Recibido / Received: 23/07/2024
Aceptado / Accepted: 27/11/2024

Para enlazar con este artículo / To link to this article:
http://dx.doi.org/10.6035/MonTI.2025.ne8.16

Para citar este artículo / To cite this article:
Peñuelas Gil, Isabel & María Teresa Ortego Antón. (2025) "Multiword expressions in culinary tourism brochures and their machine translation: does context matter?" In: Álvarez Jurado, Manuela & Francisco Luque Janodet (eds.) 2025. *La traducción del discurso turístico: desafíos actuales y emergentes / The translation of tourism discourse: current and emerging challenges*. *MonTI* Special Issue 8, pp. 452-482.

MULTIWORD EXPRESSIONS IN CULINARY TOURISM BROCHURES AND THEIR MACHINE TRANSLATION: DOES CONTEXT MATTER?

Isabel Peñuelas Gil
isabel.penuelas@uva.es
Universidad de Valladolid-CITTAC

María Teresa Ortego Antón
mariateresa.ortego@uva.es
Universidad de Valladolid-CITTAC

Abstract

Culinary tourism is a growing tourism sector, and although the literature on tourism translation is extensive, the gastronomic field remains a niche yet to be thoroughly explored (Ortego Antón & López Tirado 2024). Furthermore, machine translation (MT) is frequently used for the cross-linguistic transfer of culinary tourism brochures available on promotional websites, despite the fact that multiword expressions (MWEs) pose a challenge for MT systems. Recognizing that MT is here to stay, this paper tests whether the context of a sample of MWEs, extracted from a corpus of Spanish culinary tourism brochures, alters the product resulting from the cross-linguistic transfer to English in a selection of MT systems of different types. The results of the analysis allowed us to assess the degree to which the output was altered and to establish three behavioral patterns that can be associated with the different MT systems.

Keywords: Machine translation; Culinary tourism; Multiword expressions; Spanish-English; Corpus

Resumen

El turismo gastronómico es una modalidad turística en alza y, aunque la literatura sobre traducción turística es muy extensa, el ámbito gastronómico constituye un nicho por explorar (Ortego Antón & López Tirado 2024). Además, la traducción automática (TA) tiende a emplearse frecuentemente en el trasvase interlingüístico de los folletos turísticos gastronómicos disponibles en los sitios web promocionales, a pesar de que las expresiones multiverbales (EMV) constituyen un obstáculo para los sistemas de TA. Conscientes de que la TA ha llegado para quedarse, en este trabajo comprobamos si el contexto de una muestra de EMV, extraídas de un corpus compuesto por folletos turísticos gastronómicos en español, altera el producto resultante del trasvase interlingüístico al inglés en una selección de sistemas de TA de diferente tipología. Los resultados que se desprenden del análisis nos permitieron vislumbrar en qué grado se ve alterado el producto resultante y establecer tres patrones de comportamiento que pueden asociarse a los distintos sistemas de traducción automática.

Palabras clave: Traducción automática; Turismo gastronómico; Expresiones multiverbales; Español-inglés; Corpus.

1. Culinary tourism and translation

Tourism is one of the fundamental pillars of the Spanish economy. According to data from 2022, Spain ranks as the second-leading country in terms of annual international tourist arrivals (71.6 million international tourists per year, according to the Spanish Ministry of Industry, Commerce, and Tourism (2023)), only behind France (World Tourism Organization [UNWTO] 2023: 6). This increase in international visitors has promoted the availability of tourism translations—such as websites, brochures, guides, menus, etc.—intended for a broader non-Spanish-speaking international audience.

Beyond historical monuments and sunny beaches, food and drinks have become an essential attraction for those seeking to immerse themselves in the local culture. Additionally, it is an economic asset that benefits a sustainable economy, local consumption, and trade. In the case of Spain, a country known for its rich culinary tradition, culinary tourism plays a crucial role in promoting its cultural heritage by appreciating local culinary resources and expressing authenticity, sensory spectacle,

consumer erudition, experiences, and service co-production (Hernández Mogollón *et al.* 2015: 410).

As tourists arrive from different parts of the world, the need to translate the Spanish culinary richness into other languages arises. Translation does not only involve the transfer of words into another language but must also convey the cultural essence and emotions associated with each dish. The choice of words and descriptions can alter the perception and the culinary experience of tourists (World Tourism Organization and Basque Culinary Center 2019).

In this sector, while some translations are done by professionals, it is also common to see that small and medium-sized enterprises with limited resources use machine translation systems (MTSs). The emergence and popularization of free online MTSs in recent decades have democratized their use and consequently facilitated the communication process (Moorkens 2022: 137). MTSs are globally accessible tools, so large amounts of information that previously might have been unreadable for a large proportion of the population—due to several factors such as the scarcity of professionals with a specific language pair, monetary issues, or simply the amount of material to be translated—is now available to the majority.

However, this practice can lead to communication problems that are not always considered when machine translation (MT) is chosen and contrasts, as Álvarez Jurado (2020: 4) points out, with tourists demanding higher-quality texts.

In this background, one of the main obstacles that prevent the resulting product from resembling the quality of a human translation concerns multiword expressions (Corpas Pastor 2013) or MWEs. MWEs are extremely frequent linguistic units in natural languages (Jackendoff 1977) rooted in specific cultures, and the meaning of which is not always inferred from the sum of their components. The abundance of MWEs and the frequent use of MTSs in culinary tourism raise questions about their impact on communication in this sector.

Consequently, this study aims to determine whether the immediate context of MWEs alters the output of MTSs, and to check the influence certain MWEs have on the translation of their surroundings, as well as

how that context might alter the handling of the MWEs during the identification, segmentation, and translation phases. For this purpose, the concept of MWEs is defined (section 2) and their translation using MTSs is examined to then establish the analysis methodology: the delimitation of the analysis sample extracted from a monolingual corpus, the selection of MTSs, and the classification of errors during the cross-linguistic transfer (section 3), the analysis and results (section 4), and the conclusions (section 5).

2. MWE and machine translation

Translation has not been immune to technological advancements. As a result of progress in areas such as computational linguistics, statistics, and artificial intelligence (AI), machine translation (MT) has emerged. MT is defined as "the automatic translation of text from one natural language to another using a computer application" (ISO 2017). Although it may seem reductionist, MT is characterized by its interdisciplinarity. As Quah (2006: 57) indicates, it involves areas such as lexicography, linguistics, computational linguistics, computer science, and linguistic engineering. Moreover, machine translation is based on the premise that natural languages can be described, controlled, and encoded.

With the proliferation of MTSs in the 21st century, two prominent perspectives have emerged. In the early stages of the popularization and expansion of these systems, many professionals viewed machine translation as a threat. As Oliver (2016: 155) points out, they did so under the premise that the quality of these systems would soon reach a level sufficient to make human intervention unnecessary. In contrast, the non-expert public perceived these systems positively because the apparent lack of human intervention—a notion regularly repeated by popular press (Bowker 2023: 97)—facilitated access to information, thereby eliminating communication barriers (Austermühl & Kortenbruck 2012: 153) in a "quick, easy, and seemingly cost-free manner" (Vieira 2020: 98).

To understand the reality faced by sectors that regularly use MT, particularly the tourism sector, it is necessary to be aware of the shortcomings of both approaches. Above all, interlinguistic mediation professionals

widely accept that translation goes beyond the mere substitution of words from one language (SL) to their equivalents in another (TL). To fit the communicative situation, a translator must possess knowledge that goes beyond the language and characteristics of the specialized language. Attention must be paid to cultural factors, as Pérez Blanco and Izquierdo (2021: 148) express:

> 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: 148)

With the implementation of AI advancements applied to natural language processing (NLP), MTSs have improved compared to their predecessors, since, due to the increase of computer processing power, they use larger training corpora. However, these technologies still require human intervention to ensure the quality of their translations, so new specializations, such as pre-editing and post-editing, have emerged. The mere appearance of these new roles indicates that MT alone does not pose a threat as initially feared by some professionals. However, if non-professional users opt for MT, they will most likely use the resulting "raw translation" without professional review, leading to a product with errors.

In such cases, superficial errors, such as lack of agreement in person or grammatical number, are easily detectable if one translates into their mother tongue or an intermediate language with which they are very familiar. However, issues like phraseology, humor, double meanings, and tones (e.g., sarcasm or irony), which can completely change the meaning of a text, may go unnoticed.

Faced with this scenario, this work focuses on the analysis of specific phraseological units, referred to hereafter as multiword expressions (MWEs)[1]. They are defined, from a computational linguistic perspective,

1. There is a lack of consensus among experts regarding phraseology's subject matter (Penadés Martínez 2015: 15). These units are commonly referred to as *collocations, fixed expressions, phraseological units, multi-word units, phrasemes*, etc. in some cases

as combinations of at least two words where the speaker does not have complete freedom in choosing the components because they have been conventionalized by their use within a specific linguistic context (Peñuelas Gil 2024b: 73).

Given their multiword and cultural nature, in addition to often presenting idiomatic meaning, MWEs are complex to translate, especially for MTSs, which, as previously mentioned, only facilitate language transfer, disregarding the other aspects that go into translation. Consequently, the non-apparent meanings of expressions—i.e., when the sum of the individual elements' meanings does not equate to the MWE's meaning as a whole—, the potential double meanings or ambiguities they generate, as well as their cultural implications are often not represented in the MT output. Furthermore, MWEs are extremely recurrent, representing a significant percentage of the habitual vocabulary of native speakers of any natural language (Jackendoff 1977; Biber *et al.* 1999; Sag *et al.* 2002). This recurrence implies that the appropriate use of MWEs is essential for achieving a natural expression in any language.

These factors have made MWEs not only a fundamental element of tourism texts in all their representations but also one of the key features in the field of NLP, as they pose one of the greatest challenges in the progress of translation technologies:

> The successful computational treatment of MWUs is essential for Natural Language Processing, including Machine Translation and Translation Technology; the inability to detect MWUs automatically may result in incorrect (and even unfortunate) automatic translation and may jeopardize the performance of applications. (Mitkov *et al.* 2018: 3)

Given the growth of machine-translated tourism texts and the considerations presented above, it is not unreasonable to think that the present and future of both fields are closely linked to advancements in the processing of MWEs.

---

even interchangeably (Mitkov *et al.* 2018: 3), although this view is not universally accepted by all linguists.

2.1. The processing of multiword expressions in machine translation systems

As machine translation systems have advanced so has research on how different MT paradigms deal with MWEs. Each MT paradigm processes the linguistic information with which it is created and the input texts to be translated differently, resulting in divergent outcomes. Currently, among the most common MTSs for specific purposes, we find statistical systems (SMT), neural systems (NMT), and hybrid systems (HMT).

Both SMT and NMT require corpora for training. Similarly, current hybrid engines generally also base their architecture on the use of corpora, as they are the product of combining results from other paradigms (Costa-Jussà & Fonollosa 2015: 4). Traditionally, these engines have partially relied on SMT systems and, more recently, they have also benefited from neural technology. Therefore, all three paradigms start in most cases from corpus technology. However, despite this shared commonality, they process this data differently.

Statistical systems represented significant progress in MT due to the positive relationship between the quality of results and the effort required to create the systems. They dominated the market until the emergence of neural technology, but they still have utility in specialized fields today. These systems extract statistical information from large corpora used during the training phase of the engine (Brown *et al.* 1988: 71) and "rely on the probability (high or low) that a sentence in the source language corresponds to a translation in the target language" (Sánchez Ramos & Rico Pérez 2020: 15). Of course, this is the same principle they use for translating MWEs. However, the theoretical foundations of this approach—first based on the alignment of word pairs (Brown *et al.* 1988, 1990) and later "phrase alignment" (Zens *et al.* 2002; Koehn *et al.* 2003)—do not address the specific needs of translating MWEs.

The first case presents two obstacles: firstly, the discordance between the idiomatic nature of MWEs and the word-by-word alignment that treats all constituent elements of an MWE as individual elements; secondly, its inability to process "many-to-many" correspondences (Mitkov *et al.* 2018: 16). This second obstacle is solved in the phrase alignment model. However, these phrases, better known as n-grams, do not always equate to

MWEs and can result in combinations with limited linguistic significance. This is a problem that SMT engines with other alignment models, specifically those including syntactic information (Yamada & Knight 2001), improve upon, a trend followed in recent years. Nevertheless, as Mitkov *et al.* (2018) note:

> In the state-of-the-art PB-SMT [Phrase-Based Statistical Machine Translation] systems, the correct translation of MWUs occurs therefore only on a statistical basis if the constituents of MWUs are marked and aligned as parts of consecutive phrases (n-grams) in the training set and it is not generally treated as a special case where correspondences between source and target may not be so straightforward, i.e., it does not consist of consecutive many-to-many source-target correspondences. (Mitkov *et al.* 2018: 17)

Neural systems, conversely, emerged in the mid-2010s and presented themselves not only as one of the most innovative paradigms but also as one capable of rapid progress (Sánchez Ramos & Rico Pérez 2020: 12). They are based on artificial neural networks, a computational model that "comprises a large number of highly interconnected processing elements that work in unison to solve specific problems" (Bowker & Buitrago Ciro 2019: 45).

As these systems are relatively new, research on how they manage and translate MWEs is still limited. However, they have a potential advantage: neural systems learn while translating. They use machine learning technology that enables the neural network to identify complex patterns in its training texts and use them to learn how to translate new texts automatically (Bowker 2023: 96). From the perspective of MWEs, this implies that, over time, NMT engines could learn to translate any expression. However, to reach that point, it is necessary to use these tools appropriately, to avoid perpetuating the same errors. This is a crucial and novel notion in the field, leading to reflections such as the following:

> An NMT system might indeed produce an idiomatic translation, but this is generally because the data it has learned from contain hundreds or maybe thousands of examples of that very translation. An NMT [...] does not know it is being idiomatic, or using a cultural equivalent, when it correctly translates [an idiom]. (Kenny 2022: 39)

Although these systems have represented a step forward in terms of information processing, their way of processing texts has not changed significantly. As Kenny (2022: 39) points out, an NMT system can properly translate an idiomatic expression, but it does not do so through a conscious decision about the context and its implications. Rather, it does so because the engine infers from the training texts that it is the most probable option based on what it already knows. This implies that a failure to recognize the elements involved could affect the entire sentence containing the MWE, creating communication obstacles derived from the presence of these phraseological elements. Thus, making appropriate use of NMT for its progress and, of course, for the progress of MT concerning MWEs is essential.

Once the object of study in this work and MT systems have been defined the methodology employed is described.

3. Methodology

To identify the behavioral patterns exhibited by MTSs with respect to MWEs in the Spanish-English language pair, the sample of MWEs is selected from a monolingual corpus that has been compiled and exploited. Next, MTSs from which we will obtain the MWE translations are chosen and, finally, the procedure for data collection and analysis is explained.

3.1. Sample of analysis

To obtain real examples, the corpus is restricted to tourist brochures about gastronomy. Tourist brochures are a specific textual genre within the field of tourism (Calvi 2010, 2019), designed to promote the offerings of a region or establishment:

> Una publicación distribuida gratuitamente, la mayoría de las veces en forma de desplegable o cuadernillo y, entre sus funciones, destacan la de informar, promover una imagen y seducir y atraer al turista. En los folletos, solemos encontrar partes descriptivas, expositivas y argumentativas combinadas con partes prácticas y elementos gráficos. Estos últimos elementos (fotografías) activan la imaginación del receptor y provocan su

> interés por el producto, servicio o destino turístico. (Soto Almela 2014: 66)

Therefore, brochures have a dual purpose: to inform and to persuade the reader. As such, gastronomic tourism brochures provide details about local cuisine, describing typical dishes, local ingredients, and traditional culinary techniques. At the same time, they aim to convince the reader to visit the promoted destination, emphasizing the quality and uniqueness of its gastronomic offerings.

To achieve this, the sender needs to understand the psychological and linguistic characteristics of the addressees. Thus, the brochure acquires the properties of an advertisement consumed by a receptive audience interested in traveling and exploring new places and cultures. It could even be considered a text with a strategic act in which the sender imposes their opinion on the addressee and reader, who is the gastronomic tourist.

Additionally, visual design is a crucial aspect of these brochures. Attractive images of dishes and culinary environments are used to capture the reader's attention, along with typography and a color palette that reinforce the cultural identity of the destination. This visual approach aims not only to inform but also to evoke an emotional response in the reader, encouraging them to experience the presented culinary adventure.

Once the textual genre was defined, the corpus was compiled following a protocol similar to that used in previous studies (Seghiri 2017; Sánchez Carnicer 2022; Ortego Antón 2024a, 2024b; Peñuelas Gil 2024a, 2024b; among others). The compilation was carried out in four phases:

> — Search: We searched for texts written by companies or organizations providing tourism services that focused on the gastronomic aspect and adhered to the textual typology defined in the study. For this work, only texts available on websites as standalone documents were considered.
> — Manual download: We manually downloaded the selected texts in their original format (PDF).
> — Conversion: Texts were converted to UTF-8 TXT files to be processed by corpus management tools, using AntFileConverter (Anthony 2022).
> — Storage: Since it is a monolingual corpus, texts were only separated according to their format (PDF or TXT).

Each text was assigned a unique identifier based on the following ID: numerical identifier according to the order of download + the abbreviation of the theme (in this case "GT" as an abbreviation of "gastroturismo") + the indication of the origin of the text (abbreviation of the province or autonomous community for which it was written) + the download date (yyyymmdd). For example, the file 001GTSO20230518 is the first document in the corpus, contains relevant information about the province of Soria, and was downloaded on May 18, 2023.

The result is a virtual monolingual corpus composed of 30 brochures on gastronomic tourism in Spain—with 93,185 tokens or words, and 16,569 types or unique words—cataloged and structured for exploitation.

Next, the corpus was exploited using two corpus management programs—Sketch Engine (Kilgarriff *et al*. 2004) and AntConc (Anthony 2020)—, which allowed for the analysis and extraction of pertinent information quickly and efficiently. Despite the two programs having similar tools and functions, our decision to use both was driven solely by a preference for their respective interfaces.

The extraction procedure for MWEs began by using *WordList*, one of AntConc's basic tools that generates a list of unique words, or types, ordered by their frequency of occurrence. To ensure linguistically relevant results, a stoplist was used to ignore less relevant words such as articles, prepositions, or conjunctions. The results were reviewed one by one using Sketch Engine's *N-grams* tool, which filters and nests results, considering each term's possible combinations and their frequency.

This review process revealed the common use of expressions built around the terms *punto* and *fuego*, which appear a total of 98 times in the corpus in the context of Spanish gastronomic tourism. The MWEs detected in the corpus that include either of the proposed terms show characteristics that make them interesting for analysis:

— Some of them exhibit a degree of idiomaticity, implying that: a) their meaning cannot be directly inferred from the meanings of their components but through sociocultural knowledge of the language; b) substantial differences with their English equivalents are likely.

— Their grammatical composition represents different patterns (noun + noun, noun + prep. + noun, noun + adj., etc.), leading us to consider possible differences in the results.

However, a number of repetitions were not related to gastronomic tourism, having a historical or geographical background, thus they will not be considered. Additionally, another sentence containing *punto* was discarded as, although it references the level of doneness of food, half of the original MWE is omitted as it is inferred from the context ("Se prepara a la parrilla y se sirve generalmente poco hecho. Este punto hay a quien le produce cierto rechazo por dar la impresión de venir ensangrentado").

Considering these factors, we decided to limit the sample of analysis to the following unique MWEs, which occur in a total of 87 sentences:

	MWE	**Occurrences**
1	a punto de nieve	1
2	aguja de hacer punto	1
3	en tu punto	1
4	estar a punto	3
5	punto caramelo	1
6	punto de congelación	1
7	punto de encuentro	1
8	punto de hebra fina	1
9	punto de referencia gastronómica	1
10	punto de sal	7
11	punto deseado de cocción	1
12	punto fuerte	15
13	punto idóneo de cocción	1
14	punto óptimo de maduración	1
15	tener su punto	1
16	un punto picante	1
17	a fuego fuerte	2
18	a fuego lento	18

19	a fuego mediano	1
20	a fuego medio	3
21	a fuego muy lento	1
22	a fuego muy suave	1
23	a fuego suave	3
24	a fuego vivo	1
25	apagar el fuego	1
26	avivar el fuego	2
27	con fuego de leña	1
28	en fuego de leña	1
29	fuego de leña	1
30	fuego del dragón	1
31	poner al fuego	2
32	poner al fuego lento	1
33	puesta al fuego	3
34	retirar del fuego	5
35	separar del fuego	1

Table 1. Analysis sample

3.2. Sample of analysis

In selecting the MTSs, two basic aspects were considered. First, even though two engines might belong to the same paradigm, they are trained with different corpora (Bowker 2023: 101), resulting in varied outcomes. Second, they should be representative of systems commonly used by the general public, that is, they are well-known enough for users to consider using them. However, it is important to note that popularity varies depending on the language pair, as not all engines translate all languages, and some language pairs are only available through an intermediate language.

By reviewing the literature on popular MTSs (Carré *et al.* 2022; Rivera-Trigueros 2022; Bowker 2023) and cross-referencing the lists found, we limited the sample to four MTSs: DeepL (NMT), Google Translate (NMT),

Microsoft Translator (HMT), and Yandex Translate (SMT for the ES-EN language pair). All four engines offer direct translation from Spanish to American English without using pivot languages, providing a solid basis for comparing their pros and cons. It is worth mentioning that the systems here examined are black-box systems, that is "a system which has been trained and tuned a priori and for which we cannot access the model parameters or training data for fine-tuning or improvements" (Mehta *et al.* 2020: 1).

Additionally, they are all free, available online, and offer the ability to translate images via mobile apps and full documents via web versions, potentially enhancing their use in a tourism context.

3.3. Data collection and analysis

Once the 35 unique MWEs were identified, with 87 total occurrences, they were translated using the MTSs mentioned above. Two points should be noted in this regard:

> — The MWEs were not translated in isolation but within their contexts to increase the chances of appropriate translation. As previously mentioned, these expressions are conventionalized through use and, therefore, evolve over time and as the socio-cultural situation in which they are used changes. Thus, they sometimes present multiple meanings depending on the context.
> — Given the ever-changing field we are in, it is important to note that all translations were obtained in May 2023.

The results were 348 segments in English that were manually analyzed and categorized following a modified version of the human translation evaluation system proposed by Ortiz Boix (2016: 63-64), based on the precision and fluency of translations. This error classification system aims to analyze errors found in machine translated documentaries, thus not all categories fit our study's needs, prompting the addition of six new ones: grammatical person, tenses, meaning change, wrong suggestion, spelling, and repetition (see Table 2).

The analysis was conducted in two different phases. The initial phase strictly focused on detecting errors in the translations of the MWEs. This

phase did not include errors in words and structures outside the expressions themselves (e.g., *tu* in "tu punto fuerte" where it is not part of the expression) or errors arising from a MWE in the sentence not being interpreted correctly.

Accuracy categories	Fluency categories

– Terminology
– Mistranslation
  – Overly literal
  – False friend
  – Should not have been translated
  – Date/time
  – Unit conversion
  – Number
  – Entity
  – Grammatical person
  – Tenses
  – Meaning change
– Omission
– Addition
– Untranslated
	– Wrong suggestion
	– Spelling
	  – Capitalization
	  – Diacritics
	  – Variant
	– Typography
	  – Punctuation
	  – Unpaired elements
	– Grammar
	  – Morphology
	  – Part of speech
	  – Agreement
	  – Word order
	  – Function words
	– Unintelligible
	– Repetition

Table 2. Error categorization

The second phase, however, focused on categorizing the errors found in the context of the MWEs. In this phase, we distinguished between immediate context—mistakes found within a radius of 4-5 words from the MWE—and general context, which is limited to the sentence in which the original MWE is found.

Comparing the results from both phases allowed us to trace processing patterns in each of the MTSs. With the methodology outlined, we now present the results obtained.

4. Analysis and results

Of the 348 translations under study, 123 (35.345%) showed errors in the MWEs, while 138 (39.655%) did in their contexts. Only 89 translations (25.575%) did not present mistakes in either of the analysis phases.

In total, 454 errors were detected, of which 131 were related to MWEs and 323 to the context (149 immediate, 174 general). All errors were counted individually and all of them were considered of the same importance. Therefore, a single translation could contain multiple mistakes, and no errors were overshadowed by others. There were five categories where no mistakes were accounted for: *False friend, Number, Unpaired elements, Date/time*, and *Unit conversion*. It is worth noting that the last two categories did not have a representation in the samples used for this study, as there were no units nor temporal cues in the sample. Conversely, we see that five categories constituted a bigger challenge, and that, together, these account for 72.687% of the registered errors: *Overly literal* (145 total errors), *Terminology* (74), *Grammatical person* (58), *Function words* (29), and *Punctuation* (24).

As portrayed in Table 3, the spread of the errors of each category is quite uneven, particularly in the case of Grammatical Person, where over 98% of the errors are registered in the immediate context of the MWEs. Some of the numbers registered are also worth noting, such as the 67 instances where the translation of a MWE was overly literal or the 60 times where that same problem was detected in the general context of an expression, both of which contrast significantly with what can be observed in the immediate context.

	MWE	Immediate context	General context	Total errors
Overly literal	67	18	60	145
Terminology	25	17	32	74
Grammatical person	0	57	1	58
Function words	3	13	13	29
Punctuation	0	7	17	24

Table 3. Error distribution in the most prolific categories

A previous study (Peñuelas Gil & Ortego Antón 2024) has shown that certain error typologies are usually linked when multiple translation problems occur within a single MWE, such as terminological mistakes

that tend to be related to an overly literal translation of the expression. Considering the broader perspective and the implications of certain MWE errors from a processing point of view, we observe patterns linking these errors to the translation of context and vice versa.

To establish these patterns, a series of small changes, based on the observation of the MTSs' preexisting tendencies, were introduced systematically. These changes ranged from adding or removing full stops at the end of the samples to reordering elements within the sentence or substituting certain terms with synonyms that retained the base lexeme. In one case, further context was added since the original sentence ("Lo mejor está a punto") was too short and ambiguous. Not all the changes yielded significant improvements in the translations or fully corrected them, but they allowed us to identify a series of elements that transcended the meaning of the sentences and consistently affected certain error typologies.

4.1. Pattern 1

The most common pattern identified involved the punctuation present in the general context. During the translation phase of the study, it was noted that, in certain instances, the four MTSs selected failed to replicate the punctuation of the original, particularly the full stop at the end of the translated sentences.

Fifteen of the original sentences did not include a full stop. When translating these, the systems added, in some cases, the final punctuation. We hypothesize that this is because the systems interpret the lack of punctuation as an error. However, no apparent patterns were detected in relation to when the full stop was added, since even when the MWE and general structure of the sentence were the same, they produced different results.

During the categorization of mistakes, the original result was used. However, upon removing the incorrectly added punctuation, it was noted that elements of the translation, particularly in the immediate context of the MWE, changed. The elements more prone to change were personal pronouns, particularly when it came to the translation of the pronoun *su*

that preceded the MWE "punto fuerte", although, as it can be seen in the following table, the changes were not limited to pronouns:

Su punto fuerte: capacidad de seguir aprendiendo		
a.	DL.1.	**His** strength: ability to keep learning.
	DL.2.	**Its** strength: ability to keep learning
b.	MST.1.	**His** strong point: the ability to continue learning.
	MST.2.	**Your** strength: ability to keep learning

Table 4. Pattern 1 sample in DeepL and Microsoft Translator

It was found that these alterations in the punctuation correlated mostly to the 58 errors cataloged under the *Grammatical person* category. This addition was found in 20 of the sentences that contained a problem regarding the grammatical person of at least one of its components. However, removing the additional punctuation did not necessarily correct the mistakes in most cases due to the ambiguity of *su*.

The 20 sentences that follow this casuistic contain information regarding a series of bakers whose shops are part of a gastronomic tour. *Su* could refer to both grammatical genders, as well as singular and plural subjects, and while most of the bakers are males, the MTSs have not defaulted to male English pronouns as often as they used to, which in many cases would have corrected the initial mistake.

This pattern was present in all four MTSs; however, it seemed to be most prevalent in DeepL's case (8 times), followed by Yandex.Translate (5), Microsoft Translator (4), and Google Translate (3). The prevalence of this pattern in DeepL aligns with this system's general tendency to try to correct mistakes and use a vocabulary more apt to an English speaker's daily life, but, as it was pointed out above, these attempts at correcting the punctuation seemed to be arbitrary rather than follow logic.

4.2. Pattern 2

The second pattern was characteristic of lengthy sentences, particularly those with long or convoluted enumerations. In such cases, it was found

that the alteration of elements of the enumeration that belonged to the general context affected and corrected the MWE, even though the immediate context, which holds a higher probability of modifying the translation had remained the same. In such cases, eliminating specific elements of the enumeration seemed to solve the preexisting errors, most commonly partial or total omissions, problems regarding the grammatical person, and lack of concordance. Conversely, paraphrasing those elements or moving them within the enumeration did not provide such results.

It is worth noting that the elements whose modification changed the translation output did not follow a clear pattern within the study sample. As shown in the example below, the changes in the MWE were prompted by the removal of specific elements, whether individual or in sets. In addition, the length of the deleted fragments varies as well, although we will always be dealing with shorter and simpler sentences.

One of the most notorious and complete cases happened in the translation of the following sentence in DeepL, which we will use as an example for this pattern:

> Se hacen con ellos guisos muy sabrosos, como los callos, los morros, las patitas de cordero, las lechecillas, las tripas hechas embuchados y gordillas, las cabezas de los corderos asadas, la asadurilla hecha con el conjunto de las vísceras troceadas, la lengua de ternera al vino tinto, el hígado encebollado, las manos de cerdo, sus orejas, que al igual que las de los corderos se consumen rebozadas como tentempié en los bares, los riñones y la sangrecilla con fritada, **un punto picante** o con huevos.

Initially, DeepL provided the translation "[...] kidneys and blood with fried, spicy or with eggs". The MWE's translation ("spicy") adds part of the intended meaning but does so without making the necessary grammar changes. However, it was detected that removing fragments from the general context resulted in the more appropriate translation of "[...] kidneys and blood with fried, spicy sauce or with eggs". The fragments that resulted in such translation when removed were:

- las tripas hechas embuchados y gordillas
- que al igual que las de los corderos se consumen rebozadas como tentempié en los bares
- la lengua de ternera al vino tinto, el hígado encebollado
- la asadurilla hecha con el conjunto de las vísceras troceadas, la lengua de ternera al vino tinto,
- el hígado encebollado, las manos de cerdo, sus orejas

Nonetheless, while the new translation of the MWE was an improvement from a grammatical point of view, it cannot be considered correct from a meaning perspective. Moreover, the sentence as a whole still presented several issues that could be associated to the presence of MWE.

DeepL was not the only MTS where this took place. A similar situation was observed in both Microsoft Translator and Yandex.Translate, which had also mistranslated the MWE. However, the elements that seemed to affect the MWE varied between systems, and the modified translation these two provided was still considered overly literal.

This example did not affect Google Translate. Nonetheless, it is possible that this pattern is linked to the case that triggered this study, which considers the context of the MWEs and was specific to this MTS.

a.	Ponemos las vieiras **a punto de sal** y las pasamos por el polvo de cecina formándoles como una piel [...]
	GT **Season** the scallops **with salt** and pass them through the cecina powder, forming them like a skin [...]
b.	Ponemos las vieiras [**a punto de sal**] y las pasamos por el polvo de cecina formándoles como una piel [...]
	GT We put the scallops and pass them through the jerky powder, forming them like a skin [...]
c.	**Salamos** las vieiras y las pasamos por el polvo de cecina formándoles como una piel [...]
	GT **We salt** the scallops and pass them through the cecina powder, forming them like a skin [...]

Table 5. Google Translate examples

As shown in Table 5's section A, Google's MTS initially left "cecina" untranslated. Even though its equivalent was correct, we theorized that it was the MWE's presence that affected the translation. Initially, we considered the problem to be structural, which prompted us to paraphrase the sentence in two ways. Firstly, we tried to ensure the meaning was the same by substituting "a punto de" with "a gusto de". The output resulting from that change had the same problem, therefore it was decided to change it further so that the structure was less similar even if that implied slightly modifying the connotations of the text.

However, as seen in sections B and C of Table 5, the translation or lack thereof of "cecina" seems to be linked to the presence of the term "sal" in the original sentence. This could mean that certain errors are connected to specific lexemes, that vary between MTSs due to the use of different training corpora, which could be what is prompting this second pattern.

4.3. Pattern 3

The last major pattern identified involved formal resemblance between MWEs in the study sample and expressions in the target language. Although the constituent elements can be similar individually to those of the English expression, the meanings do not correspond. Within the scope of this study, this third pattern has been detected solely in machine translation engines that fully or partially use a statistical architecture.

Initially, the translations obtained had a clear tendency towards word-by-word translation, making the output overly literal. These results are a potential sign of processing problems during the identification and segmentation phases carried out by the MTS. As such, it was decided to introduce some of the changes already discussed in this chapter that could favor a shift during the aforementioned phases, given that it is no longer common practice to design systems that divide the source text into phrases of a set n-length. Instead, the segments vary from text to text.

The changes were approached similarly to previous methods, starting with the least invasive ones and applying them first in the general context. Therefore, the initial changes consisted of modifying the punctuation of the sentence without applying further changes to the structure. This time,

the full stop at the end of the sentences did not affect the translation of the MWE (see Table 6, section A), thus the next step was to modify the punctuation within the sentences while ensuring that the meaning had not changed.

As seen in Table 6 section B, these changes modified the way the MWE was translated. However, the results did not improve the appropriateness and clarity of the translation. Instead, the MTS relied more heavily on the individual meanings of each component of the MWE and substituted the expression with one in the target language that included similar constituents. The same results occurred when adding the stop before "esa agua se desecha..." and when that last part was omitted, as seen in section C.

a.	Una vez limpios se ponen a cocer a fuego lento para que saquen la carne del caparazón, a continuación se aviva el fuego, esa agua se desecha y se escurren bien	
	YX	Once clean, they are put to simmer so that they remove the meat from the shell, then **the fire is fanned**, that water is discarded and drained well
b.	Una vez limpios se ponen a cocer a fuego lento para que saquen la carne del caparazón. A continuación se aviva el fuego, esa agua se desecha y se escurren bien.	
	YX	Once clean, they are put to simmer so that they remove the meat from the shell. Then **fan the flame**, that water is discarded and drained well.
c.	Una vez limpios se ponen a cocer a fuego lento para que saquen la carne del caparazón, a continuación se aviva el fuego[, esa agua se desecha y se escurren bien].	
	YX	Once clean, they are put to simmer so that they remove the meat from the shell, then **fan the flame**.

Table 6. Yandex.Translate examples

This pattern was not registered as often as the previous two, as this phenomenon is more specific and seems to require the existence of MWEs with similar constituents. However, its implications are significant enough to warrant consideration. It was detected three times, once in Microsoft Translator and twice in Yandex and both cases used the same

MWE: "avivar el fuego". In both cases the MTS understood "fuego" literally, prompting the use of verbs that are only used with an open flame, not a conventional stove such as *to fan* or *to stoke*. However, the change in punctuation abstracted the meaning of the constituents, going from an expression that implied making an open flame bigger, as did the original albeit metaphorically, to an idiom that means, according to the Cambridge Dictionary "to make a dangerous or unpleasant mood or situation worse".

We attempted to recreate these results with "a fuego vivo", another MWE found in our study and that presents similar constituents. However, in this case, the MWE had been translated correctly, and, although changing the punctuation varied the general translation the MWE remained correct. This could imply that this second MWE is found unchanged in the training corpora repeatedly, unlike the previous one, which could prompt a more erratic behavior.

Having described the results and patterns obtained from the analysis phase of this study, we proceed to present the conclusions.

5. Conclusions

The proliferation of free online machine translation systems has led us to consider to what extent their use could affect communication and in what ways it might do so, particularly concerning multiword expressions (MWEs), as they are considered complex units to translate and are extremely common in natural languages, and how the context they appear in affects them. In this study, we have analyzed this impact using texts from gastronomic tourism, a field characterized by the abundance of MWEs it employs, and four of the most popular machine translation systems on the market.

Results have shown that there is a complex interaction between the content and the terminology found in the context of MWEs, as well as the form said context takes, and the way the MWE will be translated. Accuracy and fluency problems in a translation have the potential to severely impact the comprehensibility, readability and appropriateness of the text, which could have further negative implications when dealing with the target user.

The analysis revealed three main processing patterns when translating segments of culinary tourism texts. These patterns vary in general incidence and nature, but all of them seem to be formed around particular elements within the sentence—such as pronouns, enumerations, and verbal tenses—and, to a certain degree, within the MWE.

Moreover, evidence suggests specific lexemes, both in the contexts and the MWEs, possibly interfering with the translation of other parts of the sentence. As discussed in section 4, some specific translation problems, or the lack thereof, seemed to be connected to the presence of specific word families (pattern 2). For instance, "cecina" was left untranslated whenever a term including the root "sal" was included in the sentence. Beyond the influence of specific lexemes, additional factors also contribute to translation difficulties. As examined in this chapter, factors such as terminological density and sentence length (pattern 2), and punctuation (pattern 1) further complicate the translation of MWEs, as does the existence of expressions in the TL with similar constituents but different meanings (pattern 3).

The detection of such patterns not only serves a theoretical purpose. In addition to enhancing our understanding of the process followed by machine translation systems when dealing with complex structures such as MWEs, it provides a base for practical applications: understanding how texts are processed and the results they generate, allows for more precise pre-editing of the texts to be translated, and highlights factors to consider during both the writing and the post-editing phases.

In the future, it would be essential to analyze more deeply the relationship between the highlighted processing patterns of the different translation engines studied here and various linguistic and operational factors. Firstly, it is necessary to examine how the grammatical structure of MWEs influences the incidence of specific errors, as identifying error-prone grammatical patterns could guide the development of more precise and effective algorithms.

Furthermore, it would be necessary to delve into the characteristic deficiencies of each engine, such as the repetitions or incorrect suggestions to modify the source text, as understanding these specific deficiencies will

allow for the customization of improvements in each system, increasing their efficiency and accuracy in translating MWEs. Additionally, it would be necessary to verify how the terminological density of the source sentences affects the translation of MWEs, as the data seem to indicate that the terminological complexity of the sentences affects the ability of MTSs to process and adequately translate MWEs.

Finally, it would be necessary to expand the research to a greater number of MWEs and apply the analysis to other fields of knowledge, which will allow for a deeper evaluation and understanding of how MWEs are processed by current MTSs and how to refine these systems to eliminate linguistic barriers more effectively.

6. Acknowledgements

This work has been partly carried out in the framework of the Spanish R&D projects entitled "Controlled natural languages, collaborative communication, and bilingual textual production in 3.0 environments" (PID2020-114064RB-I00), coordinated by Prof. Noelia Ramón García (University of León), "Multi-lingual and Multi-domain Adaptation for the Optimisation of the VIP system" (VIP II) (PID2020-112818GB-I00), coordinated by Prof. Gloria Corpas Pastor (University of Málaga) and "App to train in neuronal machine translation postediting through gamification in professional environments" (GAMETRAPP) (TED2021-129789B-I00), coordinated by Prof. Cristina Toledo Báez (University of Málaga).

References

Álvarez Jurado, Manuela. (2020) "Adquisición y transmisión del conocimiento experto a través de la traducción de las guías turísticas de arquitectura." *Onomázein* esp. 7, pp. 1-17. https://doi.org/10.7764/onomazein.ne7.01

Anthony, Laurence. (2020) *AntConc* (Version 3.5.9) [Software]. Tokyo: Waseda University. https://www.laurenceanthony.net/software/antconc/

Austermühl, Franz & Anke Kortenbruck. (2012) "A translator's sword of Damocles? An introduction to machine translation." In: Austermühl,

Franz (ed.) 2012. *Electronic tools for translation*. London: Routledge, pp. 153-176.

Biber, Douglas *et alii*. (1999) *Grammar of spoken and written English*. Essex: Pearson Education Limited.

Bowker, Lynne. (2023) *De-mystifying Translation: Introducing Translation to Non-translators*. London: Routledge. Online version: https://doi.org/10.4324/9781003217718

Bowker, Lynne & Jairo Buitrago Ciro. (2019) *Machine translation and global research*. Leeds: Emerald Publishing.

Brown, Peter F. *et alii*. (1988) "A statistical approach to language translation." In: *Proceedings of the 12th Conference on Computational linguistics* 1, pp. 71-76.

Brown, Peter F. *et alii*. (1990) "A statistical approach to language translation." *Computational Linguistics* 16:2, pp. 79-85.

Calvi, Maria Vittoria. (2010) "Los géneros discursivos en la lengua del turismo." *Ibérica* 19, pp. 9-32.

Calvi, Maria Vittoria. (2019) "Géneros discursivos, diversidad cultural y traducción de textos turísticos." *Altre Modernità* 21, pp. 69-86. https://doi.org/10.13130/2035-7680/11633

Carré, Alice *et alii*. (2022) "Machine translation for language learners." In: Kenny, Dorothy (ed.) 2022. *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Berlin: Language Science Press, pp. 187-207.

Corpas Pastor, Gloria. (2013) "Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas." In: Olza Moreno, Inés & Elvira Manero Richard (coords.) 2013. *Fraseopragmática*. Berlin: Frank & Time, pp. 335-374.

Costa-Jussà, Marta R. & José A. Fonollosa. (2015) "Latest trends in hybrid machine translation and its applications." *Computer Speech and Language* 32:1, pp. 3-10.

Hernández Mogollón, José Manuel *et alii*. (2015) "El turismo gastronómico como experiencia cultural. El caso práctico de la ciudad de Cáceres." *Boletín de la Asociación de Geógrafos Españoles* 68, pp. 407-425.

ISO (2017). *18587 International Standard. Translation services – Post-editing machine translation output – Requirement*. https://www.iso.org/standard/62970.html

JACKENDOFF, Ray. (1997) *The architecture of the Language Faculty*. Massachusetts: MIT Press.

KENNY, Dorothy. (2022) "Human and machine translation." In: Kenny, Dorothy (ed.) 2022. *Machine translation for everyone: empowering users in the age of artificial intelligence*. Berlin: Language Science Press, pp. 23-49.

KILGARRIFF, Adam *et alii*. (2004) "The Sketch Engine." In: *Proceedings of the 11th EURALEX International Congress*. Lorient: Université de Bretagne Sud, pp. 105-116.

KOEHN, Philipp *et alii*. (2003) "Statistical Phrase-Based Translation." In: *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*. Edmonton: NAACL, pp. 127-133.

MEHTA, Sneha *et alii*. (2020) "Simplify-then-translate: automatic preprocessing for black-box translation." In: *Proceedings of the 34th AAAI conference on artificial intelligence*. New York: AAII, pp. 8488-8495.

MITKOV, Ruslan *et alii*. (2018) "Multiword units in machine translation and translation technology." In: Mitkov, Ruslan *et alii* (eds.) 2018. *Multiword units in machine translation and translation technology*. Amsterdam & Philadelphia: John Benjamins, pp. 1-38.

MOORKENS, Joss. (2022) "Ethics and machine translation." In: Kenny, Dorothy (ed.) 2022. *Machine translation for everyone: empowering users in the age of artificial intelligence*. Berlin: Language Science Press, pp. 121-140.

OLIVER, Antoni (2016). *Herramientas tecnológicas para traductores*. Barcelona: UOC.

ORTEGO ANTÓN, María Teresa. (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, María Teresa. (2024a) "The Design of Torrezno TRAD: the Semiautomatic Spanish-English Writing and Translation Aid Tool." In: Peñuelas Gil, Isabel & María Teresa Ortego Antón (eds.) *Interpreting and Translation for Agri-food Professionals in the Global Marketplace*. Berlin: De Gruyter, pp. 69-84. Online version: https://doi.org/10.1515/9783111101729-004

ORTEGO ANTÓN, María Teresa. (2024b) "Metodología para el diseño de un asistente semiautomático de redacción y de traducción de fichas

descriptivas de embutidos del español al inglés." *Cadernos de Tradução* 44:1, pp. 1-20. https://doi.org/10.5007/2175-7968.2024.e94647

Ortego Antón, María Teresa & Ana López Tirado. (2024) "La traducción español-inglés de los culturemas gastronómicos en los folletos turísticos: el caso de 'tapa'." In: Policastro Ponce, Gisella (ed.) 2024. *Traducción, discurso turístico y cultura*. Granada: Comares, pp. 61-75.

Ortiz Boix, Carla. (2016) *Implementing Machine Translation and Post-Editing to the Translation of Wildlife Documentaries through Voice-over and Off-screen Dubbing*. Unpublished doctoral thesis. Barcelona: Universitat Autònoma de Barcelona. Online version: http://hdl.handle.net/10803/400020

Penadés Martínez, Inmaculada. (2015) *Para un diccionario de locuciones. De la lingüística teórica a la fraseografía práctica*. Alcalá: Universidad de Alcalá de Henares.

Peñuelas Gil, Isabel. (2024a) "From corpora to a semi-automatic tool: the use of model lines for the development of TorreznoTRAD." In: Peñuelas Gil, Isabel & María Teresa Ortego Antón (eds.) 2024. *Interpreting and Translation for Agri-food Professionals in the Global Marketplace*. Berlin: De Gruyter, pp. 105-118. Online version: https://doi.org/10.1515/9783111101729-006

Peñuelas Gil, Isabel. (2024b) *Estudio contrastivo del tratamiento de las expresiones multiverbales del turismo gastronómico en los sistemas de traducción automática del español al inglés*. Unpublished doctoral thesis. Valladolid: Universidad de Valladolid.

Peñuelas Gil, Isabel & María Teresa Ortego Antón. (2024) "Traducción automática: ¿a punto de caramelo? La traducción de expresiones multiverbales del turismo gastrónómico." Unpublished communication. *41 Congreso Internacional AESLA, 17-19 April 2024*. València: Universitat Politècnica de València.

Pérez Blanco, María & Izquierdo, Marlén. (2021) "Developing a corpus-informed tool for Spanish professionals writing specialized texts in English." In: Lavid-López, Julia *et alii* (eds.) 2021. *Corpora in Translation and Contrastive Research in the Digital Age*. Amsterdam & Philadelphia: John Benjamins, pp. 147-173.

Quah, Chiew Kin. (2006) *Translation and Technology*. New York: Palgrave Macmillan.

RIVERA-TRIGUEROS, Irene. (2022) "Machine translation systems and quality assessment: a systematic review." *Language Resources & Evaluation* 56, pp. 593-619. http://doi.org/10.1007/s10579-021-09537-5

SAG, Ivan *et alii*. (2002) "Multiword expressions: A pain in the neck for NLP." In: *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics*. Mexico D.F.: CICLING, pp. 1-15.

SÁNCHEZ CARNICER, Jaime. (2022). *Traducción y discapacidad. Un estudio comparado de la terminología inglés-español en la prensa escrita*. Berlin: Peter Lang.

SÁNCHEZ RAMOS, María del Mar & Celia Rico Pérez. (2020) *Traducción automática: conceptos clave, procesos de evaluación y técnicas de posedición*. Granada: Comares.

SEGHIRI, Miriam. (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, pp. 43-64. https://doi.org/10.1075/babel.63.1.04seg

SEVERAL AUTHORS (Microsoft). (n.d.) *Languages*. https://www.microsoft.com/en-us/translator/languages/

SEVERAL AUTHORS (Ministerio de Industria, Comercio and Turismo). (2023) *Datos de Frontur y Egatur del INE [Press release]*. Online version: http://www.mincotur.gob.es/es-es/GabinetePrensa/NotasPrensa/2023/Paginas/En-2022-visitaron-España-71,6-millones-de-turistas-internacionales-que-realizaron-un-gasto-de-87.061-millones-de-euros.aspx

SEVERAL AUTHORS (World tourism Organization). (2023) *World tourism barometer - May 2023 (excerpt)* 21:2. Online version: https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2023-05/UNWTO_Barom23_02_May_EXCERPT_final.pdf

SEVERAL AUTHORS (World Tourism Organization & Basque Culinary Center). (2019) *Guía para el desarrollo del turismo gastronómico*. Madrid: OMT. Online version: https://doi.org/10.18111/9789284420995

SEVERAL AUTHORS (Yandex). (n.d.) "Supported languages." *Yandex Support*. Online version: https://yandex.com/support/translate-mobile/supported-langs.html

SOTO ALMELA, Jorge. (2014) *Los términos culturales en el ámbito turístico español-inglés: traducción, manipulación y recepción real en usuarios anglófonos*.

Unpublished doctoral thesis. Murcia: Universidad de Murcia. Online version: http://hdl.handle.net/10201/41596

VIEIRA, Lucas Nunes. (2020) "Machine Translation in the News: A Framing Analysis of the Written Press." *Translation Spaces* 9:1, pp. 98-122. Online version: https://doi.org/10.1075/ts.00023.nun

YAMADA, Kenji & Kevin Knight. (2001) "A syntax-based statistical translation model." In: *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pp. 523-530.

ZENS, Richard; Franz Josef Och & Hermann Ney. (2002) "Phrase-based statistical machine translation." In: Jarke, Matthias; Gerhard Lakemeyer & Jana Koehler (eds.) 2002. *KI 2002: Advances in Artificial Intelligence 25th Annual German Conference on AI, KI 2002 Aachen, September 16-20, 2002 Proceedings*. New York: Springer, pp. 18-32.

BIONOTES

ISABEL PEÑUELAS GIL holds a Ph.D. in Translation from the University of Valladolid. At the Faculty of Translation and Interpreting of the same university, she taught for five years, training future translators in areas such as translation and interpreting technologies, and terminology. She is currently a researcher in training at the same institution. She is also a member of CITTAC, and her research activities focus on translation and interpreting technologies, machine translation, terminology, and corpus studies. Additionally, she has participated in various national and international conferences, presenting papers on her areas of specialization.

MARÍA TERESA ORTEGO ANTÓN is a Full Professor of Translation and Interpreting in the Spanish Language Department at the University of Valladolid (Spain). She is a member of the CITTAC research group at the same institution and the inter-university ACTRES group at the University of León. Additionally, she collaborates with the LEXYTRAD research group at the University of Málaga and the OLST group at the University of Montreal (Canada). She is responsible for courses on interpreting, agri-food translation, and translation technologies at the Faculty of Translation

and Interpreting at the University of Valladolid. Her research includes numerous journal articles, book chapters, and monographs, primarily exploring the intersection of translation and interpreting with technological advancements. She has presented her work at numerous international conferences and actively contributes to both national and international research projects on specialized translation and interpreting.