

Recibido / Received: 20/05/200
Aceptado / Accepted: 10/08/2020

Para enlazar con este artículo / To link to this article:
<http://dx.doi.org/10.6035/MonTI.2021.13.08>

Para citar este artículo / To cite this article:

Moreno-Pérez, Leticia & Belén López-Arroyo. (2021) "Atypical corpus-based tools to the rescue: How a writing generator can help translators adapt to the demands of the market." En: Calzada, María & Sara Laviosa (eds.) 2021. *Reflexión crítica en los estudios de traducción basados en corpus / CTS spring-cleaning: A critical reflection*. *MonTI* 13, pp. 251-279.

ATYPICAL CORPUS-BASED TOOLS TO THE RESCUE: HOW A WRITING GENERATOR CAN HELP TRANSLATORS ADAPT TO THE DEMANDS OF THE MARKET¹

LETICIA MORENO-PÉREZ

leticia.moreno@uva.es
Universidad de Valladolid - ACTRES

BELÉN LÓPEZ-ARROYO

mariabelen.lopez@uva.es
Universidad de Valladolid - ACTRES

Abstract

Corpus studies have become an undisputed aid for the evolution of translation, transferring knowledge from the academia to develop tools that have invaluable helped the profession. Nevertheless, the demands of the market require translators to improve their efficiency in order to adapt to its hectic pace. The aim of this paper is to present a possible solution through the use of corpus-based tools that are usually neglected in translation: writing aids. First, the reality of the translation market will be studied to understand the current context and translators' needs. Then, we will analyze some of the existing tools derived from corpus studies available for translators, both the most and less usual. Finally, we will focus on a booming sector of the market, that of

-
1. This study was carried out within the research project *Producción textual bilingüe semiautomática inglés-español con lenguajes controlados: parametrización del conocimiento experto para su desarrollo en aplicaciones web 2.0 y 3.0*, financially supported by Ministerio de Ciencia e Innovación since 2016 (FFI2016-75672-R).



Esta obra está bajo una licencia de Creative Commons Reconocimiento 4.0 Internacional.

oenology, to exemplify how one of the less typical tools, the writing generator, may be helpful for translators in terms of cost, time, and quality given the current demands.

Keywords: Corpus-based tools; Knowledge transfer; Translation market; Writing generator; Oenology.

Resumen

Los estudios de corpus han supuesto una inestimable ayuda para la evolución de la traducción, al generar herramientas que facilitan a los profesionales su labor mediante la transferencia de conocimiento desde la investigación. Pero el mercado exige un nuevo ritmo a los traductores, quienes necesitan mejorar su eficiencia. El objetivo de este estudio es presentar como posible solución unas herramientas basadas en corpus habitualmente desatendidas en traducción: los asistentes de escritura. Tras analizar la realidad del mercado para comprender el contexto actual y las necesidades de los traductores, expondremos algunas de las herramientas derivadas de los estudios de corpus de que estos disponen, tanto habituales como atípicas; por último, nos centraremos en un sector del mercado en auge, la enología, para ejemplificar cómo el generador de escritura puede resultar de ayuda para los traductores en aspectos como costes, tiempo y calidad, de acuerdo con las demandas actuales.

Palabras clave: Herramientas basadas en corpus; Transferencia de conocimiento; Mercado de traducción; Generador de escritura; Enología.

1. Introduction

Corpus-based studies became an undisputed turning point for the evolution of pure and applied translation studies (Laviosa 2002: 4) by incorporating, among others, quantitative methods of research to describe patterns of behavior in discourse. The combination of quantitative and generalizable data with qualitative insights into dimensions of discourse has provided a fruitful alliance for researchers (Marchi & Taylor 2018: 4). However, the global market we are in needs this knowledge to be transferred in the form of useful and usable tools and aids² (Rabadán Álvarez 2008: 105) to help translators improve their efficiency in their daily profession.

2. In this paper, *usefulness* refers to “the extent to which tools (technological, conceptual or otherwise) are relevant to the actual needs of a user” (Rabadán Álvarez 2008: 106; Landauer 1995: 4), whereas *usability* is “the extent to which a product can be used by

But, as it has been reported (Sinclair 2004, Rabadán Álvarez 2005-2008; 2008, among others), usually the academic research community does not think they have to supply solutions; it is the translator who has to derive them according to the researcher's conclusions. In this sense, Rabadán Álvarez already reported that commercially available tools and aids were not as popular or widely used as they might be expected, and the reason for this is that the user feels they are not useful since they do not supply solutions to problems (2008: 105). This situation still persists nowadays as, in the long run, translators end up developing their own research protocols and custom-made strategies to satisfy the demands of the translation market. Even though these demands are in constant change, this misunderstanding between the user needs and the researcher work should necessarily be solved in this era of globalization, technology and ICTs.

To change this situation, it is necessary to implement translation methodologies that increase value for money by supplying translators with tools that satisfy the needs of the market, helping them be more accurate and quicker; this would translate into lower rates clients are willing to pay for, without diminishing the translators' job and life quality. The aim of this paper is to explore paths to reach this result by applying corpus-based research products that already exist but are not originally designed for translators, and therefore are rarely used by them: writing aids. These neglected corpus-based tools ensure a high level of accuracy and could make their work more efficient in the terms mentioned above. To prove their usefulness, first we will analyze the reality of the translation market, to see which demands might be challenging and need to be improved. Then, we will review the typical corpus-based translation tools and we will contrast them with atypical corpus-based tools to help define how the latter could be more helpful for the current market demands. Finally, we will illustrate the pertinence and applicability of a specific atypical corpus-based tool, a writing generator, in a specific field of the market with a high demand of translation, the flourishing sector of oenology.

specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (Quesenbery 2001; also Kreitzberg & Little 2009).

2. The reality of the translation market with a focus on the business sector

Since the beginning of the 21st century, researchers have used different adjectives to describe the translation market, but all of them seem to agree on its changing nature, as translation is “highly dependent on external factors” (Sosoni & Rogers 2013: 7). Back in 2005, the research group Aula.int from the University of Granada, Spain, provided an overview of the translation market taking into account past trends and future prospects, and defined the following characteristics (Aula.int 2005: 133-135):

- Global, as the growth of international trade has resulted in an increased necessity of international communication, thus translation.
- Decentralized, as there are no geographical boundaries, increasing competition between translators who find it more complicated to focus their work on a certain language combination, field of expertise, or market.
- Specialized, not only in ‘conventional’ fields, such as legal, financial, medical, etc., but also in areas that emerged later, such as localization, emerging technologies, digital contents, etc.
- Dynamic, as new fields and translation aids emerge, so translators need to adapt and update their knowledge on a regular basis.
- Virtual, as relationships between translators and clients, agencies, experts or colleagues are developed mostly online.
- Demanding, as, on the one hand, deadlines are becoming shorter and the rates lower, resulting in translators being forced to accept a higher workload than desired. On the other hand, clients’ demands in terms of quality and methodology to be followed (particular terminology, specific software, etc.) are increasingly restricting translators’ freedom and working systems.

This group’s perspective has been confirmed later on by researchers such as Dunne (2012) or Moorkens (2017), who noted how these characteristics have made the translation market evolve into a project-driven industry, in which stakeholders associate for a specific job and do not usually build strong client-provider relationships. In these terms, translation is progressively seen, especially by clients, as a commodity that a number of providers can supply

based on “availability, productivity, cost, or perceived translation quality” (Moorkens 2017: 469), leading to an even higher competition within the market.

Nevertheless, competition among translators might not be the most concerning issue arising from this trend. Moorkens (2017: 467) highlights two important external threats: the popularization of machine translation and postediting, and the emergence of non-professional translation practices. The former trend has taken over a significant part of the market (Robinson 2020: 32), as it is usually cheaper and faster, to the point that some researchers believe that “the translator’s function can be expected to shift to linguistic postediting” (Pym 2013: 487) in the future. In some concerning cases, translation clients resort to online automatic translation and postediting to save time and money leaving the quality discussion aside, even when research has shown that decisions along this line can entail economic losses (Robinson 2020; Sosoni & Rogers 2013; Hennecke 2017). As regards non-professional translation, this trend is quite diverse in nature, but involves clients resorting to untrained mediators to perform translation tasks. It was confirmed as the most concerning perceived threat for the Spanish translation market stakeholders by Rico Pérez and García Aragón (2016: 35). Some may argue that the use of non-professional translation is limited to certain fields and contexts, as is the case of crowd-sourced translation, a common one when translating [non-official] social media or audiovisual contents (Katan 2016: 379). However, this reality becomes an important issue when it affects official communications, especially in sectors where success highly depends on translation, as is the case of business.

Translation in the business sector is still in continuous expansion given the dominant paradigm of globalization, which produces an increasing number of texts requiring linguistic mediation (Morón Martín & Medina Reguera 2016: 227). However, companies tend to resort to employees with knowledge of foreign languages to translate documentation from and into other languages, instead of specifically hiring professional translators (Varona 2002; Mayoral 2006; Aguayo 2013; Albuquerque & Costa 2018). The reason behind this practice is that companies see translation as a “‘non-core’ business” (Sosoni & Rogers 2013: 7), just “a means to an end –a mediation resource to meet goals in a business transaction between client and

supplier” (Albuquerque & Costa 2018: 150), so minimizing costs by resorting to employees seems only logical. But is it impossible to make quality and economy meet? If translators make their work more cost-efficient, the distance between the numbers of businesses and the numbers of translators may narrow, and companies may stop seeing translation as a burden. Therefore, translators need the tools to help them be more efficient, and this is where the results of academic research come into play.

3. Helping translators adapt to the market: typical and atypical corpus-based tools

As mentioned in the previous section, translation efficiency does not only have to do with time and money, but also with quality. Clients, specifically those in specialized sectors, require target texts that comply with the standards and expectations of the target community. This usually involves specific terminology and genres which are known and shared by the given sector, members of which expect to receive texts—translated or not—that reflect a certain language and content. For that purpose, translators

need to be completely sure that the unit or expression they are employing in a specific target text is the best option to translate the source term or expression, not only regarding meaning but also register, style, geographical variant, etc. (Durán-Muñoz & Corpas Pastor 2020: 164).

If there exists a methodology that is able to meet these requirements, that is Corpus Linguistics, as it is founded on the real use of language. In fact, it is a widespread methodology both in translation theory and practice: according to previous research, most translators build their own corpora (comparable and parallel³) as part of their documentation and translation process (Durán-Muñoz 2012: 164; Zanettin 2013: 27); they consider the use of this methodology very effective despite the amount of time spent. Its appropriateness is also supported by the fact that a relevant number of studies have applied this methodology to the research on linguistics and translation with relevant results. However, although the quantitative nature

3. In this paper we understand comparable corpora as those multilingual corpora written in their original language, while parallel corpora are those multilingual corpora that are made up of texts in a source language and their translations.

of Corpus Linguistics improves the analysis and the translation as a product, it has to be combined or integrated with the qualitative techniques of other methodologies which consider the context. The reliability of statistical analysis, necessary for generalization, is combined with the precision and richness of qualitative analysis. As Marchi & Taylor state (2018: 9), “mixing methods is a form of triangulation” because they allow not only the researcher but also the professional translator to look into the data from many different windows (Baker 2018; Malamatidou 2018, among others).

Taking into account the aforementioned necessities of translators and the demands of the market, Corpus Translation Studies and Discourse Studies have been a fruitful source of tools that have helped not only translators, but also scientific and professional writers in a foreign language. There is a set of resources derived from corpus-based research that have been typically used by translators to improve efficiency and/or acceptability, especially:

- Lexicographical resources: dictionaries are the classic source of “equivalents and linguistic information” (Durán-Muñoz & Corpas Pastor 2020: 164). Lately, many lexicographers rely on corpora in the process of dictionary creation (Tarp & Fuertes-Olivera 2016) as an aid “for completing the dictionary structures they need when making a real dictionary” (Fuertes-Olivera 2012: 51).
- Ontologies: these knowledge bases are useful resources of terminology, more specifically in specialized fields, and many times their construction is based on a computerized analysis of a collection of texts (Carrero & Gomez 2008; Bautista Zambrana 2019). Ontologies, such as Oncoterm or EcoLexicon, are examples of framed-based terminologies created on a model of semantic analysis, based on the creation of lexical templates derived from corpus and dictionary analysis.
- Translation memories: “a specific type of dynamic parallel corpora” that have become “a standard tool of the trade” (Zanettin 2013: 20); they are used or built by translators through the alignment of parallel texts to “easily observe the original and translated segment” (Durán-Muñoz & Corpas Pastor 2020: 164) as a source to assist the translation process.

- Machine translation: “systems which rely largely on corpus-based statistical machine translation techniques” (Zanettin 2013: 20); translators can be both users and victims, as mentioned above (see section 2).
- Integrated corpus annotating and management tools: they include different tools, from POS concordance searches to semantic relations, to analyze DIY or existing specialized and reference corpora, e.g. Lancsbox, Termostat or Lextutor, to name a few.
- Online corpora: large and reliable compilations of texts that are accessible and manageable online by users, and which provide translators with “a wide range of linguistic and pragmatic information” (Durán-Muñoz & Corpas Pastor 2020: 167). COCA, BNC or Iweb are some of the most popular on-line corpora used by universities and translators containing millions of words divided by genres.
- Web crawlers: “tools that employ the Internet (the Web) as a direct source of information to launch linguistic queries or compile corpora automatically” (Durán-Muñoz & Corpas Pastor 2020: 167) and Corpus Manager, as search engines or online concordancers. Crawlers such Web BootCat or Webcorp, among others, use the web as a source of automatic corpus compilation, offering as well concordance and KWIC tools to analyze them.

The amount of commonly used tools based on corpus research proves to be large and heterogeneous. However, despite the development of such a number of tools aimed at improving the documentation process, some researchers and professionals have noted a relevant problem: translators are “still forced to consult a plethora of resources during the translation task” (Durán-Muñoz & Corpas Pastor 2020: 163). This does not seem to match the necessities of the market explained in the previous section, as the more time is spent, the less efficient the job. But a possible solution for this problem might be close at hand.

There is a different set of corpus-based tools that are most commonly used by writers in foreign languages, but that are atypical for translators, and which could be extremely useful given the current professional context: writing aids. Some of the most frequently used aids are:

- Writing assistants: these tools provide users with recommendations, mostly regarding terminology (terms, collocations, etc.) or style (grammar, discursive elements, etc.), to help them improve texts written in a foreign language so that the writing sounds natural; these tools vary in terms of operation, as some provide recommendations through the analysis of what has been written so far, while others require the user to launch a specific query. The solutions provided are based on an internal source of the software, as a dictionary or a corpus, and users must choose the option they consider to be the most adequate from the ones displayed and detailed in the interface. Examples of these are the different writing tools of Termium or the Write Assistant launched by Ordbogen.
- Templates: they are models, more specifically skeletal frameworks for given genres or text types. A writing template helps writers organize material and also helps them develop the kinds of sentence, paragraph, and structure that good writers display. Templates do not simply give writers advice on how to write; they show exactly *how to do it*. But while they provide step-by-step guidance in writing a given text type, they take for granted that the template user is fluent in the language being written (López-Arroyo & Roberts 2015: 150). In other words, the template shows how to write a specialized text type but it does not show the user how to write the language in which the text type is written (King, 2007; Supatranont 2012). More specifically, an English writing template is intended for English speakers who are called upon to write a specialized text in a genre with which they are not familiar.
- Writing generators: they are writing applications that allow users to produce full texts in a foreign language following the rhetorical particularities, norms and conventions of a given genre; they guide the user through the format of the genre in question, suggesting full semantic units and phrases, rather than terms or individual elements. The units offered to the user are based on quantitative and qualitative corpus analysis of that specific genre, so the resulting text will not only be correct in grammar, structure and format, but also reflect the particularities of the genre in the language being

used. The ACTRES research group is one of the most prolific on this matter, and has developed a few, as a generator for cheese description or for scientific abstract writing.

In general terms, the added value of these corpus-based tools is that they consist on several tools in one: in writing assistants, the translator can improve and proofread while producing the target text; templates offer a guide to write a particular text type; like writing templates, generators provide step-by-step guidance in writing a given text type. But since generators are in principle intended for non-native speakers of the language of the text, they provide guidance not only in rhetorical structure (text sections) and stock phrases and sentences found in such a text type, but also in the overall vocabulary and structures required for a given text type (López-Arroyo & Roberts 2015: 151). The translator is directly producing a proofread text with the words, style and format that the target community expects. Furthermore, since these tools require users to have a high knowledge of the languages involved, translators are a perfect profile for their proper use.

Given their characteristics, these tools do seem to fit the needs of translators in terms of efficiency. Focusing on specialized texts, as those in the business sector mentioned above, writing generators would probably be the most useful. Specialized genres tend to be more restrictive in terms of terminology and format, so in very closed specific genres, a previously reviewed set of structures would accelerate the process without losing quality. This is something that other translation aids, as translation memories, can also help with. However, generators provide solutions for some of the drawbacks of these other tools: the dubious quality (Doherty 2016: 954) and accuracy (Bowker 2005: 19) of the translations that sometimes feed translation memories should not be a problem in generators, as they are guaranteed by a thorough process of compilation and analysis. Also, the fact that the text is considered as a whole unit in writing generators eliminates the lack of “syntagmatic cohesion” (Pym, 2011: 3) and the problematic consequences of working with split segments (Pym 2013: 496) that sometimes derive from the use of translation memories. Furthermore, writing generators guarantee the representativeness of the specialized genres involved, which are usually

the best paid, so translators using this tool would be earning more money in less time.

This is supported by the fact that a writing generator partially skips the process of previous documentation/specialization of the translator in the subject matter; the tool provides an *ad hoc* corpus for a specific translation task, that is, a compilation aimed at creating a given specialized genre in a specific language pair. When a translator uses this tool, the representativeness and adequacy of the texts have already been checked, and the compilation analyzed and organized by the expert linguists who built the generator. This saves translators the significant amount of time that entails creating their own corpus for the documentation process, since it involves a building phase (including text search, quality check, origin verification, classification of texts, format conversion...) and an analysis phase (including determination of search settings, retrieval of data, selection and checking of candidates, advanced search of unaligned information...). The use of these tools also allows for learning and specializing while producing acceptable texts, that is, while getting paid.

Writing generators can be profitable for translators not only at an efficiency level, it can also be profitable at a competitive level, as translators would be able to offer their services to new markets more quickly. As an example, let us focus on the food sector. Food industries many times start producing a specific set of products, and later on decide to diversify. For example, many wineries in Spain have started producing olive oil recently. In this case, both products share the same or similar text genres, namely tasting notes (Sanz Valdivieso & López-Arroyo 2020:27). What is more, this text genre is common to other food products, such as spirits or cheese. Although each product has its own specificities, they share the same function and target audience, as well as a common rhetorical structure.

By acquiring a set of food-related writing generators as translation aids to ensure the quality and consistency, an important market niche would open for the translator, who could offer his or her services to the same or different companies with similar needs within one sector; this could be especially relevant in industries where there is a lack of specialized language professionals, as is the case of wine or oil (see section 4.1). The same situation would apply to other sectors or fields, as engineering, law, etc.

To illustrate how writing generators can help translators adapt to the reality of the translation market, in the following section we will focus on the example of the wine sector.

4. An atypical corpus-based tool applied to the translation market: a writing generator in the oenology sector

4.1. The market of oenology

In the last decades, the market of oenology has become highly relevant in international trade, and more specifically in Spain, the country we have chosen to illustrate our proposal. This boom has increased the need for multilingual communication in the field, but there are not many language mediators who have become experts. In fact, previous research has shown that translation in Spanish international trade companies in the food and oenology sectors is mostly done by workers within the companies (Ibáñez Rodríguez et al. 2010; Medina Reguera & Álvarez García 2014). However, the importance of this sector in the international market calls for an accurate management of language mediation, as there are millions in profits at risk. A more in-depth analysis of the market of oenology will help visualize its potential, both at an international and a local level.

The 2019 Statistical Report on World Vitiviniculture (International Organisation of Vine and Wine 2019) reveals meaningful figures about the wine industry in 2018: 292m hl of wine were produced worldwide; the consumption of wine reached 246m hl; and the import-export of wine reached 108m hl, which translates into €31,000m. Spain was the country with the largest number of hectares under vine, and was the fourth major grape producer in the world, with 96 per cent of its grape production dedicated to wine. As a result, it was the third major wine producer after Italy and France. Despite being the eighth wine consumer in the world (preceded by the USA, France, Italy, Germany, China, the United Kingdom and the Russian Federation, in this order), Spain did not appear on the list of main importers. Nevertheless, it was the main wine exporter worldwide, with 21.1m hl. An interpretation of the previous data leads to two revealing conclusions. Firstly, the main producers mostly consume their own wine, while the rest of the most important consumers import it. Secondly, as Spain is not

one of the biggest consumers, exporting is the way to monetize its surplus production. This places Spain at a relevant position in the market of oenology at a global scale.

Analysing the market at a local level (Salvador Insúa 2016: 429-435), more than 80 per cent of wine industries in Spain are microenterprises with very limited resources and production capacity to access the international marketplace; this fact hampers the export process, even though companies in this sector heavily depend on foreign trade to survive. Another side of the wine market in Spain is that of tourism (López-Arroyo & Fernández Antolín 2011). In addition to wine production, the offer of many wineries ranges from wine tastings to meals at their restaurants, lodging at their hotels, and even wine-based beauty treatments at their spas. These vacation packages are not only addressed to tourists at a national level, but also worldwide.

From both the international and national contexts, it can be inferred that language mediation should be a key element in the market of oenology, and more specifically in Spain, given the importance of this country in the market at a global scale. However, the specific situation of the sector in the country could encourage a reduction of costs in the commonly non-core use of translation, which would explain the aforementioned trend of resorting to employees for language mediation. But, why not investing in translation when studies confirm that one out of ten companies have suffered order or project cancellations due to the lack of foreign language expertise (Hennecke 2017: 23)? These companies are putting their income at stake, confirming the need in this sector for efficient translation that is acceptable for the target community, affordable, quick and of high quality. At this point, it can be observed that the situation coincides with that of the translation market in general; therefore, it is necessary to find a way to make translation more cost-effective, thus attractive for companies who actually need it. A writing generator for the oenology sector could help both parts meet.

4.2. *The ACTRES oenology writing generator*

As introduced in section two, a generator is a tool that helps a non-native speaker produce a given type of specialized text in another language. As main general characteristics, we can highlight: (i) they are doubly language-bound,

since they are designed to be used by speakers of another language, and therefore have to take the users' native language into account; in the case described in this paper, for instance, the generator is designed to be used by speakers of Spanish as a first language who need to write in English, which means that the generators take the user's native language as a starting point; (ii) they are genre bound; and (iii) they offer semantic units and phrases based on quantitative and qualitative corpus analysis, which guarantees that the resulting text is both correct and acceptable to the target specialized community. These characteristics make generators a reliable and accurate tool for translators to produce specialized texts in a foreign language, the main need of exporting companies, in this case.

To focus on the sector of wine, we will specifically describe the writing generator developed by ACTRES (Contrastive Analysis and Translation English-Spanish in its Spanish acronym), a research group devoted to the design of writing applications and to the construction of Controlled Natural Languages for the international promotion of products and services, namely in the food sector. This writing generator has been developed by IT engineers, expert linguists and translators following a corpus-based methodology, ensuring the use of authentic, accurate language. In the next sections we will describe how these researchers have set the foundations of the tool using Corpus Linguistics, to later explain and exemplify how the generator works, taking into account the way it can help translators.

4.2.1. Methodological framework of the generator

The ACTRES oenology generator uses comparable corpora, since they allow to describe the differences in the structure of the genres in the two languages under study and the results of that initial contrastive study are used to feed the writing generators in the following stage. These are *ad hoc* domain-specific corpora (Corpas Pastor & Seghiri 2009: 78) in English and Spanish compiled using pragmatic text selection criteria: representativeness, to ensure a representative sample of the language of expert members of the discourse community; and availability, taking into account the ease to obtain the texts constituting the corpus.

The corpus includes wine tasting notes from specific websites, such as those of the Appellations of Origin in Spain and North America, which give direct and restricted access to the information written by winery oenologists. Only the wine tasting notes included in wine tasting technical sheets released by wineries were used, so as to ensure that the writer was an expert and that the audience being addressed also consisted of experts. This methodological adjustment allowed for more parallelism between the English and Spanish texts, hence for more accurate interlingual comparison of the wine tasting notes.

The final corpus includes 750 wine tasting notes in Spanish and 716 wine tasting notes in English, which amount to 54,545 and 55,339 words respectively. The resulting corpus is not big in size, but it meets the appropriate criteria to fulfill the purpose of the study, taking into account that a smaller corpus may be called for when rhetorical tagging is used, which can only be done semi-automatically (Flowerdew 2005: 329). That is the case of this corpus. First, the files were tagged to provide pragmatic information for the texts, such as the winery or where the notes come from and the date of publication. Then, texts were labeled to identify the different rhetorical moves. The labeling process has to be necessarily manual, since it implies a process of constant decision-making. At the same time, the corpora samples are labeled and managed using certain software that has been integrated in an application suite that was especially designed for the creation of the writing generator, as the ACTRES Browser and Tagger, which will be described in the next section.

4.2.2. The generator development process: corpus analysis

The ACTRES generator specifically includes three elements of analysis that follow a top-down methodology.

The first element included in the generator is the prototypical rhetorical structure showing the moves and steps to be included when writing the genre in English. To obtain it, texts are manually labelled using rhetorical labels that help setting up the semantic units (*moves* and *steps*, according to Swales 1990, 2004) that serve as common ground to describe the prototypical semantic units in each language (Bondarko 1984), their cross-linguistic

juxtaposition and their contrast in order to obtain the prototypical structure for the genre under study for the target discourse community. This qualitative analysis is complemented by a quantitative one, following Suter's criteria (1993), to distinguish the most recurrent moves from the secondary ones by the frequency of occurrence of each rhetorical move. The most frequently recurrent moves, which range between a frequency of 40% and 100%, are considered conventional (Biber et al. 2007: 24) or compulsory (Suter 1993: 119). This category includes Suter's compulsory *high-priority* and *medium-priority* moves and steps. The moves occurring the least frequently (<40%) are deemed to be of low priority and occasional and are called *optional*. In this case, only compulsory high-priority and medium-priority moves are incorporated; that is to say, only moves with more than 40% frequency are included. This part of the process is carried out with help of three of the components of the ACTRES application suite: the Filemanager, an online tool that allows researchers to manage, store and use their corpus; the Tagger builder, an online tool that allows researchers to develop a semantic tree with all the moves and steps identified in the genre under study; and the Tagger, an on-line tool designed to manually label the texts (i.e. files) once a move and/or step has been identified, as well as to manage and store the files.

Secondly, the generator includes the lexico-grammatical patterns most frequently used in each move and step so as to solve problems on how to string words together, not only correctly and acceptably, but also idiomatically. Once the moves and steps are identified, the top-down methodology identifies those lexico-grammatical resources typical in the genre under study. More specifically, it focuses on specialized phraseology or terminological word combinations, also called *phraseological units*, which occur frequently in the technical language of wine. These include collocations, irreversible binomials, idioms, routine formulae and combinatorial patterns (Roberts 1998; Andrades 2014). The underlying assumption is that, by identifying key technical and subtechnical terms in each move and step in both language corpora, we are able to detect some relevant phraseological units that structure the information and are valuable for describing patterns of behaviour in the grammar used by the target discourse community when writing this specific specialized genre. The description and analysis of these units lead to the description of recursive lexico-grammatical patterns in

each move and step. This part of the process is carried out with help of another component of the ACTRES application suite: the Browser, where researchers can analyse and contrast rhetorical structures and obtain the frequency of occurrence of moves and steps, thus making it easy to identify the prototypical structures.

The third element included in the generator is the terminological and phraseological glossary (with examples extracted from the corpus) that will provide not only terms, but also their most common collocations. For this step, researchers also resort to the Browser, since it displays a word list and functions as a monolingual and bilingual concordancer, so the researcher obtains phraseological and terminological information that can be described and compared to identify lexico-grammatical patterns of behaviour.

And how does the information extracted from these elements help build the generator? As an example, a qualitative and quantitative analysis carried out with the first element showed, among other findings, that the central moves (those occurring between 70% and 100% of the cases) in the two languages correspond to the different tasting phases; in other words, colour, nose and palate are the compulsory moves for the construction of a wine tasting note in English and in Spanish. The results of the lexico-grammatical analysis showed different recurrent structures with different levels of complexity that will be reflected in the generator. Finally, the terminological and phraseological searches allowed describing and defining subtechnical terms and collocations in the two languages, which are stored in a dictionary connected to the generator. These and all the relevant data obtained in the analysis phase are extracted, and the resulting structure, patterns and terminology (dictionary) are linked to the generator software following the same top-down structure.

4.2.3. Operation of the generator

Technically speaking, generators are computer-friendly applications that guide the user through the writing process. The wine writing generator of ACTRES is a web-based tool that leads the user through the common rhetorical structure of the genre of wine tasting notes:

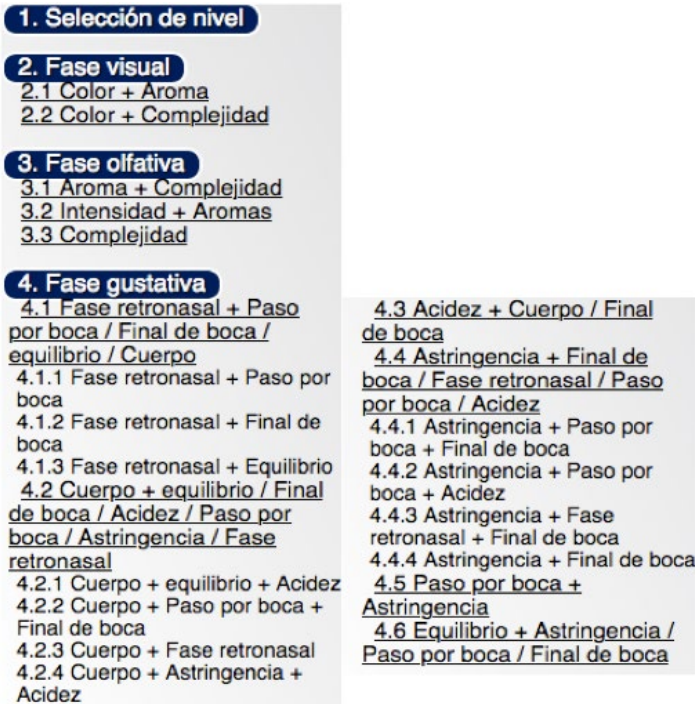


Figure 1. Generator interface: presentation of rhetorical structure.

First, the user has to choose between three levels of complexity, option 1 being the simplest, and option 3 being the most complex. Assuming that the potential users of the generator will have different levels of knowledge of the specialized language of wine tasting in English and assuming as well that, linguistically and pragmatically speaking, there are different options to express the same meaning in a given language, the generator will provide different structures for each level. Once chosen, the options offered to the user will correspond to that level of complexity throughout the process.



Figure 2. Generator interface: levels of difficulty.

Then, the generator presents moves and steps in Spanish to the user, who can then begin the writing phase. The user is offered different common structures in English for each step, so the translator can choose the steps and structures that appear in the source text. Since the structures in the generator are the prototypical ones according to the corpus, it is highly probable that the structure of the source text is among the ones suggested by the tool. Otherwise, the proposed structures are constituted by different parts that can be edited during the process.

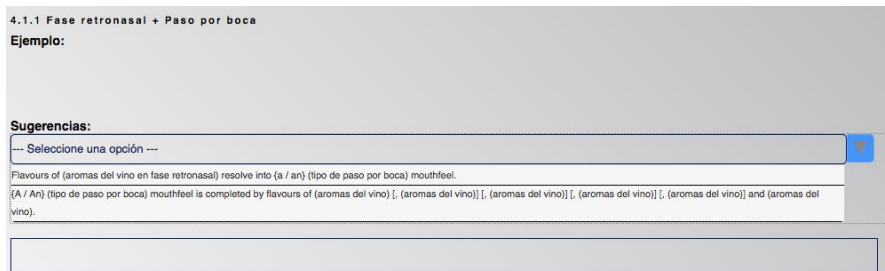


Figure 3. Generator interface: suggested structures within a step.

Each structure contains three different elements marked with different orthographic symbols:

{The wine / (nombre del vino)} is (nombre de color) in colour, {displaying / showing} aromas of (aromas del vino) [, (aromas del vino)] [, (aromas del vino)] [, (aromas del vino)] [, (aromas del vino)] and (aromas del vino).

First, there are fixed elements that form the basic skeleton of the structure and offer no changing options, although they can be altered if deemed necessary to comply with the customer's demands. Second, there are dependent elements that are required for reasons of grammar or use; they appear between keys, and the user is offered the most common options in the target language (English in this case) to choose the one that will fill that given gap by clicking on it. Finally, there are changing elements, parts in parentheses that suggest the kind of lexical information they could be filled with. When a list of elements of the same kind is usual, these changing elements are replicated one after the other and placed between square brackets, to highlight that they are optional. The suggestions of these changing elements are in the source language (Spanish) so the writer can insert the term(s) of the source text here, and the generator's dictionary will display the entries that include that term along with the English equivalents. The user can then choose the most adequate depending on the information in the source text.

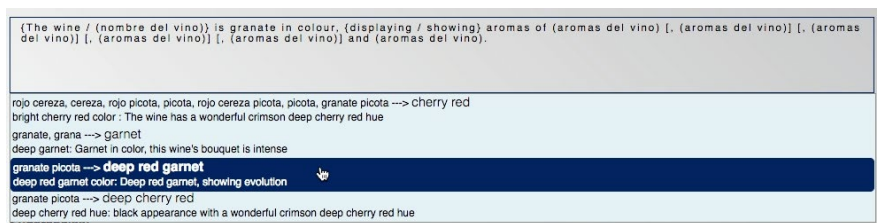


Figure 4. Generator interface: dictionary entries containing the term *granate*.

This process is repeated to fill all the moves and steps of the wine tasting note that are necessary with the lexical information desired. As a result of the thorough analysis carried out by expert linguists on the genre, the resulting text will contain the accepted structure, lexico-grammatical patterns and specific terminology of wine tasting notes in the target language and community, requiring little or no documentation/drafting on the translator's side. Furthermore, the user will only have to carry out a minimal proofreading

phase (e.g., in case there is something very specific that a given client wants to include outside the norm).

Additionally, the user can upload their own pictures (as the logo or the bottle), and other common details (website, email, address of the winery, etc.) to the generator, and download the final document in different formats (PDF, MS Word and xml). These final steps could not only save time for translators, but would also give them the chance to offer the client the full finished product, and not just the text; this would definitely add value to the translator's work in this highly competitive market.

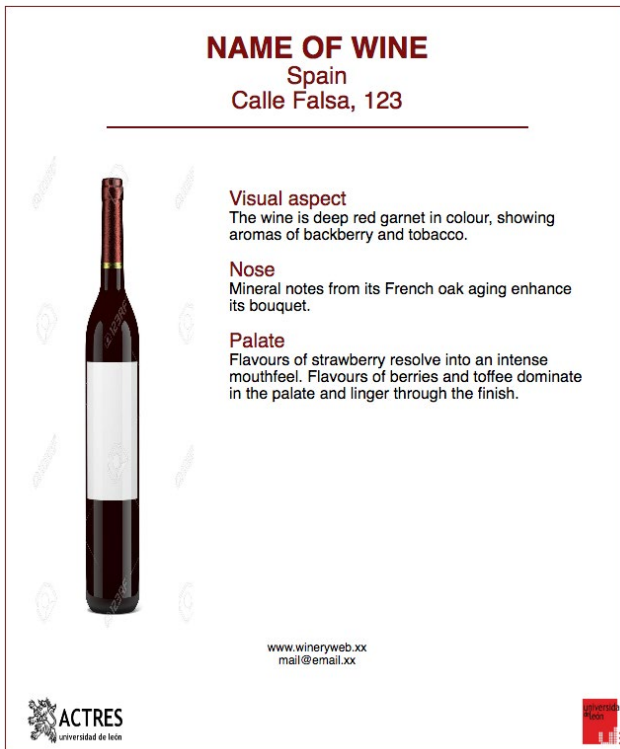


Figure 5. Preview of a finished wine tasting note.

5. Conclusions

The translation market has become a minefield for translators: the high competition makes it complicated for professionals to specialize in one field of expertise or market, so translators have to accept high workloads at low prices and despite the tight deadlines. All that competition is increased by the fact that clients, especially businesses, understand translation as a commodity that can be offered by any supplier that meets their availability, productivity, cost and perceived translation quality requirements (Moorkens 2017: 465).

In this context, Corpus Translation Studies can still help translators through the creation of tools that are adapted to the market, implying a true knowledge transfer between the profession and the academia, a process that is many times neglected in this and other areas of research. More specifically, writing aids (namely generators), are tools that can help increasing productivity, as they shorten the translation process and allow translators to handle a variety of genres from different fields of specialization; they improve costs, as specialized genres are usually better paid and using this tool translators can start producing acceptable texts from the very beginning; and they ensure quality, as they are tools built by expert linguists in the field with the information extracted from the analysis of a compilation of real texts. All this translates into high efficiency.

The generator described here is only an example, but there are more writing aids built by linguists and developed in the academia available, as the ones for the fields of tourism, medicine or public administration by the arText Project (cf. <http://sistema-artext.com>), or the other writing generators by the ACTRES group, some also related to the food industry (cheese, biscuits, herbal teas, and dried meats), some focused on other fields, such as tourist promotion, electronic products, company documentation, etc. (cf. <https://actres.unileon.es/wordpress/?lang=en>).

Although our study proposes that generators can be positive for translators to overcome some of the key problems in the translation industry, they also have a few possible drawbacks. Currently, the writing aids available are limited to certain genres and language pairs. However, if the use of generators became a trend among translators or companies, demand would

foster research in this area, so more genres and language pairs could be added to the existing sources. A second disadvantage is that users who want to use certain writing aid tools have to pay a license, as is the case of the tool described here. Although an initial investment might be seen as a burden, the aforementioned efficiency improvement in terms of time, costs and quality would help recover that investment in a short period of time. Finally, the fact that writing aids are based on corpora could help solving a common problem of specialized sources: updating “newly coined specialized units” (Durán-Muñoz & Corpas Pastor 2020: 163). If the demand existed, developers could regularly feed the generators’ dictionaries by including new texts in the corpora they are based on.

We hope to have opened a worthy path for translators, whose job is increasingly necessary and demanding, but decreasingly cost-effective. Even a slight improvement in efficiency can make the difference, since clients might stop seeing translation as a burden and start seeing it as value for money. Pedagogy about the importance of translation is still useful and necessary, but the reality calls for different approaches to make businesses and translators’ points of view meet.

References

- AGUAYO, Natividad. (2013) “El traductor-intérprete en el comercio exterior: ¿Realidad o necesidad?” *Entreculturas* 5, pp. 57-74.
- ALBUQUERQUE, Alexandra & Rute Costa. (2018) “The Satisfactory Cycle of Terminology Management in Translation-Mediated Business Communication: Problems and Opportunities.” In Gallego Hernández, Daniel & Éric Poirier (eds.) 2018. *Business and Institutional Translation: New Insights and Reflections*. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 149-164.
- ANDRADES, Arsenio. (2014) *Estudio contrastivo de unidades fraseológicas especializadas (UFE) en un corpus comparable bilingüe de contratos de derecho civil en lengua inglesa y española*. Madrid: UCM. Unpublished PhD.
- Aula.Int. (2005) “Translator Training and Modern Market Demands.” *Perspectives* 13:2, pp. 132-142.
- BAKER, Paul (2018) “Reflecting on Reflecting Research.” In Taylor, Charlotte & Anna Marchi (eds) 2018. *Corpus Approaches to Discourse: A Critical Review*. London: Routledge, pp. 281-292.

- BAUTISTA ZAMBRANA, M^a Rosario. (2019) *Terminología y ontologías: Un estudio alemán-inglés-español basado en corpus*. Granada: Comares.
- BIBER, Douglas; Ulla Connor & Thomas A. Upton (eds.). (2007) *Discourse on the Move. Using Corpus Analysis to Describe Discourse Structure*. Antwerp: John Benjamins.
- BONDARKO, Alexander. V. (1984) *Functional Grammar. A Field Approach*. Amsterdam: John Benjamins.
- BOWKER, Lynne. (2005) "Productivity vs Quality? A Pilot Study on the Impact of Translation Memory Systems." *Localisation Focus* 4 (1), pp. 13-20.
- CARRERO, Francisco; José Carlos Cortizo & José María Gómez. (2008) "Building a Spanish MMTx by Using Automatic Translation and Biomedical Ontologies." In: Fyfe, Colin; Dongsup Kim; Soo-Young Lee & Hujun Yin (eds.) 2008. *Intelligent Data Engineering and Automated Learning – IDEAL 2008*. Berlin/Heidelberg: Springer, pp. 346-353.
- CORPAS PASTOR, Gloria & Miriam Seghiri. (2009) "Virtual Corpora as Documentation Resources: Translating Travel Insurance Documents". In Beeby, Allison; Patricia Rodríguez & Pilar Sánchez-Gijón (eds.) 2008. *Corpus Use and Translating*. Antwerp: John Benjamins, pp. 75-107.
- DOHERTY, Stephen. (2016) "The Impact of Translation Technologies on the Process and Product of Translation." *International Journal of Communication* 10, pp. 947-969.
- DUNNE, Keiran J. (2012) "The Industrialization of Translation: Causes, Consequences and Challenges." *Translation Spaces* 1, pp. 143-168.
- DURÁN-MUÑOZ, Isabel & Gloria Corpas Pastor. (2020) "Corpus-Based Multilingual Lexicographic Resources for Translators: an Overview." In: Domínguez Vázquez, María José; Mónica Mirazo Balsa & Carlos Válcárcel Riveiro (eds.) 2020. *Studies on Multilingual Lexicography, (Lexicographica, Series Maior)*. Berlin: De Gruyter, pp. 159-178.
- DURÁN-MUÑOZ, Isabel. (2012) *La ontoterminografía aplicada a la traducción: Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores*. (Studien Sprachwissenschaft Und Interkulturel Kommunikation 80). Frankfurt am Main/New York: Peter Lang.
- FLOWERDEW, Lynne. (2005) "An Integrated Approach of Corpus-Based and Genre-Based Approaches to Text Analysis in EAP/ESP: Countering Criticism." *English for Specific Purposes* 24, pp. 321-332.

- FUERTES-OLIVERA, Pedro A. (2012) "Lexicography and the Internet as a (Re-) source." *Lexicographica* 28, pp. 49-70.
- HENNECKE, Angelika. (2017) "El entorno actual del mercado y la necesidad de traducción especializada en Alemania." *Cuadernos de Lingüística Hispánica* 30, pp. 19-41.
- IBÁÑEZ RODRÍGUEZ, Miguel; Jesús Bachiller Martínez & María Teresa Sánchez Nieto. (2010) "Comercio exterior y mediación lingüística en el sector vitivinícola de Castilla y León." *Hermeneus* 12, pp. 161-182.
- KATAN, David. (2016) "Translation at the Cross-Roads: Time for the Transcreational Turn?" *Perspectives. Studies in Translatology* 24(3), pp. 365-381.
- KING, Kevin B. (2007) *The Writing Template Book*. University of Michigan: USA.
- KREITZBERG, Charles B. & Ambrose Little. (2009) "Usability in Practice: Useful, Usable and Desirable. Usability as a Core Development Competence." *MSDN Magazine*. Available at: <<http://msdn.microsoft.com/en-us/magazine/dd727512.aspx>>
- LANDAUER, Thomas. K. (1995) *The Trouble with Computers: Usefulness, Usability, and Productivity*. Cambridge: The MIT Press.
- LAVIOSA, Sara. (2002) *Corpus-Based Translation Studies: Theory, Findings, Applications*. Amsterdam: Rodopi.
- LÓPEZ-ARROYO, Belén & Roda P. Roberts. (2015) "The Use of Comparable Corpora: How to Develop Writing Applications. In: Sánchez Nieto, Maria Teresa (ed.) 2015. *Corpus Based Translation and Interpreting Studies: From Description to Application*. Berlin: Frank & Timme, pp 147-156.
- LÓPEZ-ARROYO, Belén & Martín Fernández Antolín. (2011). "Estudios basados en corpus y lexicografía bilingüe: aplicaciones en un diccionario de fichas de cata". In BAZZOCCHI, Gloria, Pilar Capanaga & Sara Piccioni. (eds.) 2011. *Turismo ed enogastronomiatra Italia e Spagna. Linguaggi e territorio da esplorare*. Milán: Franco Angeli, pp. 99-116.
- MALAMATIDOU, Sofia. (2018) *Corpus Triangulation. Combining Data and Methods in Corpus Based Translation Studies*. London: Routledge
- MARCHI, Anna & Charlotte Taylor. (2018) "Partiality and Reflexivity". In TAYLOR, Charlotte & Anna Marchi. (eds.) 2018. *Corpus Approaches to Discourse: A Critical Review*. London: Routledge, pp. 1-16.

- MAYORAL, Roberto. (2006) "La traducción comercial." *Butlletí de la Associació de Traductors i Intèrprets Jurats* (diciembre), s.p. Available at: <https://www.ugr.es/~rasensio/docs/Traduccion_comercial.pdf>
- MEDINA REGUERA, Ana & Carmen Álvarez García. (2014) "La relación empresa-traducción en el sector agroalimentario andaluz." *Skopos* 4, pp. 187-206.
- MOORKENS, Joss. (2017) "Under Pressure: Translation in Times of Austerity." *Perspectives* 25(3), pp. 464-477.
- MORÓN MARTÍN, Marian & Ana Medina Reguera. (2016) "La competencia del traductor que no 'traduce': el traductor en ámbitos de internacionalización empresarial." *MonTI* 8, pp. 225-255.
- PYM, Anthony. (2011) "What Technology Does to Translating." *Translation & Interpreting* 3 (1), pp. 1-9.
- PYM, Anthony. (2013) "Translation Skill-Sets in a Machine-Translation Age." *Meta* 3, pp. 487-503.
- QUESENBERY, Whitney. (2001) "What Does Usability Mean: Looking beyond 'Ease of Use'." Available at: <<http://www.digitalspaceart.com/projects/cogweb2002v2/papers/whitney/whitney1.html>>
- RABADÁN ÁLVAREZ, Rosa. (2008) "Refining the Idea of 'Applied Extensions'." In Pym, Anthony; Miriam Schlesinger & Daniel Simeoni (eds.) 2008. *Beyond Descriptive Translation Studies: Investigations in Homage to Gideon Toury*. Antwerp: John Benjamins, pp. 103-118.
- RABADÁN ÁLVAREZ, Rosa. (2005-2008) "Tools for English-Spanish Cross Linguistic Applied Research". *Journal of English Studies* 5-6: 309-324.
- RICO PÉREZ, Celia & Álvaro García Aragón. (2016) *Análisis del sector de la traducción en España (2014-2015)*. Villaviciosa de Odón: Universidad Europea. Available at: <<http://abacus.universidadeuropea.es/bitstream/handle/11268/5057/analisis%20sector%20traduccion%2014-15.pdf?sequence=2&isAllowed=y>>
- ROBERTS, Roda P. (1998) "Phraseology and Translation." In: Fernández Nistal, Purificación and José María Bravo Gonzalo (eds.) 1998. *La traducción: orientaciones lingüísticas y culturales*. Valladolid: Universidad de Valladolid, pp. 61-77.
- ROBINSON, Douglas. (2020) *Becoming a Translator: An Introduction to the Theory and Practice of Translation*. London/New York: Routledge.
- SALVADOR INSÚA, José Antonio. (2016) *Mercado internacional del vino: intentos de modelización y estrategias territoriales de comercialización en España*. Valladolid: Universidad de Valladolid. Unpublished PhD.

- SANZ VALDIVIESO, Lucía & Belén López-Arroyo (2020). "On Describing Olive Oil Tasting Notes in English." *Fachsprache. Journal of Professional and Scientific Communication* 42.1-2, pp. 27-45.
- Several Authors (International Organisation of Vine and Wine, OIV). (2019) *Statistical Report on World Vitiviniculture*. Available at: <<http://www.oiv.int/public/medias/6782/oiv-2019-statistical-report-on-world-vitiviniculture.pdf>>
- SINCLAIR, John. (2004) *Trust the Text. Language, Corpus and Discourse*. London: Routledge.
- SOSONI, Vilemini & Margaret Rogers. (2013) Translation in an Age of Austerity: from Riches to Pauper, or not? *mTm* 5, pp. 5-17.
- SUPATRANONT, Pisamai. (2012) "Developing a Writing Template of Research Article Abstracts: A Corpus-Based Method." *Procedia Social and Behavioural Sciences* 66, pp. 144-156.
- SUTER, Hans-Jürg. (1993) *The Wedding Report. A Prototypical Approach to the Study of Traditional Text Types*. Antwerp: John Benjamins.
- SWALES, John. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- SWALES, John. (2004). *Research Genres*. Cambridge: Cambridge University Press.
- TARP, Sven & Pedro Fuertes-Olivera. (2016) "Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-Uva." *Lexikos* 26, pp. 273-296.
- VARONA, Lucema. (2002) "El traductor ante la micro y pequeña empresa pyme." In: Alcina Caudet, Amparo & Silvia Gamero Pérez (eds.) 2002. *La traducción científico-técnica y la terminología en la sociedad de la información*. Castelló de la Plana: Publicacions de la Universitat Jaume I, pp. 201-206.
- ZANETTIN, Federico. (2013) "Corpus Methods for Descriptive Translation Studies." *Procedia - Social and Behavioral Sciences* 95, pp. 20-32.

BIONOTES / BIONOTAS

LETICIA MORENO-PÉREZ holds a PhD in Professional and Institutional Translation from Universidad de Valladolid (Spain), where she has lectured topics such as English for Specific Purposes or specialized translation since 2010. Her research focuses on specialized language and its translation from

a functional approach through corpus-based studies, specifically in the field of legal and business translation. She has published and presented at international conferences on this area, and she is a member of the international research group ACTRES (Contrastive Analysis and Translation English-Spanish in its Spanish acronym), also specialized in the field. She has combined her academic work with professional translation and language training in the business sector for 10 years.

BELÉN LÓPEZ-ARROYO is an Associate Professor in ESP at the University of Valladolid (Spain). She taught legal and business translation in the School of Translation and Interpreting from 1997 to June 2013 and she currently teaches legal and business translation and Corpus Linguistics in the English Studies Degree. Her research interests include Discourse Analysis, Genre Analysis, Lexicography and Terminology Contrastive analysis and Translation. She is author of several articles and books related to contrastive analysis of scientific and professional genres and its implication for translation. In the ACTRES team she is in charge of the Rhetoric and Phraseology of Expert-to-Expert Discourse (in different areas) and its applications for developing writing aids in English for Spaniards.

LETICIA MORENO-PÉREZ es Doctora en Traducción Profesional e Institucional por la Universidad de Valladolid (España), donde ha impartido docencia en áreas como Inglés para Fines Específicos o traducción especializada desde 2010. Especializada en el campo jurídico y de los negocios, su investigación se centra en el lenguaje especializado y su traducción desde una perspectiva funcional a través de los estudios basados en corpus. Ha publicado y participado como ponente en congresos internacionales sobre este campo, y es miembro del grupo ACTRES (Análisis Contrastivo y Traducción Inglés-Español), especializado en este área. Además de su trabajo en el ámbito académico ha ejercido como traductora y formadora de lengua inglesa en el ámbito empresarial durante 10 años.

BELÉN LÓPEZ-ARROYO es Profesora Titular de Universidad en el área de Filología Inglesa en la Universidad de Valladolid (España). Impartió docencia en la Facultad de Traducción e Interpretación (Campus de Soria) desde 1997 hasta el 2013. Actualmente, imparte Traducción especializada I: derecho y economía y Lingüística del Corpus en el Grado en Estudios Ingleses

en la Facultad de Filosofía y Letras. Su investigación incluye análisis del discurso, análisis textual, terminología y fraseología contrastiva y traducción. Es autora de diversos artículos y libros en el campo del análisis contrastivo de géneros científicos y profesionales. Miembro del grupo ACTRES desde su creación, donde se encarga de la retórica y fraseología contrastiva en el discurso experto-experto en diferentes áreas y sus aplicaciones para el desarrollo de herramientas de ayuda a la escritura en inglés de géneros especializados para hispano parlantes.