

Generation of a glossary for the translation of housing purchase and sale agreements in Spain, Argentina, the United Kingdom, and the United States: a corpus-based methodology

Generación de un glosario para la traducción de contratos de compraventa de viviendas en España, Argentina, Reino Unido y Estados Unidos: una metodología basada en corpus

LORENA ARCE ROMERAL

Universidad de Valladolid. Facultad de Traducción e Interpretación. Campus Duques de Soria. Calle Universidad, s/n, 42004, Soria, Castilla y León, España.

Dirección de correo electrónico: lorena.arce@uva.es

ORCID: <https://orcid.org/0000-0003-0851-1967>

MIRIAM SEGHIRI

IUITLM, Universidad de Málaga. Facultad de Filosofía y Letras. Departamento de Traducción e Interpretación. Avenida de Cervantes, n.º 2, 29071, Málaga, Andalucía, España.

Dirección de correo electrónico: seghiri@uma.es

ORCID: <http://orcid.org/0000-0002-6817-6665>

Recibido: 30/9/2020. Aceptado: 19/10/2021.

Cómo citar: Arce Romeral, Lorena y Seghiri, Míriam, «Generation of a glossary for the translation of housing purchase and sale agreements in Spain, Argentina, the United Kingdom, and the United States», *Hermēneus. Revista de Traducción e Interpretación*, 24 (2021), pp. 87-118.

DOI: <https://doi.org/10.24197/her.24.2022.87-118>

Abstract: As a result of the integration of new technologies in the current educational and professional environment, many proposals have advocated the use of *ad hoc* corpora in translation and interpreting disciplines, due to their numerous advantages (Zanettin, 1998 and 2002; Corpas, 2008; Luque Giráldez and Seghiri, 2019 and 2021; or Sánchez Ramos 2019a, 2019b, 2019c and 2020). Since corpora are a very valuable source of grammatical, textual or terminological information, this paper illustrates a protocolized methodology for terminology extraction based on the bilingual and comparable corpus named VENTACOR in order to generate a glossary to translate housing purchase and sale agreements. The AntConc program, a corpora management tool, has been used for this purpose. The glossary will be bilingual (Spanish-English) but delimited to purchase and sale agreements written in Spain and Argentina, for the Spanish language, as well as in the United Kingdom (England and Wales) and the United States (Louisiana),

for the English language. VENTACOR is a representative corpus that has been created according to the specific design criteria and the systematized compilation protocol established by Seghiri (2006 and 2015).

Keywords: Purchase and sale agreements, housing, corpus linguistics, ICTs, translation.

Resumen: Como resultado de la integración de las nuevas tecnologías en el actual entorno educativo y profesional, muchas propuestas han abogado por el uso de corpus *ad hoc* en las disciplinas de la traducción y la interpretación, debido a sus numerosas ventajas (Zanettin, 1998 and 2002; Corpas, 2008; Luque Giráldez y Seghiri, 2019 y 2021; o Sánchez Ramos 2019a, 2019b, 2019c y 2020). Dado que los corpus son una fuente muy valiosa de información gramatical, textual o terminológica, en el presente documento se ilustra una metodología protocolizada de extracción de terminología basada en el corpus bilingüe y comparable, denominado VENTACOR, con el fin de generar un glosario para la traducción de contratos de compraventa de viviendas. Para ello se ha utilizado el programa de gestión del corpus AntConc. El glosario será bilingüe (español-inglés) pero delimitado a contratos de compraventa redactados en España y Argentina, para el español, así como en Reino Unido (Inglaterra y Gales) y Estados Unidos (Luisiana), para la lengua inglesa. VENTACOR es un corpus representativo y de calidad que se ha creado según criterios específicos de diseño y un protocolo de compilación sistemático establecido por Seghiri (2006 and 2015).

Palabras clave: Contratos de compraventa, viviendas, lingüística de corpus, TIC, traducción.

Summary: 1. Introduction; 2. The VENTACOR corpus; 3. Methodology for the implementation of a glossary based on the VENTACOR corpus, 3.1. Extraction of term candidates, 3.2. Generation of a Glossary

Sumario: 1. Introducción; 2. El corpus VENTACOR; 2.1. Compilación del corpus VENTACOR; 3. Metodología para la implementación de un glosario basada en el corpus VENTACOR, 3.1. Extracción de los candidatos a término, 3.2. Generación del glosario; 4. Conclusiones; Referencias.

1. INTRODUCTION

According to the World Tourism Organization (WTM), Spain is the second tourist destination in the 2019 (with 82.7 million tourists), followed by the United States and with the United Kingdom in ninth position. By nationality, the British visitors are the most frequent in Spain with the Americans in seventh position (*Statista*,¹ 2019).² An important part of these tourists, motivated by the excellent climate, decide to settle down in Spain. This is indicated by data³ from the Spanish Statistical Office

¹ *Statista* is a data business platform providing online data on the global digital economy, industrial sectors, consumer markets, public opinion, media, demography and macroeconomic trends. Quantitative data from 425 economic sectors in 50 countries are provided with a wide range of infographic tools. Available at <https://www.statista.com/>.

² Available at <https://es.statista.com/estadisticas/475075/numero-de-turistas-internacionales-en-espana-por-pais-de-residencia/>.

³ Available at https://www.ine.es/censos2011_datos/cen11_datos_inicio.htm.

(Instituto Nacional de Estadística, INE)⁴ which highlights an increase in the foreign population in Spain of 234.1% in the last ten years. As a result, the foreign population in Spain is around 15-20%.

Moreover, according to *Statista* (2019),⁵ North Americans are the most active tourists in the United Kingdom, with Spain taking the fifth position in this ranking. Spain also ranks fifth and the United Kingdom tops the list in terms of the number of tourists visiting the United States (Hosteltur).⁶

Therefore, it is not surprising that many Spanish tourists end up settling in these countries. Labour and economic reasons have become the main causes for international migration. Thus, the profile corresponds mainly to young people motivated by favourable working conditions (International Organization for Migration,⁷ 2020).⁸ In this sense, according to data from the INE,⁹ the number of Spanish people living abroad continues to grow. Thus, by country, the following countries stand out (data for the year 2019): Argentina with 473,519 Spanish residences; the United States with 167,426 and the United Kingdom with 152,291.

These trends in the tourism sector, together with the increase in the foreign population in the above countries, have led to a growth in the

⁴ Instituto Nacional de Estadística (INE) is a legally independent administrative autonomous institution assigned to the Spanish Ministry of Economic Affairs and Digital Transformation (Ministerio de Asuntos Exteriores y Transformación Digital). The law assigns the INE an important role in public statistic activity, expressly placing it in charge of large scale statistical operations (demographic and economic censuses, national accounts, demographic and social statistics, economic and social indicators, coordination and maintenance of company directories electoral census training, among others. Available at <https://www.ine.es/en/index.htm>.

⁵ Available at <https://www.statista.com/statistics/616972/inbound-travel-country-markets-united-kingdom-uk-international-visits/>.

⁶ Hosteltur is a Spanish media company specialized in the professional tourism sector. Available at <https://www.hosteltur.com/>.

⁷ The International Organization for Migration (IOM), established in 1951, is the leading inter-governmental organization in the field of migration and works closely with governmental, intergovernmental and non-governmental partners. With 173 member states, a further 8 states holding observer status and offices in over 100 countries, IOM is dedicated to promoting humane and orderly migration for the benefit of all. It does so by providing services and advice to governments and migrants. Available at <https://www.iom.int/>.

⁸ Available at https://publications.iom.int/system/files/pdf/wmr_2020_es.pdf.

⁹ Available at: https://www.ine.es/prensa/pere_2020.pdf.

housing purchase and sale by foreigners. The case of Spain is particularly noteworthy, since that growth consequently leads to an increase in residential tourism to meet the needs of tourists who come to Spain during the summer period or for those who decide to purchase a dwelling as their habitual residence. This is what the Spanish newspaper *El Confidencial*¹⁰ states: foreigners buy 20 out every 100 residences sold in Spain.¹¹ Specifically, according to data from the *National Bank of Spain (Banco de España)*,¹² in the year 2020, the British buyer continues to be the most active in terms of transactions carried out by foreigners, both residents and non-residents, representing 12.9% of all real estate transactions carried out in Spain by foreigners. Also remarkable is the increase in Argentina (12.7%) and the United States (11.3%), according to data from the General Council of Notaries (Consejo General del Notariado)¹³ for 2019.¹⁴

On the other hand, according to the Spanish Institute for Foreign Trade (ICEX: España Exportación e Inversiones) in 2018, Spain's accumulated investment stock in Argentina is currently over 5.8 billion euros, an amount only exceeded by US investments in the country.¹⁵ Particularly, the pillars of the Spanish economy in Argentina are, according to the Spanish newspaper *El Mundo*, energy, technology, construction and banking¹⁶.

¹⁰ *El Confidencial* is a Spanish digital newspaper of general information, specialized in economic, financial and political news, founded in 2001. Available at <https://www.elconfidencial.com/>.

¹¹ Available at https://www.elconfidencial.com/vivienda/2013-10-15/los-extranjeros-ya-compra-n-20-de-cada-100-viviendas-que-se-venden-en-espana_41404/.

¹² Banco de España is the central bank of Spain. Established in Madrid in 1782, is a member of the European System of Central Banks and is also Spain's national competent authority for banking supervision within the Single Supervisory Mechanism. Available at <https://www.bde.es/bde/es/>.

¹³ Consejo General del Notariado Spanish is a centre for statistical information on notaries. Available at <https://www.notariado.org/liferay/web/cien/estadisticas-principales/inicio>.

¹⁴ Available at <https://www.notariado.org/portal/-/en-el-primer-semestre-de-2019-la-compraventa-de-vivienda-libre-por-parte-de-compradores-extranjeros-1>.

¹⁵ Available at <https://www.icex.es/icex/es/navegacion-principal/todos-nuestros-servicios/informacion-de-mercados/paises/navegacion-principal/el-mercado/estudios-informacion/DOC2018783468.html?idPais=AR>.

¹⁶ Available at <https://www.elmundo.es/elmundo/2012/04/16/economia/1334598408.html>.

The above data evidence a real estate purchase and sale business, both nationally and internationally, which generates a growing demand for translation and writing of housing purchase and sale agreements in Spanish (Peninsular and Argentinian varieties) and English (British and American varieties). Thus, the main objective of this work consists of facilitating the processes of translation of contracts in real estate transactions between Spain, Argentina, the United Kingdom and the United States. A delimitation has been made to contracts produced in England and Wales—in the case of the United Kingdom—as well as in the State of Louisiana¹⁷ for the United States.

The importance of the documentation task in the translation and interpreting field has been evidenced by the curricular design in the Bachelor's Degree in Translation and Interpreting. An example of this is the *White Paper on the Bachelor's Degree in Translation and Interpreting (Libro Blanco del Título de Grado en Traducción e Interpretación)*, which sets the curricula in Spanish universities within the framework of the European Higher Education Area. The documentary resources for translators and interpreters are multiple and diverse. However, bilingual dictionaries are the most used sources by students (Roberts, 1990 y 1992). However, dictionaries or similar resources such as glossaries or terminology databases present words as isolated units without context and do not show how words are combined. Therefore, the only resource that can offer us such advantages is the corpus, since corpora are examples of a language (EAGLES, Expert Advisory Group on Language Engineering Standards, 1996b).

¹⁷ Due to the large number of states that make up the United States, the study will focus on the State of Louisiana. The reason that justifies this choice is the proximity between the legal systems of Spain and Louisiana, which makes it easy to establish a comparison between both regulations. Thus, the Louisiana legislation derives from the Napoleonic Code, and consequently, from Roman Law, from which Spanish legislation (and, likewise, the Argentine legal system) also comes. In other words, they are based on a system of codified laws, which is called civil law. Thus, Louisiana constitutes the exception in the United States, since the rest of the American legal systems have been based on the common law, as a legislative source. The common law, unlike civil law, is made up of a set of unwritten norms and not enacted or sanctioned; it is based, therefore, on the eminently jurisprudential law.

Since a *corpus* can be defined as a “(...) collection of naturally-occurring language text, chosen to characterize a state or a variety of a language” (Sinclair, 1991: 171), it constitutes a grammatical, discursive, lexicographic, terminological and macro source of documentation. This is especially useful for legal speeches, since it is evident that they have their own conventions, routines and recurrent structures. In fact, according to Alcaraz and Hughes (2002), the legal discourse has a distinct formalism. According to Laviosa (1998), Bowker and Pearson (2002) and Zanettin et al. (2003), the ideal type of corpus is the called *virtual corpus*.¹⁸ Virtual corpora are also easy to use and allow accessing and managing large amounts of information in seconds, if they are managed with a suitable corpus management tool for that purpose.¹⁹

To achieve the main objective of this work, we will create a flexible, low-cost and user-friendly glossary based on a representative, reliable, comparable and virtual corpus named VENTACOR. The VENTACOR corpus²⁰ comprises housing purchase and sale agreements in Spanish (Peninsular and Argentinean varieties) and English (British and American varieties).²¹

2. THE VENTACOR CORPUS

The following section describes a methodology for creating a bilingual glossary (Spanish-English) delimited to housing and purchase and sale agreements written in Spain and Argentina—for the Spanish language—as well as in the United Kingdom (England and Wales) and

¹⁸ A virtual corpus is also referred to as a corpus for specific, *ad hoc*, electronic, disposable or web purposes, among others.

¹⁹ Other works that apply corpus exploitation for terminological extraction are those of Pérez Carrasco and Seghiri (2022 and 2021) and Arce and Seghiri (2018b, 2020b y 2020c).

²⁰ For further information about the VENTACOR corpus, see Arce and Seghiri (2018a, 2019, 2020a, 2021 and 2022).

²¹ Other works that focus on the study of corpus-based real estate contracts are worth mentioning, such as those by Arce and Seghiri (2019, 2020a, 2020b and 2021), which address the study of the textual macro-structure of housing purchase and sale agreements in the United Kingdom, the United States, Spain, Argentina and Ireland. In addition, on the purchase and sale and rental in the United Kingdom and France, the work of Toledo Baez (2016, 2018 and 2019) stands out.

the United States (Louisiana) —for the English language—. This methodology has been based on the compilation of the comparable *ad hoc* corpus VENTACOR and its exploitation with AntConc.²²

2. 1. Compilation of the VENTACOR corpus

According to Seghiri (2006), for a collection of texts to be considered a corpus, it must be compiled according to specific parameters so that it can represent a state or a section of a language. Thus, the corpus must be correctly designed and documents comprising the corpus must be selected according to a protocolized methodology, which ensures the qualitative representativeness of the corpus. This methodology (*cf.* Seghiri, 2006, 2008, 2010, 2012a, 2012b, 2014, 2015, 2017a y 2017b) is divided into two differentiated steps: design criteria and compilation protocol.²³

Firstly, the clear parameters are established for the design of the corpus of housing purchase and sale agreements. Thus, it will be a *virtual* corpus (since it is integrated only by texts downloaded from the Internet), *comparable* (it includes original agreements, and not translations) and, therefore, *specialized in legal texts* and *complete*, since it is composed by full agreements. Likewise, it will be *bilingual* (English-Spanish) and delimited to housing purchase and sale agreements written in Spain and Argentina, for the Spanish language, as well as the United Kingdom (specifically in England and Wales) and the United States (Louisiana, in particular), for the English language.

Secondly, the protocolized compilation methodology is integrated by four phases: searching, downloading, text formatting and saving data. The first phase consists of *searching* for documents of the corpus on the Internet. After locating these, they must be *downloaded*. The texts can be downloaded manually or automated for groups of pages using programs that allow them to be downloaded in batches, such as GNUWget²⁴ or

²² Another similar work in which they apply this methodology for other countries such as Ireland is that of Seghiri y Arce (2021).

²³ Before proceeding to the exploitation of the VENTACOR corpus, the compilation methodology followed will be briefly addressed. For a detailed description of the design and compilation aspects of the VENTACOR corpus, see Arce Romeral y Seghiri (2018a).

²⁴ Available at: <https://www.gnu.org/software/wget/>.

GetBot.²⁵ In the third phase, *text formatting*, the documents must be edited to ASCII or plain text (.txt) format, by copying and pasting it into a plain text document. If the documents are encrypted, it is necessary to use programs such as Abby Fine Reader. The last step, *saving data*, consists of storing the documents in folders and subfolders and designing a clear code that allow texts to be stored and easily located.

Thanks to these four steps to compile the corpus (searching, downloading, text formatting and saving data) and the specific design criteria, we have obtained a corpus that can be considered representative from a qualitative point of view. The corpus, named VENTACOR, is comparable, bilingual (English-Spanish) and consists of:

- 60 agreements (and 147,825 words²⁶ or *tokens*) for the **subcorpus in Peninsular Spanish** (Spain).
- 60 agreements (and 84,140 words or *tokens*) for the **subcorpus in Argentinean Spanish** (Argentina).
- 60 agreements (and 75,253 words or *tokens*) for the **subcorpus in British English** (United Kingdom, specifically England and Wales).
- 10 agreements (and 48,070 words or *tokens*) for the **subcorpus in American English** (United States, Louisiana in particular).

To verify that the corpus is representative from a quantitative point of view, i. e., to check that the compiled corpus covers the basic terminology of this field of speciality. For this purpose, we have used the ReCor²⁷

²⁵ Available at: <http://www.getbot.com/>.

²⁶ To count the number of words we have used Word Count Tool available at: <http://wordcounttool.com/>.

²⁷ Recor is an effective solution to determine *a posteriori*, for the first time, the minimum size of a corpus or textual collection, regardless of the language or textual genre of that collection, establishing, therefore, the minimum threshold of representativeness through an algorithm (NCor) of analysis of the lexical density as a function of the incremental increase of the corpus (see Seghiri, 2006, 2014 and 2015). The programme can be ordered free of charge at <https://umapatent.uma.es/es/patent/metodo-para-la-determinacion-de-la-representa4b0>.

program. The Peninsular Spanish subcorpus begins to be representative approximately at 55 documents and 55,5000 *tokens*; the Argentinean Spanish subcorpus about 49 documents and 50,500 *tokens*; the British English subcorpus approximately at 47 documents and 60,100 *tokens*; while the American English subcorpus is representative at 1 document²⁸ and 4,807 *tokens*.

Therefore, ReCor has shown that the VENTACOR corpus is not only qualitatively but also quantitatively representative, so it is ready to be exploited and managed by a corpus management tool.

3. METHODOLOGY FOR THE IMPLEMENTATION OF A GLOSSARY BASED ON THE VENTACOR CORPUS

Once we have compiled a representative corpus, in both qualitative and quantitative terms, the next step is to exploit and manage the corpus VENTACOR to create a glossary on housing purchase and sale agreements. This aim will be achieved through the AntConc programme.

AntConc is a freeware corpus analysis toolkit for text management. It is a free, multi-platform corpus management program. This tool is one of the most popular in terms of management of comparable corpus, as it provides many functions. One of the most interesting ones is the possibility of generating word indexes and clusters (groups of words), obtaining the contexts of a word in question or an expression, among others. In addition, it allows lemmatization, choice of alphabets and obtaining a system of references in which contexts are exposed, as well as the flexible search of words, set phrases or expressions, for example. At the same time, it offers the possibility of analysing the absolute and relative frequencies of words and the forms of different words. AntConc home interface shows all its functionalities in its upper part, as shown in the following figure (see figure 1):

²⁸ The reason why the American English subcorpus is representative at 1 document is because this agreement has an *ad solemnitaten* form, as there is a unique model prescribed by the Louisiana Real Estate Commission. Thus, the said institution approved the Act 333, as an annex to the Louisiana Real Estate License Law, by virtue of which a standardized model of housing sale contracts is established. For further information, see Arce Romeral and Seghiri (2018a, 2020a and 2020b).

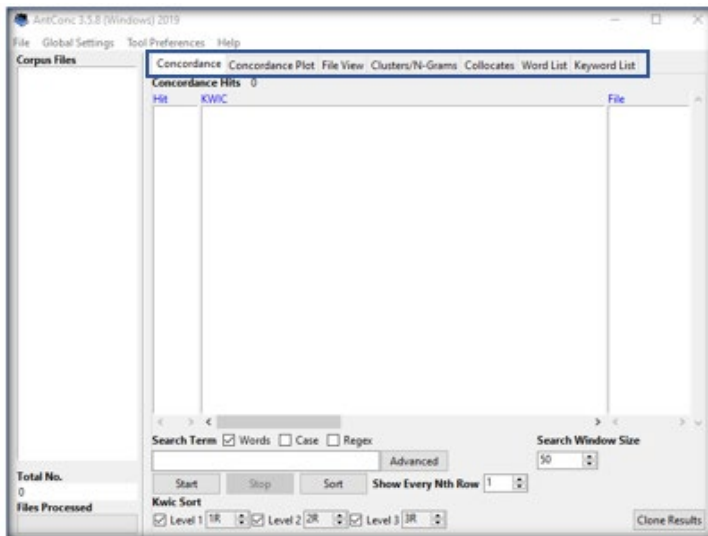


Figure 1. AntConc interface (functions)

For the implementation of the glossary, the “Word List” option will be used. The process has been divided into well-differentiated phases, namely, extractions of term candidates and translations equivalents and generation of the bilingual glossary.

3.1. Extraction of term candidates

The AntConc program allows the user to extract words or groups of words that are specialized in their own context from a frequency list (“Word List” function). This list helps, at first, to check which words are repeated most frequently and to start evaluating possible terms. To do this, it is necessary to upload to AntConc, separately, each of the subcorpus that are going to be analyzed. The corpus VENTACOR, as we have seen, is composed of four subcorpora: on the one hand, a Peninsular Spanish subcorpus, an Argentinean Spanish subcorpus, a British English subcorpus and an American English subcorpus.

So, to begin with, we must select the option “File” [1]; then “Open File [2]”, and look for the folder containing the subcorpus to be uploaded

to AntConc. Only one subcorpus can be uploaded at a time [3] (see figure 2):

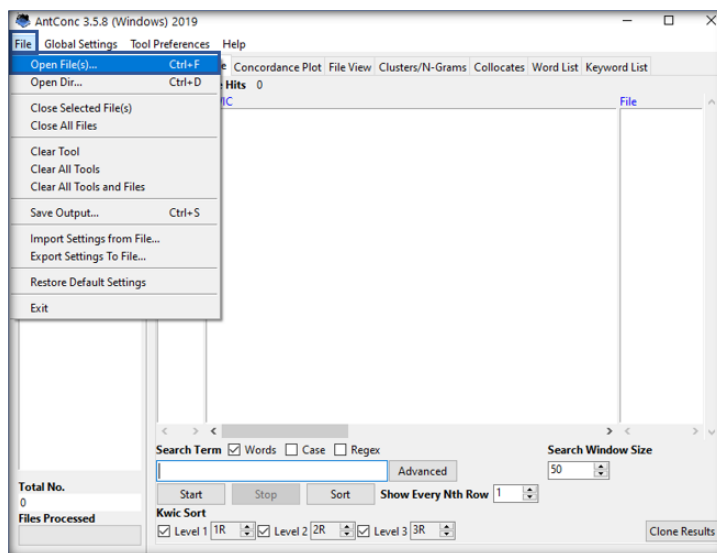


Figure 2. Corpus uploading

Uploaded the chosen subcorpus to the AntConc program, the next step is to generate a provisional word list in order to start considering possible term candidates and their equivalents. Although the process must be done for each subcorpus, due to space limitations, it will be exemplified with the Peninsular Spanish subcorpus.

Once the documents in plain text of the Peninsular Spanish subcorpus have been uploaded, the process to generate a word list is explained below. Firstly, it is necessary to select the “Word List” tab [1]. Then, the appropriate “ordering options” must be adjusted [2]. The words can be ordered by frequency or by the start or end of the word, and the ordering can be inverted. In this case, we are going to order the words from highest to lowest frequency, option that AntConc uses by default. The word list can also be generated in case-insensitive mode, where words in upper and lower case are treated the same (default) or case-sensitive, where words in upper and lower case are treated separately. For this study, the first option has been chosen, since we do not consider relevant the differentiation

between upper and lower case for the glossary. Secondly, click on “Start” [2]. The result that we obtain in this operation is a list of words ordered by the highest frequency for the Peninsular Spanish subcorpus (see figure 3):

Rank	Freq	Word
1	25153	de
2	13384	la
3	10148	el
4	9714	en
5	7218	y
6	6999	a
7	6286	que
8	6161	los
9	5438	del
10	4176	las
11	4124	se
12	3977	o
13	3331	por

Figure 3. Word list ordered by the highest frequency.
(Peninsular Spanish subcorpus)

Once each of the four subcorpora has been uploaded separately and the process of generating the list of words has been completed, four lists of words have been extracted, corresponding to each of the language subcorpus of the VENTACOR corpus, namely, Peninsular Spanish subcorpus, Argentinean Spanish subcorpus, British English subcorpus and American English subcorpus.

From the above results (see figure 3), as can be seen, the word list contains words that would not be relevant to the glossary (such as “de”, “la”, “el”, “en”, “y” or “a”, among others). Thus, in order to extract the possible term candidates that will be part of the glossary, it is necessary to filter the results of the four lists generated, in order to eliminate words that, although they have grammatical meaning, do not provide terminological information. These may be adverbs, prepositions, conjunctions, or articles, for example, but also words with very general content and that, therefore, are the most frequent words. In order to “clean up” each of the word lists, two stop-words lists will be used. Although stop-words lists, especially

useful for glossaries, are available in different languages on the Internet, they can be made manually. To do so, a list must be created manually in a plain text file (.txt) with the words to be deleted. The words of the exclusion list must be separated from each other by commas (,) or paragraph breaks (¶). In our case, we have downloaded a stop-words list in Spanish²⁹ and another in English³⁰ from the Internet. In order to be used in the AntConc program, the content of the Internet lists must be copied and pasted into a plain text document (.txt). The Spanish plain text exclusion list has been saved with the name StopwordslisSP (see figure 4). This stop-words list has been used to filter the two Spanish subcorpora, both the Peninsular variety and the Argentinean variety.

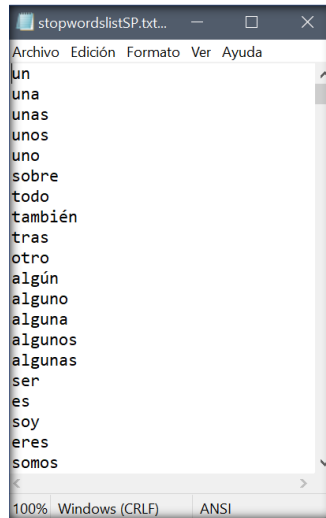


Figure 4. StopwordslisSP

Below is the stop-words list that has been used to filter the two English subcorpora, both the British variety and the American variety, copied into a .txt document (see figure 5):

²⁹ Available at <https://www.ranks.nl/stopwords/spanish>.

³⁰ Available at <https://www.ranks.nl/stopwords>.

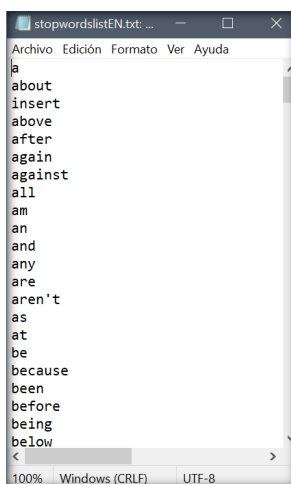


Figure 5. StopwordlistEN

Below is shown how to upload the lists and filter the lists of words of each of the four subcorpora, to extract a list of words with the terminology specific to the field of specialisation under study in this study, namely, housing purchase and sale. Firstly, go to “Tools Preferences” [1], and in the “Category” option [2], select “Word List” [3]. Secondly, in “Word List Preferences” [4], choose “Treat all data as lower case” option [5]. Next, click on “Open” [6] to upload the .txt file containing the stop-words list (see figure 6):

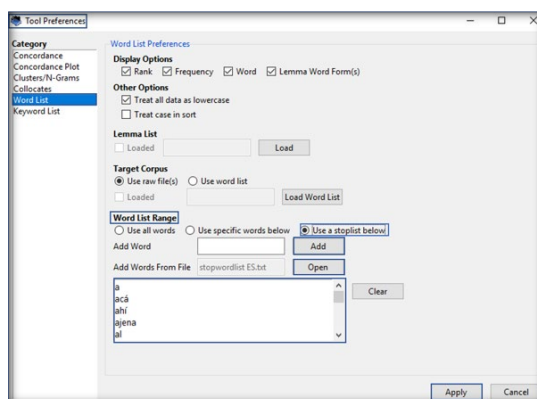


Figure 6. Steps for stop-words list uploading

As an example, the list of Spanish words has been uploaded. Next, the stop-words list will be uploaded to the box marked. To finish, click on “Apply” [7] (see figure 6).

In addition, it is necessary to consider the morphological variation, since the AntConc frequency analysis treats as different words those that have different morphological variation but the same root. Here is an example with the Peninsular Spanish subcorpus (see figure 7):

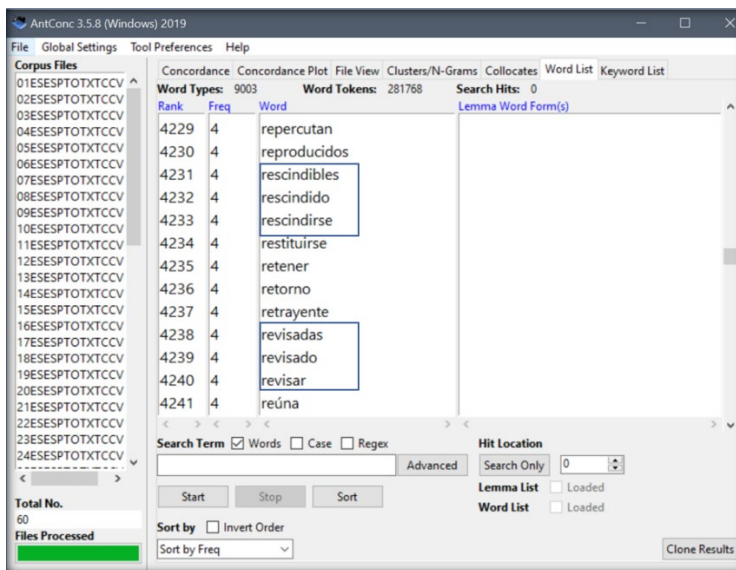


Figure 7. Word List of the Peninsular Spanish Subcorpus without lemmatization

Thus, in order for the frequency calculation function to count all the morphological variants of a word as one, it is necessary to lemmatize the VENTACOR corpus. For this purpose, two lemmatization lists are needed: one in Spanish, to lemmatize the Spanish subcorpora —Peninsular and Argentinean— and another in English, for the English subcorpora —British and American—. In this way, each word of the corpus (*token*) is related to its lemma. To do this, we need a text file where each word is related to its corresponding lemma. For AntConc, a lemma list can be

created by specifying the “lemma entry” follow by “->” and one or more “words” that should be assigned to the lemma separated by commas or tabs (with or without additional space around the words):

lemma→ MorphologicalVariationWord, MorphologicalVariationWord, etc.

See the example below:

vendedor→ vendedora, vendedores, vendedora

Lemma lists are also available for downloading online,³¹ and this option has been used in this study. So that the lemma lists downloaded from the Internet can be used with the corpus management program AntConc, it is necessary to copy their content and paste them into a plain document (.txt). These lists have also been completed manually with lemmas from the housing purchase and sale field. The list of lemmatization in Spanish is presented below (see Figure 8):

```

LemmatizationListSP.txt: Bloc de notes
-----
parte-> partes
contrato-> contratos
comprador-> compradora, compradores, compradoras
compraventa-> compraventas
vendedor-> vendedora, vendedores, vendedoras
escritura-> escrituras
entrega-> entregas
precio-> precios
caso-> casos
cuenta-> cuentas
vivienda-> viviendas
listo-> lista, listos, listas
estipular-> estipulado, estipulada, estipulados, estipuladas, estipulan, estipularán, esti
garantizar-> garantía, garantías, garantizado, garantizados, garantizada, garantizados, ga
comprar-> compra, compras, comprado, compradas, comprados, compradas, compran, comprarán
hipoteca-> hipotecas, hipotecado, hipotecada, hipotecados, hipotecadas, hipotecarán, hipo
sujetar-> sujeto, sujetas, sujetos, sujetas, sujeción
suministrar-> suministro, suministros, suministrado, suministrada, suministrados, suministra
supletorio-> supletorias, supletorios, supletorias
inscripción-> inscripciones
legal-> legales
artículo-> artículos
ley-> leyes
presente-> presentes
consumidor-> consumidores
servicio-> servicios
bloque-> bloques
préstamo-> préstamos
fondo-> fondos
inmobiliaria-> inmobiliarias
acreedor-> acreedora, acreedores, acreedoras
precarista-> precaristas
gasto-> gastos
fecha-> fechas
  
```

Figure 8. LemmatizationListSP

³¹ For the lemmatized list in Spanish, the updated lemma of the 22nd Edition of the *Dictionary of the Royal Spanish Academy (Diccionario de la Real Academia Española)* has been used, downloaded from <https://www.teoruz.com/lemario/lemario-20101017.txt>. For English, the list from the British National Corpus (BCN) has been used, extracted from: https://lexically.net/downloads/BCN_lemmafile5.txt.

In the same way, the lemmatization list in English has been downloaded from the Internet and copied into a .txt document with the name LemmatizationListEN (see figure 9):

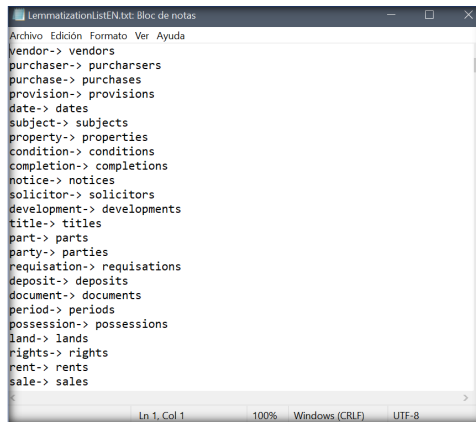


Figura 9. LemmatizationListEN

To upload a lemmatization list in AntConc, from the menu bar, in “Tools Preferences” [1], in the preferences window, select “Word List” [2]. Then, in “Lemma List Options” [3], select “Use lemma list file” [4]. Then, in the Open option, select the file with the list of lemmas in Spanish [5] and finally click on “Load” [6]. Finally, click on “Apply” [7] (see figure 10):

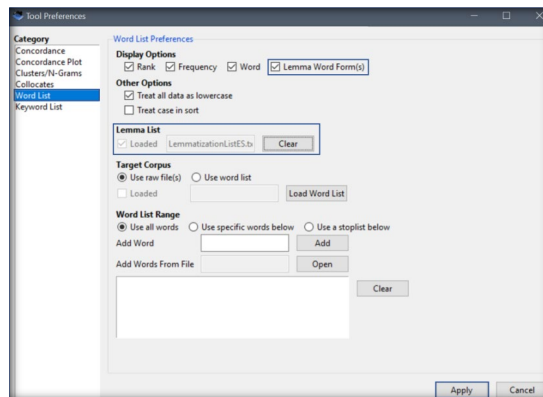


Figure 10. LemmatizationListSP uploading.

Once the Peninsular Spanish subcorpus has been lemmatized and the exclusion list has been applied, a new, definitive word list can be generated. As we can see, when the lemma list function is used, the “Lemma Word Form(s)” column will show the words in the corpus associated with each lemma. This list generated by AntConc will provide a list of term candidates for the glossary ordered by frequency (see figure 11):

Rank	Freq	Lemma	Lemma Word Form(s)
1	566	parte	parte 410 partes 156
2	490	comprador	comprador 114 compradora 340 comp
3	396	vendedor	vendedor 102 vendedora 246 vendedo
4	376	contrato	contrato 366 contratos 10
5	356	vivienda	vivienda 318 viviendas 38
6	254	compraventa	compraventa 254
7	226	presente	presente 226
8	196	escritura	escritura 196
9	170	entrega	entrega 162 entregas 8
10	156	precio	precio 148 precios 8
11	142	caso	caso 140 casos 2
12	134	cuenta	cuenta 134
13	132	gastos	gastos 132

Figure 11. Word list of term candidates ordered by frequency of the Peninsular Spanish subcorpus (I)

To save the results, first select “File” [1] and then “Save Output” [2]. Secondly, it is necessary to select the folder in which the results are to be saved [3], as well as give a name to the list [4] (see Figure 12). In this case, it has been as *WordListPENSP*, which means WordList (word list), PEN (Peninsular) and SP (Spanish).

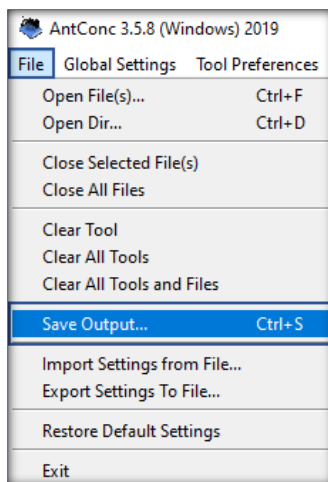


Figure 12. Word list of term candidates saving

The resulting file is a document in .txt format (see figure 13):

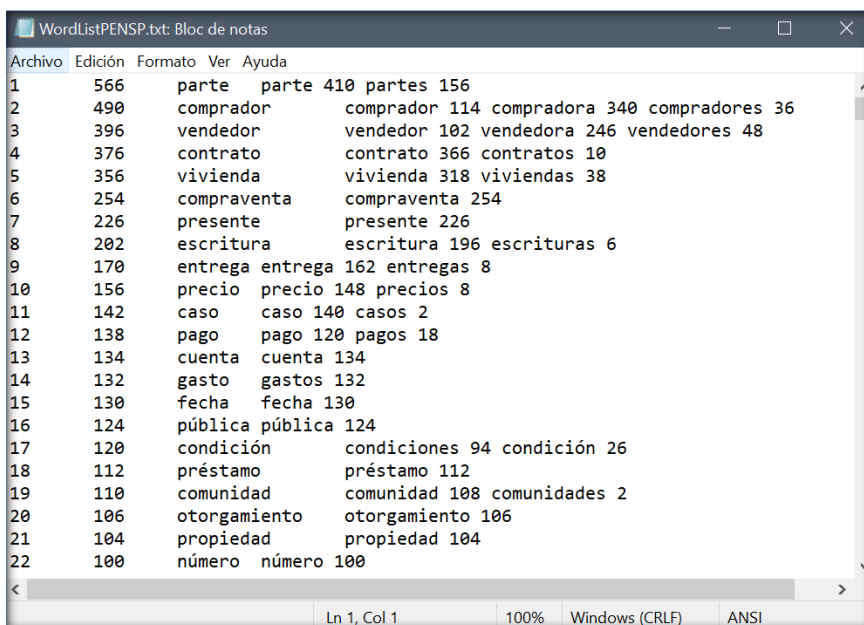


Figure 13. Word list of term candidates ordered by frequency of the Peninsular Spanish subcorpus (I)

The same methodology has to be repeated with the Argentinean Spanish subcorpus with the LemmatizationListSP. Likewise, the British English subcorpus and the American English subcorpus with the LemmatizationListEN should be lemmatized.

The term candidates of the Argentinean Spanish subcorpus are shown below (see figure 14). It has been given the name *WordListARGSP*, which means WordList (word list), ARG (Argentinean variety) and SP (Spanish).

Archivo	Edición	Formato	Ver	Ayuda
1	842	parte	partes	271 parte 571
2	637	presente		presente 637
3	626	comprador		comprador 270 compradora 356
4	588	vendedor		vendedora 394 vendedores 12 vendedor 182
5	280	inmueble		inmueble 280
6	273	precio	precio	273
7	234	contrato		contratos 12 contrato 222
8	225	dominio	dominio	225
9	216	escritura		escritura 216
10	204	días	días	204
11	194	acto	acto	194
12	188	venta		venta 188
13	184	pago	pagos	21 pago 163
14	183	operación		operación 183
15	177	podrá	podrá	177
16	173	nº	nº	173
17	168	fecha	fecha	168
18	165	provincia		provincia 165
19	164	calle	calle	164
20	161	suma	suma	161
21	160	caso	casos	34 caso 126
22	155	registro		registro 155
23	154	compraventa		compraventa 154

Figure 14. Word list of term candidates ordered by frequency of the Argentinean Spanish subcorpus

The word list containing the term candidates for the British English subcorpus is shown below (see figure 15). This list has been saved with the name *WordListUKEN*, which means: WordList (word list), UK (British variety) and EN (English).

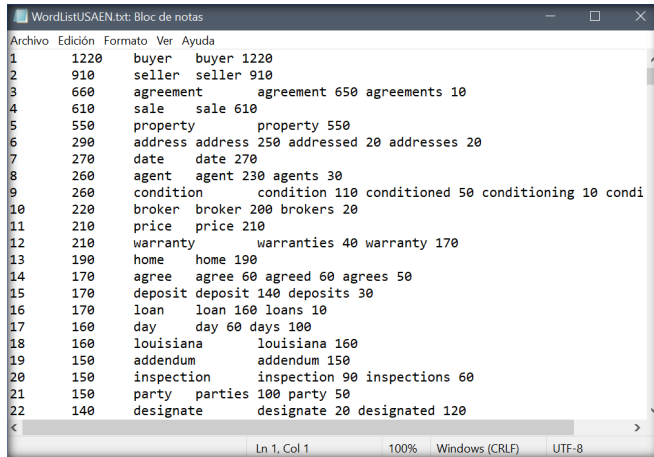


Figure 15. Word list of term candidates ordered by frequency of the British English subcorpus

The final word list for the American Spanish subcorpus is shown below (see Figure 16). This list has been given the name *WordListUSAEN*, which means: WordList (word list), USA (American variety) and EN (English).

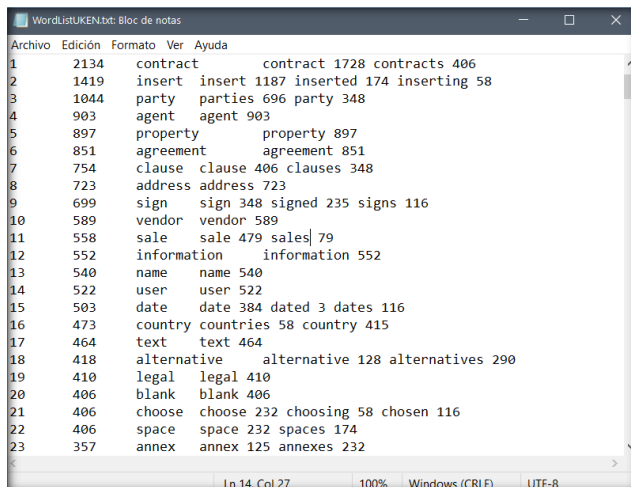


Figure 16. Word list of term candidates ordered by frequency of the American English subcorpus

After extracting the four lists of candidates for terms from the four subcorpora, the next step is the implementation of the bilingual (Spanish-English) glossary on home purchase contracts drawn up in Spain, Argentina, the United Kingdom (England and Wales) and the United States (Louisiana).

3.2. Generation of the glossary

Once the word lists with the term candidates have been extracted, in order to create the Spanish-English glossary, it is sufficient to connect each term in Spanish with its equivalent in English. At this point, although the search for the equivalent will be carried out manually, it usually finds a similar frequency in all subcorpora. Thus, once the English language equivalents have been found so as to implement the Spanish-English glossary, it is necessary to sort the Spanish term list alphabetically. To do this, select “Data” [1] in the main menu and then the option “Sort from A to Z” [2]. As a sample button, we include below the entries that begin with the letter A of the glossary (see table 1):

SPANISH		ENGLISH	
Peninsular Variety	Argentinean variety	British Variety	American Variety
Abogado	Abogado	Advocate / Solicitor / Lawyer	Advocate / Lawyer / Attorney / Attorney-at-law
Acreeedor	Acreeedor	Creditor	Creditor
Acto	Acto	Event / Occurrence	Event / Occurrence
Acta notarial	Acta notarial	Notarial certificate	Notarial certificate
Acuerdo legal informal	Acuerdo legal informal	Memorandum	Memorandum

Acuerdo / Pacto	Acuerdo / Pacto	Agreement / Contract	Agreement / Contract
Administrador	Administrador	Administrator	Administrator
Adquiriente / Comprador / Parte compradora	Adquiriente / Comprador / Parte compradora	Acquirer [not very frequently] / Buyer / Purchaser	Acquirer / [not very frequently]/ Buyer / Purchaser
Adquirir / Comprar	Adquirir / Comprar	Acquire [not very frequently] / To purchase / To buy	Acquire [not very frequently] / To purchase / To buy
Agente inmobiliario	Agente inmobiliario	Real (-) estate bróker / Broker	Real (-) estate bróker / Broker / Realtor
Anexo	Anexo	Annex / Annexe / Addendum	Annex / Addendum
Arbitraje / Arbitrio	Arbitraje / Arbitrio	Arbitration	Arbitration
Arras / Señal	Arras / Señal	Down-payment	Down payment
Arrendador	Arrendador / Locador	Lessor / Hirer / Renter	Lessor / Renter
Arrendamiento / Alquiler	Arrendamiento / Alquiler / Locación	Lease / Leasing / Leasehold / Rental / Rent / Hire	Lease / Leasing / Rental / Rent

Arrendamiento con opción a compra / Alquiler con opción a compra	Arrendamiento con opción a compra / Alquiler con opción a compra / Locación con opción a compra	Rental purchase / Rent-to-buy / Rent-to-own	Rental purchase / Rent-to-buy / Rent-to-own
Arrendatario / Inquilino	Arrendatario / Inquilino / Locatario	Lessee / Tenant / Leaseholder	Lessee / Tenant / Leaseholder
Artículo	Artículo	Section/ Article	Section / Article
Autorización	Autorización	Authorization	Authorisation

Table 1. Term list (entry A) of the bilingual (Spanish-English) glossary of housing purchase and sale agreements

Finally, the result in plain text format has been transferred to Excel format (see figure 17):

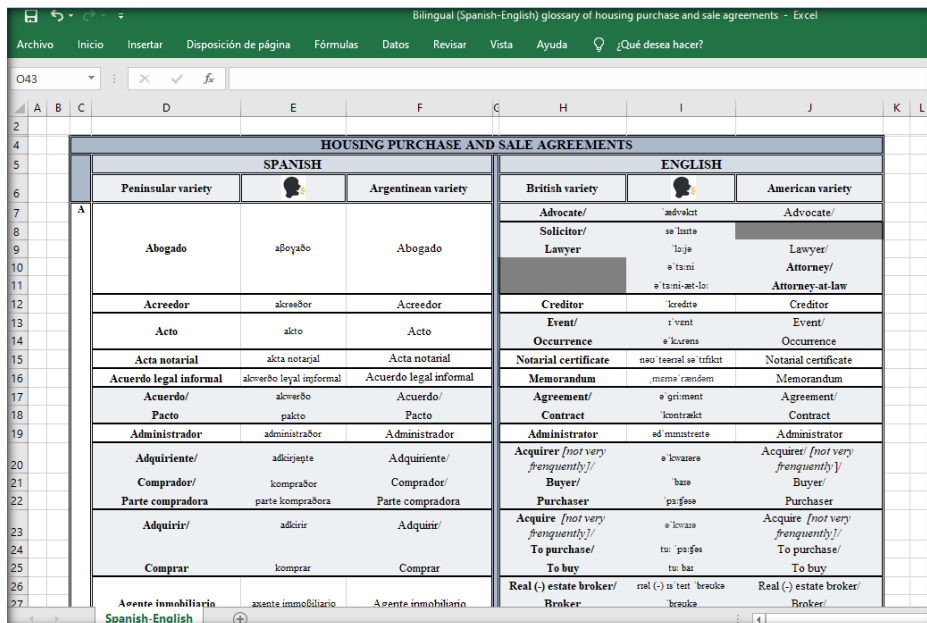


Figure 17. Bilingual (Spanish-English) glossary of housing purchase and sale agreements in .xls format

Phonetic transcription of terms in both languages³² has also been added to the glossary. This information is very useful for interpreters (see figure 17). The glossary (Spanish-English) of housing purchase and sale agreements is also available online.³³

4. CONCLUSIONS

The main objective of this paper is to prove the virtues of working with virtual and comparable corpora as a very useful tool to create glossaries for translators, since corpora represent a very valuable resource for the analysis of specialized terminological and phraseological extraction in its purely original production context. So, we have tried to prove the virtues of corpus linguistics to the documentary phase of translators. Specifically, a protocolized methodology for terminology extraction has been illustrated in order to generate a glossary on housing purchase and sale agreements.

The terminology extraction of the candidate terms for the generation of the glossary from the VENTACOR corpus has been executed with the AntConc programme and implemented in .xls format. The result has been a bilingual glossary (Spanish-English) for the translation of contracts written in Spain and Argentina, in the case of Spanish, as well as in the United Kingdom and the United States, in the case of English. This glossary is available online for consultation, which is very useful for the documentary phase for translators, and even legal interpreter or real estate agents, amongst others. This methodology of generating glossaries, due to its quick implementation and free cost, can be applied to any language and field. Thus, this methodology can be extremely useful not only for translators, but also for interpreters, who usually work against the clock.

³² For the phonetic transcription in Spanish, the Aucel program has been used, available at <http://www.aucel.com/pln/transbase.html>. In the case of English, PhoTransEdit has been used, available at http://www.photransedit.com/online/text2pho_netics.aspx.

³³ The bilingual (Spanish-English) glossary of home purchase contracts is available in .xls format at <https://cutt.ly/nsXDASa>. Likewise, on the website www.compra.ventacor.com in the section *Glosario* (Glossary).

In addition, the glossary created can also be extremely useful for real estate agents, notaries and lawyers dealing with real estate transactions with international clients.

ACKNOWLEDGMENTS

The present work is framed within the activities carried by the Research Institute of Multilingual Language Technologies (IUITLM, University of Málaga) and the LexyTrad Research Group (HUM 107, Junta de Andalucía), as well as within the framework of the projects INMOCOR (Ref. P20-00109, Junta de Andalucía), MI4ALL (AndalucíaTech), TRIAJE (Ref. UMA18-FEDERJA-067) y TorreznoTRAD (Ref. PROYEMER-2021-28).

REFERENCES

- Alcaraz Varó, Enrique and Hughes Brian (2002), *El español jurídico*, Barcelona, Ariel.
- Arce Romeral, Lorena and Miriam Seghiri (2018a), “Determinación de la representatividad cualitativa y cuantitativa de un corpus virtual de contratos de compraventa de viviendas (inglés-español)”, in Miguel Ángel García Peinado and Ignacio Ahumada Lara (eds.), *Traducción literaria y discursos traductológicos especializados*, Amsterdam, Peter Lang, pp. 309-330.
- Arce Romeral, Lorena and Miriam Seghiri (2018b), “Booth-friendly term extraction methodology based on parallel corpora for training medical interpreters”, *Current Trends in Translation Teaching and Learning E*, 5, pp. 1-46, available at http://www.cttl.org/uploads/5/2/4/3/52/43866/cttl_e_2018_1.pdf (accessed: 11/8/2020).
- Arce Romeral, Lorena and Miriam Seghiri (2019), “Diseño de plantillas de redacción y traducción al inglés (variedades británica y estadounidense) de contratos de compraventa de viviendas basadas en corpus”, in Miriam Seghiri (ed.), *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de*

traducción e interpretación y la enseñanza de segundas lenguas, Amsterdam, Peter Lang, pp. 69-106.

Arce Romeral, Lorena and Míriam Seghiri (2020a), “LEXCOR: un corpus legislativo de compraventa inmobiliaria como herramienta de traducción”, in Susana Álvarez Álvarez and María Teresa Ortego Antón (eds.), *Perfiles estratégicos de traductores e intérpretes en la transmisión de la información experta multilingüe en la sociedad del conocimiento del siglo XXI*, Granada, Comares, pp. 125-141.

Arce Romeral, Lorena and Míriam Seghiri (2020b), “Creación de una plantilla de traducción de contratos de compraventa de viviendas. Un estudio aplicado a Irlanda”, in *TSN. Transatlantic Studies Network: Revista de Estudios Internacionales*, 5, 9 (*Lazos entre Puerto Rico y España, Especial*), pp. 65-76. DOI: <https://doi.org/10.24310/TS.N.20.20.vi9.13000>.

Arce Romeral, Lorena and Míriam Seghiri (2021), “COMPRAVENTACOR: un sitio *web* para la traducción de contratos de compraventa de viviendas en España, Argentina, Reino Unido y Estados Unidos,” *Hikma: Revista de Traducción*, 20, 1, pp. 221-254, available at <https://www.uco.es/ucopress/ojs/index.php/hikma/article/view/13017> (accessed: 17/8/21). DOI: <https://doi.org/10.21071/hikma.v20i1.13017>.

Arce Romeral, Lorena and Míriam Seghiri (2022), “Creación de plantillas de redacción de contratos de compraventa de viviendas en español: variedades peninsular y argentina”, in Miguel Ángel García Peinado y Juan de Dios Torralbo (eds.), *Formas y variedades tradicionales de la traducción: literaria, jurídica y científica*. Amsterdam: Peter Lang, pp. 275-306.

Bowker, Lynne y Pearson, Jennifer (2002), *Working with Specialized Language: A Practical Guide to Using Corpora*, London. Routledge. DOI: <https://doi.org/10.4324/9780203469255>.

Corpas Pastor, Gloria (2008), *Investigar con corpus en traducción: los retos de un nuevo paradigma*, Frankfurt am Main, Peter Lang.

- EAGLES (1996b). *Text corpora Working Group reading Guide*. EAGLES Document EAG-TCWG-FR-2, Available at <http://www.ilc.cnr.it/EAGLES/corpintr/corpintr.html> (accessed: 2/8/2020).
- Laviosa, Sara (1998), “The Corpus-based Approach: A New Paradigm in Translation Studies”, in *Meta: Translator’s Journal*, 43, 4, pp. 74-479, available at https://www.researchgate.net/publication/271179_358_The_Corpus-based_Approach_A_New_Paradigm_in_Translation_Studies (accessed: 12/8/2020). DOI: <https://doi.org/10.7202/003424ar>.
- Luque Giráldez, Ángela and Míriam Seghiri (2019), “Extracción terminológica basada en corpus para la traducción de fichas técnicas de impresoras 3D”, in Gloria Corpas Pastor, Ruslan Mitkov, Maria Kunilovskaya and María Araceli Losey León (eds.), *Computational and Corpus-based Phraseology*, Geneva, Tradulex, pp. 99-106. DOI: https://doi.org/10.26615/978-2-9701095-6-3_013.
- Luque Giráldez, Ángela and Míriam Seghiri (2021), “Colocaciones y locuciones en las fichas técnicas de impresoras 3D (inglés-español): un estudio basado en el corpus 3DCOR”, in Gloria Corpas Pastor et al., *Fraseología basada en corpus*, Granada, Comares, pp. 231-251.
- Pérez Carrasco, Míriam and Míriam Seghiri (2022), “MOTOCOR Y MOTOLEX: Estudio comparativo de la terminología técnica empleada en un subcorpus de manuales de usuario y un subcorpus legislativo sobre automoción”, in Carmen Expósito Castro, Mar Ogea Pozo and Francisco Rodríguez (eds.), *Theory and Practice of Translation and Interpreting as a Vehicle for Knowledge Transfer / Théorie et pratique de la traduction et de l’interprétation comme véhicule de transfert des connaissances*, Seville, Universidad de Sevilla, pp. 139-172.
- Pérez Carrasco, Míriam and Míriam Seghiri (2021), “Estudio diacrónico de terminología técnica de manuales de usuario de automoción a través de recursos de la Real Academia Española: CORDE, CREA, CORPES XXI, NTTL y MDA”, in Matteo de Beni y Dunia Hourani, Berlín: De Gruyter (eds.), *Corpus y estudio diacrónico del discurso especializado en español*, Berlín, Peter Lang, pp. 236-292.

- Roberts, Roda P. (1990), “Translation and the Bilingual Dictionary”, in *Meta*, 35, 1, pp. 74-81. DOI: <https://doi.org/10.7202/003962ar>.
- Roberts, Roda P. (1992), “Translation pedagogy: Strategies for improving dictionary use”, in *Traduction, Terminologie et Rédaction*, 5, 1, pp. 49-76, available at https://www.research_gate.net/publication/265487615Translation_Pedagogy_Strategies_for_Improving_Dictionar_y_Use (accessed: 1/8/2020). DOI: <https://doi.org/10.7202/037106ar>.
- Sánchez Ramos, María del Mar (2019a), “Corpus paralelos y traducción especializada: ejemplificación de diseño, compilación y alineación de un corpus paralelo bilingüe (inglés-español) para la traducción jurídica”, *Lebende Sprachen*, 64, 2, pp. 269-285. DOI: <https://doi.org/10.1515/les-2019-0015>.
- Sánchez Ramos, María del Mar (2019b), “La incorporación de la traducción colaborativa en la didáctica de la traducción” in *Hikma, Estudios de traducción*, 18,1, pp. 261-281. DOI: <https://doi.org/10.21071/hikma.v18i1.11403>.
- Sánchez Ramos, María del Mar (2019c), “Rethinking professional translation roles: the localisation of mobile applications”, in *Sendebarr, Revista de Traducción e Interpretación*, 30, pp. 121-139.
- Sánchez Ramos, María del Mar (2020), *La documentación digital y léxico en la traducción e interpretación en los servicios públicos (TISP): fundamentos teóricos y prácticos*, Berlin, Peter Lang. DOI: <https://doi.org/10.3726/b16632>.
- Seghiri, Míriam (2006), *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*, doctoral thesis, University of Malaga, available at <https://riuma.uma.es/xmlui/bitstream/handle/10630/2715/16754888.pdf?isAllowed=y&sequence=1> (accessed: 15/8/2020).
- Seghiri, Míriam (2008), “Creating virtual corpora step by step”, in Purificación Sánchez Hernández, Pascual F. Pérez-Paredes, Pilar Aguado Jiménez, and Raquel Criado Sánchez (eds.), *Researching and*

Teaching Specialized Languages: New Contexts, New Challenges, Murcia, Editum, pp. 435-449.

- Seghiri, Míriam (2010), “Metodología de diseño y compilación de un corpus representativo de seguros turísticos”, in Rafael López-Campos Bodineau, María del Carmen Balbuena Torezano, and Manuela Álvarez Jurado (eds.), *Traducción y modernidad: textos científicos, jurídicos, económicos y audiovisuales*, Córdoba, Universidad de Córdoba, pp. 59-70.
- Seghiri, Míriam (2012a), “Creating electronic corpora: design, compilation protocol and representativeness”, in Eva Parra-Membrives, Miguel Ángel García Peinado, and Albrecht Classen (eds.), *Aspects of Literary Translation: Building Linguistic and Cultural Bridge in Past and Present*, Tübingen, Narr Verlag, pp. 379-382.
- Seghiri, Míriam (2012b), “El corpus comparable para la didáctica de la traducción judicial inversa”, in Susana Cruces Colado, Maribel del Pozo Triviño, Ana Luna Alonso, and Alberto Álvarez Lugiés (eds.), *Traducir en la frontera*, Granada, Atrio, pp. 815-860.
- Seghiri, Míriam (2014), “Too big or not too big: Establishing the minimum size for a legal *ad hoc* corpus”, *Hermes Journal*, 53, pp. 85-98, at <https://tidsskrift.dk/her/article/view/20981/18497> (doc. accessed: 4/8/2020). DOI: <https://doi.org/10.7146/hjlc.v27i53.20981>.
- Seghiri, Míriam (2015), “Determinación de la representatividad cuantitativa de un corpus *ad hoc* bilingüe (inglés-español) de manuales de instrucciones generales de lectores electrónicos (Establishing the quantitative representativeness of an E-Reader User’s Guide *ad hoc* corpus [English-Spanish])”, in María Teresa Sánchez Nieto (ed.), *Corpus-based Translation and Interpreting Studies: From Description to Application*, Berlin, Frank & Timme, pp. 125-146.
- Seghiri, Míriam (2017a), “Corpus e interpretación biosanitaria: extracción terminológica basada en bitextos del campo de la Neurología para la fase documental del intérprete [Corpora and medical interpreting:

terminology extraction based on bitexts for the interpreter's documentation process]”, *Panacea*, 18 (46), volumen monográfico de interpretación en el ámbito biosanitario, pp. 123-132, available at https://www.researchgate.net/publication/322551541_Corpus_e_inte_rpretacion_biosanitaria_extraccion_terminologica_basada_en_bitextos_del_campo_de_la_Neurologia_para_la_fase_documental_del_int_erprete (accessed: 20/8/2020).

Seghiri, Míriam (2017b), “Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español / español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores [Methodology for elaborating bilingual and bidirectional glossaries (English-Spanish/Spanish-English) based on corpora to translate TV User Manuals]”, *Babel*, 63, 1, pp. 43-64.

Seghiri, Míriam and Arce Romeral, Lorena (2021). *La traducción de contratos de compraventa inmobiliaria: un estudio basado en corpus aplicado a España e Irlanda*, Amsterdam, Peter Lang. DOI: <https://doi.org/10.3726/b17692>.

Sinclair, John (1991), *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.

Toledo Báez, María Cristina (2016), “La macroestructura de los contratos de compraventa inmobiliarios en Inglaterra / Gales y en España: estudio contrastivo previo a una traducción jurídica”, in Miguel Ángel García Peinado, José Manuel Oliver Frade, and Alfonso Corbacho Sánchez (eds.), *La traducción especializada. Vertientes y modalidades (aplicaciones teóricas y prácticas)*, Granada, Atrio, pp. 329-350.

Toledo Báez, María Cristina (2018), “El contrato de arrendamiento de viviendas en Francia y España: análisis y traducción”, in Miguel Ángel García Peinado, and Ignacio Ahumada Lara (eds.), *Traducción literaria y discursos traductológicos especializados / Literary Translation and Specialized Translation Discourses*, Frankfurt am Main, Peter Lang, pp. 250-265.

- Toledo Báez, María Cristina (2019), “Traducción y reformulación intralingüística: mejora de legibilidad de textos institucionales”, in Daniel Gallego Hernández (ed.), *Nuevos estudios sobre traducción para el ámbito institucional y comercial*, Frankfurt am Main, Peter Lang, pp. 183-206.
- Zanettin, Federico (1998), “Bilingual Corpora and the Training of Translators”, in *Meta*, 43, 4, pp. 616-63. DOI: <https://doi.org/10.7202/004638ar>.
- Zanettin, Federico (2002), “DIY Corpora: The WWW and the Translator” in Belinda Maia, Johann Haller, and Margherita Ulrych (eds.), *Training the Language Services Provider for the New Millennium*, Porto, Universidade do Porto, pp. 239-248.
- Zanettin, Federico; Bernardini, Silvia, and Stewart, Dominic (eds.) (2003), *Corpora in Translator Education*, Manchester, St. Jerome.