



Matej Bel University, Banská Bystrica, Slovakia
Has been issued since 2014
ISSN 1339-6773
E-ISSN 1339-875X

Quality Translations of Legal Texts in the Context of the Methods of Corpus Analysis

Zuzana Tuhárska

Matej Bel University, Banská Bystrica, Slovak Republic
PhD
E-mail: Zuzana.Tuharska@umb.sk

Abstract

The aim of the article is to present possibilities of the corpus linguistics methods as a means of increasing the quality of the output of the translation process in the context of legal translation. Firstly, the specifications of a legal text as a text type in the context of translation will be presented; secondly, the competences necessary to master legal translation will be explained; thirdly, the benefits provided by the corpus analysis method will be summarized. Based on selected empirical data obtained through the corpus analysis, the potential of the corpus analysis methods will be illustrated in practice in order to explain how it can be used to increase the quality of the translation process and its product.

Keywords: corpus linguistics, corpus analysis method, specialized language, legal translation.

Introduction

Legal texts are amongst the most frequently translated types of text not only in the context of the legislative framework at European level. These texts are characterized by their very specific features in several aspects. The correct reception and interpretation of such texts are inevitable prerequisites for achieving acceptable quality of their translation; several competences are required to master these skills. Therefore, legal texts need to be paid special attention in the context of translation. We believe that the method of corpus linguistics is one of the tools which can definitely be used to improve the quality of the translated text. In this article we will present possibilities of the corpus linguistics method as a tool for increasing the quality of the output of the translation process in the context of legal translation.

Legal texts as a specific group of specialized text from the perspective of translation and interpreting

The specificity of legal texts, which qualifies them as a subgroup of specialized text, can be derived from several facts. In order to present the legal text from the perspective of translation studies, it is important to explain its nature from the point of view of understandability (we can only translate or interpret contents which we actually understand). Legal texts are infamous for their overall low readability, i.e. they are generally considered hard to understand. It is not only due to the specific terminology, but also to the stylistic peculiarities exceeding the lexical level. The fact that legal terms are often difficult to understand for a layman and there is a risk of misinterpretation due to its de-terminologization (compare Molnárová, 2013. p. 13). This can result in legal consequences in these cases as certain parts of such texts are perceived incorrectly. From the point of view of a translator and/or interpreter, mostly aspects such as terminological inconsistencies and frequent zero equivalence, resulting from incompatibility of the two legal systems in question are perceived as a problem.

Theoretical background determining the quality of legal translation

All of the above mentioned factors reflect the quality of translation or interpreting of legal texts. In order to increase the quality of translation or interpreting of legal texts, it is important to

recognise which **competencies** are required to successfully master the translation process in the given field, and also detect the **starting point** leading to adoption of such competencies.

The quality of translation or interpreting of (not only) legal texts is dependent on linguistic, translating and interpreting, but also professional competences of the translator/interpreter related to the given field. This premise is in accordance with the following statement: “...*the legal language needs to be perceived as a phenomenon which extends over the law, linguistics, and translation studies; during translation, it is important to keep in mind that features of the text result from its interdisciplinary nature*” (Đuricová, 2012, p. 10). The translator’s ability to master the language as their working tool is the very basic prerequisite for successful handling of any text. A measure of linguistic competence can be perceived as the translator/interpreter’s ability to select appropriate expressive means for a given text type (morphological and syntactical, lexical, stylistic, etc.) on individual language levels and smoothly harmonize them in a balanced manner whose attributes will thus correspond with the relevant text requirements. As for the **translation and interpreting competence**, it is based primarily on mastering the principles of translation and interpreting, and their specificities relating to legal texts. The ability to apply theoretical postulates in practice as well as being able to deal with possible discrepancies between the theory and the real-life situations complement this competence. **Specialized skills** cover mainly familiarity with the law at least to such extent that the translator/interpreter is capable of correct understanding of the text which enables them to translate/interpret it without meaning shifts.

To ensure that the translation will be of acceptable quality, practical, theoretical, and empirical knowledge must be applied correctly. To achieve this high quality, the above mentioned competences need to be confronted with this knowledge. All three competence types can be perceived within a unified **practical basis** which provides information on the translator’s ability to apply translation and interpreting procedures in practice, but also about their ability to apply their theoretical knowledge in specific situations. The practical knowledge is related to one’s actual ability to use a specialized language pertaining to a specific field. However, besides the fact that this ability results (or should result) from the theoretical principles (which should be reflected in the practice), the experimental base plays an important role. It can only be achieved through practice in processing of specialized text. The importance of **theory** could be inherently understood from the previous text. Mastering the theory is important for a translator to be able to apply it in practice (i.e. the theory can only be applied if we are familiar with it, otherwise there would be nothing for us to apply at all). As for improving the quality of legal translation in the context of both theoretical and practical knowledge, the following question emerges: what theoretical phenomena and practical skills should be paid attention to? Here it is important to shift our attention to **empirical knowledge**, i.e. detection of characteristic features of a given text type in a systematic and objective way. The effort to objectively determine the typical, i.e. recurring (the quantitative viewpoint) and specific (the qualitative viewpoint) attributes of a specialized legal text enables the translator/interpreter to focus on the phenomena which might help them to extend their practical and empirical knowledge. One approach to achieving this task is the method of corpus analysis. The next section of the article will describe it in more detail.

Benefits of the corpus analysis method

When focusing attention on the empirical knowledge as one of the starting points for improving the quality of translation or interpreting of legal text, the key question remains: how can the corpus analysis method help us achieve our goals? The method provides several benefits*; we have selected three of them to provide a deeper insight. Any corpus is a reflection of an actual language reality. In creating a corpus, real texts taken from practice are usually processed. i.e. statements spontaneously created by language users, instead of being intentionally (artificially) created in order to, e.g. demonstrate certain grammatical phenomena. Since the man is a producer (creator) of the language, i.e. every language has been created by man, it is logical to assume that such language reflects its author to some extent. Therefore, only naturally originating statements can provide a source of information on the language and its users (compare Höhn, 2011). Corpora

* Compare e.g. Šuchová (2009, 285): “Working with a text corpus pertaining to a specific language provides us with a quality and quantitatively sufficient database or language units which can be used for the empirical research of language use.”

also provides the opportunity to perform systematic analysis based on criteria correlating with a given research goal which helps minimize the subjectivity. This also implies another opportunity – to obtain empirical output. The use of the corpus analysis method usually requires the corpus to be in an electronic form due to the vast amount of data that needs to be processed in order to achieve statistical relevance. An example of a piece of software designed for the corpus analysis is *AntConc*. The version *AntConc 3.4.4w (Windows) 2014* was used to obtain the data analyzed in this article.

Selected aspects of corpus analysis

In order to demonstrate the possibility of applying the corpus analysis method can be used to improve the quality of the translation process and its product, provides several examples of an analysis obtained through the *AntConc* software. We are examining the selected issue from the point of view of translation studies which determine the goals of our analysis. The researched material consists of two parts: the *Abgabenordnung* (hereinafter referred to as AO) German legal document which is in force in Germany, and its equivalent called *Daňový poriadok* (hereinafter referred to as DP) is in force in the Slovak Republic. The size of AO and DP is 57.870 and 44.090 tokens respectively.

The Word List function

One of the functions which is able to provide information useful in improving the quality of the translation process is the *Word List* function. Frequency values pertaining to each individual lexemes allows for basic orientation in the corpus. The most frequently occurring lexical units are presented in Table 1.

Table 1. Function: *Word List*

Function: <i>Word List</i>							
Occurrence of autosemantic expressions							
DP (Word Tokens: 44090, Word Types 4985 → 8.844)				AO (Word Tokens: 57870, Word Types 6051 → 9.563)			
	Lexem	Freq.	Rank		Lexem	Freq.	Rank
1	<i>dane</i>	1018	3		Abs.	486	15
2	<i>správca</i>	490	12		<i>Finanzbehörde</i>	332	27
3	<i>daňového</i>	437	13		Gilt	253	33
4	<i>daňový</i>	396	14		Absatz	239	39
5	<i>č.</i>	374	15		Satz	238	40
6	<i>zákona</i>	328	16		!Art	136-32	54
7	<i>predpisov</i>	237	18		Steuer	126	55
8	<i>ods.</i>	202	19		Nr.	125	57
9	<i>znení</i>	195	20		Personen	125	58
10	<i>správcu</i>	194	21		Steuern	120	60
11	<i>daňovej</i>	187	24		Beteiligten	109	65
12	<i>odseku</i>	183	25		Vorschriften	109	66

Besides specifications such as whether the language is flexive or analytic*, the software provides information on the word family and semantic fields which calls for attention. From the point of view of translation studies, it is interesting to study the equivalence of frequently occurring autosemantic expressions in both language parts of the single corpus. The usage of conventional means of expression typical for the text type can be contrasted against each other in the given language. Full equivalence indicates that literal translation is possible, whilst partial equivalence indicates that broader co-textual relationships will have to be taken into consideration, thus possibly generating translation difficulties. However, it also provides information on how these can be overcome as thanks to the corpus analysis we can observe the conventional use of language in any given type of text. Possible equivalence are highlighted by colours in Table 1, partial

*Compare with Tuhárska (2014, 201).

equivalence is distinguished by italics. Both parts of the corpus clearly show the frequent occurrence of the following expressions: *dane* – *Steuer/Steuern*, *č. – Nr.*, *predpisov* – *Vorschriften*, *ods. – Abs.*, *odseku* – *Absatz*. After verifying the idea in other sources, we see that the expression *správca dane* can be partially compatible with the German expression *Finanzbehörde**

The Concordance function

Another function of the *AntConc* software which allows the researcher to focus attention on specific word units (verbal) and their integration into the co-feature of the text[†] is the *Concordance* function. This function generates concordances[‡] with the selected expression. Use of wildcards (* – no character or several characters; + – zero or one character? – one more character; | - and also) and the *Words* tool (searching specific word forms, case sensitivity off; *Case* (searching various morphological versions of a lexeme, case sensitivity on); *Regex* (identification of words and word formation products containing a component inserted in the search field); combination of the *Word* and *Case* functions (identification of a particular word form with case sensitivity on) – all of these increase precision of the analysis results. Exact values allow the analyst to observe flexive and word forming potential of a selected expression or its root. An example of the application of this function is illustrated in Table 2.

Table 2. *Concordance I* function

<i>Concordance I</i> function		
Expression, wildcard [function]	Examples from corpus	Freq
zdan* dan* *daň* [Words]	daň, dane, daní, dani, zdaniteľných, daniach, daňou, zdanania	1214
daň* [Words]	daň, daňou, daňovým daňových, daňového, Daňový, daňový, Daňovému...	1635
daň [Words]	daň, Daň	98
daň [Words, Case]	Daň	96
Daň [Words, Case]	Daň	2
daň [Case] [§]	daň, daňový, daňových, daňového, daňovým, daňovou ...	1517

Filtering particular morphological forms and/or the word formation products leads to an even more detailed analysis of the language material, e.g. the *Concordance* function through identifying frequently occurring combinations with other word units. The most frequently occurring lexical units containing the “dane” component are presented in Table 3.

* In the Slovak legislation, the expression *správca dane* (English: tax administrator) is defined as the tax office, customs office or municipality (Source: Act No. 563/2009 Coll. on tax administration as amended, Article 1, Section 4 (1); <http://www.zakonypreludi.sk/zz/2009-563> (26-4-2016) Based on this, it can be stated that the German expression *Finanzbehörde* [Slovak: *daňový úrad*] is a hyponym of the Slovak expression *správca dane*.

† *Co-text* is defined as language environment; the text occurring in the immediate vicinity of the researched means of expression. In contrast, *context* represents the non-language environment -- the communication situation.

‡ Concordance is defined as a set of co-texts pertaining to a key word. Co-texts of certain length grouped around the selected key word are extracted from the corpus (compare Lemnitzer – Zinsmeister 2006, p. 196).

§ Since the *-daň-* component lacks word formation potential in Slovak, the same values would be returned if we used the *Regex* function.

Table 3. *Concordance II* function

Concordance II function		
Expression, wildcard [function]	Examples from corpus	Freq
správ* dan* [Words]	správě daní, správu daní, správou daní, Správca dane, správcom dane, Správcom dane, správcu dane...	884
pl?t dan* [Words]	platenie dane, platenia dane, PLATENIE DANE, platení dane, zaplatenia dane, zaplatením dane, nezaplatením dane, platiteľom dane, splatnej dane, splatnosti dane, platbu dane, splátku dane...	90
urč* dan* [Words]	určovania dane, určovaní dane, určenie dane, Určenie dane, URČENÍ DANE, určenia dane	31
rát da* [Words]	vrátenie dane, vráti daňovému, vrátenie daňového, vrátiť daňovému, vrátenia dane...	20
vyrub* dan* [Words]	vyrubeníu dane, vyrubenie dane, vyrubenej dane	19

The practical applicability of data obtained this way in order to enhance the quality of the translation process depends on the values pertaining to the occurrence frequency of particular word forms, lexemes, or phrases. In studying of the morphological and semantical level, it is important to pay due attention to these. It helps us to select a suitable equivalent in the target language and also allows us to verify the proposed selection in parallel with the target language corpus. The translator should focus mainly on emerging discrepancies, whether qualitative* or quantitative†, and examine them in detail, e.g. using explanatory dictionaries or other expert literature sources‡.

Collocates and Clusters functions

The *AntConc* software provides other functions to further analyse the word forms, lexemes or phrases selected through the previously described processes such as *Clusters* and *Collocates*§. Due to limited space and the fact that these functions were already discussed elsewhere (see Tuhárska 2013 and 2014), we will only comment on them briefly and summarize why they are useful to us.

The output of the *Collocates* function reflect the diversity of the possibilities of combining the examined expressions with other neighbouring elements from a statistical point of view. Here it is possible to identify direct co-textual units by generating two-component collocates with the examined expression as the core, i.e. we identify which expressions usually occur right before (1L) or after (1E), the examined expression. The volume of the co-text observed can be extended as needed. The outcomes of such an investigation can empirically reflect the relevance of the particular word phrases based on a generally valid statistical connection between the collocate elements and the data on occurrence frequency obtained from the corpus.

As for the *Clusters* function of the *AntConc* software, clusters of three or more elements can be generated to observe their occurrence frequency. In general, the *Clusters* function is useful for

* Qualitative viewpoint is defined as the extent to which we understand the morphological and/or semantic (in)consistency of means of expression in the source and target languages. Discrepancies on this level can potentially generate translation problems and call for solution through suitable translation procedures.

† The quantitative viewpoint is related to the (in)consistencies in occurrence frequency of equivalent language means in the source and target languages.

‡The process was illustrated in Table 1, therefore we will not further elaborate on this issue.

§ To explain the term collocate, we will draw on the following definition from specialized literature: A collocate is an expression which usually consists of an arbitrary and conventional connection of two components (e.g. blonde hair). In a collocate, a collocation base as the semantically independent element can be found (hair) and the collocator (blonde) as the semantically dependant element can be found (Engelberg – Lemnitzer 2001, p. 391).

identifying multielement clusters with interesting occurrence frequency which is widely applicable in practice. Data related to clustering values also provide information on the diversity and unification in the way of expressing information through selected language means on a syntagmatic level. The occurrence intensity of clusters provides information on the extent of conventionalization in using a language in a certain text type. The more frequently the clusters occur in the corpus, the higher their conventionalization and typization, therefore predictability. In this context it is important to focus not only on the frequency, but also the length of the clusters. The stability of the quantitatively larger (longer) language structures correlates with the extent of their conventionalization and their potential for typization of the given text. In other words: If the same procedures of expression keep occurring in the corpus consistently even regarding larger clusters, it can be assumed that their stability is high and the language structures can be predicted from the quantitative viewpoint. This fact can be put to use in practice. Learning the complex language structures creates a precondition for learning to use phrases used in a foreign language even despite the fact that these are not inevitably identical with the structures occurring in the source language.

Conclusion

As for the research goals pertaining to the issue, the following can be stated: The two approaches 1) moving from the empirical to the theoretical, and 2) moving from the conventions to the standard pose two viewpoints of the same issue. Both approaches are connected with increasing of effectiveness in the translation practice. On one hand, empirical data can be used as a basis to formulate theoretical starting points applied in the translation process. On the other hand, particular corpus data (mainly with focus on the frequency information) can help the translator and/or interpreter to focus on stable (conventional) means of expression used in the given text type. These can serve as a basis for the creation of standards applicable to translation of rather less frequent, yet parallel clusters of language units. The article deals primarily with the corpus analysis method, however, it is important to point out the importance of the role of the translator and/or interpreter. Everything depends on their experience which has impact on all of the above mentioned components. Since the translation process is usually related to human activity and their active cognitive approach, the empirical (objective) corpus analysis method can be perceived as a counterpart to the subjective work with the handling of the text, the translation as a process represents (e.g. the translator has to consider all specialized, semantic, and textual criteria, see Štefaňáková, 2014, p. 150). The dialectic relationship between these two elements can be a very effective tool. The results of a corpus analysis can be stimulating to the translator and/or interpreter and help them focus on the data which are important for them from an empirical point of view. Their careful consideration in conjunction with the translator's and/or interpreter's knowledge base can have a positive impact on their translating practice. Therefore, it can be concluded that mutual complementarity of these components has a powerful potential to significantly improve the quality of the translation.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0226-12.

References:

1. Molnárová, E. (2013): Spoločensko-politická lexika z kontrastívneho aspektu. Banská Bystrica : Vydavateľstvo UMB – Belianum, FHV.
2. Ďuricová, A. (2012): Quo vadis slovenský právny jazyk v procese globalizácie a internacionalizácie? In: Ďuricová, A. (ed.): Od textu k prekladu VII. Praha : JTP. p. 6-22.
3. Höhn, E. (2011): The sociology of culture in the context of cultural regional development. In: Universities in Central Europe, 20 years after. Volume 1 : Transformations and stakes. Bruxelles : Bruylant. p. 225-235.
4. Abgabenordnung. online: https://www.gesetze-im-internet.de/bundesrecht/ao_1977/gesamt.pdf (26.4.2016)

5. Zákon č. 563/2009 Z. z. Zákon o správe daní (daňový poriadok) a o zmene a doplnení niektorých zákonov, Čl 1, § 4, ods. 1, online: <http://www.zakonypreludi.sk/zz/2009-563> (26.4.2016)
6. Šuchová, J. (2009): Korpusová lingvistika vo výučbe nemeckého jazyka (na poli frazeológie). In: Forlang. Cudzie jazyky v akademickom prostredí. Košice : TUKE.
7. Tuhárska, Z. (2014): Od konvencií k normám – od empirie k praxi. In: Ďuricová, A. (ed.): Od textu k prekladu IX. Praha : JTP, p. 198-211.
8. Lemnitzer, L. – Zinsmeister, H. (2006): Korpuslinguistik. Tübingen : Gunter Narr Verlag.
9. Tuhárska, Z. (2013): Možnosti skúmania odborného právneho textu prostredníctvom korpusovej analýzy. In: Ďuricová, A. (ed.): Od textu k prekladu VIII. Praha : JTP, p. 174-185.
10. Engelberg, P. Lemnitzer, L. (2001): Einführung in die Lexikographie und Wörterbuchbenutzung. Tübingen: Stauffenburg.
11. Štefaňáková, J. (2014): K niektorým znakom nemeckého právneho jazyka a aspektom práce s terminológiou so zreteľom na rôzne právne systémy nemecky hovoriaceho priestoru. In: Aplikované jazyky v univerzitnom kontexte – didaktika, terminológia, preklad. Zvolen : Technická univerzita vo Zvolene, p. 144-155.
12. Paprotté, W. (2002): Korpuslinguistik. In: Müller, H. M. (Hrsg.). Arbeitsbuch Linguistik. Paderborn : Verlag Ferdinand Schöningh, p. 364-380.
13. Scherer, C. (2006): Korpuslinguistik. Heidelberg : Universitätsverlag WINTER.