

María Teresa Ortego Antón

# The design of TorreznoTRAD: The semiautomatic Spanish-English writing and translation aid tool

**Abstract:** The internationalization of the agri-food sector in Spain has led to an exponential increase in writing and translation services from Spanish to English. In this socio-economic context, the methodology used to build TorreznoTRAD is described. TorreznoTRAD is a corpus-based tool that assists in the writing and translation from Spanish to English of *torrezno* product cards. TorreznoTRAD relies in the results of exploiting a virtual comparable (Spanish-English) corpus (C-MARMEAT). This corpus was compiled, annotated, and exploited following the protocol proposed by Seghiri (2017) and Ortego Antón (2019). Once C-MARMEAT was compiled, the Spanish subcorpus as well as the English subcorpus were rhetorically tagged with the help of OpenTagger (Sanjurjo-González & Andaluz-Pinedo, 2021) to determine the rhetorical structure composed of moves and steps (Biber et al., 2007, p. 23-24). Next, model lines, defined as typical sentences where the content and format are standard (Pérez Blanco & Izquierdo, 2021), were established and enriched with a bilingual Spanish-English terminological database with different categories such as additives, allergens, nutritional elements, packaging, ingredients, materials, origin, etc. Finally, the semiautomatic writing aid tool was developed to guide Spanish speakers through the writing process in English. Thus, it will automatically generate *torrezno* product cards to satisfy the needs of translators and writers who draft specialised texts about this product for the agri-food sector.

---

**Acknowledgements:** This work has been carried out in the framework of the national R&D project entitled “Controlled natural languages, collaborative communication, and bilingual textual production in 3.0 environments” (PID2020-114064RB-I00), coordinated by Dr. Noelia Ramón García (University of Leon), and partially within the following projects: “TorreznoTRAD” (PROYEMER-2021-028), coordinated by Dr. María Teresa Ortego Antón and funded by the Emerging Projects Grants (Call 2021) of the University of Valladolid, “Multi-lingual and Multi-domain Adaptation for the Optimisation of the VIP system” (VIP II) (Ref. PID2020-112818GB-I00, Spanish Ministry of Science and Innovation), led by Prof. Gloria Corpas Pastor (University of Malaga) and “App para entrenar en posesición de traducción automática neuronal mediante la gamificación en entornos profesionales” (GAMETRAPP) (TED2021-129789B-I00, Spanish Ministry of Science and Innovation), led by Prof. Cristina Toledo Baez (University of Malaga).

---

**María Teresa Ortego Antón**, CITTAC, University of Valladolid, e-mail: mariateresa.ortego@uva.es

**Keywords:** Translation, comparable corpora, semi-automatic tool, English, Spanish

## 1 Translation and the agri-food sector

One of the pillars of the Spanish economy is the agri-food industry and, more specifically, the meat industry, which has become the most relevant one in terms of turnover and direct employment (MAPA, 2022, p. 4–5). In addition to this, the main economic engines of rural regions located in Southern Sparsely Populated Areas (SSPA) are meat companies, which have a family structure, a small size, and a farming tradition. In order to market their products abroad, they require writing and translation services from Spanish into English.

Although one may think that neural machine translation systems could be the solution to meet the requirements of small and medium sized companies, genres are characterized by different patterns depending on the target culture. Hence, the promotion of a given product needs to be sensitive to cross-cultural differences to guarantee that target texts satisfy the standards and expectations of the target community, “not only regarding the meaning but also register, style, geographical variant, etc.” (Durán Muñoz & Corpas Pastor, 2020, p. 164). One of the best tools to shed light on such similarities and differences are corpora, which have prominent roles in translation and contrastive studies.

Aware that agri-food translation is a field of knowledge still little explored in comparison with other domains, the ACTRES research group<sup>1</sup> is developing several corpus-based Spanish/English contrastive studies. This group analyzes pragmatic annotation in the agri-food corpus CLANES (Rabadán et al., 2021a), the promotional texts on the food industry (Rabadán et al., 2021b), promotional texts on cheese (Labrador and Ramón, 2020), texts on wine (Moreno Pérez & López Arroyo, 2021), olive oil (Sanz Valdivieso & López Arroyo, 2022), herbal teas (Pérez Blanco and Izquierdo, 2020, 2022), or dried meat product cards (Ortego Antón, 2019, 2021; Fernández Nistal, 2020; Ortego Antón & Fernández Nistal, 2020). Results enable the development and implementation of applications that can be integrated into linguistic tools based on natural language processing that will assist agri-food companies to overcome language barriers.

---

<sup>1</sup> ACTRES (Contrastive Analysis and Translation English-Spanish) is an interdisciplinary research group led by Prof. Rosa Rabadán and Prof. Noelia Ramón (University of Leon, Spain): <https://actres.unileon.es/wp/> (30 November 2022).

## 2 The development of semiautomatic writing aid tools

Translation efficiency does not only have to do with time and money, but also with quality. Although it could be thought that translation could be threatened by two factors: the popularization of machine translation and postediting, and the emergence of non-professional translation practices (Moorkens, 2017, p. 467), translators should be aware that successful interlinguistic communication depends not only on the use of English as a *lingua franca*, but also on the following factors:

Successful communication will depend not only on the accurate transmission of relevant subject-specific information within the professional domain, but also on compliance with cultural conventions, both at the big and small cultural levels. To this end, acceptable usage language, plus an awareness of genre conventions, are paramount. (Pérez Blanco & Izquierdo, 2021, p. 148).

Hence, comparable corpora are paramount in Translation Studies because they guarantee that target texts satisfy the standards and expectations of the target community (Bowker, 2002; Laviosa, 2002; Beeby et al., 2009; Corpas Pastor & Seghiri, 2017; Sánchez Ramos, 2019; Seghiri & Arce Romeral, 2021; Ortego Antón, 2022, among others). Consequently, comparable corpora-based tools have multiple advantages for users, as they allow them

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 (Moreno Pérez & López-Arroyo, 2021, p. 259–260).

Although writing aid tools are intended for non-native speakers, they provide guidance not only in rhetorical structure and stock phrases and sentences, but also in the overall vocabulary and structures required for a given text (López Arroyo & Roberts, 2015, p. 151). Thus, translators directly produce a proofread text with the words, style, and format that the target community expects, so they fit the needs of the translator in terms of efficiency and provide solutions that machine translation cannot solve in its own.

Thus, in this paper the methodology used to build TorreznoTRAD is described. TorreznoTRAD is a corpus-based writing aid that assists translators and technical writers, among others, in transferring *torrezno* product cards from Spanish into English. It relies on the results of analyzing a comparable (Spanish-English) virtual

corpus (C-MARMEAT) in order to identify typical patterns in Spanish and English related to this genre, so technical writers and translators can directly produce a proofread text with the terminology, the register, and the macrostructure that the target community expects.

### 3 C-MARMEAT compilation and exploitation

Following the protocol first used by Seghiri (2017) and applied in several studies (Ortego Antón, 2019, 2020, 2023; Fernández Nistal, 2020; Seghiri & Arce Romeral, 2021, or Sánchez Carnicer, 2022, among others), C-MARMEAT is a virtual unidirectional Spanish-English comparable corpus compiled following four steps:

- 1) Searching for texts on the websites of renowned companies from different countries which sell pork belly, *torreznos*, and marinated meats. The collection of texts published on the Internet ensures authenticity and, at the same time, has allowed us to select texts belonging to a wide variety of authors.
- 2) Downloading the texts manually in HTML or XML format.
- 3) Formatting texts to TXT UTF8 so that they could be processed by corpus management software.
- 4) Storing texts in the C-MARMEAT folder, which was divided into three subfolders, HTML for XML files, and two more subfolders for TXT texts in English (EN) or in Spanish (ES) (see Figure 1).

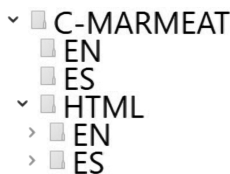


Figure 1: A sample of the structure.

In addition, the files have been named using an ID (e.g., 001MMwsMS211013ES.txt) composed of the number of the file (001, 002, etc.), the abbreviation for marinated meats (MM), the indication that the texts have been extracted from the web (ws), the abbreviation of the company they come from (e.g., MS for Moreno Sáez), the date they were downloaded (aamddd) and the language (EN or ES).

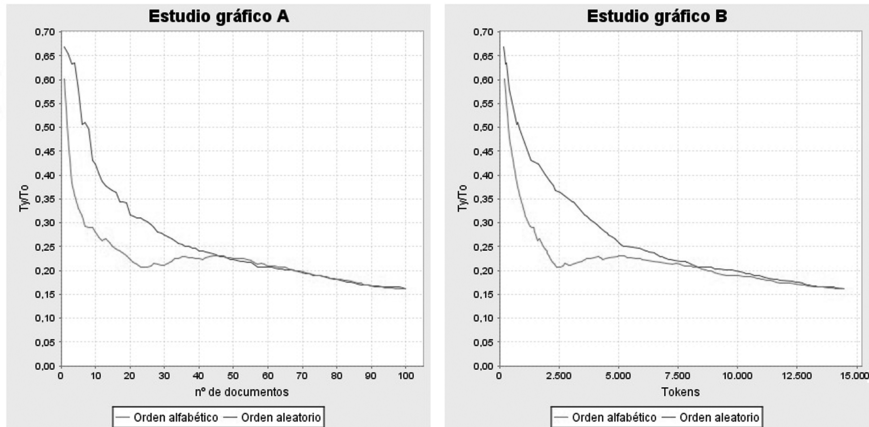
As a result, we have a virtual comparable (Spanish-English) corpus composed of 100 texts about marinated meat product cards originally written in Spanish (14042 tokens) and 100 original texts written in English (26391 tokens). C-MARMEAT is qualitatively representative because of the previously detailed procedure used to compile it. The difference in the number of words in each language in C-MARMEAT

is due to the fact that, in English, product features, packaging, recycling, and preparation and use are detailed, whereas in Spanish this information is much more synthetic. This phenomenon was also found in previous contrastive studies (Ortego Antón, 2019, 2021).

To conclude the process, the quantitative representativeness has been checked with the ReCor software (Seghiri, 2006; Corpas Pastor & Seghiri, 2010), which calculates the minimum number of words that the corpus must include to be representative in terms of the basic terminology in this genre.<sup>2</sup> ReCor provides two charts: A and B. In chart A, the horizontal axis represents the number of documents whereas the vertical axis shows the quotient of types divided by the number of tokens. In addition, there are two functions, the red one for the files listed alphabetically and the blue one for the files listed randomly. When both functions are steady, the corpus achieves the quantitative representativeness. Simultaneously, a second chart (B) is generated, which shows the minimum number of tokens needed.

Regarding the Spanish subcorpus of C-MARMEAT, quantitative representativeness is achieved with 70 documents and 13000 tokens, as shown in Figure 2.

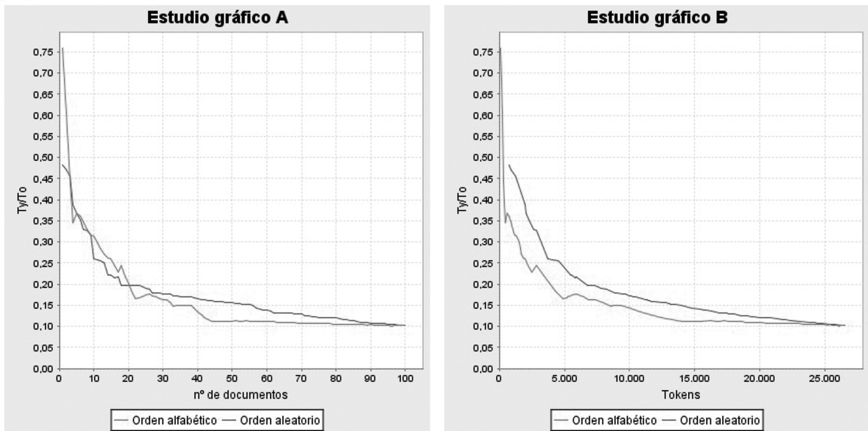
Representatividad del Corpus



**Figure 2:** Quantitative representativeness of the Spanish subcorpus of C-MARMEAT calculated with ReCor.

<sup>2</sup> More information about ReCor can be found at <https://rsoftuma.uma.es/es/software/recor/> (30 November 2022).

Representatividad del Corpus



**Figure 3:** Quantitative representativeness of the English subcorpus of C-MARMEAT calculated with ReCor.

Besides this, in the English subcorpus of C-MARMEAT, more documents are needed to be lexically representative (80) and more words, 25000 tokens, as observed in Figure 3.

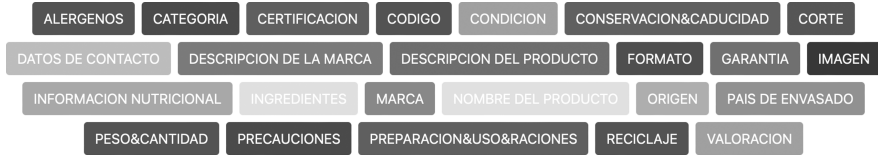
Data gathered from C-MARMEAT was examined in a multilevel analysis to develop TorreznoTRAD. Their analysis enabled us to establish the prototypical rhetorical structure, the model lines, and the bilingual terminological database, in line of previous ACTRES-led research results (López Arroyo & Roberts, 2015; Labrador & Ramón, 2020; Ortego Anton, 2019, 2020, in press; Pérez Blanco & Izquierdo, 2020, 2022, among others).

## 4 The rhetorical structure

To establish the prototypical rhetorical structure – defined as “the hierarchical organization of a text [involving] the various sections and subsections of a text, moves and steps” (López Arroyo & Roberts, 2015, p. 155) – the methodology proposed by Biber et al. (2007) was followed. These authors consider that textual genres are characterized by a series of rhetorical components called “move”, which is defined as “a discursal or rhetorical unit that performs a coherent communicative function” (Biber et al., 2007, p. 23). In turn, these moves can be divided into several steps, whose function is “to achieve the purpose of the move to which it belongs” (Biber et al., 2007, p. 24). The analysis proposed by these authors makes

it possible to identify the linguistic characteristics of the moves; it provides a description of the typical structural and distributional characteristics of each move; it offers data on its relative position in relation to other moves; and it makes it possible to develop a certain textual genre.

Hence, we established a set of tags associated with the moves (see Figure 4).



**Figure 4:** Set of tags used to rhetorically tag C-MARMEAT.

Next, C-MARMEAT texts in both languages were rhetorically tagged with the aid of OpenTagger (Sanjurjo-González & Andaluz-Pinedo, 2021).<sup>3</sup> OpenTagger is an open-source linguistic tagger written in the JavaScript programming language which allows users to include any type of information to the different paragraphs, sentences, or words that compose a text.

Once both subcorpora of C-MARMEAT were tagged, the percentages of moves in each of the annotated subcorpora (Spanish and English) were checked and compared, as well as the occurrence of such moves. The data resulting from this comparison allowed us to develop a rhetorical structure prototype in each of the working languages as shown in Table 1. The frequency of occurrence is represented with stars: from the compulsory nature shown by five stars (\*\*\*\*\*) to the little frequency labelled with one star (\*).

**Table 1:** Prototypical rhetorical structure of marinated meat cards.

SPANISH	FREQ.	ENGLISH	FREQ.
1. <i>Nombre del producto</i> (*****)	100%	1. Image (*****)	95%
2. <i>Imagen</i> (*****)	98%	2. Name of the product (*****)	93%
3. <i>Descripción del producto</i> (*****)	70%	3. Category (***)	39%
4. <i>Categoría</i> (***)	47%	4. Code (*)	12%
5. <i>Conservación, preparación y precauciones</i> (***)	43%	5. Format (*)	15%
6. <i>Peso y cantidad</i> (****)	72%	6. Preparation, use & servings (****)	76%
7. <i>Formato</i> (***)	41%	7. Origin (**)	36%

<sup>3</sup> Opentagger: <https://actres.unileon.es/opentagger/> (21 December 2022).

Table 1 (continued)

SPANISH	FREQ.	ENGLISH	FREQ.
8. <i>Ingredientes</i> (***)	58%	8. Warning	12%
9. <i>Alérgenos</i> (***)	47%	9. Use by (***)	41%
10 <i>Preparación, uso y raciones</i> (**)	41%	10. Product description (*****)	81%
11. <i>Código</i> (**)	27%	11. Weight (****)	65%
12. <i>Valoración</i> (*)	12%	12. Nutritional values (****)	66%
13. <i>Información nutricional</i> (**)	37%	13. Allergens (**)	32%
		14. Ingredients (****)	73%
		15. Recycling (**)	27%
		16. Review (*)	17%
		17. Brand description (*)	11%
		18. Contact (**)	35%

Having established the prototypical rhetorical structure in Spanish and in English, the most frequently used lexicogrammatical patterns in each move were semi-automatically detected to solve problems in how to string words together, not only correctly and acceptably, but also idiomatically.

## 5 The model lines

Model lines can be defined as “typical sentences and parts of sentences found in a given text type where the content and format are fairly standard” (López Arroyo & Roberts, 2015, p. 157). They are found in both writing templates and translation-based writing applications.

In C-MARMEAT, when establishing the prototypical sentences, we agree with Pérez Blanco and Izquierdo (2021, p. 157) that phraseological and lexicogrammatical patterns, even though formally diverging, are functionally equivalent. On the basis of this similarity, with the aid of AntConc 3.5.8. (Anthony, 2019), which is a freeware corpus analysis toolkit for concordancing and text analysis, all instances of each move were examined in the English subcorpus of C-MARMEAT. To find the instances, we searched for the tags, i.e. “</INGREDIENTES>”, and the content tagged was shown (see Figure 5).

Each occurrence was examined and the most frequent model lines for the moves were identified. This examination was manually handled, paying attention to the phraseology whose content matched the communicative function of the chunk where it was found. Results show three different patterns following the methodology used in Ortego Antón (2023).



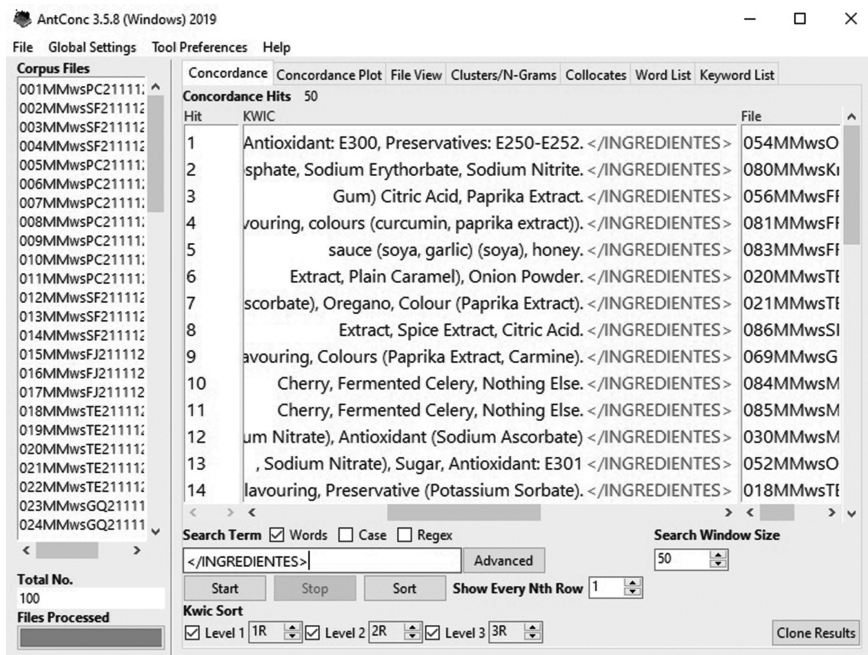


Figure 5: Example of search with AntConc 3.5.8. (Anthony, 2019).

- compulsory patterns, which are represented with curly brackets, making it necessary to insert a word or group of words from a selection list;
- optional patterns, which are emphasized with brackets to show that the information included in them is optional and can be omitted;
- selection between two options delimited with braces and separated by a slash.

For instance, the move “Ingredients” has three different model lines which are shown accompanied by some examples to help translators and technical writers transfer the content into English:

- [{Ingredients / Ingredients list}]{(allergens in CAPITALS)}]: (INGREDIENTE) [(CANTIDAD)] [(INGREDIENTE)]; [Flavouring: (INGREDIENTE) [(CANTIDAD)] [(INGREDIENTE)]].
  - Example 1: Pork (97%), Demerara sugar (Sugar, Cane Molasses), Sugar, Cinnamon, Salt, Ph Regulator (E500), Preservatives (E252, E250), Antioxidant (E301).
  - Example 2: Ingredients: British Pork (95%), Sugar, Cornflour, Salt, Honey Powder (Maltodextrin, Honey), Fennel Seeds, Soya Sauce (Water, Soy Beans (Soya), Wheat, Salt), Garlic Powder, Maltodextrin, Onion Powder,

- Aniseed, Beetroot Powder, Stabiliser: Guar Gum; Cinnamon, Colour: Paprika Extract; Flavourings, Cloves, Black Pepper.
2. [{Ingredients / Ingredient list}.] [Pork] {Cured / Rubbed} with: (INGREDIENTE).
    - a. Example 1: Ingredients. PORK CURED WITH: WATER, SALT, SUGAR, SODIUM PHOSPHATES, SODIUM ERYTHORBATE, SODIUM NITRITE
    - b. Example 2: Ingredients. Cured With: Water, Salt, Sugar, Sodium Phosphate, Sodium Erythorbate, Sodium Nitrite.
  3. (NOMBRE DE PRODUCTO) is made by curing (PIEZA) with (INGREDIENTE).
    - a. Example 1: Pancetta is made by curing pork belly with salt & pepper.

These model lines function as controlled language choices for the Spanish-speaking user to consider during their production of the English texts to ensure the idiomaticity, the grammatical accuracy, and genre acceptability (Pérez Blanco & Izquierdo, 2021, p. 159–160). When a term has to be chosen, TorreznoTRAD also incorporates a terminological database.

## 6 The terminological database

TorreznoTRAD, the semiautomatic writing and translation aid tool, includes a terminological database with terms and equivalents. First, term candidates were automatically extracted from the Spanish subcorpus of C-MARMEAT using TermoStat Web 3.0. (Drouin, 2003) and then, terms were validated by applying the criteria proposed by L’Homme (2020, p. 72–75). This procedure was repeated in the English subcorpus of C-MARMEAT and equivalents were manually established following the methodology proposed by Ortego Antón (2019, p. 180). Terms and their equivalents were gathered in several Microsoft Excel spreadsheets organized by semantic fields: actions, adjectives, allergens, kitchen gadgets, categories, cooking, additives, cuts, ingredients, nutritional values, unit of measures, etc., accompanied by the part of speech, the equivalents, and an example of use (see Figure 6).

Once we have completed the linguistic data, we have developed the computing tool with the aid of computing engineers.

	ES	EN	GRAM	EX
2	aditivos	additives	N	This product contains additives
3	antioxidantes	antioxidants	N	This product contains Antioxidants
4	apio	celery	N	Contains celery
5	aromas artificiales	artificial flavours	N	Free from artificial flavours
6	azúcares	sugar	N	Sugar free
7	cerdo	pork	N	Contains pork and its derivatives
8	colorantes artificiales	artificial colours	N	Free from artificial colours
9	conservantes	preservatives	N	This product contains Preservatives
10	dióxido sulfúrico	sulfur dioxide	N	May contain traces of sulphur dioxide
11	fenilalanina	phenylalanine	N	May contain phenylalanine
12	frutos secos	nuts	N	May contain nuts
13	gluten	gluten	N	Gluten free
14	lactosa	dairy	N	Dairy free
15	leche	milk	N	May Contain Milk

**Figure 6:** Screenshot of terms belonging to the semantic field of allergens.

## 7 Examples

The previous linguistic work at language level has been transferred to the computational level using a variety of programming language. HTML5 is a mark-up language used in the design of the the graphical user interface. Accordingly, three cascading style sheets (CSS3) were used to achieve a modern-looking interface that adapts to screen size, using the Bootstrap toolkit, and to personalize general views and the sidebar menu with `main.css` and `sidebar-menu.css` respectively. In addition, AngularJS (1.5.0) is a JavaScript-based front-end web framework for client-side model-view control. This program is useful to develop single-page applications. MongoDB (2.6.11) is a NoSQL document-oriented database that stores JSON-like documents. Finally, Node.js (4.2.6) is a server-side JavaScript environment that has been used to query the MongoDB database program, and to create API REST, which is a protocol for data exchange on the Internet. We have used it to retrieve query hits in a JSON format. It also enables image storage.

Once logged onto the home page, TorreznoTRAD is a dynamic web-based software with a user-friendly interface that guides translators and technical writers through the writing and translation of *torrezno* product cards. They get suggestions for drafting that they will have to complete, either with the help of the terminological database or by themselves. As the user progresses, the writing is saved, which enables a preview of the output at any time, whether or not the text is fully edited. Users navigate between the different moves (see Figure 7).

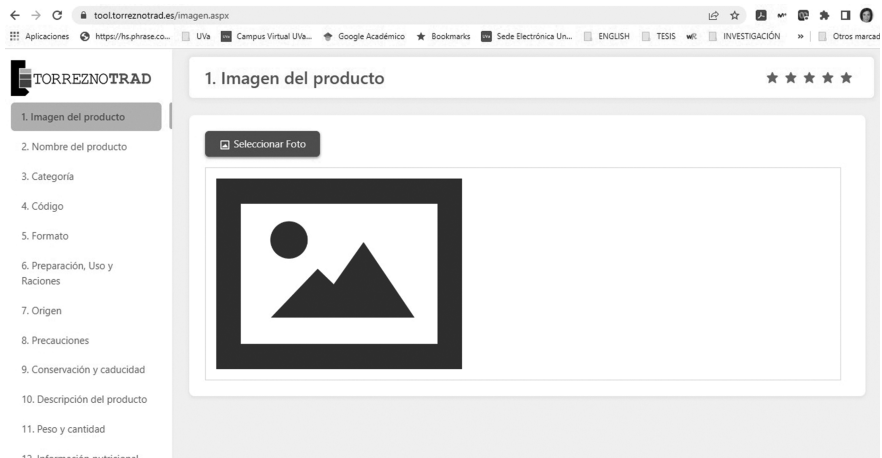


Figure 7: Menu with the structure of the marinated meat product card.

For example, when the move “14. Ingredientes” is selected and the “Sugerencias” button is pressed, a pop-up window appears where the three model lines are shown so that the user must choose one of them by clicking on “Añadir”. Besides, when scrolling through each of the model lines, an example appears as shown in Figure 8.



Figure 8: Model lines of the “Ingredients” move.

The tool offers different elements that guide the user with different colored buttons, that is, red indicates that a word or group of words is an optional text fragment, dark grey is used for selecting options from the terminological database, and orange shows two or more options from which the user must choose one.

Therefore, the user completes the content of the marinated product card following the instructions provided by the different windows. When users have to choose a lexical unit from the terminology database, it appears in dark grey. Clicking on “INGREDIENTE” displays a window in which the user types the first characters of the term in Spanish, and the English equivalents are offered (see Figure 9).



**Figure 9:** Example of a drop-down menu for choosing a lexical unit.

When the user has completed all the fields, the preview button at the left menu provides the final version of the English marinated meat product card, in DOCX format, adapted to the linguistic and cultural conventions of English discourse.

To sum up, corpus-based research has enabled the development of writing aids which are useful and efficient because they rely on the extraction and analysis of real data from comparable corpora. Among the existing writing aids, Torrezno-TRAD is a tool that can help increase the productivity of agri-food companies, as it shortens the translation process, since translators and language professionals are not generally trained nor specialized in agri-food genres.

## 8 Conclusions

The aim of this chapter has been to explain the development and functioning of an English specialized writing and translation tool for Spanish professionals. In particular, the tool is customized to meet the needs of a very specific group of

users in the agri-food sector manufacture and promotion, that is, local meat companies grouped under Asociación de Fabricantes de Torrezno de Soria.

Besides, the development of linguistic applications such as TorreznoTRAD makes it possible to transfer research results to the productive framework so that small and medium sized companies may reach the international market, and translators and technical writers can increase productivity. All in all, dried meat companies can offer quality information on the Internet.

Moreover, the methodology used for design and development of TorreznoTRAD could be easily reproduced in other studies dealing with different genres in this field or other pair of languages, so new resources could be built in.

From an academic perspective, future multilingual translators and technical writers must be trained in the particularities of the different specialized genres so that they are not limited to transferring the content to another language but can also ensure quality and acceptable texts that suit the communicative situation of the target culture. Thus, it is important to transfer these results to specialized translation courses.

Finally, a neural machine translation module is currently being added to TorreznoTRAD to supplement data which have not been included in the terminological database. At the same time, multilingual writing aids for other areas of the agri-food sector will continue to be developed.

## References

- Anthony, L. (2019). *AntConc (Version 3.5.8) [Computer Software]*. Tokyo, Japan: Waseda University. Electronic version: <http://www.laurenceanthony.net/software> (17 <http://www.laurenceanthony.net/software> (17 March 2019).
- Beeby, A., Rodríguez Inés, P. & Sánchez Gijón, P. (2009). *Corpus Use and Translating*. Amsterdam/Philadelphia: John Benjamins.
- Biber, D., Connor, U. & Upton, T. A. (2007). *Discourse on the Move. Using Corpus Analysis to Describe Discourse Structure*. Amsterdam/Philadelphia: John Benjamins. DOI:10.1075/scl.28
- Bowker, L. (2002). *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.
- Corpas Pastor, G. & Seghiri, M. (2010). Size Matters: A Quantitative Approach to Corpus Representativeness. In R. Rabadán, T. Guzmán & M. Fernández (Eds.), *Lengua, traducción, recepción: en honor de Julio César Santoyo* (pp. 111–145). León: Universidad de León, Área de Publicaciones.
- Corpas Pastor, G. & Seghiri, M. (2017). *Corpus-based Approaches to Translation and Interpreting. From Theory to Applications*. Wien: Peter Lang.
- Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, 9(1): 99–117.

- Durán Muñoz, I. & Corpas Pastor, G. (2020). Corpus-Based Multilingual Lexicographic Resources for Translators: an Overview. In M. J. Domínguez Vázquez, M. Mirazo Balsa y C. Valcárcel Riveiro (Eds.), *Studies on Multilingual Lexicography* (pp. 159–178). Berlin: De Gruyter.
- Fernández Nistal, P. (2020). Los corpus como herramienta de traducción para los traductores del siglo XXI: el caso del chorizo ibérico de bellota. In S. Álvarez Álvarez & M. T. Ortego Antón (Eds.), *Perfiles estratégicos de traductores e intérpretes. La transmisión de la información experta multilingüe en la sociedad del conocimiento del siglo XXI* (pp. 143–160). Granada: Comares.
- Labrador, B. & Ramón, N. (2020). Building a second-language writing aid for specific purposes: Promotional cheese descriptions. *English for Specific Purposes*, 60, 40–52. DOI: 10.1016/j.esp.2020.03.003.
- Laviosa, S. (2002). *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam/ New York: Rodopi.
- L’Homme, M. C. (2020). *Lexical Semantics for Terminology. An Introduction*. Amsterdam/Philadelphia: John Benjamins.
- López-Arroyo, B. & Roberts, R. P. (2015). The Use of Comparable Corpora: How to Develop Writing Applications. In M. T. Sánchez Nieto (Ed.), *Corpus Based Translation and Interpreting Studies: From Description to Application* (pp. 147–156). Berlin: Frank & Timme.
- MAPA. 2022. *Informe anual de la industria alimentaria española periodo 2021–2022*. <[https://www.mapa.gob.es/es/alimentacion/temas/industria-agroalimentaria/20220728informeanualindustria2021-2022t22ok\\_tcm30-87450.pdf](https://www.mapa.gob.es/es/alimentacion/temas/industria-agroalimentaria/20220728informeanualindustria2021-2022t22ok_tcm30-87450.pdf)> (28 November 2022).
- Moorkens, J. (2017). Under Pressure: Translation in Times of Austerity. *Perspectives*, 25(3): 464–477.
- Moreno-Pérez, L. & López-Arroyo, B. (2021). Atypical corpus-based tools to the rescue: How a writing generator can help translators adapt to the demands of the market. *MonTI*, 13, 251–279.
- Ortego Antón, M. T. (2019). *La terminología del sector agroalimentario español-inglés en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos*. Berlin: Peter Lang.
- Ortego Antón, M. T. (2020). Las fichas descriptivas de embutidos en español y en inglés: un análisis contrastivo de la estructura retórica basado en corpus. *Revista Signos*, 53(102): 170–194.
- Ortego Antón, M. T. (2021). E-DriMe: A Spanish-English frame-based e-dictionary about dried meats. *Terminology*, 27(2): 330–357. DOI: 10.1075/term.20013.ort.
- Ortego Antón, M. T. (2022). *La investigación en tecnologías de la traducción. Parámetros de la digitalización presente y la posible incidencia en el perfil de los futuros profesionales de la comunicación interlingüística*. Berlin: Peter Lang.
- Ortego Antón, M. T. (2023). Building corpus-based writing aids from Spanish into English: the case of GEFEM. In M. Izquierdo & Z. Sanz-Villar (Eds.), *Corpus use in cross-linguistic research: Paving the way for teaching, translation and professional communication* (pp. 218–235). Amsterdam/Philadelphia: John Benjamins.
- Ortego Antón, M. T. & Fernández Nistal, P. (2020). Estudio contrastivo de la terminología de embutidos en inglés y en español con ParaConc y tICorpus a partir del corpus paralelo P-GEFEM y del comparable C-GEFEM. In M. Seghiri (Ed.), *El uso de los corpus lingüísticos como herramienta pedagógica para la enseñanza-aprendizaje de lenguas, traducción e interpretación* (pp. 23–48). Bern: Peter Lang.
- Pérez Blanco, M. & Izquierdo, M. (2020). A multi-level contrastive analysis of promotional strategies in specialized discourse. *English for Specific Purposes*, 58, 43–57.
- Pérez Blanco, M. & Izquierdo, M. (2021). Developing a corpus-informed tool for Spanish professionals writing specialized texts in English. In J. Lavid-López, C. Maíz-Arévalo & J. R. Zamorano-Mansilla (Eds.), *Corpora in Translation and Contrastive Research in the Digital Age* (pp. 147–173). Amsterdam/Philadelphia: John Benjamins.

- Pérez Blanco, M. & Izquierdo, M. (2022). Engaging with customer's emotions: a case study in English–Spanish online food advertising. *Languages in Contrast*, 22(1): 43–76.
- Rabadán, R., Pizarro, I. & Sanjurjo-González, H. (2021a). Authoring support for Spanish language writers: A genre–restricted case study. *RESLA*, 34(2): 677–717.
- Rabadán, R., Ramón, N. & Sanjurjo-González, H. (2021b). Pragmatic Annotation of a Domain-Restricted English–Spanish Comparable Corpus. *Bergen Language and Linguistics Studies*, 11(1): 209–23. DOI: 10.15845/bells.v11i1.3445.
- Sánchez Carnicer, J. (2022). *Traducción y discapacidad. Un estudio comparado de la terminología inglés-español en la prensa escrita*. Berlin: Peter Lang.
- Sánchez Ramos, M. M. (2019). 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): 269–285.
- Sanjurjo-González, H. & Andaluz-Pinedo, O. (2021). OpenTagger: A flexible and user–friendly linguistic tagger. In: *Proceedings of the 56th Linguistics Colloquium*. Berlin: Peter Lang.
- Sanz Valdivieso, L. & López Arroyo, B. (2022). The phraseology of wine and olive oil tasting notes: A corpus based semantic analysis. *Terminology*, 28(1): 37–64.
- Seghiri, M. (2006). *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*. Thesis dissertation. Málaga: Universidad de Málaga. <<http://hdl.handle.net/10630/2715>> (10 September 2012).
- Seghiri, M. (2017). 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. *Babel*, 63(1): 43–64. DOI: 10.1075/babel.63.1.04seg
- Seghiri, M. & Arce Romeral, L. (2021). *La traducción de contratos de compraventa inmobiliaria: un estudio basado en corpus aplicado a España e Irlanda*. Berlin: Peter Lang.