

MULTIMODAL COMMUNICATION IN THE 21ST CENTURY: PROFESSIONAL AND ACADEMIC CHALLENGES. 33rd Conference of the Spanish Association of Applied Linguistics (AESLA), XXXIII AESLA CONFERENCE, 16-18 April 2015, Madrid, Spain

## Extraction of terminology and phraseology towards the design of instructional resources for legal translation

Tamara Rubio Donat, Miguel Ángel Candel-Mora\*

*Universitat Politècnica de València, Camino de Vera, Valencia 46022, Spain*

---

### Abstract

Computer-assisted translation (CAT) technology has experienced a dramatic growth in recent years and its use has spread to specialties like legal translation, not so much for their performance and productivity but for their quality control functions. The aim of this work is to propose a protocol to extract frequent terminology and phraseology in order to increase CAT tools performance, improve quality control features and help to produce resources with teaching purposes in the translation of contracts.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Scientific Committee of the XXXIII AESLA CONFERENCE

*Keywords:* legal translation, terminology management, terminology resources

---

### 1. Introduction

Computer-assisted translation achieves its utmost performance in the processing of highly repetitive texts, such as user manuals, help guides and software interfaces; it also provides the opportunity of team work and the addition of specific terminological and phraseological glossaries in order to obtain higher efficiency and productivity. However,

---

\* Corresponding author. Tel.: +34 96 3877000 ext. 75341

*E-mail address:* [mcandel@upvnet.upv.es](mailto:mcandel@upvnet.upv.es)

these translation technologies have experienced a dramatic growth in recent years and their use has spread to specialties like technical translation, legal translation and audio-visual translation, not only for the degree of repetition in their texts, but also for their quality control functions.

In spite of the almost complete integration of the computer assisted translation tools in the professional environment, from the point of view of the academic research, it should be noted the lack of studies that reveal the real degree of integration, as well as the efficiency that these technologies can provide to the work of the translator (Candel, 2010).

Therefore, the aim of this research is to contribute to the literature on translation technology studies with an exhaustive study of the efficiency of these tools in a specialized and common environment in the professional sector of the translation of sale contracts and later on, extract the most frequent terminology and phraseology in these documents in order to optimize the efficiency and quality control features of translation memories and reuse those resources with teaching purposes.

Considering the expectations created by software manufacturers on these computer-assisted translation tools regarding suitability and the degree of efficiency obtained with their use, this paper proposes an in-depth analysis to verify CAT tools performance to translate contracts and, later on compare the results with the analysis of a corpus of real estate sale contracts. On the one hand, this analysis is carried out to verify the degree of internal repetition in the texts (one of the premises on which computer-assisted translation tools are based); and on the other hand, to analyze the efficiency of these tools to conclude that, besides the degree of repetition, the performance these tools is based on the translation of a large volume of text as well as on the use of terminological and phraseological resources integrated into the translation flow.

### *1.1. Methodology*

Firstly, a specialised corpus of legal texts made up with sale contracts was compiled. Then, different statistical analysis, frequency calculations and analysis of repetitions were conducted on this corpus with professional translation tools (SDL Trados Studio) and with corpus linguistics tools to identify and extract partial matches, expressions and frequent subsegments (PhraseMiner). Once the most frequently repeated segments were identified and extracted from the corpus, they were added to the computer-assisted translation system and reused towards the design of teaching resources for the training of future translators specialized in legal translation.

The following section revises the relevant literature on the confluence of the research topics of this work: legal translation, sale contracts and technology applied to translation. This work concludes with the presentation of the most relevant results and the projection of the results of the analysis, as well as with the proposal of future research lines and the application of the results to other disciplines.

## **2. Legal translation and computer-assisted translation tools**

Nowadays, the translation activity requires the use of computer-assisted translation tools and translators are required to incorporate this technology in their workflow. Regardless of the manufacturer, the translation process with these tools can be divided into three general stages: analysis, translation and revision (Candel, 2010).

In the first stage of analysis and processing of the text, the system provides the percentage of translation that can be recovered from the translation memory (TM) and the degree of internal repetition in the text compared to the TM in use. At the same time, this stage can be divided into two parts: the preliminary analysis of the text, in which the TM is consulted and statistics about total, partial and no-coincidences are obtained; and the preliminary analysis of the terminology, in which lists of lexical units from the text are obtained and arranged depending on their frequency - from the most frequent terminological units and from the units which are not included yet and which can be added as terms. Frequently, after this first analysis, and depending on the system used, there is a pre-translation stage in which the total number of full matches found in the TM and the lexical units incorporated in the terminological database are automatically translated.

During the translation phase, the task of translating is usually done from a word processor-like interface assisted by the suggestions of translations offered by the TM and by tools such as the terminology management system, the translator's glossaries or the Internet.

Finally, in the revision stage, the translation is checked, the target text undergoes a verification stage with the terminology database, and the automated quality control processes are performed in order to detect the degree of terminological consistency between the text, the glossary and the pair of texts.

Thus, the ideal texts whose features would allow a better use of the translation memories are those with repetitions, such as user guides, software help guides and web pages, for which these tool were initially conceived.

Translation memories are also useful in those texts with high internal repetition of the terminology or phraseology and those with typical repetitions such as annual reports, agreements, minutes, etc. (Valli, 2012). Two types of repetitions can be distinguished: internal repetitions, which can be found within the same text, and external repetitions, which occur throughout different texts belonging to the same text typology.

In the case of the legal documents, there is a high repetition index, both internal and external, since their technical and semi-technical term density is highly recurrent (Mayoral, 2007; Rubio and Candel, 2013). Moreover, it is a text typology which tends to follow very fixed structures, using even templates in order to create new legal documents (Borja, 2000; Ortega Arjonilla, 1999).

Hence, due to these common features of legal texts, it is fully justified the use of computer-assisted translation tools with this kind of texts.

### 3. Analysis and Discussion

The corpus used in this work comprises 29 legal texts which belong to the genre of real estate sales and rent contracts, with a total of 37,483 words. The texts came from two different sources: internet search of sample and template texts and authentic models and texts from law firms.

The tools with which the analysis and the exploitation of the corpus were performed are SDL Trados Studio and PhraseMiner. The computer-assisted translation tool SDL Trados Studio allows to make a preliminary analysis of the documents to translate and with this analysis the translator obtains information about the internal repetitions in the document analyzed, the number of words, as well as the number of segments, number of characters and their degree of coincidence or matching with the translation memory of previously translated segments, classified by the percentage of coincidence: partial or full.

PhraseMiner is a tool for internal identification and extraction of partial matches, expressions and terminology. The results obtained are compatible with computer-assisted translation systems and they allow the elaboration of teaching resources. PhraseMiner works with MSWord macros and includes modules which allow a wide range of functions such as analysis and terminology extraction, internal partial matches extraction, comparison and extraction of subsegment.

The exploitation of the corpus was carried gradually, starting from common phraseology and lists of frequent words, to collocations and specialized terminology.

For the exploitation of the corpus with teaching purposes for the training of translation of contracts, the extraction was proposed from the widest perspective and aimed at achieving homogeneity, translation quality and improvement of the translator's productivity. Therefore, not only the most common specialized terms are included, but also usual collocations, typical phraseology and lexical bundles or groups of words which, even not forming an established phraseological unit, are taken into account due to their frequency of use. The criteria followed for the extraction can be summarized in three categories:

- *belonging criterion*: every element which has been extracted from the corpus of sale contracts belongs to this domain;

- *economy criterion*: priority has been given to the occurrence frequency and to the detection of possible difficulties for the future translator;

- *teaching criterion*: as a guide for the translator, some terms are included in order to show the word-formation process of usual terms in the specialized area.

Table 1. Phraseology and Lexical bundles

Frequency in corpus	Phraseology and Lexical bundles	
128	Ambas partes	- llevan a cabo - celebraron - renuncian
42	A la parte vendedora	
12	A instancias de cualquiera de las parte	
12	A la finalización del acuerdo	
11	Según acredita	
9	A favor de la parte	
8	Conforme viene establecido	
4	A cuenta del precio total	

Table 1 shows a selection of the typical phraseology extraction of the sale contracts, and which, ultimately, gives ease to this kind of texts. For the translators’ training, this phraseology has a dual purpose: to reveal the high frequency of use and the composition of the elements of the established expression.

Table 2. Doublets

Frequency in corpus	Doublets
27	Lugar y fecha
24	Gastos e impuestos
20	Libre de cargas y gravámenes
14	A los juzgados y tribunales
12	Firman y rubrican
9	Libre y espontáneamente
5	Cesión y subarriendo
5	Daños y perjuicios

Doublets or binomic expressions (Table 2) are, according to Borja (Borja, 2005), one of the phenomena which more contribute to the “magical poetry of the legal language and which have disappeared from other genres but remain in the legal language”, possibly they are quite obscure for the future translators since they are archaic and established structures.

Table 3. Collocations

collocations	
contrato	privado de compraventa
	inscribible
	de leasing
	de compraventa
	de arrendamiento
conforme	a
	viene establecido
	establece
copia de	contrato de compraventa
	contrato de arrendamiento

Table 3 shows the most frequent collocations useful to apprehend the typical features of a determined kind of text or speciality field, as according to Alonso Ramos (Alonso Ramos, 2005), learners make more grammar mistakes, due to their unawareness of the proper collocation in the proper moment.

Table 4. Specialised terminology

Frequency in corpus	Term
538	parte compradora
420	presente contrato
234	Escritura Pública
140	parte vendedora
96	Comunidad de Propietarios
76	contrato de arrendamiento
64	opción de compra
54	Registro de la Propiedad
50	mayor de edad
50	contrato de compraventa
38	arrendamiento de vivienda
30	cuenta corriente
28	otorgamiento de la escritura
20	libre de cargas
13	aval bancario

This last table (Table 4) shows a sample of the 234 elements extracted that belong to the technical and semi-technical recurrent terminology in these documents.

#### 4. Conclusions

After the initial analysis of the corpus and the extraction of terminology and phraseology, it can be concluded that from a total of 37,483 words, 25% are repetitions both at lexical and phraseological level, and also both at internal and external level, that is, within the same document and in all the texts of the corpus. Thus, the training of future translators achieves a twofold benefit: the awareness of the existence of typical repetitions within legal texts, as well as the students' acquaintance with real examples extracted from the texts object of translation

On the other hand, in translator training, the preliminary work of documentation and terminology and phraseology extraction has a crucial role with this double purpose, professional and didactic.

#### References

- Borja Albi, A. (2000). *El texto jurídico inglés y su traducción al español*. Barcelona: Ariel.
- Borja Albi, A. (2004). La investigación en traducción jurídica. In García Peinado y Ortega Arjonilla (Eds.), *Panorama actual de la investigación en traducción e interpretación* (pp. 415-426). Granada: Atrio.
- Borja Albi, A. (2005). ¿Es posible traducir las realidades jurídicas? Restricciones y prioridades en la traducción de documentos de sucesiones británicas al español. In Monzó, E & Borja, A. (Ed.), *La traducción y la interpretación en las relaciones jurídicas internacionales* (pp. 65-90). Castellón: Publicacions de la Universitat Jaume I.
- Candel Mora, M. A. (2010). Las memorias de traducción para la enseñanza de la traducción especializada. In Alarcón Navío, E. (Ed.), *La traducción en contextos especializados. Propuestas didácticas* (pp. 267-276). Granada: Atrio.
- Mayoral Asensio, R. (2007). Comparación de los contratos en inglés y en español como ayuda al traductor. *Papers Lextra*, 3, 55-61.
- Ortega Arjonilla, E. (1999). Algunas reflexiones sobre la traducción en y para la administración de justicia desde una perspectiva académica y profesional. In Feria García, M. C. (Ed.), *Traducir para la justicia* (129-152). Granada: Editorial Comares.
- Rubio Donat, T. & Candel-Mora, M.A. (2013) Análisis de la optimización del entorno de trabajo con tecnología de traducción asistida aplicada a la traducción de contratos de compraventa. In Ortega Arjonilla, E. (Ed.) *Traducir la Cultura. Vol.3. De barreras culturales en la traducción jurídica, económica y judicial* (151-162). Granada: Comares.
- Valli, P. (2012). A study on concordancing: EU policy areas as translation problems. In Zibatow, L., Petrova, A. & Ustaszewski, M. (Eds.), *Translation Studies: old and new types of translation in theory and practice*. Frankfurt: Peter Lang Publishing.