



**How to cite this article:**

Umaña Corrales, O.C. (2023). Análisis bibliométrico del enfoque de la lingüística de corpus en estudios de terminología y lexicología en la categoría linguistics de Web of Science. *MLS-Educational Research*, 7(2), 134-151. 10.29314/mlser.v7i2.1598.

**BIBLIOMETRIC ANALYSIS OF THE CORPUS LINGUISTICS  
APPROACH TERMINOLOGY AND LEXICOLOGY STUDIES IN THE  
WEB OF SCIENCE CATEGORY *LINGUISTICS***

**Olga Clemencia Umaña Corrales**

Center for Industrial Automation (SENA) / ESAP

[olga.umana.c@gmail.com](mailto:olga.umana.c@gmail.com) · <http://orcid.org/0000-0003-2596-7798>

**Abstract.** Corpus linguistics makes it possible to analyze, describe and unveil the functioning of language, as well as to reorient its study based on the exploration of its actual use. The aim of this article is to present a descriptive bibliometric study to identify trends in the implementation of corpus linguistics in the most relevant publications in Terminology and Lexicology in the Linguistics category of the Web of Science (WoS) database between 2012 and 2021. Bibliometric elements and text-mining techniques are used to account for the most relevant authors, the institutions with the most publications and the most productive journals. Indicators of productivity, collaboration and scientific leadership are also described and plotted using the VOSviewer tool. The results show that there has been an exponential increase in the productivity of research based on corpus linguistics in the last decade and that Spain, with the University of Granada, and Belgium, with Ghent University, lead this productivity. It was also possible to determine that the most relevant authors are Hoste, Lefever, Rigouts Terryn, Faber, Rojas and Tercedor-Sánchez, and that the independent journal Terminology leads the number of publications in the area. In addition, through the Tree of Science (ToS) tool, it was possible to determine that automatic term extraction, corpus methodology, frame semantics, and lexicology and terminology work in specialized fields are the areas with the greatest research prospects.

**Keywords:** Corpus linguistics, Terminology, Lexicology, Bibliometric analysis.

**ANÁLISIS BIBLIOMÉTRICO DEL ENFOQUE DE LA LINGÜÍSTICA  
DE CORPUS EN ESTUDIOS DE TERMINOLOGÍA Y LEXICOLOGÍA  
EN LA CATEGORÍA *LINGUISTICS* DE WEB OF SCIENCE**

**Resumen.** La lingüística de corpus permite analizar, describir y develar el funcionamiento de la lengua, así como reorientar su estudio a partir de la exploración de su uso real. El objetivo de este artículo es presentar un estudio bibliométrico descriptivo para identificar las tendencias sobre la implementación de la lingüística de corpus en las publicaciones más relevantes en Terminología y Lexicología en la categoría Linguistics de la base de datos Web of Science (WoS) entre 2012 y 2021. Se utilizan elementos bibliométricos y técnicas de minería de texto para identificar los autores más relevantes, las instituciones con más publicaciones y las revistas más productivas. También se describen los indicadores de productividad, colaboración y liderazgo científico y se grafican mediante la herramienta VOSviewer. Los resultados muestran que se ha presentado un aumento exponencial en la productividad de las investigaciones basadas en lingüística de corpus en esta última década y que España, con la Universidad de Granada, y Bélgica, con la Universidad Ghent, lideran dicha productividad. También se determinó que los autores más relevantes son Hoste, Lefever, Rigouts Terryn, Faber, Rojas y Tercedor-Sánchez y que la

revista independiente Terminology encabeza el número de publicaciones en el área. Adicionalmente, al identificar los estudios más actuales mediante la herramienta Tree of Science (ToS), se estableció que la extracción automática de términos, la metodología de corpus, la semántica de marcos y el trabajo de lexicología y terminología en ámbitos de especialidad son algunas de las áreas con mayor perspectiva de investigación.

**Palabras clave:** Lingüística de corpus, Terminología, Lexicología, Análisis bibliométrico.

## Introduction

Corpus linguistics consists of a series of procedures and methods implemented to study the actual use of language from compiled texts and by means of computer technologies. The importance of the development of corpus linguistics lies in the fact that it has the potential to reorient some theories of language, facilitate the elucidation of its features, and provide a more detailed description of its structure, functions and lexical repertoires, among others. Corpus-based studies use data derived from corpora in order to explore theories or hypotheses, especially those already established in the current literature, for the purpose of validating, refuting, or refining them (McEnery and Hardie, 2011).

Disciplines such as Terminology and Lexicology have been closely related to the increased use of the corpus linguistics approach. For years, traditional dictionaries contained made-up examples, without a natural context and were compiled primarily on the basis of intuition and introspection by dictionary compilers (Hanks, 2012). Then, the use of corpora in lexicography changed this situation as corpus analysis offers lexicographers the possibility to build dictionaries based on empirical data and authentic data (Hanks, 2012).

Considering the importance and advantages of the corpus-based approach to unveil the complexities of the study of language and the fact that, to date, there has been no review of the evolution of the relationship between this approach and the areas of Lexicology and Terminology, this bibliometric analysis was developed in order to present the most relevant documents and their contributions in this regard. For this purpose, a search was made in the Web of Science (WoS) database using the *Linguistics* category by means of an equation that included the concepts of interest, which ensures the inclusion of the most important specialized journals and the identification of the articles that deal with the subject matter.

In addition, network analyses were performed to determine research productivity, evolution, and visibility, as well as the scientific activity and impact of the sources. The VOSViewer tool was also used to plot the data obtained. Although the time period is restricted to 10 years, the analysis of the information provides some of the most important sources of data, offers quantitative data and shows behaviors of the subject, so it is hoped that the results presented here will contribute to the orientation of further research in the areas analyzed.

## Method

The documents that formed the corpus for the bibliometric analysis were retrieved using the *Linguistics* category of the Web of Science (WoS) database, in a time window between 2012 and 2021. A search was made under the concept of canonical equation, i.e., "that which combines two or more concepts, with at least one of them needing to be represented by two or more synonyms" (Codina, 2020, p. 5).

Thus, the following search equation<sup>1</sup> was established for document extraction: (lingüística de corpus OR corpus OR metodología corpus OR corpus linguistics OR corpora OR corpus method\*) AND (terminología OR lexicología especializada OR lexicografía especializada OR terminografía OR terminology OR specialized lexicology OR specialized lexicography OR terminography) in the TOPIC field (*title, abstract, author keywords, and Keywords Plus*) and typologies *Article, Review, Early access, Proceedings papers*.

Table 1 shows the description of the indicators analyzed on the implementation of corpus linguistics in studies in the areas of Terminology and Lexicology.

**Table 1**  
*Description of the indicators analyzed*

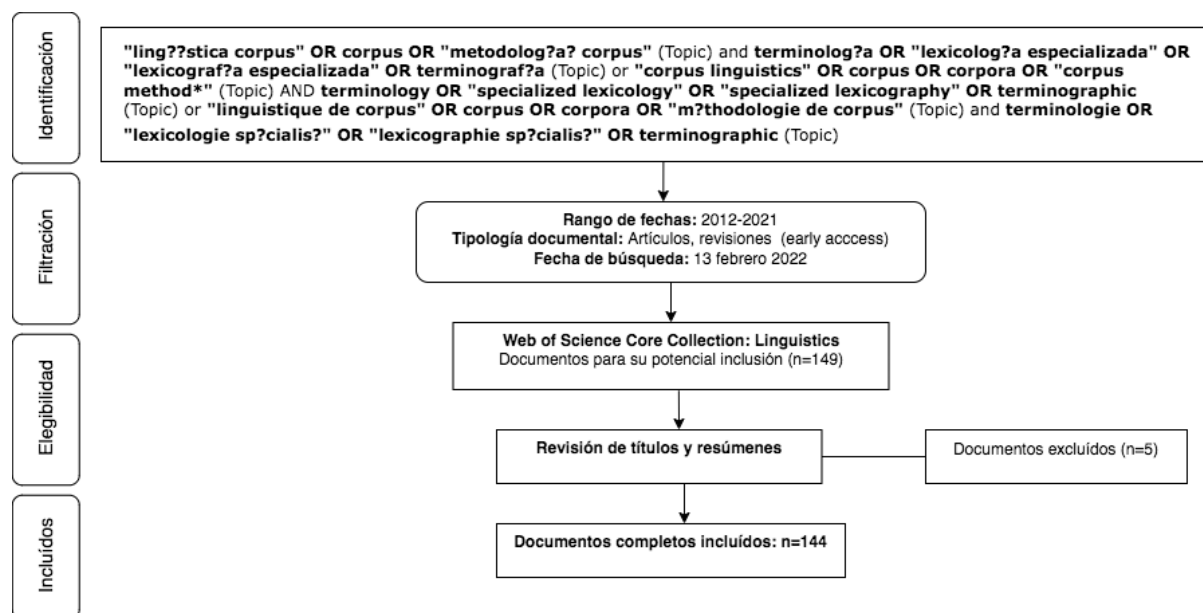
Indicator	Description
Behavior of scientific production	Reveals regularities and trends. The Price model was used to evaluate the rate of growth of scientific information in the area of interest.
Productivity of authors and countries	It shows whether a smaller number of authors gathers a larger amount of scientific production. For this purpose, Lotka's law was used, which shows that there is an unequal distribution since most of the articles are concentrated in a small number of highly productive authors, and a negative relationship with respect to their productivity of plus or minus equal to two.
Production by magazines	Establishes the source journals of scientific production and their visibility and impact indicators. The Bradford model is used, which establishes that there is a highly unequal distribution in the production of articles in journals because most of the articles are concentrated in a small number of journals.
Co-authorship network	It consists of a representation of the system that arises from the collaborative relationships between authors researching in a certain area of knowledge.
Collaboration patterns	Indicates how the authors relate to the writing process and the degree of openness of the research.
Scientific leadership	It marks the authors, countries and institutions that lead the participation in research and, therefore, the production of documents.
Keyword network	It shows the names of the main descriptors in the documents reviewed to facilitate the analysis of the thematic focus and research areas from the creation of the clusters.
Research perspectives	Reveals the most recently published papers in the area that allow to determine the future of studies in the area.

Figure 1 shows the selection process that resulted in 144 documents. The variables author, institution, country and keywords were normalized because they are the basis for generating bibliometric indicators. This search strategy made it possible to retrieve 149 documents that underwent a metadata normalization process; 5 documents were eliminated because they did not meet the inclusion criteria of the present study, especially in terms of subject matter (*corpus christi, corpus callosum*, for example).

<sup>1</sup> Truncators or masks were used to broaden the search in case of plurals and accents (example: terminología and the truncator \*(asterisk) to broaden the search for the root of a word (example: method\*)).

**Figure 1**

*Planning the document search and selection process*



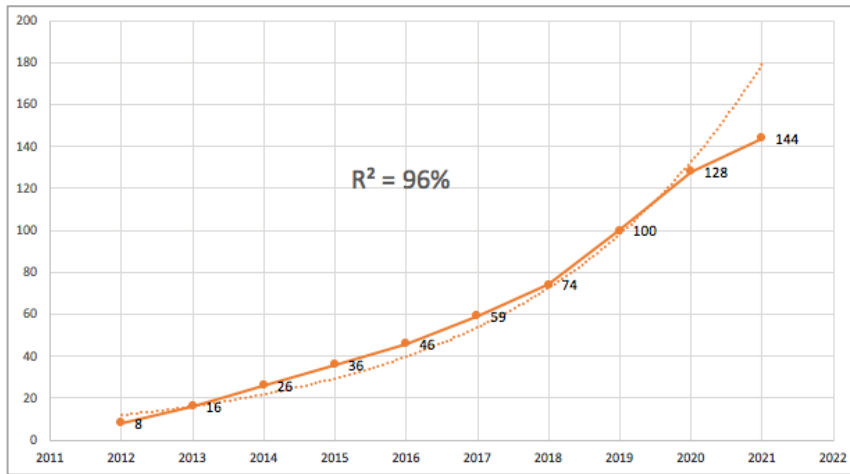
## Results

### *Annual production performance*

The cumulative scientific production was determined using Price's exponential model with an average interannual variation rate, i.e. the relative variation compared to the initial value of the variable, of 11% and a goodness-of-fit index, i.e. the discrepancy between the observed values and the values expected in the study model, of 96%.

Figure 2 shows that the subject of interest for this study has an exponential growth trend in terms of publications in the established time period, i.e., 136 between 2012 and 2021.

**Figure 2**  
Cumulative scientific production 2012-2021

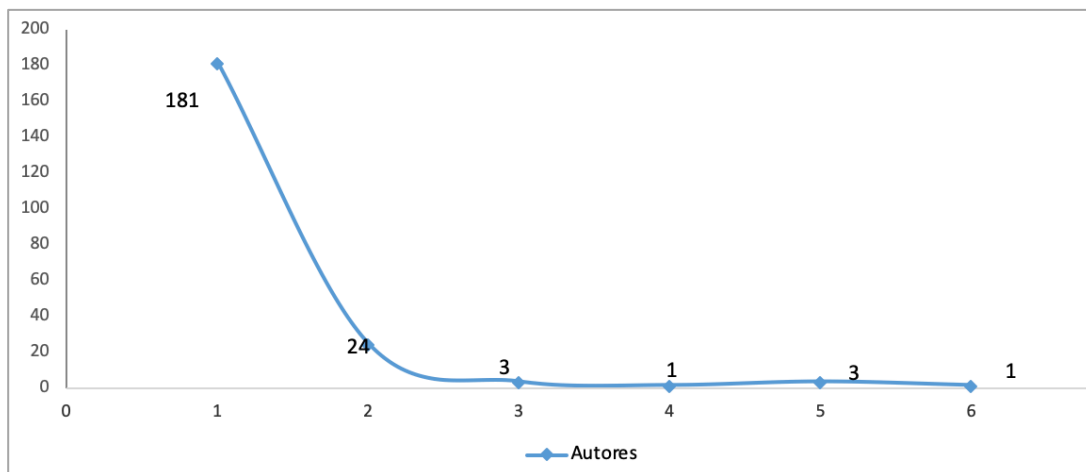


It can be observed that between the intervals from 2012 to 2013 (8 publications in each year), from 2014 to 2016 (10 publications in each year) there was no growth in production. The increase in productivity started to occur in 2017 with 3 more publications, 2018 with 2 more, 2019 with 11 more, and 2020 with 2 more. In contrast, 2021 was the only year whose publications decreased by 12. It was also established that between 2018 and 2019 there was the highest average growth in scientific productivity (73 %), while between 2020 and 2021 there was the lowest (-43 %).

**Scientific productivity leadership by author**

This indicator was obtained by applying Lotka's Law, which describes the quantitative relationship between authors and the frequency of their contributions in a given field over a period of time. Figure 3 shows that the production ratio of the 213 authors involved in the 144 papers retrieved is as follows: 181 authors contributed to 1 paper, 24 authors contributed to 2 papers, 3 authors contributed to 3 papers, 4 authors contributed to 4 papers, 5 authors contributed to 5 papers and 1 author contributed to 6 papers.

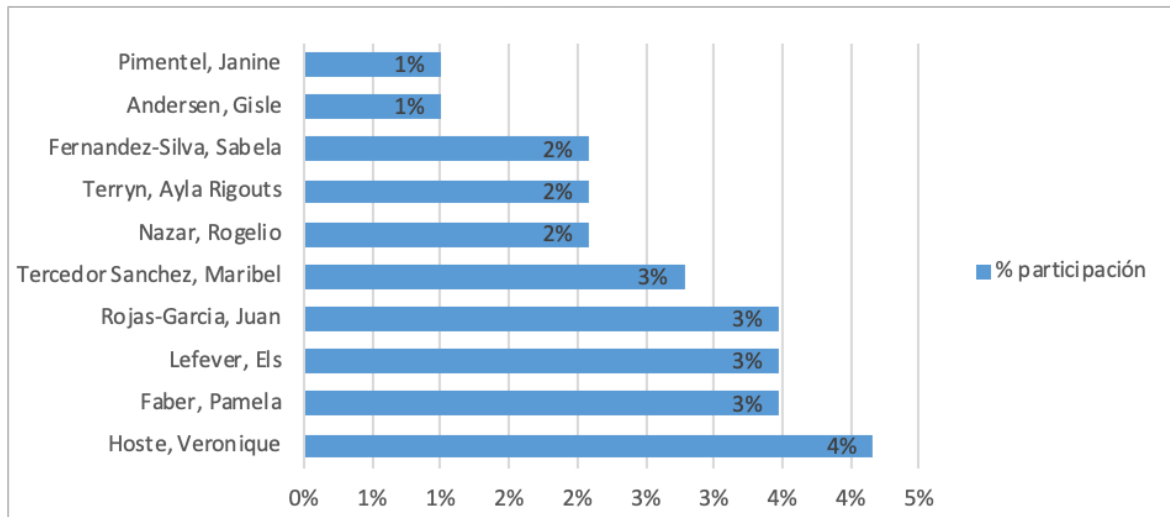
**Figure 3**  
Authors' productivity



Thus, with the applied inverse distribution model, the production nuclei are identified and the elite researchers are highlighted through their contributions. It can be established that the 8 most productive authors, with a number greater than 3 publications, participate in 34 documents and compile 23% of the scientific production for the universe of this research, i.e., the closer the authors are to the X axis, the greater the productivity in the subject. Figure 4 shows the percentage of productivity of the most specialized authors in the subject.

**Figure 4**

*Percentage of authors' productivity*

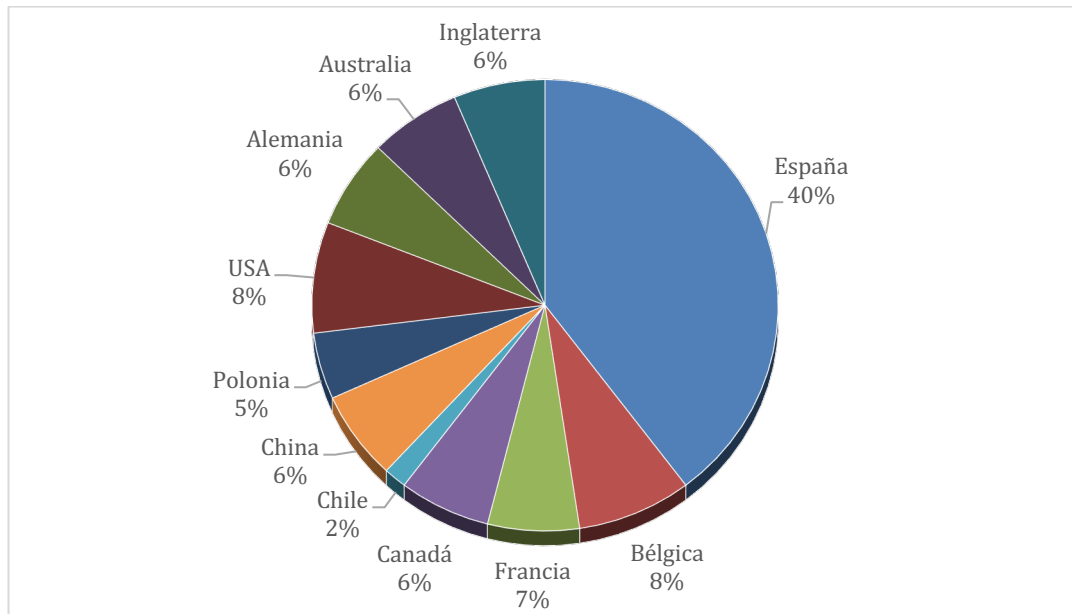


In this case, Veronique Hoste, with contributions in 6 papers and 40 citations, Pamela Faber, with participation in 5 papers and 37 citations, Els Lefever, with participation in 5 papers and 13 citations, Juan Rojas-García with participation in 5 papers and 4 citations and Maribel Tercedor-Sánchez with participation in 4 papers and 13 citations are the authors who lead the productivity with contributions above 3 papers. Rogelio Nazar, Ayla Rigouts Terryn and Sabela Fernández-Silva have participated in 3 documents each, and 9, 2 and 1 citation, respectively.

#### ***Scientific productivity leadership by country***

Scientific leadership by country is determined by the corresponding author who is the main contact and determines the institutional affiliation and, therefore, the nation to which he/she belongs. This factor also makes it possible to establish the scientific capabilities of a country in the research context. In this study, 25 Spanish and 5 Belgian institutions account for a total of 47% of the productivity in the subject of this analysis.

In addition, the productivity percentage of the countries was identified according to the number of institutions. The prevalence of the English language was also evident with 78.5% of the publications, followed by the Spanish language with 15%. Figure 5 shows the 12 countries contributing more than 3 documents in the area studied.

**Figure 5***Percentage of productivity by country according to institutions*

With respect to the 144 basic documents of this study, Spain is consolidated as the leading country in productivity with 57 documents published by 25 institutions (40 %), followed by Belgium with 10 documents published by 5 institutions (8 %). This is followed by France with 6 papers published by 4 institutions (7 %), Canada with 5 papers published by 4 institutions (6 %), Chile with 5 papers published by 1 institution (2 %), China with 5 papers published by 4 institutions (6 %), Poland with 5 papers published by 3 institutions (5 %), the US with 5 papers published by 5 institutions (8 %), Germany with 4 papers published by 4 institutions (6 %), Australia with 4 papers published by 4 institutions (6 %) and England with 4 papers published by 4 institutions (6 %).

### ***Scientific productivity leadership by institution***

Considering that contemporary universities have three inherent functions: teaching, research and extension, it is important to make visible their leadership in scientific research productivity. This indicator is also determined through the institutional affiliation of the author of the correspondence and allows establishing the scientific capabilities of an institution to make contributions in different areas of the research context. This study found 89 institutions that have published papers related to corpus linguistics. Table 2 presents the institutions with a number of publications greater than 3, the country to which they belong and the number of documents they contribute.

**Table 2**  
*Scientific leadership by institution*

<b>University</b>	<b>Country</b>	<b>Doc.</b>	<b>%</b>
Granada University	Spain	15	10 %
Ghent University	Belgium	6	4 %
Pontifical Catholic University of Valparaíso	Chile	5	3 %
Córdoba University	Argentina	4	3 %
Polytechnic University of Valencia	Spain	4	3 %
Valladolid University	Spain	4	3 %
Vigo University	Spain	4	3 %
Hong Kong University	China	3	2 %
Paris University	France	3	2 %

The University of Granada (Spain) leads with 15 published papers (10%), followed by the University of Ghent (Belgium) with 6 published papers (4%). Authors contributing more than 5 papers belong to these two institutions.

This is followed by the Pontificia Universidad Católica de Valparaíso (Chile) with 5 published papers, followed by the Universidad de Córdoba (Argentina), the Universidad Politécnica de Valencia (Spain), the Universidad de Valladolid (Spain), and the Universidad de Vigo (Spain), each with 4 published papers (3% each). They are followed by the University of Hong Kong (China) and the University of Paris (France), each with 3 published papers (2% each). Finally, the remaining 80 institutions contribute 1 or 2 published articles (1%) to the total data.

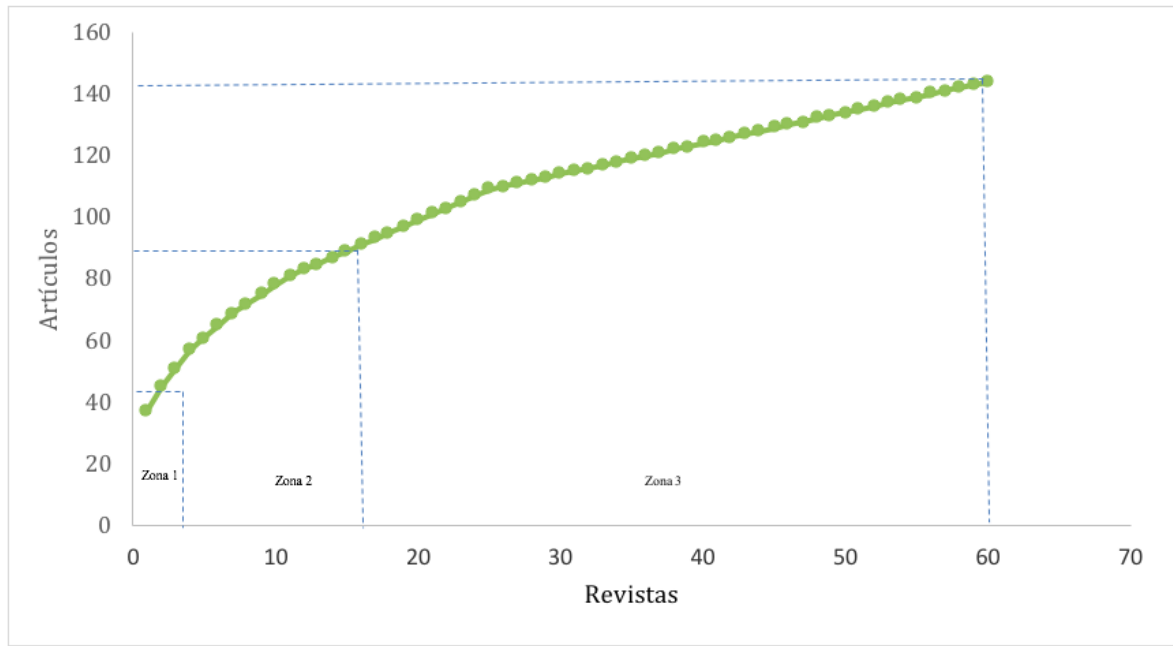
### ***Scientific productivity leadership by journals***

The Bradford dispersion model made it possible to identify the most relevant periodicals and to observe the location of the journals most used by researchers in the topic of interest in the core zone, the intermediate zone or the peripheral zone, which show the productivity of the journals from highest to lowest. Figure 6 shows the logarithmic distribution of the 60 journals that published articles on the subject of interest of this study.



**Figure 6**

*Distribution of magazines by zones according to the Bradford model*



The application of the aforementioned model in this analysis made it possible to establish that in Zone 1 (core) there are 2 journals with 45 articles, which account for 31% of the total number of publications in the sample. Meanwhile, in Zone 2 (intermediate) there are 15 journals with 48 published articles (33%) and in Zone 3 (periphery) there are 43 journals with 51 published articles (35%).

Table 3 shows the characteristics of the 2 journals located in Zone 1. These are *Terminology*, published by John Benjamins Publishing Company and *Onomázein*, published by the Pontificia Universidad Católica de Chile. The academic level of the papers published in both journals is guaranteed by the objective review of external international judges, recruited from the international community of specialists.

**Table 3**

*Description of Core Zone magazines*

Magazine	Number of documents	% 144	Quartile category: <i>Linguistics WoS</i>	JIF 2020	H-index
Terminology	37	26 %	Q2	0.826	25
Onomázein	8	6 %	Q2	0.419	12

According to the Scimago website, *Terminology* is an independent cross-cultural and interdisciplinary journal. It focuses on the discussion of (systematic) solutions, not only to the linguistic problems encountered in Translation, but also, for example, to the (monolingual) problems of ambiguity, reference and development in multidisciplinary communication. In the *Linguistics and Language* category, *Terminology* is currently ranked Q2.

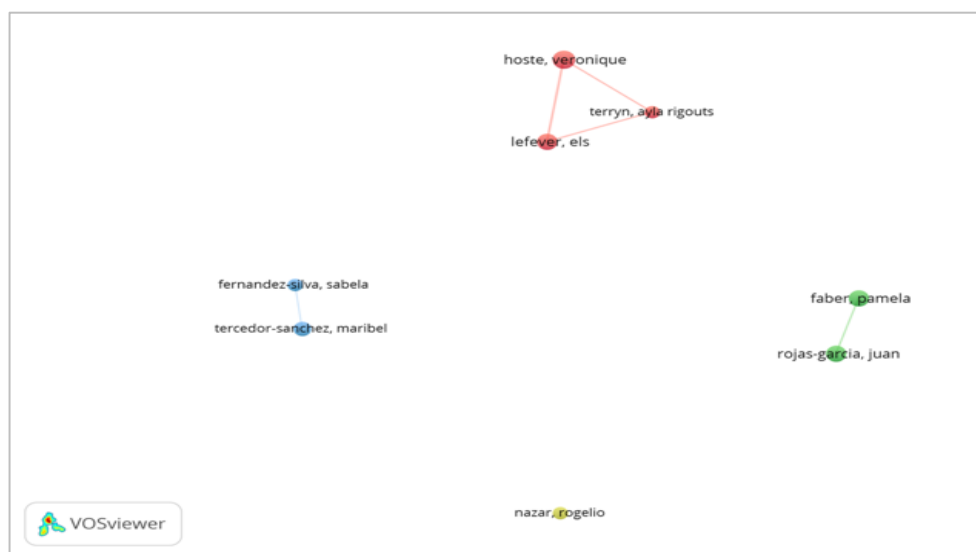
According to the Scimago website, *Onomázein –Journal of Linguistics, Philology and Translation–* welcomes previously unpublished papers originating from scientific research in the different branches of Theoretical and Applied Linguistics, in Classical, Indo-European, Romance and Hispanic Philology, as well as in Translation Theory and Terminology, and relevant studies in indigenous languages. In the *Linguistics and Language* category, *Onomázein* is currently ranked Q2.

### ***Collaboration networks between authors and between institutions***

From the 213 authors identified in the 144 documents, those with at least 3 collaborations were selected for the construction of the co-authorship network; this resulted in 8 authors that formed 4 clusters. Next, a co-authorship matrix is constructed in which the times that these top authors worked together are identified.

Figure 7 shows that the red cluster consists of authors Veronique Hoste, Ayla Rigouts Terryn and Els Lefever, linked to Ghent University. The blue cluster is formed by Sabela Fernández (Universidad Católica de Valparaíso) and Maribel Tercedor-Sánchez (Universidad de Granada). The green cluster is formed by Pamela Faber and Juan Rojas-García (University of Granada) and, finally, Rogelio Nazar of the Catholic University of Valparaíso with a solo contribution.

**Figure 7**  
*Co-authorship network*



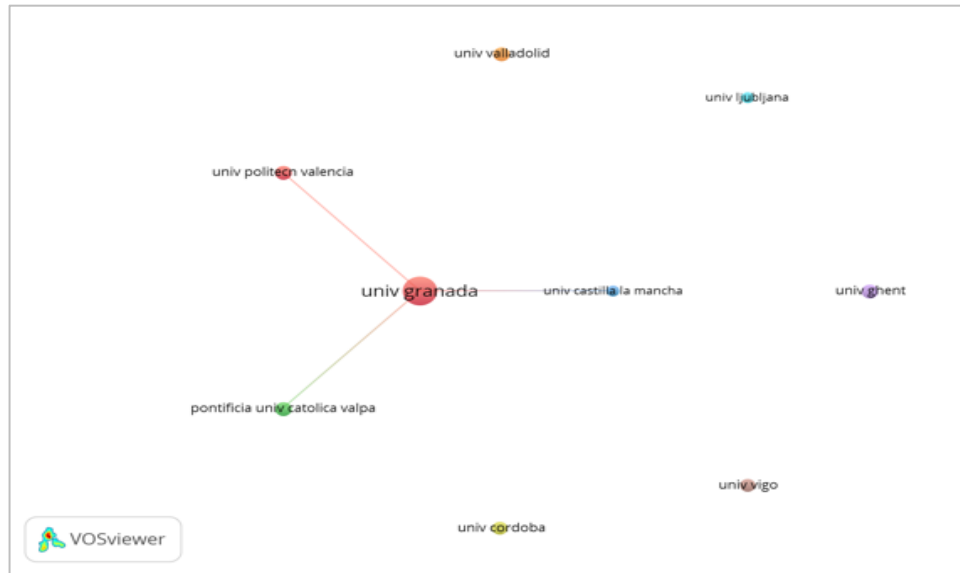
*Note.* minimum bond strength of the items: 0. Of the 213 authors, 8 met the threshold (3 papers); standardization method: association strength; attraction: 1; repulsion: -2; grouping resolution: 1,0. Result: items: 8; clusters: 4; Links: 5; *Total Link strength*: 15.

Of the 116 institutions participating in the publications of this analysis, 9 were identified by means of a minimum cut-off point of 3 papers per institution (citation was not included) to construct the institutional collaboration network. Figure 9 shows that the University of Granada leads the production of this topic in the *Linguistics* category and, therefore, is established as the point of institutional collaboration with the Universities of Castilla de la Mancha and Valencia, both in Spain, and the Pontificia Universidad Católica de Valparaíso in Chile.

In addition, there was joint participation in 20 documents and leadership of 15 institutions.

**Figure 8**

*Collaboration network between institutions*



*Note:* minimum bond strength of the items: 0. Of the 116 institutions, 9 met the threshold (3 documents); standardization method: *association strength* ; attraction: 4; repulsion: -5; grouping resolution: 1,0 Result: items: 9; clusters: 8; Links: 3.

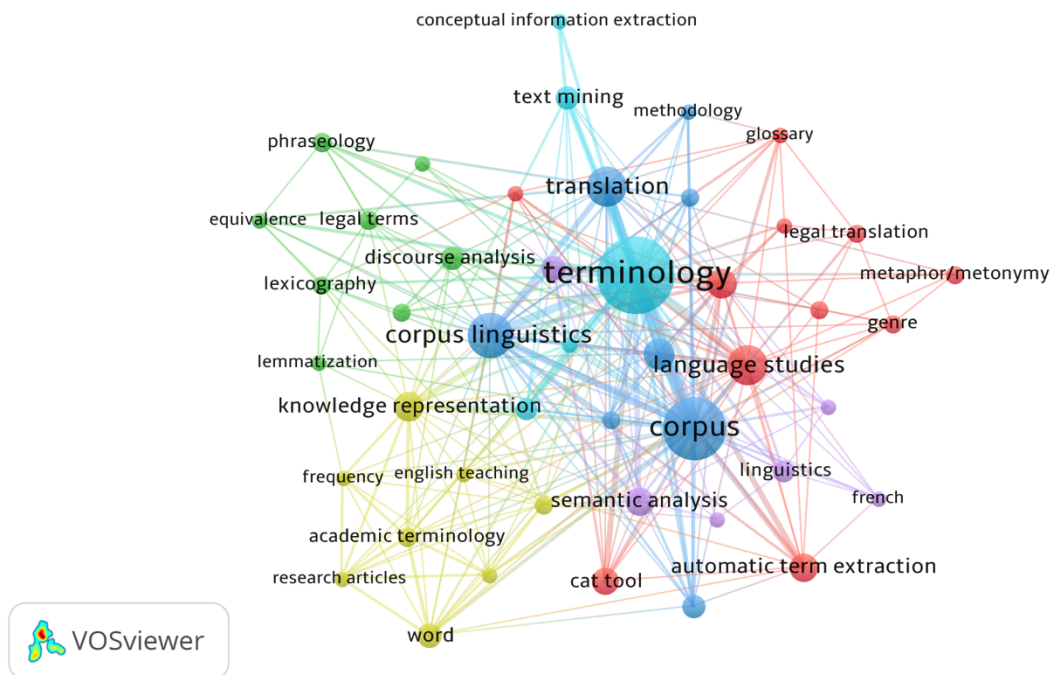
Notably, the article "Pragmatic borrowing" by author Andersen Gisle of the Norwegian School of Economics has the highest number of citations in the entire dataset (45 citations in total). This article explores the notion of pragmatic borrowing, that is, the incorporation of pragmatic and discursive features of a source language into a target language. The study illustrates how pragmatic functions are transferred cross-linguistically, through notions such as functional stability, adaptation, narrowing, broadening and change. It also illustrates the extent of borrowing of set phrases and colloquialisms, focusing especially on expletives, interjections, and English discourse markers that have recently appeared in Norwegian.

***Cooccurrence networks***

In the sample of 144 articles, 634 keywords were obtained, which were normalized to 453 after creating and applying a list of terms or thesaurus. To simplify the representation of the knowledge structures, only those keywords whose frequency was  $\geq 3$  were considered (a lower threshold would have generated a very long list of keywords). Before creating the co-word network, the keywords *named river* and *wine testing notes* were manually removed because they were related to the word corpus, but in a different field than linguistics.

Figure 9 shows the 6 clusters obtained. The interpretation of the map took into account the number of keywords within each thematic group, the number of occurrences of each keyword, their interrelation and their spatial location.

**Figure 9**  
*Keyword co-occurrence clusters*



Note: minimum bond strength of the items: 0. Of the 634 keywords (author + keyword plus), 453 met the threshold (3 occurrences); normalization method: association strength; attraction: 1; repulsion: -3; grouping resolution: 1,0.

The colors indicate clusters of keywords with some kind of relationship between them according to the association obtained through the VOSviewer program. An analysis was also made of the thematic focus of each cluster based on the concepts conveyed by its key words. Table 4 shows the clusters and the thematic focus of each cluster, their keywords, the number of occurrences.

**Table 4**  
*Clusters and thematic approaches*

Cluster	Keyword	Occurrences	Thematic focus
Cluster 1. Red	automatic term extraction	10	Translation
	cat tool	9	
	comparable corporate	4	
	español	11	
	genre	4	
	glossary	3	
	language studies	20	
	legal translation	4	
	metaphor/metonymy	4	
	natural language processing	3	
standardization	3		
Cluster 2. Green	discourse analysis	7	Translation studies
	equivalence	3	
	eu terminology	3	
	legal terms	5	
	lemmatization	3	
	lexicography	4	
	phraseology	5	
	science	4	
Cluster 3. Blue	corpus	49	Specialized translation
	corpus linguistics	25	
	lexicology	12	
	medical terminology	7	
	medical translation	4	
	methodology	3	
	spanish	4	
	translation	20	
Cluster 4. Yellow	academic terminology	5	Didactics of languages
	collocation	4	
	engineering english	3	
	english teaching	3	
	frequency	3	
	knowledge representation	11	
	research articles	3	
	word	8	
Cluster 5. Purple	distributional semantics	3	Terminology
	french	3	
	grammar	3	
	linguistics	6	
	semantic analysis	10	
	term extraction	8	

Cluster 6. Light blue	conceptual information extraction	3	Terminotics
	FunGramKB	6	
	ontology	7	
	terminology	4	
	text mining	7	

### **Research perspectives**

Using the methodology proposed by Robledo, Osorio and López (2014), the retrieved documents were uploaded to the *Tree of Science* (ToS) web platform in order to classify the articles according to their position in the tree, an analogy used by the aforementioned authors to determine the following three groups:

- The root with the theoretical referents of the subject, i.e., the classic authors who laid the foundations in this field of study and who are cited more frequently than the rest of the authors. These articles are dated between 1991 and 2010.
- The trunk with structural items based on the root items, but with a more elaborate theoretical framework. These items have dates between 2012 and 2019.
- The leaves with those articles that show the different current perspectives on the subject, using the findings of the root and stem articles. These items are dated between 2013 and 2021.

Table 5 shows this last category as it is the most relevant for the present study. The leaf articles are also characterized by having as reference the writings that make up the roots and the trunk.

**Table 5**  
*Most current articles on the subject*

<b>Author</b>	<b>Article</b>
Rojas-García, J. (2021).	Extraction of Terms Semantically Related to Colponyms: Evaluation in a Small Specialized Corpus.
Kwong, OY. (2021).	User-driven assessment of commercial term extractors.
Rojas-García, J. (2020).	Application of Topic Modelling for the Construction of Semantic Frames for Named Rivers.
Ortego-Anton, MT. (2021).	e-DriMe A Spanish-English frame-based e-dictionary about dried meats.
Terryn, AR. (2021).	HAMLET Hybrid Adaptive Machine Learning approach to Extract Terminology.
Unzalu, IZ. (2021).	[en] Current challenges in the development and learning of the oral and written academic registers in Basque.
Polyakova, O. (2021).	An integrated approach to the higher education terminology in Spanish-Russian university texts.
Trigo, ES. (2021).	The terms manifestation (fr) and manifestation (es) in biomedical journal articles: a corpus-based research.
Rodríguez, CIL. (2020).	Predicative frames for the concept SIGN AND SYMPTOM in Spanish Medical Texts.
San Martín, A. (2020).	Present and future of the terminological knowledge base EcoLexicon.

Hoste, V. (2019).	The trade-off between quantity and quality. Comparing a large crawled corpus and a small focused corpus for medical terminology extraction.
Rieder-Bunemann, A. (2019).	Capturing technical terms in spoken CLIL A holistic model for identifying subject-specific vocabulary.
Cardenas, BS. (2019).	Eliciting specialized frames from corpora using argument-structure extraction techniques.
Santos, IG. (2019).	The economy is sick - l'economie est malade. The chronology of the crisis through terminology.
Terryn, AR. (2019).	Validating multilingual hybrid automatic term extraction for search engine optimisation: the use case of EBM-GUIDELINES.
Perinan-Pascual, C. (2018).	A framework of analysis for the evaluation of automatic term extractors.
Ghazzawi, N. (2018).	Automatic extraction of specialized verbal units A comparative study on Arabic, English and French.
Costa, LA. (2018).	Explicit term variation in Brazilian lexicography: proposal for its representation in the micro structure of the Brazilian Lexicography Dictionary.
Perinan-Pascual, C. (2018).	DEXTER: A workbench for automatic term extraction with specialized corpora.
Gagne, AM. (2016).	Opposite relationships in terminology.
Nazar, R. (2016).	Distributional analysis applied to terminology extraction First results in the domain of psychiatry in Spanish.
Hanouille, S. (2015).	The efficacy of terminology-extraction systems for the translation of documentaries.
Lefever, E. (2014).	HypoTerm Detection of hypernym relations between domain-specific terms in Dutch and English.
Silva, SF. (2013).	The influence of the disciplinary field on terminological variation: A corpus-based study in the interdisciplinary domain of fishing.

*Note:* Own elaboration.

Once the most recent articles were identified, a simple data mining process was performed using the article titles and a word cloud was created using the Voyant tool. This was done in order to determine the topics that are currently being worked on and that, in turn, lay the groundwork for future research. Figure 10 shows the terms with the highest frequency in each title.





As for the results of scientific productivity leadership by author, they were in line with the applied distribution model of Lotka's Law, which establishes an inverse relationship in which a few authors specialize in a field of knowledge and, therefore, concentrate the greatest volume of publications, while many authors will publish very few publications.

It is noteworthy that the results of scientific productivity reflect the coordinated efforts of institutions and academics in the search for interdisciplinary and increasingly detailed descriptions of linguistic phenomena. It was found that the total production is represented in 60 scientific journals with a participation of 213 authors. Although for the purposes of this study the quantitatively most outstanding data are mentioned, it is important to note that, apart from Spain and Belgium, 13 other countries contribute three or more documents to the publications per country.

In terms of productivity by country, Spain justifies its leadership, since the University of Granada is among the top ten; it has also been ranked first in translation and interpretation studies for several years, and has programs for teaching languages such as Portuguese, Italian, Danish, Dutch, Czech, Polish, Romanian, Bulgarian, Russian, Modern Greek, Hebrew, Arabic and Turkish. It also houses the only Russian Center that the Russkiy Mir Foundation maintains in Spain. The University of Granada is one of the best public universities in Spain and is ranked 494th in the QS *Academic Ranking of World Universities 2023*. For its part, Belgium is in second place in this leadership, with Ghent University in the lead. This institution ranks 74th in this list of more than 2500 research institutions worldwide in 2022 and is the highest ranked Belgian university in the *Academic Ranking of World Universities*.

In relation to the leadership of scientific productivity by journals, the present analysis has correspondence with the hypothesis of Bradford (1934), who postulated that most of the articles on a specialized subject could be published by a few journals especially dedicated to that subject, in conjunction with certain frontier journals and other general or dispersion journals (Urbizagástegui Alvarado, 2015). In this case, *Terminology* and *Onomázein* are at the forefront of publications in the area of interest of this analysis.

According to its website, *Terminology* pays special attention to new and developing subject areas such as knowledge representation and transfer, software tools, expert systems and terminology databases. *Terminology* covers general (theory and practice) and specialized fields (LSP), such as Physics, Biomedical Sciences, Technology, Engineering, Humanities, Management, Law, Arts, Business Administration, Commerce, Corporate Identity, Economics, Methodology and any other area where Terminology is essential to improve communication. *Onomázein* is aimed primarily at specialists and is intended to serve as an effective vehicle for scientific exchange among researchers in the linguistic sciences.

One of the already published papers that has similar findings was by Liao and Lei, who in 2017 developed a bibliometric analysis of the WoS SSCI (*Social Science Citation Index*) using the *Linguistics OR Language Linguistics* category, between the years 2000 and 2015. Its purpose was to know the number of documents implementing corpus methodology. The results showed that the production of publications related to the corpora had increased considerably during those 15 years. In addition, it was noted that while traditional scientific powers, such as the United States, play a leading role in this area, countries such as China also have an impact in the field. The most important result was related to the fact that corpora have permeated a wide range of research areas in linguistics and have changed, at least in terms of methodology, these areas.

Even when this analysis only includes a ten-year interval, the results obtained through text mining tools and biometric analysis techniques were consistent. Further research could be

developed by increasing the volume of information and using additional tools to expand the results and reveal other trends associated with the topic studied.

### References

- Ardanuy, J. (2012). *Breve introducción a la bibliometría*. <http://diposit.ub.edu/dspace/bitstream/2445/30962/1/breve%20introduccion%20bibliometria.pdf>
- Cobo, M. J. (2011). Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools. *Journal of the American Society for Information Science and Technology*, 1382-1402. <https://doi.org/10.1002/asi.21525>
- Codina, L. (2020). *Interfaces de búsqueda en bases de datos académicas: Análisis comparativo de Scopus, WoS, Google Scholar y Microsoft Academic*. <https://www.lluiscodina.com/interfaces-de-busqueda/>
- Galves, C. (2018). El campo de investigación del Análisis de Redes Sociales en el área de las Ciencias de la Documentación: un análisis de co-citación y co-palabras. *Revista General de Información y Documentación*, 28(2), 455-475. <http://dx.doi.org/10.5209/RGID.60805>
- Hanks, P. (2012). *Corpus evidence and electronic lexicography*.
- Liao, S., & Lei, L. (2017). What we talk about when we talk about corpus: A bibliometric analysis of corpus-related research in linguistics (2000-2015). *Glottometrics*, 38, 1-20.
- Maltrás Barba, B. (2003). *Los indicadores bibliométricos. Fundamentos y aplicación al análisis de la ciencia*. Trea S.L.
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Merton, R. K. (1968). The Matthew Effect in Science: The reward and communication systems of science are considered. *Science*, 159(3810), 56-63.
- Pulgarín, A., Carapeto, C., & Cobos, J. M. (2004). Análisis bibliométrico de la literatura científica publicada en Ciencia. *Revista hispano-americana de ciencias puras y aplicadas (1940-1974)*. *Inf. Res.*, 9(4).
- Robledo, S.; Osorio, G.; Lopez, C. (2014). Networking en pequeña empresa: una revisión bibliográfica utilizando la teoría de grafos. *Revista Vínculos*, 11(2), 6-16. <https://revistas.udistrital.edu.co/index.php/vinculos/article/view/9664>
- Urbizagástegui Alvarado, R. (2016). El crecimiento de la literatura sobre la ley de Bradford. *Librarianship Research: Archivonomía, Bibliotecología e Información*, 30(68), 51-72. <https://doi.org/10.1016/j.ibbai.2016.02.003>

**Date received:** 29/08/2022

**Revision date:** 22/09/2022

**Date of acceptance:** 29/11/2022