USING CORPORA IN SCIENTIFIC AND TECHNICAL TRANSLATION TRAINING: RESOURCES TO IDENTIFY CONVENTIONALITY AND PROMOTE CREATIVITY[1]

Clara Inés López Rodríguez*
Universidad de Granada

Abstract: Since the first Corpus Use and Learning to Translate (CULT) Conference in Bertinoro (Italy) in 1997, the usefulness of corpora for translators and trainee translators has been highlighted. From an initial approach where translators compiled ad hoc corpora in their hard drive for a subsequent study with lexical analysis software, there emerged a new trend towards the use of the Internet as corpus. In this second approach, the Web is perceived as a huge corpus which is accessed by means of online tools which produce monolingual wordlists and concordances from texts available from the Internet or pre-existing corpora, or by means of bilingual or multilingual concordancers displaying aligned texts from international institutions' parallel corpora. Bilingual concordancers and translation memories are widely used by translators and trainee translators because of the immediate translation solutions they offer, but these tools can restrain creativity by offering conventional solutions and eliminating layout and multimodal elements in texts. The aim of this article is to describe the exploitation of quality corpora in a scientific and technical translation course, focusing on

---

* Clara Inés López-Rodríguez: PhD in Translation and Interpreting (2001), and BA in English Language and Literature from the University of Granada (UGR). Diploma in English Studies (University of Portsmouth). Tenured professor (scientific, technical and multimedia translation) in the Department of Translation and Interpreting of the University of Granada. Granada, Spain. E-mail: clarailr@ugr.es

texts on health translated from English into Spanish, and on terminological variation as a reflection of creativity in language.

Keywords: Parallel corpora. Monolingual. Bilingual and multilingual concordancers. Sketch Engine. Conventionality. Creativity. Terminological variation.

EL USO DE CORPUS EN LA FORMACIÓN DE TRADUCTORES CIENTÍFICOS: RECURSOS PARA IDENTIFICAR CONVENCIONES LINGÜÍSTICAS Y PARA PROMOVER LA CREATIVIDAD

Resumen: Desde el primer congreso CULT (Corpus Use and Learning to Translate), celebrado en Bertinoro (Italia) en 1997, se ha destacado el potencial de los corpus en soporte electrónico para traductores y estudiantes de traducción. El traductor ha pasado de compilar corpus ad hoc en el disco duro de su ordenador para analizarlos con herramientas de análisis léxico, a utilizar Internet como un gran corpus. En este último enfoque el traductor interactúa con herramientas en línea que permiten generar listas de frecuencia o líneas de concordancia monolingües a partir de textos publicados en la Red o corpus existentes, o bien recursos que generan concordancias bilingües o multilingües a partir de corpus paralelos de instituciones internacionales. El uso extendido de concordancias bilingües y memorias de traducción entre traductores y estudiantes de traducción tiene la ventaja de la inmediatez, pero puede frenar la creatividad del traductor, al ofrecer soluciones de traducción bastante convencionales y suprimir elementos de formato y multimodales de los textos. En este artículo describiremos cómo trabajar con corpus en el aula de traducción científica, centrándonos sobre todo en textos médicos que se han traducido del inglés hacia el español, así como en el fenómeno de la variación denominativa, que es una manifestación de la creatividad del lenguaje humano.

Palabras clave: Corpus paralelos. Herramientas para generar concordancias monolingües, bilingües o multilingües. Sketch Engine. Normalización. Creatividad. Variación denominativa.

1. Introduction: conventionality and creativity in language and translation

Creativity is an important element in cognitive processes such as communication, problem-solving, and translation. Translation Studies have paid attention to creativity during the cognitive processes involved in translation (Kussmaul, 1991, 2000, Bayer-Hohenwarter 2013, Ruschiensky, 2015), and to its relevance in specialized translation (Tercedor *et al.*, 2009).

Creativity is responsible for some linguistic phenomena in monolingual or multilingual communication as indicated by Kenny (94): "Linguistic creativity can reside in the creation of new words, the novel use of existing words, the use of unconventional syntax or punctuation, the manipulation of typographical conventions, and so on". For instance, primary term formation (Sager, 1997)[2] is a creative operation in the source language to coin new terms for new concepts and ideas. Other linguistic phenomena such as doublespeak[3] and euphemization are also creative uses of language aimed at deceiving or minimizing harsh aspects of reality, respectively. A case in point is the expression "collateral damages" to refer to the deaths of civilians during a war. The introduction of new ways of describing events are the result of voluntary or involuntary manoeuvres that distort reality, as shown in Tymoczko's (2013) exploration of the role of ideology in the creative aspects of translation.

Creativity also underlies the use of untypical collocations, a tendency recognized by Mauranen (44): "translations […] display collocational patterns that deviate from those found in comparable target language originals". Although these combinations are possible in the target language, they are "rare or absent from actual target language texts" (Mauranen, 44). The use of untypical

---

[2] Sager (1997: 27) defines primary term formation as "the process of terminology creation that accompanies concept formation as a result of scientific and technological innovation or change in a linguistic community".

[3] Doublespeak is language that deliberately disguises, distorts, or reverses the meaning of words.

collocations can also be the result of interference from the source language. In any case, untypical collocations in translated texts contribute to the introduction of novel and creative expressions in the target language.

Much of language use is driven by creativity but also by patterns and conventions. In this sense, Hanks (85) affirms:

> The creative potential of language is undeniable, but the concordances to a corpus remind us forcibly that in most of our utterances we are creatures of habit, immensely predictable, rehearsing the same old platitudes and the same old clichés in almost everything we say. If it were not so, language would become unworkable.

Linguistic conventions associated to genres ease the production and comprehension of texts. Formulaic language, term standardization, and the repetition of terms in some genres such as patents, research articles and abstracts are a safeguard against ambiguity, polysemy and misunderstandings in specialized communication. However, as opposed to the prescriptive approach of the General Theory of Terminology (Wüster, 1979), the current trend in Terminology is to consider terminological variation as an inherent feature in specialized communication (Cabré, 1999, Temmerman, 2000, Fernández Silva and Kerremans, 2009, Tercedor, 2011, Faber 2012, *inter alia*).

The ability to see patterns and regularities both within a language and across languages, and to select those which are adequate and appropriate in a certain context is one of the basic principles for corpus use and learning to translate. According to Zanettin (109):

> ...translators need to be able to see patterns and regularities both within a language and across languages. To produce

adequate texts in the target language, a translator is faced
with the task of finding 'the right words' and the 'right way'
to put things for a certain type of context.

Translators look for patterns and, at the same time, struggle
with words, and stretch language to its limits in many translation
processes requiring a considerable amount of creativity: the
translation of neology in scientific texts, the translation of poetry,
the transcreation of names of characters and places in videogames,
etc. Translators and human beings in general move along this
continuum of creativity and conventionality everyday, something
that can be observed by means of corpora.

Creativity is... concerned with spontaneous, imaginative use of
language, while conventionality refers to the idea, so powerful in
the British linguistic tradition... that much language use is routine.
Electronic corpora have made such phenomena increasingly
observable (Stewart, 74-75)

Conventionality and creativity have drawn the attention of
corpus linguists (Hanks, 1996, Kenny, 1998, 2000, Stewart, 1998,
inter alia) because both conventionality and creativity are closely
related in language: "Routine is not such a bad thing [...] It is
what allows the creative use of language to be identified as such"
(Kenny, 515). Corpus linguistics can facilitate the identification
of patterns in texts (pointing to frequent and standard uses of
language), as well as "rare" instances which may indicate creative
uses of language. These "rare" examples have been termed *hapax
legomena* (i.e., words that appear only once in a corpus).

Considering that both conventionality and creativity are key
in specialized translation, and that corpora can shed light on the
textual product of these processes, the aim of this paper is to
describe how to exploit corpora in the Scientific and Technical
classroom, and to study their usefulness in exploring conventional
and creative solutions to translation tasks. We draw on our
experience in the research and development projects *VariMed*

and *CombiMed*,[4] which focus on the cognitive and communicative motivation for term variation in medicine, and the teaching innovation actions *Communication and European Citizenship (English-Spanish): multimodal resources on Health and the Environment* [http://www.ugr.es/~clarailr/comunicadania/] and *Tradusaluda: audiovisual resources for the promotion of health in Europe: accessible subtitling and translation* [http://tradusaluda. wordpress.com].

2. A brief history of translator trainee's use of corpora

Corpus Use and Learning to Translate (CULT) has experienced a fruitful evolution since its beginnings in the 1990s. Initially, trainee translators searched ready-made corpora (Brown Corpus, British National Corpus, etc.), and they compiled ad hoc corpora in their hard drive for a subsequent analysis with lexical analysis software such as Monoconcord or Wordsmith Tools. These ad hoc corpora were called DIY corpus by Zanettin (2002, 2012) or disposable corpora (Varantola, 2003). One application of DIY corpora to translator training was López-Rodríguez (2001), who described some corpus tasks in which students compiled newspaper comparable corpora in English and Spanish as a way of keeping up with current affairs, and gaining familiarity with cultural references and typographical conventions of newspaper style guides. The teacher suggested search structures to retrieve in English texts some translation problems not solved in bilingual dictionaries, for example, the translation of forms of address (Mr, Lord, Ms...), quotations and verbs of speech (Figure 1). These suggestions led students to analyze instances of cultural references, text conventions and translation strategies in those DIY corpora.

---

[4] Some of the results of those projects are published in an online medical resource called VariMed [http://varimed.ugr.es/].

Figure 1: Search structures to retrieve forms of address and quotations in a DIY corpus (adapted from López-Rodríguez 219-220)[5]

SOURCE TEXT: • Mr, Mrs, Ms, Lord, Dr, QC, prof • ,"		TARGET TEXT: • Lord, Dr, prof, D. • ",
1	g of whole people will also remain banned. Lord Sainsbury, the Science Minister with re	
2	después de que el responsable de Ciencia, Lord Sainsbury, admitiera en una entrevista	
3	ña después de que el ministro de Ciencias, Lord Salisbury, admitiera en una entrevista	
4	ufrimiento entre la población civil. Ayer, lord Judd, el jefe británico de una delegaci	
5	Russian military," said delegation leader Lord Judd. Western nations have expressed c	
1	ious if this will be Al Gore's last stand as well," said Bush spokesman Dan Bartlett.	
2	Nd well-being of all those who have been detained," Mr Annan said. The corpses believ	
3	lice and the ambulance crews are dazed with shock," he said. Confirming that seven pe	
4	break up the crowd and it actually was tear gas," Mr Ramsey said. "It did disperse	
5	lly have become more violent since the ceasefires," Jim McDaid told the National Ass	
1	elo en los 7.000 años próximos puede ser inevitable", afirma el profesor Howard Conw	
2	se halla sólo a 75 kilómetros". "Eso está muy lejos", replicó el embajador. Era la úl	
3	artes no pueden llegar a un acuerdo en esta ocasión", decía un escueto comunicado del	
4	le una respuesta directa a menos que me la pregunte", añadió. "Utilizo palabras como	
5	condiciones de llegar a un acuerdo en este momento", ha asegurado Crowley.	
6	S no podemos decir a los Estados cómo deben hacerlo", advirtieron desde la Comisión.	
7	Imas semanas vamos a trabajar en ponerlo todo junto", explicó. "Esto se supone que pe	
8	Opeo basta para convertirle a uno en un anti alemán", argumentó Chevènement. El ex pr	
9	Fein responsabiliza a Mandelson de la "peor crisis", según palabras de Adams, que at	
10	estará al frente. "Va a haber un cambio de guardia", manifiestan las fuentes consult	
11	"No creemos que vaya a salir algo nuevo de allí", admitieron las fuentes, "en est	
12	tarán inferiores al objetivo establecido por el BCE", destacó el portavoz de Bruselas	
13	ra siempre en Alemania. Me veo viviendo en Londres", comentó el director ejecutivo d	
14	n figuras internacionales extremadamente respetadas", defendió el ministro británico	
15	republicana y sus autores están hoy en el Gobierno", señaló David Jones, portavoz de	
16	tuguesas islas Azores, "no se tomó ninguna decisión", puntualizó el jefe de la diplom	
17	abría problemas. Contra los niños no tenemos nada", manifiesta Meimaj. Doce meses d	
18	ast. "El Acuerdo de Viernes Santo está hecho trizas", denunció el presidente del Sinn	
19	jico a manejar una silla de ruedas, si no a caminar", declaró Merger, según una cita	
20	iles. "Aquí todo el mundo viene a ver a esta gente", se queja el jefe de MSF, "es	
21	e la cumbre: "Se han logrado avances significativos", insistía Clinton con tono res	
22	tuguesas islas Azores, "no se tomó ninguna decisión", puntualizó el jefe de la diplom	
23	abría problemas. Contra los niños no tenemos nada", manifiesta Meimaj. Doce meses d	
24	ast. "El Acuerdo de Viernes Santo está hecho trizas", denunció el presidente del Sinn	
25	jico a manejar una silla de ruedas, si no a caminar", declaró Merger, según una cita	
26	iles. "Aquí todo el mundo viene a ver a esta gente", se queja el jefe de MSF, "es	
27	e la cumbre: "Se han logrado avances significativos", insistía Clinton con tono res	

[5] Source: Author's personal archives (permission to use granted by author)

Apart from DIY corpora, other studies on Corpus Use and Learning to Translate, and by extension to Teaching and Language Corpora (TALC), have focused on the use of multilingual parallel corpora (Barlow, 2002, Ruiz Yepes 2012), and of learner corpora (López-Rodríguez, Robinson and Tercedor, 2007, López-Rodríguez and Tercedor, 2008, Kübler, 2008, Castagnoli *et al.*, 2011, *inter alia*). Learner corpora were the basis for projects such as MeLLANGE (Multilingual eLearning in LANGuage Engineering).[6] MeLLANGE created corpus-based eLearning teaching content for translation by building a Learner Translator Corpus (LTC) with an annotation scheme for errors. López-Rodríguez, Robinson & Tercedor (2007) proposed tagging activities involving the recognition of potential translation problems, the classification of mistakes/errors in students' translation assignments, their evaluation according to adequacy/appropriateness, and the selection of the best translation option.

The new millennium brought about a new trend towards the use of the Internet as a corpus. Web crawling was a milestone since it enabled initiatives such as WebCorp (Renouf, 2003), WacKy corpora (Baroni and Bernardini 2004a, Bernardini, Baroni and Evert, 2006, Baroni *et al.*, 2009) and WebBootCaT (Baroni and Bernardini, 2004b). In this approach the Web is perceived as a huge corpus which is accessed by means of online tools. These tools produce monolingual wordlists and concordances from either texts available from the Internet (e.g. WebCorp or Corpus of Global Web-Based English) or pre-existing corpora accessed by means of the Sketch Engine corpus query system (Lexical computing Ltd., n.d.).

These technical advances and the fact that building a corpus from scratch is time-consuming have changed the way corpora are exploited in the translation classroom. Trainee and professional translators alike are reluctant to spend time manually compiling a corpus when many corpora and bi/multilingual tools based on corpora are readily available online (see section 3), and when

---

[6] http://mellange.eila.jussieu.fr/

translation memories (both offline and cloud-based) can provide an instant equivalent.

In spite of the availability of online corpora, there are not many translators who search in monolingual corpora and corpora preloaded in systems such as Sketch Engine (Buendía and López-Rodríguez, 2013) or who build ad hoc corpora for specific translation assignments or terminology purposes using BootCaT technology (Baroni & Bernardini 2004b, Baroni *et al.* 2006). BootCaT technology was initially developed by the University of Bologna, and is present in their BootCaT toolkit [http://bootcat.sslmit.unibo.it/] as well as in Sketch Engine's WebBootCat. Gatto (140) summarizes the usefulness and relevance of BootCaT as follows:

With BootCaT, the process of corpus compilation is, in principle, the same as in the manual creation of DIY corpora, but rather than having the linguist first querying the web, then downloading relevant results, and finally performing the necessary format changes and archiving procedures to create his/her corpus, it is the software that performs most of these tasks together in a few minutes. […] This is a very interesting feature from the point of view of its contribution to the changing face of corpus linguistics. By making the creation of ad hoc temporary corpora and easily achievable goal, BootCaT really brings the ideal notion of the web as a sort of virtual multilingual multipurpose corpus on demand a bit closer to reality.

Bowker and Barlow (2008) show the strengths and limitations of bilingual concordancers and translation memory systems, and conclude that corpora provide better translation solutions. The equivalents suggested by translation memories normally contain "oversimplified syntax, monotonous rhythm, and a lack of diversity" (Bowker, 180). In a similar vein to Bowker & Barlow's suggestion that new tools should combine the best features of both bilingual concordancers and translation memories, in this paper we present different resources in order to take advantage of the best of both worlds, and to display conventional and non-conventional translation solutions that are appropriate for the target culture. We claim that corpora, together with other multimedia resources

(photographs, audio files, videos, etc.), can boost creativity in translation (Tercedor *et al.*, 2009). These resources have been used in a course on Scientific and Technical Translation in the Faculty of Translation and Interpreting of the University of Granada (Spain), and are described in the following sections.

3. Exploiting on-line bilingual concordancers and resources based on translation memories in Scientific and Technical Translation

Translators frequently use on-line bilingual and multilingual tools based on parallel corpora or translation memories because of the immediate translation solutions they offer, although these resources have some drawbacks that will be described later on. Linguee is the most widely used bilingual concordancer and translation help tool amongst translators. Linguee combines a dictionary and a search engine in order to retrieve words and expressions in bilingual texts crawled from bilingual websites, EU documents and patent specifications. Other resources also based on bilingual "big data" from the Internet are gaining ground, and they offer solutions in different language combinations (Table 1).

Table 1: On-line bilingual concordancers

Linguee: http://www.linguee.es/
Reverso Dictionaries (Context): http://context.reverso.net/traduccion/
WeBiText: http://www.webitext.com/
YouAlign: http://www.youalign.com/AlignDocs.aspx

Reverso Context, for instance, is a contextual dictionary providing bilingual examples taken from previously translated texts; the concordancer is linked to online dictionaries (monolingual, bilingual, and thesaurus) and tools to assist users with conjugations, grammar, and spelling.

Both YouAlign and WeBiText are based on an alignment tool developed by Terminotix Inc., a company specializing in computer-aided translation. YouAlign creates online bitexts from the source and target texts which users upload in the platform; the system can also generate files in order to feed translation memories, websites and terminology and phraseology search engines. WeBiText is a bilingual concordancer that retrieves translations of words and expressions in pre-defined websites used as corpora, and it allows users to search their own sites. Although some of the pre-defined sites have been pre-processed to speed up the response, the system can be slow at times.

Another interesting tool for translators and editors is Term-minator, which allows the use of the web as a gate to search words in the WWW in the context of dictionaries, encyclopaedias, translations, multimedia, etc. [http://www.term-minator.eu].

Moreover, there are bilingual and multilingual tools based on European Union documents (Table 2) that can display aligned texts from EU parallel corpora. For example, the Multilingual Display of EUR-Lex can simultaneously show legal and administrative documents of the European Union in several EU official languages. With the help of these tools, different translation tasks can be carried out. Based on the bilingual display of EUR-Lex, the teaching innovation action *Communication and European Citizenship (English-Spanish): multimodal resources on Health and the Environment* [http://www.ugr.es/~clarailr/comunicadania/] includes some activities in which students extract EU terminology, elaborate glossaries, and translate.[7]

---

[7] The project proposes some tasks on Directive 93/42/EEC concerning medical devices (task 1, available from: http://www.ugr.es/~clarailr/comunicadania/media/Salud_2.pdf), Directive 2010/75/EU on industrial emissions (integrated pollution prevention and control), and on Directive 98/8/EC concerning the placing of biocidal products on the market (tasks 1 and 2, available from: http://www.ugr.es/~clarailr/comunicadania/media/Medio_9.pdf).
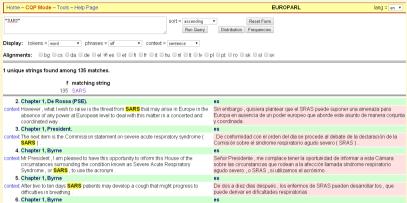
Table 2: On-line bilingual concordancers based on European Union texts

Multilingual display of European Union Legislation of EUR-Lex:
http://eur-lex.europa.eu/content/guided-tour/multilingual-display.html
OPUS: the open parallel corpus: http://opus.lingfil.uu.se/
Sketch Engine Parallel corpora (DGT, Spanish; EUROPARL7, es;
OPUS2 Spanish, etc.): http://www.sketchengine.co.uk/

In OPUS, compiled by Jörg Tiedemann, users can search in one or several corpora through its multilingual search interface, and they can also download translation memories (TM)[8] in TMX (Translation Memory eXchange) format. Figure 2 shows a query to look for potential translations for the English acronym SARS (Severe Acute Respiratory Syndrome) in the EUROPARL corpus, which contains proceedings of the European Parliament. The results retrieved indicate a tendency towards interference and the use of cognates: "síndrome respiratorio agudo severo" is used instead of the more idiomatic "síndrome respiratorio agudo grave", whose acronyms in Spanish texts are SRAS and SARS.

---

[8] Some downloadable TM for Specialized Translation are: a) DGT: A collection of translation memories provided by the Joint Research Centre; b) EMEA; c) European Parliament Proceedings; d) European Constitution.

Figure 2: Searching SARS in OPUS: the open parallel corpus[9]

OPUS also provides translation memories under the "Search & Download resources" menu of its website. The TM available for English and Spanish are shown in Figure 3.

Figure 3: Translation memories available for English and Spanish in OPUS: the open parallel corpus[10]

It is possible to download and export these translation memories to CAT tools such as Trados Studio. Although it takes some time to download DGT[11] —a collection of translation memories provided by the Joint Research Centre of the Commission's Directorate General for Translation—, and to convert the TMX format to the native format of Trados Studio, SDLXLIFF, the quality and richness of this resource outweighs this inconvenience. In terms of term accuracy, the concordancer implemented in Trados Studio provides a conventional and appropriate translation into European Spanish for the term "medical device" on the basis of the DGT-Translation Memory: "producto sanitario". Oddly enough, Trados Studio pre-translates the segment "Medical devices" as an identical (100%) match which is not very precise, "Aparatos médicos", ignoring the results of the concordancer (Figure 4). In any case, "medical device" is often translated into Spanish as "dispositivo médico" (Mexico, Chile, and Venezuela) or "aparato médico", instead of the normalized term "producto sanitario", which is the one used by experts in medical and pharmaceutical texts, and dictionaries in Europe, and therefore, it is considered more appropriate for a specialized context.

---

[11] DGT-Translation Memory (DGT-TM) is a Translation Memory of the Acquis Communautaire, i.e. the body of European legislation, including all the treaties, regulations and directives adopted by the European Union (EU) and the rulings of the European Court of Justice [https://ec.europa.eu/jrc/en/language-technologies].

Figure 4: Using the DGT-TM in Trados Studio[12]

The Sketch Engine corpus query system allows access to EU institutional texts and their translations into several languages. In the case of Spanish, translators can search expressions in the following corpora: DGT Spanish, EUROPARL7 Spanish or OPUS2 Spanish. The latter includes texts from the European Medicines Agency, and can be a good resource for the translation of patient information leaflets and summaries of product characteristics as illustrated in Figure 5. Again, conventionality and cognates *(creatinine clearance / aclaramiento de la creatinina)* prevail over the use of more idiomatic options such as "depuración de la creatinina" or "capacidad de eliminación de la creatinina". Similar examples of conventionality and normalization are found in cognates such as *rare diseases / enfermedades raras* or *evidence-based medicine / medicina basada en la evidencia,* which are appropriate terms in Spanish.

---

[12] Source: Author's personal archives (permission to use granted by author)

Figure 5: Bilingual concordance around "creatinine clearance" generated by Sketch Engine from the OPUS2 parallel corpus[13]

3.1 Drawbacks of on-line bilingual concordancers and resources based on translation memories

On-line bilingual concordancers and resources based on translation memories are not always accurate from the point of view of terminology, and the language they include does not always sound natural to a native speaker. This applies mainly to widely-spoken languages encompassing different dialectal varieties. The examples provided in the second column of concordancers are normally translations, and even some of them are in fact post-edited machine translations. For instance, the results for the English into Spanish search "cardiopulmonary resuscitation" are "reanimación cardiopulmonar", "resucitación cardiopulmonar", "CPR" or "resucitación cardio pulmonar". Health professionals in Spain would use "reanimación cardiopulmonar" and other options not retrieved by these bilingual systems: "reanimación cardíaca", "reanimación cardiorrespiratoria", "RCP" or "RCR". On the

---

[13] Source: Author's personal archives (permission to use granted by author)

bright side, the most frequent result in bilingual concordancers is "reanimación cardiopulmonar".

Consequently, the results of these on-line resources should be critically analyzed and contrasted with other resources such as dictionaries, monolingual corpora, and multimodal resources. On-line bilingual concordancers do not provide conceptual information about terms, and eliminate layout and multimodal elements originally included in texts. Besides, sentence alignment is not always correct.

Another drawback of bilingual and multilingual concordancers is that they can restrain creativity by offering conventional solutions based on cognates that do not always sound natural in the target language. For instance, the cognate "estetoscopio" (stethoscope) is the only option retrieved by bilingual resources while in Spain, Chile and Colombia the preferred term would be "fonendoscopio". When a resource only provides a cognate to the detriment of other idiomatic solutions, the documentation process of translators with bilingual concordancers may lead to over-standardization in language in the long-term: "translators may actually over-standardize, creating consistency in places where the use of variants was deliberate and well reasoned." (Bowker and Hawkins, 2006 cited in Fernández Silva and Kerremans, 2009). Language and specialized discourse are dynamic, and include different words to highlight a special facet of the concept (cognitive motivation) or due to geographical, social, historical or communicative reasons. In this sense over-standardization does not reflect the dynamic nature of language, and the existence of denominative variation (Alarcón, López-Rodríguez, and Tercedor, 2016). Therefore, other tools and strategies are needed so that students can develop their 'use of corpora to translate' competence (Rodríguez-Inés 2009).

4. Exploiting monolingual resources (wordlists and concordancers) and Sketch Engine in the Scientific and Technical classroom

This section proposes some monolingual wordlists and concordancers that can complement the resources described above, and details the procedure that guides students in the process of compiling and analyzing corpora with Sketch Engine. The ultimate goal is that they develop documentation and translation competence, and that they are able to recognize patterns and instances of creativity in both source and target texts.

4.1 Monolingual wordlists and concordances

In the same way as trainee translators are told to look up words in monolingual dictionaries, they should receive training in the use of monolingual resources based on corpora, which enable the recognition of patterns and unusual words and collocations by means of wordlists and concordances. Some of these resources are valid for different languages (Table 3) and are based on a Web As Corpus approach.

Table 3: Resources for monolingual queries in different languages

Sketch Engine: http://www.sketchengine.co.uk/ (EnTenTen, EsTenTen, etc.)
WebCorp: http://www.webcorp.org.uk
Querying Internet Corpora (Centre for Translation Studies, Leeds)
http://corpus.leeds.ac.uk/internet.html
Aranea Project No Sketch Engine: http://ucts.uniba.sk

A description of Sketch Engine is provided in sections 3 and 4.2. WebCorp is particularly interesting because concordances are associated with a link to the text where they appear, and it is

possible to generate a wordlist for each text. Serge Sharoff from the Centre for Translation Studies in Leeds has compiled Internet monolingual corpora in 14 languages, and has developed a query system (Querying Internet Corpora) to retrieve concordances and collocations; frequency lists for some of these corpora are also provided. Finally, *Aranea. A Family of Comparable Gigaword Web Corpora*, prepared by Vladimír Benko, allows guest access to an interface quite similar to Sketch Engine.

Some monolingual corpora only available in English are shown in Table 4. Most of them represent general language, with the exception of Exemplar, which displays scientific terms in the context of academic journals published by Springer. Monolingual corpora in Spanish are listed in Table 5.

Table 4: Resources for monolingual queries in English

Exemplar Words in context: http://www.springerexemplar.com/
BNC - simple searches online: http://www.natcorp.ox.ac.uk/
Corpora compiled by Mark Davies (Brigham Young University) http://corpus.byu.edu/
Corpus of Global Web-Based English (GloWbE): http://corpus2.byu.edu/glowbe/

Table 5: Resources for monolingual queries in Spanish

CREA (Corpus de Referencia del Español Actual): http://corpus.rae.es/creanet.html
Corpes XXI (Corpus del Español del siglo XXI):
http://www.rae.es/recursos/banco-de-datos/corpes-xxi
Corpus del Español (Mark Davies): http://www.corpusdelespanol.org/

The Spanish Royal Academy provides free access to two reference corpora which include oral and written texts produced in

different Spanish-speaking countries: a) a corpus of Spanish dating from 1975 to 2004 (CREA corpus), and b) a corpus of contemporary Spanish dating from 2001 onwards (CORPES XXI). Mark Davies has compiled a diachronic corpus of Spanish of 100 million words whose contemporary section will expand to two billion words.

4.2 The use of Sketch Engine

A methodology for teaching students to use Sketch Engine in a course of Scientific and Technical translation was suggested in Buendía and López (2013). In this study, an experiment was designed to discover whether the use of automatically compiled corpora, more specifically, WaCky corpora and WebBootCat accessed through the Sketch Engine interface offered more advantages to translation students than the use of paper and on-line lexicographic and terminographic resources, and the selection of parallel texts from the web without using lexical analysis tools. The answers to questions 20 to 37 indicated that the majority of students agreed on the usefulness of Sketch Engine for translation, and the reliability of texts offered by the system. For instance, 81% of students insisted on the fact that they would use Sketch Engine in translation tasks under exam conditions because this tool enabled them to rapidly retrieve and compile a specialized corpus of more reliable texts than the ones offered by Google. They also mentioned that it allowed them to analyze texts more easily through the use of concordances, and to acquire expert knowledge on the subject field. All of them agreed that the best way to make the most of Sketch Engine was to use it in combination with Google and quality dictionaries (64-67).

In our course in Scientific and Technical Translation of the Degree in Translation and Interpreting of the University of Granada (Spain), students carry out a translation and terminology project based on Sketch Engine (SkE). SkE provides a 30-day free license, extensive documentation, and has recently included some video

tutorials [https://www.sketchengine.co.uk/sketch-engine-video-tutorials/]. Once students have been introduced to the program and its main functions (Word List, Concordance, WordSketch, Thesaurus, Sketch-Diff, and WebBootCat to compile DIY corpora) in a two-hour session, they have to follow these stages:

1. Compile with WebBootCat both an English and a Spanish corpus on one health topic. Altogether the corpora should contain around 400,000 tokens.
2. Before selecting the "seed terms" (i.e. relevant terms within the specialized domain which will be used by the system to crawl the Web in search for relevant texts), choose 4 reliable texts for each language and identify keywords. Save these texts and their reference in your computer.
3. Reflect upon what "seeds" will provide more suitable results.
4. Introduce your seeds in English (English corpus). Remember that multiword lexical units should go between quotation marks: "cost of living". Use spaces between "seeds".
5. Follow the steps indicated by Sketch Engine and accept the default options. Unmark the check boxes of inappropriate texts.
6. Once the corpus has been compiled, search for terminology and interesting linguistic phenomena (frequent collocations, unusual patterns, polysemy...) using the potentialities of Sketch Engine (concordances, word-sketches, thesaurus…)
7. Save the compiled corpus (Corpus - Download corpus-Plain text).
8. Prepare a presentation highlighting relevant findings. This presentation should be complemented with the following material:
   - Preliminary texts used to extract seeds (pdf/html) and spreadsheet with a bibliographic reference.
   - Comparable corpora compiled with Sketch Engine (around 400,000 tokens).
   - An academic report including most frequent terms,

frequent collocates, related terms, keywords, most frequent verbs, interesting findings, or insights on how to exploit your corpus.
- A bilingual glossary in Excel with 20 terms.
- References.

Furthermore, the use of Sketch Engine can shed light on linguistic patterns and regularities. Figure 6 represents the wordsketch for CPR *(cardiopulmonary resuscitation)*, which highlights the verbal structures into which this acronym fits *(perform/administer/learn/ begin/teach CPR)* and some frequent collocations such as *hands-only/bystander/infant CPR*.

Figure 6: Wordsketch for the acronym CPR (cardiopulmonary resuscitation)[14]

The analysis with Sketch Engine can also point to priority areas of certain institutions and organizations as indicated in the keywords and thesauri generated from corpora based on their texts. For example, based on a corpus of nearly 1.5 million tokens including

---

[14] Source: Author's personal archives (permission to use granted by author)

texts on Health from the European Union Directorate General of Health and Food Safety (section Public Health), a list of keywords and terms can indicate their main concerns in relation to health (Figure 7): diabetes, mental diseases, cancer, HPV, reproductive and sexual health, HIV or obesity.

Figure 7: Keywords and terms from a corpus of texts published by the Directorate General of Health and Food Safety[15]

Corpus size	1,435,681 tokens
Number of texts	75
Medium	Written texts
Subject	Health
Text types	Reports, leaflets, legislation, FAQs, Commission programmes (e.g. Health for growth prog. (2014-2020))…
Authorship	Varied and declared
Language	English
Publication date	2000-2015

Corpus *EU Health* - stat

Counts	
Tokens	1,435,681
Words	1,040,183
Sentences	34,016
Paragraphs	24
Documents	75

---

[15] Source: Author's personal archives (permission to use granted by author)

Keywords		Score	F	RefF	Terms		Score	F	RefF
☐ diabetes	W	1,302.48	3,679	72	☐ cancer screening	W	200.39	435	6,705
☐ dementia	W	736.20	1,155	7	☐ epidemiological report	W	200.01	286	16
☐ perinatal	W	341.13	541	8	☐ marketing authorisation	W	171.13	259	785
☐ cancer	W	322.50	2,703	360	☐ perinatal health	W	123.94	179	191
☐ alzheimer	W	312.29	465	3	☐ national diabetes	W	103.85	149	119
☐ hpv	W	279.11	437	7	☐ annual epidemiological report	W	102.00	145	0
☐ births	W	227.46	593	61	☐ cervical cancer	W	97.98	337	18,241
☐ colorectal	W	212.05	303	0	☐ medicinal product	W	95.39	144	812
☐ neonatal	W	202.24	367	20	☐ diabetes prevalence	W	92.25	133	199
☐ fig	W	200.58	879	153	☐ gestational age	W	91.27	152	2,223
☐ eurocat	W	185.58	265	0	☐ cross-border healthcare	W	84.44	121	134
☐ deaths	W	177.85	817	164	☐ screening programme	W	84.24	127	815
☐ cancers	W	174.54	266	5	☐ prevalence rate	W	83.51	132	1,469
☐ autism	W	174.37	259	3	☐ colorectal cancer	W	81.35	222	11,847
☐ diagnoses	W	173.38	291	13	☐ patient registry	W	78.90	114	248
☐ cervical	W	170.13	545	92	☐ injecting drug	W	76.14	121	1,561
☐ hiv	W	164.79	881	203	☐ neonatal mortality	W	75.74	111	452
☐ gestational	W	163.06	261	9	☐ case definition	W	75.14	110	432
☐ congenital	W	160.93	357	41	☐ marketing authorisation holder	W	71.83	102	48
☐ fetal	W	158.97	236	3	☐ authorisation holder	W	71.47	102	109
☐ cp	W	156.05	1,268	347	☐ cervical cancer screening	W	71.32	107	772
☐ obesity	W	146.71	280	25	☐ birth weight	W	67.41	145	6,660
☐ emf	W	133.81	263	28	☐ colorectal cancer screening	W	67.16	100	684
☐ morbidity	W	119.75	330	69	☐ injecting drug use	W	64.90	95	461
☐ ecdc	W	115.74	415	112	☐ p p	W	63.19	92	391

5. Other resources to promote creativity in Scientific and Technical Translation

Since corpora focus on the linguistic context of words and on conventional language to the detriment of other creative and multimodal elements, which are key to communication, translators should check and contrast the results retrieved with corpora with other resources. They should inquire into the causes of denominative variation in monolingual lexicographical and terminographical

resources (both paper and on-line) such as those endorsed by professional associations or those which include visual information.

VariMed [http://varimed.ugr.es/] is a visual terminological resource on medicine which focuses on the cognitive and communicative motivations of denominative variation, and their reflection in the terminological variants of a concept. As opposed to the conventional translation solutions offered by bilingual corpora and translation memories (i.e. "cáncer cervical" is the only Spanish translation for "cervical cancer"), VariMed includes other terms frequently used by experts and laypeople: *cancer of the cervix, malignant neoplasm of cervix uteri* (in English)*,* and *cáncer cervico-uterino, cáncer de cérvix, cáncer de cuello uterino, carcinoma de cérvix, neoplasia del cuello uterino* (in Spanish). The inclusion of images in VariMed also facilitates the understanding of the cognitive and metaphorical motivation behind the coining of some terms. For example, the entry for *stress-induced cardiomyopathy* —a heart condition in which there is a temporary weakening of the left ventricle, usually as the result of severe emotional or physical stress— includes an image and several denominations for this concept in English and Spanish. The image represents the shape of the heart when this syndrome occurs, and can explain the motivation for: a) the terms *ampulla cardiomyopathy / miocardiopatía de la ámpula* because *ampulla* is the Latin word for a vessel used to hold oil or wine, and b) the terms *Tako-Tsubo syndrome, Takotsubo's cardiomyopathy / síndrome de takotsubo, miocardiopatía de takotsubo,* based on the Japanese name for an octopus trap.

Figure 8: Denominative variation in VariMed, an online medical resource [http://varimed.ugr.es][16]

imagen (subido por cristina)
especialización **baja** representatividad **media** semejanza **abstracta**

- ampulla cardiomyopathy (en)
- stress-induced cardiomyopathy (en)
- Tako-Tsubo syndrome (en)
- Takotsubo's cardiomyopathy (en)
- cardiomiopatía de tako-tsubo (es)
- cardiomiopatía por estrés (es)
- discinesia apical transitoria (es)
- disfunción ventricular transitoria (es)
- enfermedad de tako-tsubo (es)
- miocardiopatía de la ámpula (es)
- miocardiopatía de takotsubo (es)
- síndrome de tako-tsubo (es)
- síndrome de takotsubo (es)
- síndrome del corazón roto (es)

Students can also consult other documentation resources such as EuroVoc and EcoLexicon in order to grasp the conceptual background of terms and to facilitate knowledge representation. Eurovoc [http://eurovoc.europa.eu/] is a multidisciplinary thesaurus in 24 languages covering the activities of the European Union, created as an indexing tool. EcoLexicon [http://ecolexicon. ugr.es/en/index.htm] represents the conceptual structure of the specialized domain of the Environment in the form of a visual thesaurus which is enriched through multimodal resources such as definitions, concordances, contexts and other linguistic resources, and the depiction of concepts provided by images, animations, etc.

Finally, audiovisual materials published in multimedia channels such as Youtube, MOOCs, TED conferences and the like can also trigger creativity since they present information through different

---

[16] Source: Author's personal archives (permission to use granted by author)

channels and semiotic signs, thus contributing to the representation and acquisition of specialized knowledge (Tercedor, López-Rodríguez & Prieto 2014, López-Rodríguez & Tercedor *in press*). For example, the project *TRADUSALUDA, Audiovisual resources for the promotion of Health in Europe: accessible subtitling and translation* [https://tradusaluda.wordpress.com/] illustrates how images, videos and accessible subtitling can contribute to the representation, comprehension and acquisition of medical knowledge. The interplay of several channels of information enables understanding specialized concepts from different points of view, and going beyond the single perspective which is lexicalized in standardized terms.

6. Conclusions

Both translation memories and online bilingual and multilingual corpora are usual tools amongst trainee and professional translators because immediateness is more important than quality in certain translation assignments. Corpus analysis in the Scientific and Technical Translation classroom seems to indicate that denominative variation and creativity processes are tamed by conventionality in translation workflows where the use of translation memories and set inventories of terms coined by international institutions are common practice. As opposed to the standardized and narrow solutions offered by translation memories and bilingual resources, we have claimed that quality corpora (either ready-made or automatically compiled), if properly exploited, can foster creativity and help to identify better strategies for an appropriate translation solution. Obviously, the solutions when translating popular science or science adverts will require a higher degree of creativity and terminological variation than the translation of institutional texts or conventional genres such as patents, summaries of product characteristics, medical abstracts, etc. We have also proposed an inventory of useful corpora resources for translators.

In any case, resources should be used with a critical eye, and adapted to context. The exploration of conventional and creative translations for specialized terms can help trainee translators to discover those contexts in which translation options are limited for the sake of univocity and genre constraints, and those in which terminological variation is the result of creative processes and different ways of viewing and understanding specialized concepts. Moreover, in a context of multimodal texts and technologies in which the verbal component is just one amongst other sensory channels, the use of corpora and translation memories should be combined with other multimodal resources to activate creativity, improve comprehension and perceive the complexity and nuances of scientific terminology.

References

Alarcon-Navío, E., López-Rodríguez, C.I. & Tercedor-Sánchez, M. "Variation dénominative et familiarité en tant que source d'incertitude en traduction médicale." *Meta* 61.1 (2016).

Barlow, M. "ParaConc: Concordance software for multilingual parallel corpora." *Proceedings of the Third International Conference on Language Resources and Evaluation. Workshop on Language Resources in Translation Work and Research,* (2002): 20-24.

Baroni, M., Bernardini, S., Ferraresi, A. & Zanchetta, E. "The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora." *Journal of Language Resources and Evaluation* 43.3 (2009): 209-226.

Bayer-Hohenwarter, G. "Triangulating translational creativity scores: A methodological study in translation process research." *Tracks and Treks in Translation Studies: Selected papers from the EST Congress*. C. Way, S. Vandepitte, R. Meylaerts & M. Bartłomiejczyk (eds.), Leuven 2010. Amsterdam/ Philadelphia: John Benjamins, (2013): 63-85.

Baroni, M. & Bernardini S., eds. *WaCky! working papers on the web as corpus*. Bologna: GEDIT, 2004a.

_____. "BootCaT: Bootstrapping corpora and terms from the web." *Proceedings of LREC*, Lisbon (Portugal), 2004b. (http://sslmit.unibo.it/∼baroni/publications/lrec2004/bootcat_lrec_2004.pdf).

Baroni, M., Kilgarriff, A., Pomikalek, J. & Rychly, P. "WebBootCaT: a web tool for instant corpora." *Proceedings of Euralex 2006,* 123-132. Alessandria: Edizioni dell'Orso (2006): 123-132.

Bernardini, S., Baroni M. & Evert, E. "A WaCky introduction." In: M. Baroni & S. Bernardini (eds.). *WaCky! working papers on the web as corpus,* 1-32. Bologna: GEDIT (2006): 1-32.

Bowker, L. "Translation memory and 'text'." In: L. Bowker (Ed.). *Lexicography, Terminology, and Translation. Text-based studies in honour of Ingrid Meyer*. Ottawa : University of Ottawa Press (2006): 175-187.

Bowker, L. & Barlow, M. "A comparative evaluation of bilingual concordancers and translation memory systems." *Topics in Language Resources for Translation and Localisation*. E. Yuste Rodrigo (Ed.), Amsterdam/Philadelphia: John Benjamins (2008): 1-22.

Bowker, L. & Hawkins, S. "Variation in the Organization of Medical Terms: Exploring Some Motivations for Term Choice." *Terminology* 12.1 (2006): 79-110.

Buendía, M. & López-Rodríguez, C.I. "The Web for Corpus and the Web as Corpus in Translator Training." *New Voices in Translation Studies* 10 (2013): 54-71. (http://www.iatis.org/images/stories/publications/new-voices/Issue10-2013/articles/article-buendia-2013b.pdf).

Cabré, M. T. *Terminology Theory, Methods and Applications*. Amsterdam/Philadelphia: John Benjamins, 1999.

Castagnoli, S., Ciobanu, D., Kübler, N., Kunz, K. & Volanschi, A. "Designing a Learner Translator Corpus for Training Purposes." *Corpora, Language,*

Teaching, and Resources: From Theory to Practice. N. Kübler (Ed.). Bern: Peter Lang (2011): 221-248.

Faber, P., ed. *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin/Boston: Mouton De Gruyter, 2012.

Fernández Silva, S. & Kerremans, K. "Studying terminology and variation in specialised texts: corpus requirements and tools." *Journée d'études CVC-TERMISTI,* Brussels, 24th April 2009.

Gatto, M. *Web As Corpus: Theory and Practice. Corpus and Discourse*. London/New Delhi/New York/Sydney: Bloomsbury Academic, 2014.

Hanks, P. "Contextual Dependency and Lexical Sets." *International Journal of Corpus Linguistics* 1.1 (1996): 75-98.

Kenny, D. "Creatures of Habit? What translators Usually Do with Words." *Meta* 43.4 (1998): 515-523.

Kenny, D. "Lexical Hide-and-Seek: looking for creativity in a parallel corpus." *Intercultural Faultliness. Research Models in Translation Studies*. M. Olohan (Ed.), Manchester: St Jerome Publishing (2000): 93-104.

Kübler, N. "A Comparable Learner Translator Corpus: creation and use." *Proceedings of the Comparable Corpora Workshop of the LREC Conference*, P. Zweigenbaum (Ed.), May 31 2008, Marrakech.

Kussmaul, P. "Creativity in the Translation Process: Empirical Approaches." *Translation Studies: The State of the Art.Proceedings of the First James S. Holmes Symposium on Translation Studies*. K.M. van Leuven-Zwart & T. Naaijkens (Eds.), Approaches to translation studies 9. Amsterdam/Atlanta: Rodopi (1991): 91-101.

Kussmaul, P. "A Cognitive Framework for Looking at Creative Mental Processes." In: M. Olohan (Ed.). *Intercultural Faultlines: Research Models in Translation Studies I Textual and Cognitive Aspects*. Manchester/Northampton, St. Jerome, 2000. 57-70.

Lexical Computing Ltd. (n.d.). Sketch engine. Available at: http://www.sketchengine.co.uk/

López-Rodríguez, C. I. "Training translators to learn from news report corpora: the case of Anglo-American cultural references." *Training the Language Services Provider for the New Millenium.* Maia, B., J. Haller & M. Ulrych (Eds.), Porto: Faculdade de Letras Universidade do Porto (2002): 213-222.

López-Rodríguez, C. I. & Tercedor, M. I. "Corpora and students' autonomy in Scientific and Technical Translation training." *JoSTrans (Journal of Specialised Translation)* 9 (2008): 2-19. (http://www.jostrans.org/issue09/art_lopez_tercedor.pdf).

_____. (in press). "Multimodalidad y accesibilidad en recursos para el fomento de la salud." In *e-AESLA*. Instituto Cervantes. (http://cvc.cervantes.es/lengua/eaesla/).

López-Rodríguez, C. I, Robinson B. & Tercedor, M. I. "A learner generated corpus to direct learner-centered courses." *Translation and Meaning, Part 7.* M. Thelen & B. Lewandowska-Tomaszczyk (Eds.), Maastricht: Zuyd University, Maastricht School of International Communication (2007): 197-211.

Mauranen, A. "Strange Strings in Translated Language. A Study on Corpora." In: M. Olohan (ed.). *Intercultural Faultlines*. Manchester/Northampton: St Jerome Publishing, (2000) 119-142.

Mauranen, A. "Universal Tendencies in Translation." *Incorporating corpora: the linguist and the translator*. G. M. Anderman & M. Rogers (Eds.), Clevedon: Multilingual Matters (2008): 32-48.

Renouf, A. "WebCorp: providing a renewable data source for corpus linguists." *Extending the scope of corpus-based research: new applications, new challenges*. S. Granger & S. Petch-Tyson (Eds.), Amsterdam: Rodopi (2003): 39-58. (http://rdues.bcu.ac.uk/publ/WebCorp.pdf).

Rodríguez-Inés, P. "Evaluating the process and not just the product when using corpora in translator education." *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. A. Beeby, Rodríguez-Inés, P. & Sánchez-Gijón, P. (Eds.), Amsterdam: John Benjamins (2009): 129-149.

Ruiz Yepes, G. "Parallel Corpora In Translator Education." *Redit. Revista Electrónica de Didáctica de la Traducción y la Interpretación* 7 (2011): 65-80.

Ruschiensky, C. "Competence and Creativity in Translation: Multilingual Perspectives." (Unpublished PhD Thesis, Concordia University, 2015). (http://spectrum.library.concordia.ca/979835/1/Ruschiensky_MA_S2015.pdf).

Sager, J. C. "Term formation." *Handbook of Terminology Management*. S. E. Wright & G. Budin (Eds.), 1.0. Amsterdam: John Benjamins (1997): 25-41.

Stewart, D. "Conventionality, creativity and translated text: The implications of electronic corpora in translation." *Intercultural faultlines: Research models in translation studies*. M. Olohan (Ed.), Manchester: St Jerome Publishing (2000): 73-91.

Temmerman, R. *Towards New Ways of Terminology Description: The Sociocognitive-approach*. Amsterdam/Philadelphia: John Benjamins, 2000.

Tercedor-Sánchez, M. I., Alarcón-Navío, E., Prieto Velasco, J. A. & López-Rodríguez, C. I. "Images as part of technical translation courses: implications and applications." *JoSTrans (Journal of Specialised Translation)* 11 (2009): 143-168. (http://www.jostrans.org/issue11/art_tercedoretal.pdf).

Tercedor-Sánchez, M. I. "The cognitive dynamics of terminological variation." *Terminology* 17.2 (2011): 181-197. DOI:10.1075/term.17.2.01ter.

Tercedor-Sánchez, M. I., López-Rodríguez, C. I. & Prieto Velasco, J. A. "También los pacientes hacen terminología: los retos del proyecto VariMed." *Panace@, Rev. de Medicina, Lenguaje y Traducción* 15.39 (2014): 95-102.

Tirkkonen-Condit, S. "Unique items —over- or under-represented in translated language?" *Translation Universals. Do They Exist?* A. Mauranen & P. Kujamäki (Eds.), Amsterdam/Philadelphia: John Benjamins (2004): 177-185

Tymoczko, M. "Translation, ideology, and creativity." *Linguistica Antverpiensia, New Series – Themes In Translation Studies*, 0.2 (2013). (https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/74).

Varantola, K. "Translators and disposable corpora." *Corpora in translator education*. F. Zanettin, Bernardini, S. & Stewart, D. (Eds.), Manchester: St. Jerome (2003): 55-70.

Wüster, E. *Introduction to the General Theory of Terminology and Terminological Lexicography*. Wien: Springer, 1979.

Zanettin, F. "DIY Corpora: the WWW and the Translator." *Training the Language Services Provider for the New Millenium*. B. Maia, J. Haller & M. Ulrych (Eds.), Porto: Faculdade de Letras Universidade do Porto (2002): 293-248.

_____. "Parallel words designing a bilingual database for translation activities." *Corpora in Language Education and Research. A Selection of Papers from TALC 94*. University of Lancaster. T. McEnery & Wilson A. (Eds.), (1994): 99-111.

_____. *Translation-Driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome Publishing, 2012.