Highlighting Readability Issues by the Content Analysis of a Legal Text, for Translation Purposes. Case Study

Alina Buzarna-Tihenea (Gălbează)
Lavinia Nadrag
"Ovidius" University of Constanta, Faculty of Letters, Romania
lnadrag28@yahoo.comm
alina_buzarna84@yahoo.com

Abstract

In a globalized world, linguistic harmonization leads to complex translation issues. Globalization contributed to the intricacy of the economic, social and political background, and triggered urgent requirements regarding the harmonization of legislation, which, in turn, entailed the necessity to tackle legal terminology and terminological differences. In the first part of the paper, we have briefly presented some features of the technical language and several issues caused by its intricacy. The second part of the work analyzes a specialized text in terms of lexical density, dealing with aspects such as text statistics, frequency and top words, frequency of word structures, as our main purpose has been to contribute to an enhanced understanding of specialized texts. Both jurists and translators should approach legal concepts carefully when transposing them into the target-language, in order to avoid misunderstandings. Moreover, multicultural knowledge and flexibility are only some of the requirements of a faithful legal translation.

Key words: English legal language, translation, specialized terminology, readability, lexical density
J.E.L. classification: Z13

1. Introduction

Since law is deeply influenced by the society where it functions and it cannot be strictly explained and defined as a set of organized rules, not only does a better understanding of a particular legal culture imply the focus on issues such as legal terms and norms, but also on the wide range of distinct customs and social concepts typical of the respective legal system, an interdisciplinary approach being therefore imperiously necessary. For instance, the translator may deal with terms that have different meanings in different law branches in the target language or that express several different meanings and sometimes s/he needs either specialized knowledge in the legal field or the assistance of a specialist for a better understanding and appropriate translation of the respective terms. Moreover, the translator may encounter specialized terms with no corresponding legal meanings or interpretations in the target language and, in such circumstances, s/he can use either neologisms or retain the word as it is (in the source language) (Onufrio, 2007, p. 4), accompanied by a translator's note.

2. Theoretical background

Since different legal systems have been created against the background of diverse cultures (in this respect, the translation process being understood as a form of cultural interaction, where the translator functionally replaces cultural elements and changes the source text based on the cultural norms and concepts of the target language) (see Botezat, 2011, p. 229) in order to meet the necessities of different nations, this led to inadvertences and incongruities of a great number of legal terms and of their understanding, in various national systems. Consequently, many translation theorists agree that "absolute equivalence" is nothing more than a utopia, i.e. something that cannot

be achieved. In this sense, for example, when discussing the fact that every legal system is deeply influenced by its country's historical, cultural and socio-economic background, being equipped with its own conceptual system, the scholar Susan Sarcevic asserts that "legal terminology of different legal systems is, for the most part, conceptually incongruent" (Sarcevic, 1989, p. 278), because law and legal language are social constructs and the quest for perfect identical concepts in many languages is dangerously misleading and even impossible (Heikki, 2006, p. 122).

Having in view that all languages function according to certain standardized rules – which can vary in terms of registers – the translator's mission is even more difficult. As Duszak (1997, p. 9) argues, "Recent insights into academic writing have shown considerable variation in text characteristics across fields, languages and cultures. [...] Among the most notable differences are field-and culture-bound disparities in global organization schemata of texts". For instance, technical language is dominated by special characteristics, such as the use of short sentences, specialized vocabulary (typical of the debated/ presented topic), nouns referring to actions, long and complex noun phrases, simple and direct structures, rare adverb use, impersonal, formal style, avoiding figures of speech, metaphorical meanings, colloquial words and phrases, etc. (see Eggins 1994; Martin, 2000; Alcaraz Varó, 2000; Duque García, 2000). According to Duszak (1997, p. 2), all these features draw "the image of a dehumanized language of science, and likewise the image of a dehumanized writer (...) uniformity of academic writing styles was taken for granted and was accounted for in terms of objectivised research standards". Moreover, as stated by Alcaraz Varó (2000, pp. 138-139), all these features – to which high semantic density, frequent use of impersonal forms and structures, impersonal, objective style are also added – distinguish this register from the others. Among others, these rules play two opposing roles, i.e. guiding second language speakers for a better understanding and adequate use of specific language items triggering the acquisition of native-like fluency, but also hindering natural communication.

3. Research methodology

Since this paper aims to contribute to an enhanced understanding of specialized texts, the practical section, which is a corpus-based research inspired by Teubert and Čermáková's work (2007, pp. 65-77), analyzes a specialized corpus in terms of lexical density, dealing with aspects such as text statistics, frequency and top words, word length, syllable count, frequency of word structures.

For the purpose of our research, excerpts from a representative legal text, i.e. *articles 1-11* of the *United Nations Convention on the Carriage of Goods by Sea,* 1978 (*Hamburg Rules*), has been chosen in order to perform the analysis focused on lexical density. This Convention regulates the international transport of goods, in an endeavor to homogenize and balance the legal background for the dispatch of freight by sea.

It is noteworthy that this research is a continuation of a previous analysis published in 2016 (see Nădrag and Buzarna-Tihenea (Galbeaza), "Aspects of Legal Translation in Contracts of Carriage", "*Ovidius University Annals, Economic Sciences Series"*, Ovidius University Press, XVI(1), 2016, pp. 35-40) and it also completes another article that will be sent for publication together with the present work, in the same journal (i.e. Nădrag and Buzarna-Tihenea (Galbeaza), "Translation Issues in the Legal Field of the European Union. Case Study on Specialized Terminology").

The analysis of the chosen legal corpus has been performed by means of three specialized software, i.e. *Analyze My Writing* (see https://www.analyzemywriting.com/index.html), *Seoscout* (see https://seoscout.com/tools/keyword-analyzer?lang=en#analysis) and *Text Analyzer* (see http://www.roadtogrammar.com/textanalysis/), which include a series of text analysis tools providing automatically valuable pieces of information regarding lexical density – which deals with measurements of the structure and intricacy of human communication, i.e. linguistic complexity (Halliday 1985) – and content analysis – word count, number of characters, syllables, sentence count, average sentence length, average number of syllables per word, lexical density, lexical diversity, readability.

By analyzing the quantitative data furnished by the three specialized software, some important features of the legal specialized language will be highlighted, having in view that the corpus-based techniques play an essential part in quantifying them. The quantification of linguistic phenomena

contributes to the identification of those language items often encountered in the analyzed register, which will allow the formulation of statistical suppositions with respect to specialized language use. Therefore, by correlating these data provided by the specialized software with the lexical profile of the legal language, certain specialized language patterns will be identified.

4. Findings

As already mentioned in the previous section, the practical part of our paper focuses on several aspects related to the lexical density and readability of the analyzed corpus (i.e. text statistics/ content analysis, frequency and top words, word length, syllable count, frequency of word structures). The results of the analysis are highlighted in Tables 1, 2, 3, 4 and 5 below.

The data on the lexical density of the analyzed excerpts, calculated by dividing the number of lexical items (i.e. nouns, adjectives, verbs and adverbs) by the total numer of words, have been generated by *Analyze My Writing*, which revealed the following in terms of the predominant parts of speech identified in the excerpt:

Table no. 1 Lexical Density: Parts of Speech

Part of Speech	Percentage
Nouns	30.54%
Adjectives	4.3%
Verbs	8.92%
Adverbs	2.34%
Prepositions	19.28%
Pronouns	1.17%
Auxiliary Verbs	4.17%

Source: authors' own processing using Analyze My Writing

Table 1 shows that almost one third of the words in the anayzed excerpts are nouns (30.54%), followed by verbs (8.92%), adjectives (4.3%) and adverbs (2.34%); in total, almost half of the words (46.1%) are lexical items. Having in view that the less dense a text, the easier to understand it, the lexical density shown by the specialized software suggests that the analyzed corpus is of upper-intermediate difficulty. As far as the grammatical or the functional items are concerned, whose main role is to connect lexical items (see Halliday 1985), prepositions hold the greatest share (19.28%), followed by auxiliary verbs (4.17%) and pronouns (1.17%). Overall, functional items represent almost a quarter (24.59%) of the total number of words in the analyzed excerpt, which highlights their importance in content creation.

It should be noted that lexical density is directy connected to the individual's age, education level, communication style, context, creativity and even medical condition (Yoder 2006) and influences text readability and one's understanting of the respective text (To et. al. 2013). Moreover, it affects the memorability and retentivity of the message conveyed by the text (Perfetti 1969).

In its turn, *Seoscout* provided data on content analysis (see Table 2) and readability (see Table 3), which, among others, underlined a lexical density of 46% (almost identical to the one shown by *Analyze My Writing*), a reading ease index of 48% and a grade level of 12.5, which suggests that the difficulty level of the corpus is upper-intermediate and that it can be understood by experienced readers.

In addition, according to the data provided by *Text Analyzer*, the corpus corresponds to the C2 (Advanced) level of "The Common European Framework of Reference" and to the 8+ IELTS level (i.e. very good user of English) (see Tables 2 and 3 below).

Table no. 2 Content Analysis

Word Count	2908
Character Length	16252
Letters	13078
Sentences	121
Syllables	4421
Average Words/Sentence	24.0
Average Syllables/Word	1.6
Lexical Density	46%
Lexical Diversity	17%
Reading ease	48%
Grade level	12.5

Source: authors' own processing using Seoscout

Table no. 3 Readability

Reading Ease	48%
Grade Level	12.5
Gunning Fog	16.2
Coleman Liau Index	10.7
Smog Index	12.1
Automated Reading Index	11.8

Source: authors' own processing by Seoscout

The upper-intermediate difficulty level of the analyzed text is also revealed by the data provided by *Seoscout* in terms of word length, frequency and top words, frequency of two-word structures (see Tables 4, 5 and 6 below):

Table no. 4 Frequency and Top Words

Keyword	Words	Uses	Uses (%)
Carrier	1	79	2.7%
Goods	1	47	1.6%
Article	1	45	1.5%
Carriage	1	42	1.4%
Loss	1	29	1.0%
Contract	1	26	0.9%
Sea	1	25	0.9%
Delay	1	25	0.9%
Damage	1	24	0.8%
Convention	1	23	0.8%
Delivery	1	22	0.8%
Provisions	1	20	0.7%
Liability	1	18	0.6%
Actual	1	17	0.6%
Paragraph	1	16	0.6%

Source: authors' own processing using Seoscout

Table no. 5 Word Length

Word Length (characters)	Word count	Frequency
2	711	24.5%
3	517	17.8%
4	305	10.5%
7	266	9.2%
5	255	8.8%
6	204	7%
8	202	7%
1	153	5.3%
9	134	4.6%
10	94	3.2%
11	33	1.1%
12	12	0.4%
14	7	0.2%
13	5	0.2%
15	3	0.1%

Source: authors' own processing using Seoscout

Table no. 6 Frequency of Two-Word Structures

Keyword	Words	Uses	Uses (%)
loss damage	2	24	0.8%
delay delivery	2	20	0.7%
damage delay	2	19	0.7%
contract carriage	2	18	0.6%
carriage sea	2	18	0.6%
actual carrier	2	16	0.6%
provisions convention	2	9	0.3%
carriage goods	2	8	0.3%
bill lading	2	8	0.3%
article transport	2	7	0.2%
servants agents	2	7	0.2%
carrier liable	2	7	0.2%
goods sea	2	6	0.2%
provisions paragraph	2	6	0.2%
liability carrier	2	6	0.2%
carrier servants	2	6	0.2%
limits liability	2	6	0.2%

Source: authors' own processing using Seoscout

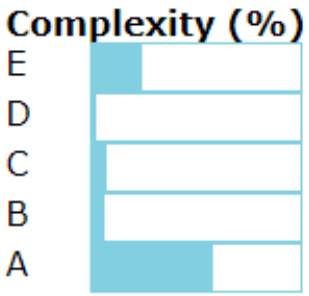
The medium difficulty level of the text is also highlighted by the average number of words pe sentence (24) and the average number of syllables per word (1.6). The shortest word has only one letter ("a") and the longest word has fifteen letters (i.e. "notwithstanding"), used only 3 times. There are 306 words made of one, two and three letters (usually determiners, prepositions or auxiliary verbs, e.g. "a", "on", "to", "by", "in", "or", "the", "has", "any", "are"), 187 words made of four, five, six and seven letters (e.g. "this", "that", "port", "means", "order", "person", "actual", "purpose", "carrier"), 105 words made of eight, nine and ten letters ("contract", "optional", "applicable", "discharge", "character", "evidencing", "Convention", "uniformity"), and only fifteen words made of eleven, twelve, thirteen, fourteen and fifteen letters ("contracting", "regulations", "nationality", "respectively", "nevertheless", "consequences", "international", "circumstances", "responsibility", "interpretation", "notwithstanding"). It should be noted that the great majority of

the words consisting of five or more letters are technical terms that belong to the legal field.

As far as collocates – "the words which occur in the neighborhood of your search word" (Duan, 2007, p. 26) – and clusters – "words which are found repeatedly in each other's company [which] represent a tighter relationship than collocates" (Scott, 1996: 35) – are concerned, they were identified in concordance lines and helped us highlight the meanings and uses of the tackled terms.

The classification of words provided by *Text Analyzer* also confirms the above-mentioned findings about the difficulty level of the corpus. Figure 1 shows that more than half of the words in the excerpt are simple (A level), while almost one quarter are complex (E level):

Figure no. 1. Shares of Word Complexity

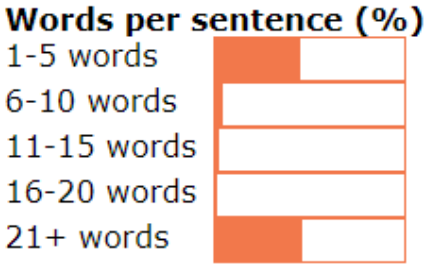


Source: author's own processing using Text Analyzer

According to the above-mentioned specialized software, A level (simple) words include mostly prepositions ("on", "by", "to", "for", "of", "in", "with", "against", "from", "over", "between", "without", "under", "at", "through", "into"), determiners ("the", "a", "this", "any", "some", "an", "that"), conjunctions ("and", "or", "if"), auxiliary verbs ("have", "has", "been", "are", "is", "be", "had", "does", "were", "do"), modal verbs ("must", "could", "may", "would", "can") pronouns ("they", "it", "them", "he", "itself", "himself") but also nouns ("person", "name", "part", "place", "case", "life", "party", "days"), adjectives ("different", "general", "good"), adverbs ("well", "probably", "far") and verbs ("live", "take", "made", "need", "took", "report", "put", "said", "make", "held", "used"). It should be noted that these words belong to general English and their meaning is easily accessible to readers. At the opposite pole, there are the complex words (E level), which include mostly nouns ("desirability", "convention", "provisions", "purposes", "payment", "bearer", "contracts", "legislation", "nationality", "holder", "charterer", "charter-parties", "shipments", "interpretation", "uniformity", "disposal", "reference", "discharge", "claimant", "freight", "claims", "tort", "consignee", "pursuance", "shipper", "prejudice", "recourse", "proceedings", "carriage", "liability", "delivery"), verbs ("conclude", "follows", "relates", "undertakes", "surrender", "includes", "constitutes", "govern", "covers", "occurs", "mitigate", "proves", "founded", "acted", "referred", "deemed", "insert", "entrusted", "aggregate"), adjectives ("optional", "pursuant", "applicable", "diligent", "consecutive", inherent", "reasonable", "equivalent", "attributable", "payable", "aforesaid", "non-contractual", "separate", "contrary", "statutory", "recoverable", "liable", "judicial", "competent") and adverbs ("solely", "recklessly", "respectively", "reasonably", "expressly", "explicitly"). All these E level words are specialized (legal) terms that can be properly understood by an experienced reader with knowledge in the legal field. The important share held by these terms in the excerpt (approximatively one quarter) highlights the overall difficulty level of the corpus.

Figure 2 below, processed by *Text Analyzer* also shows the contrast that characterizes the corpus, as almost half of the excerpt is made of 1–5-word sentences and the other half – by 21+ word sentences, which encumbers its readability. In order to improve the readability of the text, the sentences should be shorter and complex phrases should be avoided.

Figure no. 2. Shares of Sentences per Length



Source: author's own processing using Text Analyzer

The data furnished by the three specialized software tools in terms of readability, content analysis and lexical density highlighted several differences between the standard language and the specialized register, the latter being characterized by the preponderance of specialized lexical items (complex words), longer and encumbering sentences, upper-intermediate/ advanced readability levels, higher grade levels (which involves the fact that the reader should have some knowledge in the legal field in order to understand the legal text properly).

5. Conclusions

In a globalized world and in a continuously harmonizing Europe, where European legal instruments are mandatorily translated into more and more official languages, the lack of, and the vain efforts to achieve linguistic harmonization has led to complex and costly translation issues. Therefore, those working both in the legal field and in other specialized languages should not treat legal translations as an approximating process, focusing merely on the linguistic side. They should pay close attention to the understanding and transposition of legal concepts from the source-language system into the target-language system, with the subsequent employment of elements of comparative law. Thus, a translator should be equipped with flexibility, attention to details, multicultural knowledge and interdisciplinary abilities when s/he tackles the intricate network of legal terminology and translation.

6. References

- Alcaraz Varó, E. 2000. *El ingles professional y academic (Professional and Academic English)*. Madrid: Alianza Editorial.
- *Analyze My Writing,* [online]. Available at: <https://www.analyzemywriting.com/index.html> [Accessed 15 March 2020].
- Botezat, O. 2011. "The Complexity of Legal Translation: Social and Cultural Bounds Aspects". In Ridley, H., Farber, M. and Hull, S. eds. *Proceedings of the 7th World Congress on the Advancement of Scholarly Research in Science, Economics, Law and Culture. 12th March 2011, The 1st IISHSS International Conference on Law and Social Order, Constanta,* Vol. II. New York: Addleton Academic Publishers.
- Duque García, M. M. 2000. "*Manual de estilo. El arte de escribir en inglés científico- técnico (Stylebook. The art of writing in scientific-technical English").* Madrid: Paraninfo.
- Duszak, A. 1997. *Intellectual styles and cross-cultural communication*. Berlin: Mouton de Gruyter.
- Eggins, S. 1994. *An introduction to systemic functional grammar*. London: Printer.
- Halliday, M. 1985. "Chapter 5", "Chapter 6". In *Spoken and Written Language*. Deakin University..
- Heikki, M. 2006. *Comparative legal linguistics*, Hampshire: Ashgate Publishing Limited.
- Martin, J. R. 2000. *Grammar meets genre: Reflections on the Sydney school*. Inaugural lecture at Sydney University Arts Association.
- Nădrag, L. and Buzarna-Tihenea (Galbeaza), A. 2016. "Aspects of Legal Translation in Contracts of Carriage". "*Ovidius University Annals, Economic Sciences Series"*, Ovidius University Press, XVI(1), pp. 35-40.

- Onufrio, M. V. 2007. "Harmonization of European Contract Law and Legal Translation: A Role for Comparative Lawyers", *InDret, Revista para el Analisis del Derecho*, Barcelona: University of Palermo.
- Perfetti, C. A. 1969. "Lexical density and phrase structure depth as variables in sentence retention". *Journal of Verbal Learning and Verbal Behavior*. Elsevier BV. 8(6): 719–724.
- Sarcevic, S. 1989. "Conceptual Dictionaries for Translation in the Field of Law". *International Journal of Lexicography*, 2(4), pp. 277–293.
- *Seoscout, "Keyword Analysis, Content Optimization + Word Count Tool",* [online]. Available at: <https://seoscout.com/tools/keyword-analyzer?lang=en#analysis> [Accessed 15 March 2020].
- Teubert, W. and Čermáková, A. 2007. *Corpus Linguistics. A Short Introduction*, Bloomsbury Academic.
- *Text Analyser, "Text Content Analysis Tool"*, *UE Using English.com*, [online]. Available at: <https://www.usingenglish.com/resources/text-statistics.php> [Accessed 15 March 2020].
- Text Analyzer, *RoadToGrammar.com*, [online]. Available at: <http://www.roadtogrammar.com/textanalysis/> [Accessed 15 March 2020].
- To, V., Fan, S. and Thomas, D. P. 2013. "Lexical density and Readability: A case study of English Textbooks". *The International Journal of Language, Society and Culture*. 37(7) pp. 61–71.
- "United Nations Convention on the Carriage of Goods by Sea, 1978 (Hamburg Rules)", articles 1-11, [online]. Available at: <https://www.uncitral.org/pdf/english/texts/transport/hamburg/hamburg_rules_e.pdf> [Accessed 15 March 2020].
- Yoder, P. 2006. "Predicting Lexical Density Growth Rate in Young Children With Autism Spectrum Disorders". *American Journal of Speech-Language Pathology*. 15(4), pp. 362–373.