has more examples than “breach” and “infringe”.

According to the source corpora, “breach” is very British whereas “violate” is overwhelmingly American and “infringe” is mixed. Furthermore, contrary to conventional materials and surprisingly from this corpus study, “breach” appears to imply wilful intention from the forceful adjectives found unlike the others. This should be reflected in translators work and they should always check the “cultural” identity of their TL readership and gear the TT to its specific cultural expectations.

Regarding their grammatical patterning, they all have set structures which translators need to imitate in their English TL versions. As for their uses, “breach” has its fixed expressions as explored above and translators should reproduce each individual, translation equivalent e.g. breach of contract, breach of the peace.

Furthermore, legal concepts such as “breach of trust” and “breach of fiduciary” may have to be rendered applying Bakers translational techniques of simplification and explicitation, in order to accurately reflect the SL concept in the TL version.

“Violate” is used in many contexts particularly in US corpora and “infringe” is used considerably for IP law which should be relayed in TT.
Although, the Spanish and French corpora contain fewer examples than the BOE, it has been interesting to examine their similar lexico-grammatical patterns and examples plus identify the similar and different uses of the Spanish and French equivalent verbs.

For example, the Spanish verb, “incumplir” is certainly the equivalent verb for “breach” and the verbs “violar” and “vulnerar” seem to fulfil similar functions to the English verb “violate”, although “vulnerar” is employed more with “rights” (derechos).

The French verb “violer” is used in a similar way as its Spanish equivalent for rape crimes and international issues, such as “violating international airspace”, and for both general and specific human, personal and social rights. However, it is clearly shown how “violer” can be employed in many legal areas, which legal translators and lawyer-linguists should take into account when translating this verb into English. Furthermore, it is demonstrated that the French equivalent verbs for “breach” only reflected to a limited extent what the standard and traditional translation tools state, as the verb “enfreindre” is hardly mentioned in the CCF, and the verbs “rompre” and “manquer” could be effectively translated not only by breach, but often by other more appropriate verbal structures as explained above. Also, from the CCF, the standard French verbs used for “infringe” like ”contrefaire” in relation to IP rights strangely enough did not figure in any of the matching lines, but the verb “violer” was used instead for copyright matters.

It would appear that in many cases, the basic contents and examples of the standard translation materials and the opinions of the Spanish translator and French lawyer-linguist differ, to varying degrees to what was actually recorded in the particular corpora explored for the purposes of this investigation. This surely reinforces the fact that both lawyers-linguists and legal translators should be aware not only of the benefits that corpora offer, but also recognize the importance and necessity to implement such technological tools in their translation assignments. This would in turn, improve the overall quality and accuracy of the documents that they are entrusted to translate, while working between different languages and legal systems.

Additionally, all the cited examples could be used as future references for legal translators or at least borne in mind while contemplating the most suitable translation equivalent for particular cases. Although, I was unable to access a bilingual parallel or comparable corpus albeit their usefulness for this investigation, it is hoped that the information and results produced from the four listed corpora clarify such issues under examination.

Finally, corpus studies are constantly developing and have serious implications for translators and legal-linguists, however, although extremely useful and important, “the corpus remains one of the linguists’ tools, to be used together with introspection and elicitation techniques. Wise linguists, like experienced craftsmen, sharpen their tools and recognise their appropriate uses” (Aijmer 1991: 313), they should also be used in conjunction with conventional materials, plus the professional experience and knowledge of competent translators which would provide a balanced framework from which translators could produce more accurate, fluent and idiomatically correct TT.
Acknowledgments

I would like to express my gratitude to my supervisor Sophia Butt for her concise and constructive advice throughout the process of completing this research and furthermore my thanks to my family and friends for their personal support and encouragement.

REFERENCES


