



Specialized terminology, instrumental competence, and corpus management tools / Terminología especializada, competencia instrumental y herramientas de gestión de corpus

Candelas Bayón Cenitagoya

Universidad de Alcalá

<https://orcid.org/0000-0001-5111-1999>

Resumen: Desde sus inicios en los años 40 y 50, y su posterior auge en los 80, la lingüística de corpus se ha erigido como un campo de investigación interesante para multitud de disciplinas, incluyendo los Estudios de Traducción (Corpas Pastor, 2001). Relacionada con los Estudios de Traducción, la Traducción e Interpretación en Servicios Públicos (TISP), entendida como la disciplina que posibilita la comunicación entre proveedores y usuarios de servicios públicos, realiza una labor crucial en la integración de la población migrante. Uno de los ámbitos de investigación más prolíficos dentro de la TISP es la traducción jurídico-administrativa, dentro la cual puede enmarcarse el ámbito laboral, dado que presenta muchos retos, sobre todo de naturaleza terminológica. En este sentido, el presente artículo pretende analizar dos herramientas de gestión de corpus siguiendo los criterios de Urbano Mendaña (2017), en un intento de subrayar la relevancia de la documentación en el proceso traductor y la aplicación de la lingüística de corpus a ámbitos especializados como el laboral.

Palabras clave: herramientas de gestión de corpus; traducción especializada; materiales de documentación; Traducción e Interpretación en Servicios Públicos (TISP)

How to cite this article / ¿Cómo citar este artículo?

Bayón Cenitagoya, C. (2024). Specialized terminology, instrumental competence, and corpus management tools. *FITISPos International Journal*, 11(1), 220-238. <https://doi.org/10.37536/FITISPos-IJ.2024.11.1.402>

Abstract: Since its start in the 40s and 50s and its re-emergence in the 80s, corpus linguistics has established itself as an interesting research field for a variety of disciplines, such as Translation Studies (Corpas Pastor, 2001). Related to it, Public Service Interpreting and Translation (PSIT), as the discipline that deals with communication between service providers and users, also plays a role in migrants' societal inclusion. One of the most prolific settings for research within PSIT is that of legal-administrative translation, within which the employment field might be located, since it is a setting that poses many challenges for language professionals, primarily of terminological nature. In this vein, this article aims at analyzing two corpus management tools according to Urbano Mendaña's (2017) criteria, in an attempt to highlight the relevance of the documentation process in translation and the applicability of corpus linguistics to specialized domains, like the employment setting.

Keywords: corpus management tools; specialized translation; research materials; Public Service Interpreting and Translation (PSIT)

1. Introduction

1.1 *Justification and objectives*

The Marxist approach of dialectical materialism, that aims to provide an explanation of the world in terms of its material characteristics, employment, considered as material condition and class identifier, has been the subject of several studies. Disciplines such as Philosophy, Economics and Sociology have been influenced by this approach, as can be seen in Adorno's (2018) and Lefebvre's (1991) application of dialectical materialism to Sociological Theory.

The focus on the material conditions allows an exploration of the relationships and dynamics that impact factors such as living conditions, demographics, migratory flows, and policies derived from societal changes and necessities.

From a linguistic and translational perspective, demographics and migratory flows are also relevant since they lay the foundation for different language and culture dynamics. In our interconnected and pluralist societies, the study of every process that has to do with how they relate to each other, how dynamics such as cultural hegemony happen or how we exist in the world, should be considered as key.

The material conditions that propitiate migratory flows and their consequences, as well as those that migrants face upon arrival have been explored by a variety of academics (Colectivo Ioé, 2000; Benito et al., 2003; Leal & Alguacil, 2012; Molpeceres Álvarez, 2012). As Gottdiener (1993) puts it "[e]very mode of social organization produces an environment that is a consequence of the social relations it possesses" (p.132). From the perspective of Translation Studies, the Sociological Turn (see Wolf, 1999; Hermans, 2007a; Latour, 2005; Tyulenev, 2009; Valero Garcés & Gauthier Blasi, 2010, among others) has enabled a look at the material conditions of the professionals of Translation and Interpreting but it has also fostered reflection about how translation as a discipline and practice relates to its surroundings. As Boukhelef and Babou (2021, p. 78) emphasize, "the social nature is one of the fundamental natures of translation".

Studies that explain the factors for inclusion usually include inequality and exclusion based on language, culture, or country of origin (Tezanos & Tezanos Vázquez, 2006; Subirats i Hume et al., 2005). On the one hand, the notion of employment as a means for insertion in the receiving society of the migrant person has been discussed by scholars such as Pina et al. (2015), Suárez Suárez et al. (2017), and Jaraiz Arroyo and González Portillo (2019). On the other hand, the notion of language as a means for insertion has been studied by authors such as Iglesias-Martínez and Estrada (2018). Joining these two notions together, the question would arise of what happens when linguistic and cultural barriers intersect with the material approach.

Therefore, and in line with what has just been said, the objectives of the study are threefold:

- First, to highlight the relevance of the employment field as a sub-branch of legal-administrative translation, and to relate its relevance with the purpose of Public Service Interpreting and Translation.
- Secondly, to draw a relationship between specialized legal translation, corpus linguistics, terminology, and instrumental translation competence, and to highlight the relevance of the corpus linguistics approach to overcome terminological challenges.
- The third objective of the present paper is the comparison of two corpus management tools in their performance with a legal setting corpus containing the most frequent terminology of the employment field within the Spanish system. The tools that will be analyzed are AntConc and Sketch Engine.

2. Theoretical framework

In order to contextualize the topic, a brief overview of Public Service Interpreting and Translation (PSIT) is provided, mainly regarding the context of its practice in Spain and the settings it entails as a discipline. In the subsequent section, the notions of PSIT, migrant population inclusion, and the employment field are detailed. Finally, section 2.3 deals with specialized translation, terminology, corpus linguistics and its applicability to Translation Studies.

2.1. Some considerations on PSIT

According to Valero Garcés (2014), PSIT had not yet been professionalized in Spain and other southern countries. Studies such as the one by Rojo (2015) also point in that direction. As Valero Garcés (2014) explains, the job market of Public Service Translation and Interpreting is varied, and it includes a wide array of roles and settings, as well as different levels of training of the translators and interpreters. This scholar also notes that both PSIT as a discipline as well as its scope are difficult to define. In this vein, it is interesting the fact that different names are given to the professionals carrying out the job within this discipline, for example, according to the setting or profession they deal with, i.e. Healthcare Interpreter, Legal Interpreter, Cultural Interpreter, etc. (Valero Garcés, 2014).

In terms of scope, this type of practice of translation and/or interpreting is defined as “a form of communication that takes place in any multicultural society where speakers of different languages must directly communicate with one another and where those who know both languages must act as intermediaries” (Valero Garcés, 2014, p. 27). Although settings might be an open category, it might be relevant to bring forth Wadensjö’s (1998) famous definition of PSIT, where she lists some of the contexts in which it might take place: social services, immigration offices, police settings, health care centers or schools. To this classification, we might add the legal-administrative setting, where translations and interpretations regarding the employment field are inscribed.

2.2. *PSIT, inclusion, and employment*

In this section we will provide a brief explanation supporting the choice of the employment setting as the setting for the present study, linking it to the notion of social inclusion for the migrant population.

In this regard, the International Organization for Migration (IOM, 2019) defines the notion of “employment migration” as the “[m]ovement of persons from one state to another, or within their own country of residence, for the purpose of employment”. On the other hand, social inclusion is defined by scholars like García Martínez (2006, p.238) as the process of adaptation of a migrant person to the receiving culture, influenced by living and material conditions. García Martínez (2006, p. 239) underlines how the incorporation of the migrant population to the receiving society’s employment market and economic system is key for their inclusion in the host society. In this line, other authors, such as Tezanos and Tezanos Vázquez (2006), include employment conditions as one of the 5 categories that are relevant for inclusion.

The link between PSIT and social inclusion is strong, since many aspects that are specific to the inclusion process coincide with the areas of action of PSIT as a discipline, which have been mentioned in the previous section. On the other hand, the public services’ user receives complex and varied information when it comes to legal-administrative texts, and certain administrative traditions, like the Spanish one, are known for their bureaucracy (Pérez et al. 2021). As Costa and Ewert (2014) explain, linguistic, cultural, economic, and bureaucratic factors influence the inclusion of migrant population in a host society.

In this regard, PSIT would mainly deal with linguistic and cultural factors influencing inclusion of the migrant population. It seems important to consider that PSIT professionals deal with both the host societies’ languages and cultures but also with that of the different subsystems existing within them. This is, within PSIT, professionals might work with employment documents in each language, while bearing in mind the culture gap associated with it, but they would also deal with the differences existing in between the languages and cultures of the employment systems from different societies.

This is why the employment setting is considered a relevant sub-setting within legal-administrative translation and PSIT. In the following section we develop on corpus linguistics and its application to specialized translation, as legal-administrative translation is considered to be, as well as translation competence.

2.3 *Specialized translation, corpus linguistics, and terminology*

Translation of specialized texts entails a series of challenges, such as: building thematic knowledge, dealing with specialized terminology, or following the textual genre conventions. In this vein, corpus linguistics might be of help for translators approaching this type of texts. According to Sinclair (1991, p. 171), a corpus is a “[c]ollection of naturally occurring language text, chosen to characterize a state or variety of a language”.

There are different studies that deal with the applicability of corpus linguistics to translation studies (Corpas Pastor 2001, 2004a, 2004b, 2007, 2008; Corpas Pastor & Seghiri, 2016; Vigier Moreno, 2016) but this is an area in need of further explanation.

According to Sánchez Ramos (2017), the bigger the lexical knowledge, the quicker and higher the quality of the translation, which will result in the job being more beneficial for translators. In this same line, the same author also establishes that the use of translation

memories (TM), corpora and terminological extraction tools are part of the documentation process of the specialized translator (2017). Similar remarks have been established by research groups like the PACTE Group (2009, 2018) or the compendium of competences for translation developed by the EMT Network (2022).

In fact, according to Valero Garcés (2014, p. 73) “it can be agreed that [Public Service] interpreters need training and must master a number of techniques to provide and ensure quality work”. The same can be applied to translators in Public Service settings, of whom it is generally required to have thematic competence, terminological and writing conventions knowledge. In this vein, Valero Garcés (2014) states a series of contrastive objectives that translators are recommended to have, namely:

- Knowledge of differences and similarities regarding writing conventions in both source and target language.
- Vocabulary knowledge regarding a variety of aspects such as semantic fields, neologisms, idiomatic phrases, and the like.
- Syntactic knowledge in terms of differences in the expression of denial, exclamation, obligation, inquiry...
- Knowledge regarding the elaboration of the text.
- Practice in the translation of specialized texts and knowledge of textual field-based translation challenges, such as problems caused by the importance of terminological elements or the importance of the document (in the case of legal, economic, medical texts, and so forth).
- Ability to translate the style of the text.
- Ability to identify, understand and translate dialectical variations of their working languages.
- Ability to identify, understand and translate idiolects or individual linguistic variations.

As can be observed in Table 1, references to some kind of instrumental or research competence are made by scholars such as Kelly (2002), the PACTE Group (2000; 2001) or Nord (1991). Therefore, it can be argued that it is relevant to understand how tools that will help with the research process work, such as Sketch Engine and AntConc. References to some kind of knowledge or competence that encompass the understanding of the specialized domain, or the mastery of terminology are made by the PACTE Group within its thematic competence, as well as Kiraly’s (1995) inclusion of linguistic and cultural knowledge in both languages and the specific knowledge of the specialized domain; Nord (1991) also makes reference to this aspect with her factual competence, while Bell’s classification (1991) define it as subject-area knowledge.

Table 1. Comparison of scholar's outlooks on translation competence. Source: Valero Garcés (2014) and Sánchez Ramos (2020). Elaborated by the author.

Valero-Garcés (2014)	Kelly (2002)	PACTE Group (2000, 2001)	Kiraly (1995)	Nord (1991)	Bell (1991)
Syntactic knowledge/ Vocabulary knowledge	Communicative and textual sub-competence	Communicative competence in both languages	Linguistic knowledge of L1 and L2	Linguistic competence in first and second language	Target language knowledge/ Source language knowledge
	Cultural sub-competence	Extralinguistic competence	Cultural knowledge of L1 and L2	Cultural competence	
		Transference competence	Ability to formulate the TT and evaluate its adequacy as a translation of the ST	Transference competence: reception of ST, production of TT, research and documentation	Reception (reading)/ Production (writing)
	Instrumental professional subcompetence	Instrumental competence		Technical competence: documentation, information storage	
	Psychophysiological subcompetence	Psychophysiological competence			
	Strategic competence	Strategic competence			
Knowledge on textual field-based translation challenges/ Knowledge of writing conventions					Text-type knowledge
					Contrastive knowledge
	Thematic sub-competence		Knowledge of the specialized domain of the translation	Factual competence: knowledge of specialized domains	Subject-area knowledge
Ability to translate style					
Ability to identify, understand and translate idiolects or individual linguistic variations					
	Interpersonal subcompetence		Awareness of the situational factors involved in a translation task		

In order to achieve these competences, corpus linguistics emerges as a discipline that is yet to be explored in its applicability to the translation process and one of its main points of interest for this discipline is the development of both thematic terminological and technological skills for professionals and students. Nonetheless, it seems important to mention the age-old adage regarding whether corpus linguistics is a discipline in itself or a methodology of linguistic studies. However, and regardless of what the solution of said open debate might be, several scholars have noted and written about the applicability of using corpus linguistics within translation studies and more specifically about its applicability for legal translation.

For example, Corpas Pastor (2001) highlights the usability of corpus linguistics by stating that it has fostered empirical methods and real data observation. In this vein, the scholar determines that corpus linguistics have had a great impact in other research methodologies or disciplines such as terminology and translation and interpreting. Another application according to the author might be teaching, also in the domain of translation, as a way of enhancing the development of translation competences. In fact, she defines computerized corpus as a means of documentation for the translator as fast, reliable, and economical.

Regarding specialized translation of legal texts, Vigier Moreno (2016, p.103) defines corpus linguistics as “one of the most – if not the most –popular resources for the study of translation-related phenomena”. This author also highlights the lack of legal training of the translation students, and establishes that, when said legal training is provided in specific legal translation courses, the focus is placed on developing thematic competence (i.e., basic legal knowledge) and does not put much emphasis on technological competence. Vigier Moreno’s study is of special interest because it suggests a three-step protocol for corpora compilation that includes the following steps:

1. Source-text documentation. In this first step, the translator must read texts that are similar to the source text they have to translate. This is carried out to ensure that they understand the nature of the type of text in general of the text they have to translate in particular. Within this step it is also beneficial to identify similar texts within the target system or language that can be included in the corpus compilation process. According to Vigier Moreno this step activates the thematic and instrumental competences of the translators.
2. Compilation process. This step takes place after the download and research for appropriate material for the corpora. Vigier Moreno recommends free software to use for corpus compilation and he also mentions the importance of converting the files to a .txt format.
3. Corpus analysis. This is the final step that takes place after the two previous ones and can take place when all documents are stored and converted to appropriate file type.

Regarding terminology, different scholars (Ciapuscio, 2003; Merlo Vega, 2004; Vargas Sierra, 2010; Vigier Moreno, 2016) have mentioned its relationship with specialized translation in general, and with legal language and its translation in particular (Alcaraz Varó, 1994; Vitalaru, 2012; Borja Albi & García Izquierdo, 2016; Sanz de la Rosa, 2023). In this vein, terminology would constitute an aspect of thematic and linguistic competences and could be studied by translators through the application of corpus linguistics research methodologies. For example, Durán Muñoz (2011) analyzes different terminological resources for translators, resulting in a thorough classification, from which we include the following:

- Portals and resources compilations for translators. From general and specialized dictionaries (i.e. Lexicool), to linguistic and terminological portals (e.g. Terminology Collection, TERMCAT) or glossaries searchers (e.g. Glossarist).
- Databases to consult complex terminological issues, e.g. EcoLexicon.

Moreover, an application of the terminological analysis of texts is recommended by authors such as Cabré and Feliu (2001). Others, like Vitalaru (2019) determine the relevance of the creation of specialized bilingual materials for both professionals and students. In this line, works such as those of Bayón Cenitagoya (2021) and Sanz de la Rosa (2023) highlight the usefulness of domain-specific glossary creation in the legal-administrative setting using a corpus methodology. While Bayón Cenitagoya's (2021) study is ground in a penitentiary setting, Sanz de la Rosa (2023) works with the terminology of witness statements.

In order to carry out corpus analyses and terminological extractions, corpus management tools are useful. In this vein, the aim of the present study is to compare the features of two corpus management tools, namely AntConc and Sketch Engine. As means for the comparison, the classification of features proposed by Urbano Mendaña (2017) will be employed. Said classification studies features of corpus management tools in terms of:

- 1) general features
- 2) technical requirements and specificities
- 3) functionalities
- 4) manageability and usability

3. Methodology

The methodology followed consisted of 1) compiling a corpus with texts belonging to the employment field in Spain, for which the aforementioned steps suggested by Vigier Moreno (2016) were followed, 2) uploading the corpus to both AntConc and Sketch Engine, 3) using the features and tools within both AntConc and Sketch Engine with the same corpus, and 4) comparing the characteristics of both corpus management tools. Said comparison was carried out according to the categories specified by Urbano Mendaña (2017), which were mentioned in the previous section.

Regarding the steps suggested by Vigier Moreno (2016), the first one, about "source-text documentation", was not carried out in its entirety, since, for the purpose of the present study we do not have a source text that must be translated. However, simulating a potential case scenario in which a translator would have to work with a Spanish source text belonging to the employment field, we selected topic-related texts. Thus, the corpus is composed of the Estatuto de los Trabajadores and examples of work contracts available and retrievable from the Servicio Público de Empleo Estatal (SEPE, Public State Service of Employment) website. This means the corpus was composed by a total of 4 texts, namely:

- Estatuto de los Trabajadores: Royal Decree-law 2/2015, of the 23rd of October, approving the revised text of the Law of the Statute of Workers' Rights.
- Contracts:
 - *Temporal* (fixed-term)
 - *Indefinido* (permanent or indefinite)
 - *Prácticas* (apprenticeship)

Secondly, the second step included in Vigier Moreno’s classification was carried out. Here, the first distinction between tools can be noted. Sketch Engine allows users to compile the corpus using a .pdf format directly. On the other hand, AntConc requires previous compilation with a different software. Then, the third step suggested by Vigier Moreno was conducted, for which the different features included in each one of the tools was employed to analyze the uploaded corpus. Finally, in accordance with the last step of our methodology, both tools were compared according to Urbano Mendaña (2017) features. The results of the third and fourth step are presented in the following section.

4. Analysis and results

We will start this section providing a brief description of both management tools. Thus, according to its creator, Anthony, L. (2014), AntConc is “a freeware corpus analysis toolkit for concordance and text analysis”. On the official website (Kilgariff, 2013), Sketch Engine is defined as a “tool to explore how language works” operating with an algorithm that “analyzes authentic texts of billions of words (text corpora) to identify instantly what is typical in language and what is rare, unusual or emerging usage”. Other applications suggested are text mining and analysis. Furthermore, on the official website, Sketch Engine presents itself as a tool useful for a wide array of professionals: linguists and lexicographers, translators, terminologists, text analysts, teachers and students, publishers, and historians.

In terms of the features analyzed according to Urbano Mendaña’s (2017) classification, the results were the following:

1. General features.

a. Sketch Engine has 7 tools for corpus analysis, namely: concordance, n-grams, wordlist, keywords, word sketch, word sketch difference, and thesaurus.

- Concordance (Figure 1). This feature searches words, phrases, tags, documents, text types or corpus structures and displays the results in context.

Figure 1. Concordance tool in Sketch Engine.

	Details	Left context	KWIC	Right context
1	doc#0 76 TÍTULO III.	De la negociación colectiva y de los	convenios colectivos	76 C,
2	doc#0 82 Artículo 91.	Aplicación e interpretación del	convenio colectivo	83 Sec
3	doc#0 85 Disposición adicional décima.	Cláusulas de los	convenios colectivos	referidas al cumplimiento de la edad ordinaria de
4	doc#0 iones legales y reglamentarias del Estado.	Por los	convenios colectivos	Por la voluntad de las partes, manifes
5	doc#0 : menos favorables o contrarias a las disposiciones legales y	convenios colectivos	antes expresados.	Por los usos y cost.
6	doc#0 amente de los derechos reconocidos como indisponibles por	convenio colectivo	.	Sección 2.a Derechos y deberes laborales:
7	doc#0 nto, durante el segundo, de la jornada máxima prevista en el	convenio colectivo	de aplicación en la empresa, o, en su defecto, de l	
8	doc#0 La retribución será la establecida para estos contratos en el	convenio colectivo	de aplicación.	En defecto de previsión con
9	doc#0 es ni exceder de un año.	Dentro de estos límites los	convenios colectivos	de ámbito sectorial estatal o autonómico, o en su
10	doc#0 le ámbito sectorial estatal o autonómico, o en su defecto, los	convenios colectivos	sectoriales de ámbito inferior podrán determinar s	
11	doc#0 ningún caso podrá exceder de un mes, salvo lo dispuesto en	convenio colectivo	.	El puesto de trabajo deberá permitir la c
12	doc#0 tribución por el tiempo de trabajo efectivo será la fijada en el	convenio colectivo	aplicable en la empresa para estos contratos o en	
13	doc#0 dad y características de estas personas.	Mediante	convenio colectivo	de ámbito sectorial estatal, autonómico o, en su de
14	doc#0 ámbito sectorial estatal, autonómico o, en su defecto, en los	convenios colectivos	sectoriales de ámbito inferior, se podrán determir	
15	doc#0), se considerará la jornada a tiempo completo prevista en el	convenio colectivo	de aplicación o, en su defecto, la jornada máxima	
16	doc#0 as, así como el modo de su distribución según lo previsto en	convenio colectivo	.	De no observarse estas exigencias, el coi

- N-Grams (Figure 2) produces a frequency list of sequences of tokens. N-Grams are also known as multi-word expressions (MWE).

Figure 2. N-grams tool in Sketch Engine.

N-gram	Frequency ?	N-gram	Frequency ?
1 el convenio colectivo	28	26 convenio colectivo de ámbito	3
2 en el convenio	20	27 según el convenio	3
3 en el convenio colectivo	19	28 que por convenio colectivo	3
4 convenio colectivo de	17	29 según el convenio colectivo	3
5 en convenio colectivo	15	30 trabajo según el convenio	3
6 convenio colectivo o	14	31 que por convenio	3
7 por convenio colectivo	12	32 convenio colectivo sectorial	3
8 el convenio colectivo de	10	33 por el convenio	2
9 un convenio colectivo	10	34 previsto en convenio	2
10 convenio colectivo de aplicación	10	35 por convenio colectivo sectorial	2
11 de un convenio	9	36 previsto en convenio colectivo	2
12 del convenio colectivo	8	37 que el convenio	2

- Wordlist. This tool generates a frequency list of various kinds, such as nouns, verbs, adjectives and other parts of speech; words tags, lemmas and other attributes; words beginnings, endings, words containing certain characters.
- Keywords (Figure 3). This tool is used to extract terminology for translation and interpreting; extract single- and multi-word units typical from a text or corpus or which define its content or topic; to compare two corpora/documents by identifying what can be singled out in the first corpus in comparison to the second one. The extraction is based on simple math (keyness score).

Figure 3. Keywords tool in Sketch Engine.

Term	Term	Term
1 convenio colectivo	18 reducción de jornada	35 tiempo de trabajo
2 persona trabajadora	19 apartado anterior	36 comisión representativa
3 centro de trabajo	20 párrafo anterior	37 comisión negociadora
4 contrato de trabajo	21 condición de trabajo	38 asociación empresarial
5 autoridad laboral	22 hora complementaria	39 cuota empresarial a la seguridad social
6 periodo de consultas	23 delegado de personal	40 disposición transitoria
7 relación laboral	24 texto refundido	41 comunidad autónoma
8 tiempo parcial	25 representante de los trabajadores	42 legislación vigente
9 representante legal	26 negociación colectiva	43 organización sindical
10 comité de empresa	27 contrato formativo	44 plazo máximo
11 puesto de trabajo	28 oficina pública	45 oficina pública dependiente
12 disposición adicional	29 grupo profesional	46 año b

- Word Sketch (Figure 4) summarizes the word's grammatical and collocational behavior.

Figure 4. Word Sketch tool in Sketch Engine.

The screenshot shows the Word Sketch tool interface for the term "convenio". The search bar contains "ESlaborterm" and the term "convenio" is identified as a noun with 232 occurrences. The interface is divided into five main panels, each with a title and a list of related terms and example sentences:

- modifiers of "convenio"**: colectivo (los convenios colectivos), sectorial (los convenios colectivos sectoriales), aplicable (en el convenio colectivo aplicable), nuevo (un nuevo convenio), propio (convenio propio), vigente (convenio colectivo vigente), mismo.
- verbs with "convenio" as object**: regular (tramitación que los convenios colectivos regulados en esta ley), negociar (negociar un convenio), suscribir (los convenios colectivos suscritos), existir (exista otro convenio), referir (Cláusulas de los convenios colectivos referidas al cumplimiento de).
- verbs with "convenio" as subject**: suceder (convenio colectivo que sucede), establecer (Los convenios colectivos establecerán medidas para facilitar), poder (los convenios colectivos podrán establecer).
- "convenio" and/or ...**: acuerdo (los acuerdos y convenios), contrato (paclado en convenio colectivo o contrato individual), disposición (disposiciones legales y convenios).
- prepositional phrases**: ... en "convenio", ... de "convenio", "convenio" de ..., "convenio" en ..., ... por "convenio", "convenio" para ..., ... según "convenio", "convenio" a ..., ... a "convenio", ... para "convenio".

- Word Sketch Difference. This tool compares contrasting collocations.
- Thesaurus automatically generates a list of synonyms or words belonging to the same semantic fields.

b. AntConc presents 9 tools, including Key-Word-In-Context (KWIC), Plot tool, file tool, cluster, n-grams, collocate, word list, keyword list, and wordcloud. In this section, the characterization of the tools is extracted from Aparicio Boussif's (2014) guide to use AntConc.

- Plot Tool shows the results in a barcode format so that the user can see the position in which the results appear in target texts.
- File Tool allows to see the text of individual files so that the results obtained from other AntConc features can be examined in more detail.
- Collocate (Figure 5) is a feature that generates an organized list of the collocations presented next to the searched term, which allows to find non-sequential language patterns.

Figure 5. Collocate tool in AntConc

The screenshot shows the Collocate tool interface in AntConc. The search query is 'convenio'. The results table is as follows:

File	Left Context	Hit	Right Context
1 TEMPOR...	se considerará la jornada a tiempo completo prevista en el	convenio	colectivo de aplicación o, en su defecto, la jornada
2 BOE-...	uvieran legitimados para formar parte de la comisión negociadora del	convenio	colectivo de aplicación a la misma. En el supuesto
3 BOE-...	apartado 3 y las letras b) a e) del apartado 4. 6. El	convenio	colectivo de aplicación para las empresas contratistas y subcontratistas
4 BOE-...	el ámbito funcional de los convenios colectivos y sobre el	convenio	colectivo de aplicación a una empresa, así como la
5 PRACTIC...	se considerará la jornada a tiempo completo prevista en el	convenio	colectivo de aplicación o, en su defecto, la jornada
6 BOE-...	la inaplicación de las condiciones de trabajo, presentes en el	convenio	colectivo de aplicación, cuando dicha inaplicación afecte a centros
7 INDEFINI...	se considerará la jornada a tiempo completo prevista en el	convenio	colectivo de aplicación, o , en su defecto, la jornada
8 PESCADO...	Si el	convenio	colectivo de ambito sectorial permite en los contratos fijos
9 BOE-...	mismos de la imposibilidad de suscribir en dicho ámbito un	convenio	colectivo de los previstos en este título III, debida
10 BOE-...	grado de discapacidad y características de estas personas. e) Mediante	convenio	colectivo de ámbito sectorial estatal, autonómico o, en su
11 BOE-...	no podrá ser superior a nueve diarias, salvo que por	convenio	colectivo o, en su defecto, acuerdo entre la empresa
12 BOE-...	nales de trabajo efectivo de promedio en cómputo anual. 2. Mediante	convenio	colectivo o, en su defecto, por acuerdo entre la
13 BOE-...	erjuicio de su concreción en el momento del llamamiento. 3. Mediante	convenio	colectivo o, en su defecto, acuerdo de empresa, se
14 BOE-...	de trabajo, fijada de acuerdo con el artículo anterior. Mediante	convenio	colectivo o, en su defecto, contrato individual, se optará

- Cluster: adjacent word groups.
- N-Gram (Figure 6) analysis the corpus in its entirety, looking for N-sized clusters which allows to find common expressions.

Figure 6. N-Grams tool in AntConc.

The screenshot shows the N-Gram tool interface in AntConc. The search query is 'convenio'. The results table is as follows:

Type	Rank	Freq	Range
1 convenio colectivo	1	107	5
2 convenio indique	4	2	2
3 convenio sectorial	4	2	1
4 convenio aplicable	9	1	1
5 convenio artículo	9	1	1
6 convenio conculca	9	1	1
7 convenio cuenten	9	1	1
8 convenio deberá	9	1	1
9 convenio decaerán	9	1	1
10 convenio entrará	9	1	1
11 convenio especial	9	1	1

- Word List counts all the words in the corpus and presents them in order of frequency.
- Keyword List (Figure 7) extracts words that appear unusually frequently in target corpus in comparison to the reference corpus. Therefore, they can be said to be characteristic of the target corpus.

Figure 7. Keywords tool in AntConc.

KWIC		Plot	File View	Cluster	N-Gram	Collocate	Word	Keyword	Wo
Keyword Types 316/4603		Keyword Tokens 52711/74222		Page Size 100 hits				1 tc	
	Type	Rank	Freq_Tar	Freq_Ref	Range_Tar	Range_Ref	Keyness (Likelihood)		
1	contrato	15	485	0	5	0	304.128		
2	empresa	18	447	0	5	0	280.237		
3	trabajadores	19	445	0	5	0	278.980		
4	artículo	21	421	0	5	0	263.897		
5	trabajador	24	307	0	5	0	192.311		
6	ley	25	281	0	5	0	175.997		
7	caso	28	249	0	5	0	155.926		
8	personas	29	247	0	5	0	154.672		
9	social	30	246	0	5	0	154.045		
10	laboral	33	193	0	5	0	120.819		
11	derecho	34	186	0	5	0	116.432		

- Key-Word-In-Context (KWIC) shows the context of the keywords.
- Wordcloud (Figure 8) forms a wordcloud where the size of the word is related to its frequency.

Figure 8. Wordcloud tool in AntConc.



2. Technical requirements

a. Sketch Engine requires the use of a computer and access to Internet connection. As mentioned before, it also accepts different document formats to compile the corpus and does not require the uploading of a reference corpus, as it already has integrated ones. This tool allows for its use through license or with a 30-day free trial.

b. AntConc also requires a computer with access to Internet connection. As opposed to Sketch Engine, it only allows uploading of documents in a .txt format. The reference corpus must also be uploaded by users although there is a demo corpus in English language. Contrary to Sketch Engine, AntConc is free to use.

3. Functionalities

a. Sketch Engine: Allows uploading a stopword (non-words) list and for certain functions (e.g. Wordlist) non-words are excluded automatically. Keyword function (useful for glossaries). Keyword function includes multi-word terms. Allows download in XLS.

b. AntConc: Allows to upload a stopword list. Keyword function (useful for glossaries). Allows saving of results in txt. Search options: case, regex (regular expressions).

4. Manageability and usability

a. Sketch Engine's interface is seemingly very intuitive, and the way in which results from corpus analysis by any of the tools is shown is clear and understandable. The corpus compilation is fast, and, as previously mentioned, it already has access to reference corpora.

b. AntConc allows for a faster compilation than Sketch Engine, since the user only needs to upload the documents to use them as corpus, whereas in Sketch Engine previous steps are required. However, it must be taken into account that the documents must have previously been converted to .txt format. Another aspect is that a reference corpus must be uploaded in order to be able to use certain tools, such as the keyword one. AntConc also allows access to the files uploaded as corpus within the software, with a feature named File View, which might be useful to make consultations.

Table 2. Comparison of corpus management tools Sketch Engine and AntConc according to Urbano Mendaña's (2017) features. Source: Elaborated by the author

	Sketch Engine	AntConc
General features	7 tools: concordance, n-grams, wordlist, keywords, word sketch, word sketch difference, and thesaurus.	9 tools: Key-Word-In-Context (KWIC), Plot tool, file tool, cluster, n-grams, collocate, word list, keyword list, and wordcloud
Technical requirements	Computer Internet connection Different document formats No reference corpus required License/ 30-day free trial	Computer Internet connection .txt format Requires reference corpus Free
Functionalities	Stopword (non-words) list allowed Non-words are excluded automatically Keyword function (useful for glossaries) Keyword function includes multi-word terms Download in XLS.	Stopword list allowed Keyword function (useful for glossaries) Allows saving of results in txt. Search options: case, regex (regular expressions)
Manageability and usability	Intuitive interface Fast compilation Clear presentation of results (WordSketch) Reference corpus incorporated	Faster compilation Keyword function requires a reference corpus for the File View function

5. Conclusion and recommendations

In synthesizing the findings presented throughout this paper, we now attempt to summarize the main insights that emerge from our exploration. By weaving together the threads of our inquiry, this conclusion seeks to offer a holistic perspective, delineating both the achievements realized and the avenues yet to be explored in the dynamic landscape of corpus linguistics.

In the first place, it should be noted that Sketch Engine is a bit easier and more intuitive to use but it requires a license, while AntConc is free. One of the reasons why we consider Sketch Engine to be more comfortable for users is because it reads more file types and already has a wide reference corpus in Spanish (esTenTen18) and other languages (e.g. English). Another reason might be the fact that the tools KWIC, File View and Collocate in AntConc are all on the Concordance tool in Sketch Engine.

It might also be worth noting that the Word Sketch function of Sketch Engine can be very useful once the most frequent words of the specialized corpus have been identified, since it can help the translator to see the most usual collocations of a certain lemma in a very visual and organized way. Something similar might be achieved in AntConc with the Cluster and Collocate Tools when the Search Option 'Regex' is activated.

Thirdly, both tools allow for Keyword selection based on keyness score, which can be considered a very useful option for term extraction for glossaries, as noted in Bayón Cenitagoya (2021) and Sanz de la Rosa (2023).

Another feature to be taken into consideration is multi-word terms identification. In this regard, Sketch Engine is able to select multi-word terms automatically. This can be done in AntConc by using N-Grams without a search (general search).

Therefore, it might be argued that both tools have similar functions and features, presenting minor differences already mentioned (reference corpus, file type, and so forth) as well as a different interface. The decision of which one to employ will depend on the user's preference and access to a reference corpus to be uploaded, as well as on the possibility of paying for a license or accessing the 30-day free trial in the case of Sketch Engine. The fact that both tools use the keyness score system to extract terms is considered a positive characteristic and one that will help users in their research process.

Finally, the relevance of corpus linguistics for specialized Translation Studies should be highlighted once more. As already mentioned, corpus-oriented approaches might be said to pose a clear advantage for carrying out terminological work, which is one of the challenges of legal translation. In order to overcome it, it is important for translators to have a strongly developed instrumental and/or technical competence of tools such as the ones analyzed in the present study. Said instrumental or technical competence should encompass knowledge of the different existent research methodologies, as well as of the most adequate tools to work with for each task. It is also relevant to mention further research that could be carried out in this direction, such as the analysis of more corpus management tools or the analysis of the differences and similarities of the applicability and functioning of these types of tools for different settings of PSIT.

References

- Anthony, L. (2014). AntConc (4.2.0) [Program]. Tokyo, Japan: Waseda University. <http://www.antlab.sci.waseda.ac.jp/>.
- Aparicio Boussif, J. (2014). *AntConc (Windows, Macintosh OS X y Linux)*.
- Bayón Cenitagoya, C. (2021). *Sistemas penitenciarios en España y Reino Unido (Inglaterra y Gales). Comparación, terminología y propuesta de glosarios* [Master's dissertation]. Universidad de Alcalá.
- Benito, M. J. R., Girón, M. Á., Gil, C. G., & Parés, A. S. (2003). Las mujeres inmigrantes en Sevilla: salud y condiciones de trabajo. Una revisión desde la medicina social. *Revista Española De Investigaciones Sociológicas*, 104, 123. <https://doi.org/10.2307/40184571>.
- Borja Albi, A., & García-Izquierdo, I. (2016). Web-based tools and resources for legal translators: the JudGENTT translation-oriented glossaries for criminal courts translators. *In ONOMÁZEIN*, 33, 226-250. <https://doi.org/10.7764/onomazein.33.15>
- Boukhelef, F., & Babou, A. (2021). The social nature of translation: The shift from textual to contextual. *Altralang Journal*, 3(1). <https://doi.org/10.52919/altralang.v3i01.104>
- Cabré, M. T., & Feliu, J. (Eds.). (2001). *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Universitat Pompeu Fabra.
- Ciapuscia, G.E. (2003). *Textos especializados y terminología*. Universitat Pompeu Fabra.
- Colectivo Ioé (Pereda, C., Actis, W., & de Prada, M. A.). (2003). La sociedad española y la inmigración extranjera. *Papeles de Economía Española, FUNCAS*, 98, 16-31.

- Corpas Pastor, G. (2001). Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada. *TRANS, Revista de Traductología*, 5, 155-184. <https://doi.org/10.24310/TRANS.2001.v0i5.2916>
- Corpas Pastor, G. (2004a). La traducción de textos médicos especializados a través de recursos electrónicos y corpus virtuales. In L. González & P. Hernández. (Eds.), *Las palabras del traductor*. Proceedings of the II Congreso Internacional "El español, lengua de traducción", 20 y 21 de mayo, Toledo 2004. Bruselas: Comisión Europea/ESLETRA. 137-164.
- Corpas Pastor, G. (2004b). Localización de recursos y compilación de corpus vía Internet: aplicaciones para la didáctica de la traducción médica especializada". In V. García Yebra & C. Gonzalo García (Eds.), *Manual de documentación y terminología para la traducción especializada* (pp. 223-506). Colección "Instrumenta Bibliologica". Arco/ Libros.
- Corpas Pastor, G. (2007). Lost in specialised translation: the corpus as an inexpensive and under-exploited aid for language service providers. *Translating and the Computer 29. Proceedings of the Twenty-ninth International Conference on Translating and the Computer*, 29-30 November 2007. Aslib/IMI. 1-21.
- Corpas Pastor, G. (2008). *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Peter Lang.
- Corpas Pastor, G., & Seghiri, M. (2016). *Corpus-based approaches to translation and interpreting: from theory to applications*. Peter Lang.
- Costa, G., & Ewert, B. (2014). Cities of migration: The challenges of social inclusion. In Ranci, C., Brandsen, T., & Sabatinelli, S. (Eds.), *Social Vulnerability in European Cities: The Role of Local Welfare in Times of Crisis* (pp. 134 -159). Palgrave Macmillan UK. https://doi.org/10.1057/9781137346926_5.
- Durán Muñoz, I. (2011). Recursos electrónicos para la búsqueda terminológica en traducción: clasificación y ejemplos. *Revista tradumàtica: traducció i tecnologies de la informació i la comunicació*, 9, 137-146. <https://doi.org/10.5565/rev/tradumatica.13>
- EMT Network. (2022): Competence Framework. *European Commission*. https://commission.europa.eu/system/files/2022-11/emt_competence_fw_k_2022_en.pdf.
- Erlenbusch-Anderson, V., (trans.), O'Kane, C., (trans.), & Adorno, T. (2018). Theodor W. Adorno on 'Marx and the Basic Concepts of Sociological Theory': From a Seminar Transcript in the Summer Semester of 1962. (1 ed.) *Historical Materialism*. <https://doi.org/10.1163/1569206X-00001619>.
- García Martínez, J. A. (2006). Migraciones, inserción laboral e integración social. *Revista de Economía Mundial*, 14, 231-249.
- Gottdiener, M. (1993). A Marx for our time: Henri Lefebvre and the production of space. *Sociological Theory*, 11(1),. 129-134. <https://doi.org/10.2307/201984>
- Hermans, T. (2007a). Translation, Irritation and Resonance. In M. Wolf & A. Fukari (Eds.), *Constructing a Sociology of Translation* (pp. 57-75). John Benjamins.
- Iglesias-Martínez, J., & Estrada, C. (2018). Birds of passage? The social integration of refugees in Spain. *Revista Iberoamericana de Estudios de Desarrollo= Iberoamerican Journal of Development Studies*, 7(1), 144-167. https://doi.org/10.26754/ojs_ried/ijds.279

- International Organization for Migration (2019). *Labour Migration*. <https://www.iom.int/labour-migration>.
- Jaraiz Arroyo, G., & González Portillo, A. (2019). The impact of local inclusion policies on disadvantaged urban areas: perceptions in the case of Andalusia. *Investigaciones regionales = Journal of Regional Research*, 44, 47-62.
- Kilgarriff, A., et al. (2013). Sketch Engine [Online program]. <http://www.sketchengine.co.uk/>.
- Leal, J., & Alguacil, A. (2012). Vivienda e inmigración: las condiciones y el comportamiento residencial de los inmigrantes en España. *Anuario CIDOB de la Inmigración*, 126-156.
- Latour, B. (2005). *Reassembling the Social. An Introduction to ActorNetwork-Theory*. Oxford University Press.
- Lefebvre, H. (1991). *The Production of Space*. Blackwell Publishing.
- Merlo Vega, J. A. (2004). La documentación en el proceso de la traducción especializada. In C. Gonzalo García & V. García Yebra (Eds.), *Manual de documentación y terminología para la traducción especializada* (pp.3 09-336). Arco/Libros.
- Molpeceres Álvarez, L. (2012). Situación laboral de las mujeres inmigrantes en España. *Cuadernos de Relaciones Laborales* 30(1), 91-113. https://doi.org/10.5209/rev_CRLA.2012.v30.n1.39116
- PACTE. (2009). Results of the validation of the PACTE Translation Competence Model: Acceptability and decision making. *Across Languages and Cultures*, 10(2), 207-230. <https://doi.org/10.1556/Acr.10.2009.2.3>
- PACTE. (2018). Competence levels in translation: working towards a European framework. *The Interpreter and Translator Trainer*, 12(2), 111-131. <https://doi.org/10.1080/1750399X.2018.1466093>
- Pérez, M., Ávila, D., García, S., & Ayala, A. (2021). Bureaucracy as border. Barriers to social rights in Spain. In D. della Porta & E. Steinhilper (Eds.), *Contentious Migrant Solidarity: Shrinking Spaces and Civil Society Contestation*. Routledge. <https://doi.org/10.4324/9781003083429>.
- Pina, Á., Corluy, V. & Verbist, G. (2015). Improving the Labour Market Integration of Immigrants in Belgium. *OECD Economics Department Working Papers*, 1195. <https://doi.org/10.1787/5js4hmbt6v5h-en>.
- Royal Decree-law 2/2015, of the 23rd of October, approving the revised text of the Law of the Statute of Workers' Rights. *Official State Gazette*, 255, of the 24/10/2015. <https://www.boe.es/eli/es/rdlg/2015/10/23/2/con>.
- Sancho Pascual, M. (2013). Dimensión lingüística de las migraciones internacionales. Introducción: lengua y migraciones en el mundo de la globalización. *Lengua y migración*, 5(2), 5-10.
- Sánchez Ramos, M. M. (2017). Metodología de corpus y formación en la traducción especializada (inglés-español): Una propuesta para la mejora de la adquisición de vocabulario especializado. *Revista de lingüística y lenguas aplicadas*, 12, 137-150. <https://doi.org/10.4995/rlyla.2017.6969>
- Sánchez Ramos, M. M. (2019). Corpus paralelos y traducción especializada: ejemplificación de diseño, compilación y alineación de un corpus paralelo bilingüe (inglés-español) para la traducción jurídica. *Lebende Sprachen*, 64(2), 269-285. <https://doi.org/10.1515/les-2019-0015>

- Sánchez Ramos, M. M. (2020). *Documentación digital y léxico en la Traducción e Interpretación en los Servicios Públicos (TISP): Fundamentos teóricos y prácticos*. Peter Lang.
- Sánchez Ramos, M. M. (2022). Documentation in specialised contexts: A quasi-experimental corpus-based study in public service interpreting and translation studies. *Translation and translanguaging in multilingual contexts*, 8(1), 30-48. <https://doi.org/10.1075/ttmc.00083.mar>
- Sanz de la Rosa, A. (2023). Creating glossaries with legal terminology: The text genre witness statement. In C. Valero Garcés (Ed.), *Traducción e interpretación en los servicios públicos (TISP) en transición* (pp. 183-192). Universidad de Alcalá.
- Sinclair, J. (1991). *Corpus, concordance, and collocation*. Oxford.
- Suarez Suarez, L., Pacheco Suarez, Y., & Hernandez Arencibia, L. (2017). Insertion to employment with a gender equity approach. *Estudios del desarrollo social: Cuba y América Latina*, 5(2), 191-202.
- Subirats i Hume, J. (Dir.), Gomá Carmona, R., & Brugué Toruella, J. (Coords.). (2005). *Análisis de los factores de exclusión social*. Fundación BBVA. Institut d'Estudis Autònoms.
- Tezanos, J.F., & Tezanos Vázquez, S. (2006). La cuestión migratoria en España. Tendencias en inmigración y exclusión social. *Sistema*, 190-191, 9-39.
- Tyulenev, S. (2009). Why (not) Luhmann? On the applicability of the social systems theory to Translation Studies. *Translation Studies* 2(2). <https://doi.org/10.1080/14781700902937706>
- Urbano Mendaña, M. (2017). *Fortalezas y debilidades de las herramientas de gestión de corpus comparable aplicadas al proceso de traducción*. Congreso internacional AIETI 7.
- Valero Garcés, C., & Mancho-Varés, G. (Eds.). (2002). *Traducción e Interpretación en los Servicios Públicos: Nuevas necesidades para nuevas realidades / New Needs for New Realities*. Universidad de Alcalá.
- Valero Garcés, C., & Gauthier Blasi, L. (2010). Bourdieu and Public Service Interpreting and Translation: Towards a social theory in PSIT. *MonTI 2trans*.
- Valero Garcés, C. (2011). Nuevos caminos con buenos maestros en traducción e interpretación. En homenaje a Valentín García Yebra. *Mutatis Mutandis*, 4(2), 155-171. <https://doi.org/10.17533/udea.mut.10160>
- Valero Garcés, C. (2014). *Communicating Across Cultures. A Coursebook on Interpreting and Translating in Public Services and Institutions*. University Press of America.
- Vargas-Sierra, C. (2010). Combinatoria terminológica y diccionarios especializados para traductores. En M. Ibáñez Rodríguez (Ed.), *Lenguas de especialidad y terminología* (pp. 17-46). Comares.
- Vigier Moreno, F.J. (2016). Teaching the use of ad hoc corpora. *Language and Law / Linguagem e Direito*, 3(1), 100-119.
- Vitalaru, B. (2019). Organización temática de terminología jurídica para traductores: proceso de elaboración de una ontología del proceso penal en español-rumano-inglés. *Hermēneus. Revista de Traducción e Interpretación*, 21, 463-514. <https://doi.org/10.24197/her.21.2019.463-514>
- Wadensjö, C. (1998). *Interpreting as Interaction*. Longman.
- Wolf, M. (1999). Zum 'sozialen Sinn' in der Translation. Translationssoziologische Implikationen von Pierre Bourdieus Kultursoziologie. *Arcadia* 34(2), 262-275.