

---

## Visualizing Studies on Language Typology: A Bibliometric Analysis over 30 Years

Danlin Yang<sup>1</sup>

### Abstract

Language typology is a branch of linguistics that explores languages' classification and structural patterns, providing insights into their underlying ontological characteristics. Over the past few decades, this field has gained significant attention from the global academic community, driven by a growing interest in understanding linguistic diversity and universality. Despite its rapid development, a comprehensive analysis of research trends in this area remains limited. Therefore, this study aims to fill this gap by examining the evolution of language typology research from 1993 to January 2023. Utilizing the bibliometric tool CiteSpace, this study seeks to identify key research hotspots and emerging trends within the field. The research methodology involves a detailed bibliometric analysis, which allows for mapping scholarly publications and identifying significant contributions, collaborations, and citation networks. By analyzing data over 30 years, this study offers a longitudinal perspective on the progress and changes in the focus of language typology research. The findings reveal several notable trends, including a shift towards interdisciplinary approaches and the integrating of computational tools to analyze linguistic data. The research highlights emerging areas such as typological universals, linguistic complexity, and the interaction between language structure and cognition. Based on these results, the study recommends further exploration of under-researched languages and applying advanced computational techniques to deepen our understanding of typological patterns. Furthermore, it suggests fostering more international collaboration to address linguistic diversity across regions and language families. This study provides valuable insights for linguists interested in language typology research's current state and future directions.

**Keywords:** *Language Typology, CiteSpace, Bibliometric Analysis*

### A. Introduction

Typology is the classification of language or components of languages based on shared formal characteristics. The study aims to identify crosslinguistic patterns and correlations between these patterns (Aikhenvald, 2024). Based on the need for cross-linguistic historical comparisons, traditional typology originated in the 19th century. Since traditional typology categorizes languages mainly on the basis of morphological features, it is also called morphological typology. Since the 1960s, contemporary typology has emerged. Compared with traditional typology, contemporary typology focuses more on the examination of inflectional types (Ying & Chen, 2024). In the 1960s, Greenberg's theory of sequential commonality pioneered contemporary language typology (Greenburg, 1957&1963). Through nearly half a century of research, contemporary language typology is "linguistic commonality and language typology", which no longer aims at classifying languages, but mainly explores the commonalities of human languages through cross-linguistic comparisons, and explores the laws of "all variations". The purpose and orientation of the study is very different from that of traditional typology.

---

<sup>1</sup>Ocean University of China, China. [yangdanlin@stu.ouc.edu.cn](mailto:yangdanlin@stu.ouc.edu.cn)

Under the guidance of different theories, the hypotheses and commonalities proposed by language typology studies have been expanded. A large number of studies have put forward theoretical hypotheses from a language typology perspective, but there are relatively few review articles on the overall development of language typology research. In order to prepare for grasping the hot spots and development trends of foreign second language pragmatics research, this paper intends to visualize and analyze foreign language typology research articles with the help of CiteSpace to sort out the research results in this field and provide reference for domestic language typology research and practice.

The general objective of this study is to analyze the development of research in the field of language typology from 1993 to January 2023, with the aim of providing insights into key trends, contributions, and future directions in language typology research. Specifically, the study seeks to identify major trends in foreign language typology research through bibliometric analysis, uncover research hotspots and emerging areas within the field, and analyze shifts in research focus as well as significant contributions over the past three decades. Additionally, it aims to map collaboration patterns among scholars in the field at an international level. Finally, the study provides recommendations for future research, particularly in under-explored languages and interdisciplinary approaches to deepen the understanding of typological patterns.

*Visualizing Studies on Language Typology: A Bibliometric Analysis over 30 Years* is grounded on several critical assumptions and addresses an urgent need in linguistics. The primary assumption of this research is that language typology, as a field that investigates structural similarities and differences across languages, has undergone significant transformation over the past three decades. This transformation has been shaped by evolving theoretical perspectives, technological advancements, and an increasingly globalized research community. Another assumption is that a bibliometric approach can effectively capture these changes by mapping publication trends, key contributors, and emerging areas of interest. The urgency of this study stems from the rapid expansion of linguistic data and the growing complexity of the field, which requires a systematic review to synthesize past research and identify future directions. As new languages are increasingly documented and computational tools for analyzing linguistic data become more sophisticated, there is a pressing need to visualize and understand how language typology research has progressed and where it is heading. This is particularly important as linguistic diversity continues to dwindle, and the opportunity to document under-studied languages diminishes. Additionally, there is a strong academic and practical interest in exploring the evolution of language typology to highlight key contributions and address the gaps in current research. This study contributes to advancing the field and encourages more targeted, data-driven, and interdisciplinary investigations by providing a comprehensive analysis of trends, research hotspots, and collaborative networks. The findings will also help scholars and institutions allocate resources more efficiently and guide future research in the most critical areas for the ongoing development of language typology.

## **B. Methods**

### **1. Research Design**

This study employs a bibliometric research design, focusing on analyzing academic literature within the field of language typology over 30 years from 1993 to January 2023. The bibliometric approach enables the identification of key trends, research hotspots, and emerging frontiers in the field. CiteSpace, a widely used bibliometric tool, is utilized to visualize and analyze patterns in co-citations and co-occurrence networks, providing a comprehensive overview of the field's development.

## **2. Research Procedure**

The research begins with the construction of a relevant database. The Social Science Citation Index (SSCI) within the Web of Science (WoS) Core Collection was selected as the primary source of data. A search query was executed using the subject term "language typology" with the filters "English" and "article," under the categories of "language linguistics" or "linguistics." This process retrieved 1667 articles published from 1993 to January 2023. After removing duplicates through CiteSpace's built-in de-duplication function, 1658 valid documents were included for further analysis. These documents formed the basis for identifying research trends and summarizing the field's progress.

## **3. Data Collection Techniques**

Data collection involved downloading relevant articles from the Web of Science database. Articles were selected based on predefined criteria, including publication language (English) and document type (article). After extraction, the data were cleaned and prepared using CiteSpace to remove any duplicate records, ensuring the reliability of the dataset for subsequent analysis.

## **4. Data Analysis Techniques**

CiteSpace 6.2.R2 was used to conduct both co-citation and co-occurrence network analyses. The time slicing was set to five-year intervals ("Years per Slices") to provide a clear view of the evolution of language typology research over time. The analysis also incorporated the "pruning the merged network" function to refine the visualization, enhancing the clarity and concentration of the co-citation and co-occurrence maps. This method allowed for the identification of key authors, influential papers, and thematic clusters, providing insights into the intellectual structure and emerging trends in language typology research.

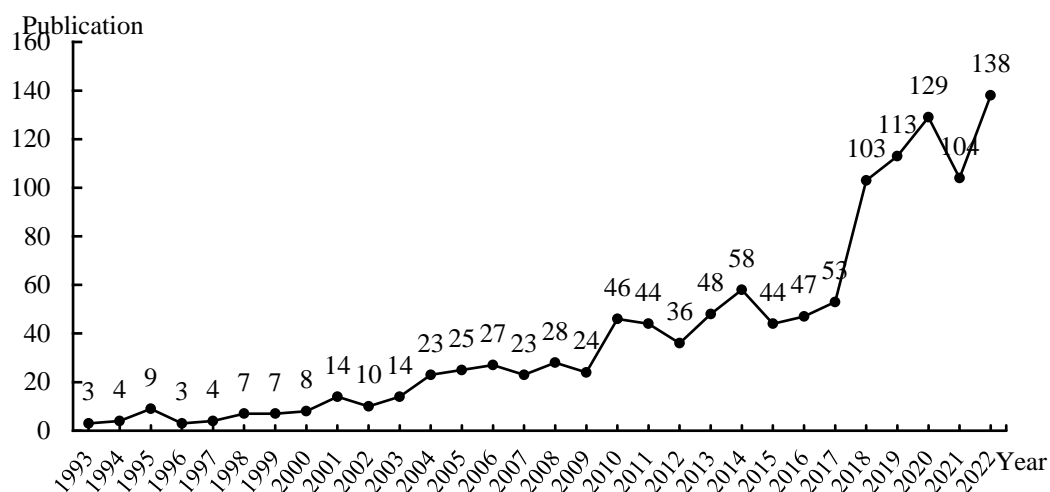
## **C. Findings and Discussion**

### **1. Temporal Distribution of Language Typology Researches**

The number of papers issued and annual publications in a field visually represents the level and development of academic research in this field. As of the date of data collection, there was no literature published in 2023, so it is not presented in the graph. From 1993 to 2022, the number of foreign language typology studies has shown an overall dynamic upward trend and good growth. By comparing the average annual number of publications, this paper argues that foreign language typology studies in the past 30 years can be divided into the following stages.

The first stage is the slow period (2002 and before). Except for 2002, when the annual number of articles was 10, the annual number of articles in this period was mostly in single digits, reflecting that the studies on the topic of "language typology" abroad were still in the initial development stage, and the research results were relatively few. The second stage, the period of steady growth (2003 - 2007). After entering the second stage, the annual number of foreign language typology research articles surged for the first time in 2004, with an annual average of 17 articles, more than double the number in 2003. between 2003 and 2007, the annual number of articles showed an overall upward trend, with an annual average of 19 articles, showing a fluctuating growth, indicating that the language typology studies began to attract the attention of scholars at this stage. The third stage, the period of rapid growth (2008-2017), was the first time that the annual number of articles published in foreign language typology studies exceeded 30 in 2009, and the increase in the number of articles in this period was particularly obvious in 2017. The average annual number of publications in this period was 61.4, with an average increase of 223.18% compared with the previous period, indicating that the studies on language typology in this decade were in an active growth stage and the academic community increasingly valued the related studies. The last stage is the explosion period (2018 - 2022).

Entering the fourth stage, the annual number of articles published in foreign language typology research surged in 2019, exceeding 150, nearly doubling compared with 2018; the average annual number of publications in this period was 173.2, and the average number of articles published in this stage increased by 182.08% compared with the previous stage. The overall upward trend and fluctuating growth of the annual number of articles published during 2018-2022 indicate that the heat of language typology research rose greatly in this stage.



**Figure 1** The Temporal Distribution of Language Typology Research (1993–2022)

## 2. Keywords

Keywords serve as a reflection of the core themes and research content of academic articles. In this study, we aim to visualize and analyze the keywords present in foreign language typology literature to identify and understand the key research hotspots. By examining keyword patterns, we can gain valuable insights into the most frequently studied topics and emerging trends in the field of foreign language typology research over the past three decades. This analysis helps pinpoint the areas that have attracted the most scholarly attention, thereby offering a clearer picture of the field's current focus and future directions.

### *Keywords Co-occurrence*

Co-occurring network analyses were conducted using keywords, including "node type," to identify patterns in recent research trends. The results revealed that popular topics in language typology have shifted towards emerging areas of interest, reflecting the field's evolving focus. This analysis highlighted key themes and concepts that have gained significant attention in recent years, providing insights into the current research priorities and the intellectual landscape of language typology. These findings underscore the importance of understanding co-occurrence patterns to map the ongoing development of this field.

**Table 1** Parameters used in keywords co-occurrence analysis in CiteSpace

<b>Time span</b>	Form 1993 Jan. to 2023 Jan.
<b>Time slice</b>	5 years
<b>Pruning</b>	Pruning the merged network
<b>Node type</b>	keywords
<b>Term Source</b>	title/abstract/author keywords/keywords plus

The keyword co-occurrence knowledge graph analysis of foreign second language discourse studies finally generated 350 nodes with 1423 connected lines (see Figure 1). The nodes in Figure 1 represent keywords, and the size of the node indicates the frequency of the keyword, and the higher the frequency of the keyword, the larger the node. The linkage between

nodes represents the connection between keywords. Table 2 ranks the keywords of foreign language typology studies by the frequency of keywords above 20. From Table 2, we can see that, excluding the subject term “language typology”, the core keywords of foreign bilingual pragmatic research are language, typology, English, language contact, syntax, acquisition, semantics, grammar, word order, linguistic typology, speech and agreement etc.



**Figure 2** Keywords Co-occurrence in Language Typology (1993-2022)

The figure represents the co-occurrence network analysis of key research topics in language typology from 1993 to 2023, generated using CiteSpace. The size of each keyword reflects its frequency of occurrence in the literature, while the lines connecting the nodes indicate the relationships between these topics. Major topics such as "language," "typology," "word order," "grammar," "language contact," and "acquisition" appear as the most prominent research areas, highlighting their central role in language typology studies. The large size of these keywords suggests that they have been recurring focal points over the past three decades. Additionally, the network shows how these topics are interconnected, with "language contact" being linked to "grammar" and "typology," indicating research on how linguistic interactions influence grammatical structures and typological classifications. The color coding of the nodes and links corresponds to different time periods, with blue and green representing earlier research (1998-2012), and yellow and light green indicating more recent trends (2018-2023). Keywords like "agreement," "syntax," and "semantics" seem to have gained more attention in recent years, suggesting shifts in research focus. This visual also highlights that traditional topics, such as "word order" and "grammar," remain vital, while newer areas like "language contact" and "acquisition" have emerged as growing interests, especially in sociolinguistic and psycholinguistic studies. Overall, the image provides a clear overview of the evolving trends in language typology research, mapping both long-standing and emerging themes over the past 30 years.

**Tabel 2** Top 30 High-frequency Keywords in Language Typology

Ranking	Keywords	Count	Ranking	Keywords	Count
1	language	159	16	verbs	28
2	typology	135	17	motion events	27

Ranking	Keywords	Count	Ranking	Keywords	Count
3	English	93	18	optimality theory	27
4	language contact	59	19	discourse	27
5	syntax	54	20	language change	26
6	acquisition	53	21	movement	25
7	semantics	43	22	constructions	25
8	grammar	42	23	perception	25
9	word order	41	24	information structure	25
10	linguistic typology	40	25	phonology	24
11	speech	39	26	morphology	24
12	agreement	38	27	manner	22
13	language typology	34	28	Spanish	20
14	languages	31	29	marking	20
15	stress	29	30	contact	20

The table provides an overview of the most frequently occurring keywords in language typology research, revealing key areas of focus and emerging trends. The most common keywords are "language" (159 occurrences), "typology" (135 occurrences), and "English" (93 occurrences), indicating that broad studies on linguistic structure and classification dominate the field. Other significant topics include "language contact" (59 occurrences), "syntax" (54 occurrences), and "acquisition" (53 occurrences), reflecting strong interest in how languages influence one another, the structural aspects of language, and language learning processes. Additionally, keywords such as "semantics," "grammar," and "word order" emphasize the importance of meaning, linguistic rules, and sentence structure in typological research. Moderately frequent terms like "speech," "agreement," and "stress" suggest attention to phonological and morphological features, while more specialized topics such as "verbs," "motion events," and "optimality theory" highlight specific linguistic frameworks and phenomena being explored. Research on language evolution ("language change"), syntactic structure ("movement"), and sentence organization ("information structure") also feature prominently. The appearance of "Spanish" as a keyword indicates that language-specific studies, particularly involving Spanish, play a notable role in this research area. Overall, the table underscores a comprehensive focus on both foundational linguistic concepts and newer, specialized topics within the field of language typology.

### ***Clustered Keywords***

This study conducted a keyword clustering analysis to generate a knowledge map of foreign language typology research (see Figure 2). The clustering analysis yielded results with a Modularity value (Q) of 0.5, which is greater than the benchmark of 0.3, indicating a well-defined cluster structure. Additionally, the Weighted Mean Silhouette (S) score was 0.7905, surpassing the threshold of 0.7, demonstrating strong cohesion within the clusters and high credibility of the clustering results. The Density (D) value was calculated at 0.0233, further supporting the robustness of the clusters. Figure 2 illustrates the analysis of the main thematic clusters in foreign language typology research over the past three decades.



**Figure 3** Keywords Clusters (1993-2022)

The figure represents a knowledge map generated through keyword clustering analysis of language typology research over the past three decades, using the bibliometric tool CiteSpace. The main clusters, indicated by large keywords such as "linguistic typology," "semantics," "agreement syntax/grammar," "language acquisition," and "word order," highlight the dominant themes in the field. These areas have attracted significant attention from researchers, representing major focal points in the study of language typology. The colorful clusters depict the grouping of related concepts, with different colors corresponding to various time periods from 1993 to 2023. For instance, the green clusters seem to focus on syntax and morphology. The timeline at the top illustrates how research focus has evolved, with keywords like "language acquisition" and "agreement syntax" becoming prominent in specific years. The Modularity Q score of 0.5 and Weighted Mean Silhouette value of 0.7905 indicate that the clusters are well-structured and internally coherent, meaning each cluster represents distinct and significant research themes. Additionally, smaller clusters, such as those on "asian englishes," "language contact," and "sociolinguistic typology," reveal more specialized areas of research. Overall, the visualization provides a comprehensive overview of how language typology research has developed, identifies key trends, and highlights emerging areas of interest that may warrant further exploration.

### 3. Citation

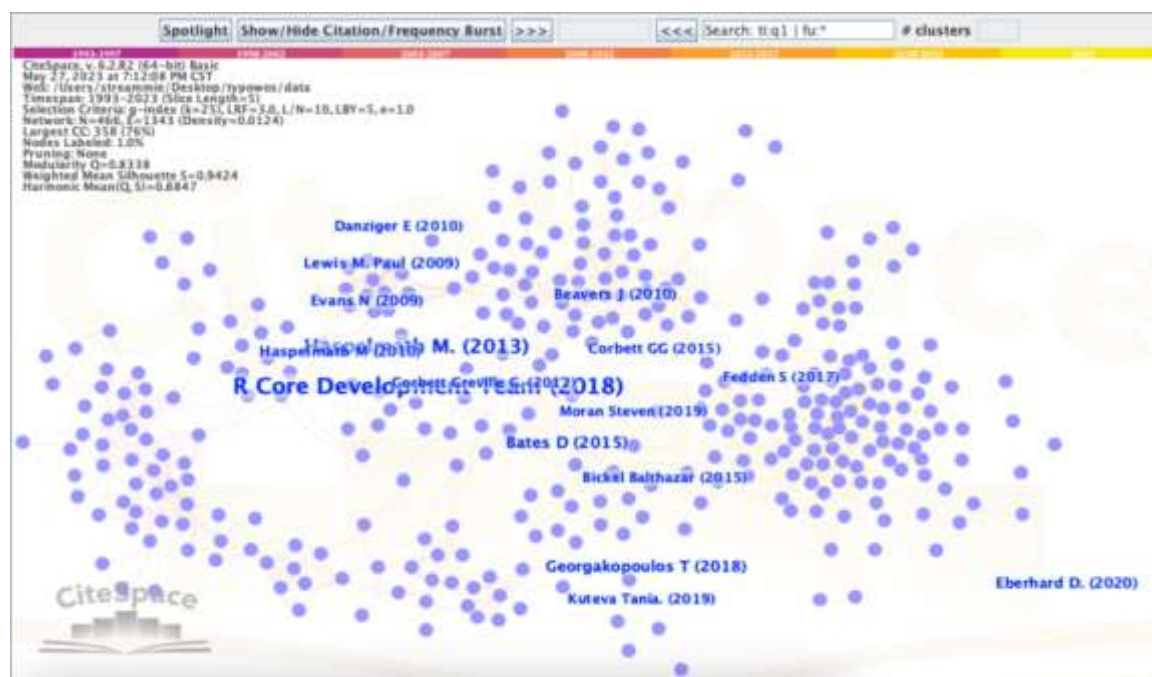
The development of a new subject must draw on the body of knowledge already accumulated in related fields. The development of a single research article cannot be based entirely on its own content. The paper must reference prior writings in the field or other pertinent fields and research findings from other fields. Research articles that are published in journals typically represent the cutting edge of particular fields, and the references they cite serve as their knowledge foundation. We were able to group references frequently mentioned in the language typology field and locate co-citation clusters using specific computer software. This phase is crucial for determining the foundational understanding of the language typology research, and it could be accomplished by visualizing the knowledge of the language typology research journal articles (Li et al, 2017).

Co-citation analysis can visualize the information in a research field's literature and distill the core fundamentals of understanding the field (Liu and Chen, 2019). Through co-citation analysis of the literature, this study explores the overall lineage and key paths of foreign language typology research development in the past two decades to grasp the core frontiers.

Literature co-citation refers to two pieces of literature being cited by one literature at the same time. This study explores the knowledge base and research frontiers of foreign language typology research through co-citation analysis of the obtained data. Excluding technical literature, it can be seen from Table 3 that the most frequently cited literature is Martin Haspelmath's article Occurrence of Nominal Plurality, published in 2013. Haspelmath is a comparative linguist who studies the diversity of the world's grammatical and lexical systems and tries to understand what is universal about languages. Occurrence of Nominal Plurality explains a map that shows the extent to which plural markers on full nouns are used in a language (full nouns are nouns which are not pronouns; see chapter 35 for the plural of personal pronouns) (Haspelmath, 2013).

The second most frequently cited article in the literature is by Thanasis Georgakopoulos (2018) in *Language and Linguistics Compass* "The semantic map model: State of the art and future avenues for linguistic research". This paper attempts to give a comprehensive overview of the model, including presenting the steps involved for building the maps and discussing different types of maps with their advantages and disadvantages which focuses on the kinds of linguistic generalizations captured (Georgakopoulos and Polis, 2018).

The third most frequently cited article in the literature is The typology of motion expressions revisited by John Beavers et.al in *Journal of Linguistics*. This study offers an innovative viewpoint on the options languages have for encoding directed motion events. It shows that, contrary to popular belief, purported verb-framed languages may prefer satellite-framed patterns when appropriate contextual support is available, which is contrary to the assumptions made by two- and three-way typologies. Finally, the authors explain the appeal of previously proposed two- and three-way typologies by pointing out that they capture the encoding options (Beavers et.al, 2010).



**Figure 4** Mapping of Co-cited Literature (1993-2022)

The figure represents a co-citation network map generated through CiteSpace, visualizing influential publications and their relationships in the language typology research field from 1993 to 2023. Each node in the network corresponds to a published work, with larger nodes representing highly cited papers, indicating their significant influence on the field. Prominent nodes such as R Core Development Team (2018), Haspelmath M (2013), Corbett GG (2015),

and Bates D (2015) reflect foundational works that have been extensively cited, showing their major contributions to the development of language typology. The clusters of nodes reflect groups of studies that are frequently co-cited together, indicating shared themes or research focuses. For example, the dense cluster around Haspelmath (2010) suggests that this work is central to many studies, likely contributing to theories or methodologies widely applied in language typology. The edges, or lines connecting the nodes, represent co-citation links, meaning these papers are often cited together in subsequent studies. Thicker lines indicate stronger co-citation relationships, suggesting that these works are frequently referenced in the same contexts. The connections between nodes such as Corbett (2012) and Bickel (2015) imply a strong thematic or methodological linkage in their contributions. The timeline at the top shows how the research has evolved, with earlier works (e.g., Evans N (2009) and Lewis M. Paul (2009)) laying the groundwork for more recent influential studies, such as Eberhard D (2020). The relatively high Modularity Q score of 0.8338 indicates a well-structured network with distinct clusters, and the high Weighted Mean Silhouette score of 0.9424 suggests good internal cohesion within these clusters, further reinforcing the reliability of the co-citation relationships visualized.

**Tabel 3** Top 10 Most Co-cited Researches (1993-2022)

Ranking	Title	Author & Year of Publication	Total Frequency of Co-citation
1	A Language and Environment for Statistical Computing	R Core Development Team (2018)	51
2	Occurrence of Nominal Plurality	Haspelmath M. (2013)	31
3	Fitting Linear Mixed-Effects Models Using lme4	Bates D (2015)	16
4	The semantic map model: State of the art and future avenues for linguistic research	Georgakopoulos T & Polis S (2018)	13
5	Ethnologue: Languages of the World	Lewis M. Paul (2009)	12
6	The typology of motion expressions revisited	Beavers J (2010)	12
7	The myth of language universals: Language diversity and its importance for cognitive science	Evans N (2009)	12
8	Comparative concepts and descriptive categories in crosslinguistic studies	Haspelmath M (2010)	11
9	What is the most spoken language?	Eberhard D. (2020)	11
10	Morphosyntactic complexity: A typology of lexical splits	Corbett GG (2015)	10

The table highlights the top 10 most co-cited studies in language typology research from 1995 to 2022, showcasing the most influential works based on their frequency of co-citation. The most frequently co-cited work is "A Language and Environment for Statistical Computing" by the R Core Development Team (2018), with 51 co-citations, indicating the critical role of statistical tools like R in analyzing linguistic data. Haspelmath M. (2013) and Bates D. (2015) follow closely, with significant contributions to nominal plurality and mixed-effects models, respectively, reflecting the importance of methodological innovations in typology research. Works by Beavers J. (2010), Evans N. (2009), and Haspelmath M. (2010), which explore topics

such as motion expressions, language universals, and linguistic diversity, emphasize key thematic focuses in the field. Additionally, foundational resources like Ethnologue by Lewis M. Paul (2009) and the semantic map model by Georgakopoulos T & Polis S (2018) underscore the importance of linguistic documentation and theoretical frameworks. Recent studies, such as Eberhard D. (2020) on the world's most spoken languages, show the field's continued evolution and growing interest in new areas. The table illustrates the prominence of methodological advancements and the focus on linguistic diversity in shaping current language typology research.

## D. Conclusion

This study employed scientific bibliometric tools to analyze the development, research hotspots, and core themes within foreign language typology research. The analysis yielded several key findings. Firstly, the overall trajectory of foreign language typology research shows a rising trend, though there have been fluctuations in the number of publications each year. Secondly, the field has increasingly embraced new technologies, such as the R programming language, to explore linguistic phenomena from fresh perspectives. Thirdly, the core topics of this research field include language contact, syntax, acquisition, semantics, grammar, speech, motion events, and discourse, indicating the breadth and depth of focus areas within language typology. In light of these findings, future research in language typology should prioritize further refinement and diversification of research methods. It is recommended that scholars continue to leverage advanced tools, such as corpora and computational linguistics methods, to deepen the analysis and understanding of typological patterns across languages. Expanding methodological approaches will enhance future studies' robustness and contribute to the evolution of language typology as a dynamic and interdisciplinary field.

## References

- Aikhenvald, A. (2024). Linguistic typology in action: how to know more. *Theoretical Linguistics*, 50(1-2), 23-31.
- Beavers, J., Levin, B., & Tham, S. W. (2010). The typology of motion expressions revisited. *Journal of linguistics*, 46(2), 331-377.
- Chen, C. M., & Li, J. (2016). *CiteSpace: Text Mining and Visualization in Scientific Literature*. Capital University of Economics and Business Press
- Georgakopoulos, T., & Polis, S. (2018). The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass*, 12(2), e12270.
- Greenberg, J. H. (1957). The nature and uses of linguistic typologies. *International Journal of American Linguistics*, 23(2), 68-77.
- Greenberg J. (1963). *Some universals of grammar with particular reference to the order of meaningful elements*. Stanford University Press
- Haspelmath, M. (2013). 34 Occurrence of Nominal Plurality. *The world atlas of language structures online*.
- Li, X., Ma, E., & Qu, H. (2017). Knowledge mapping of hospitality research— A visual analysis using CiteSpace. *International Journal of Hospitality Management*, 60, 77-93.
- Liu, Z., & Chen, D. (2019). Research Progress and Trend of Online Continuing Education Abroad in Recent 20 Years: Based on a Visualization of CiteSpace. *Tsinghua Journal of Education* (04), 123-132.
- Ying, X. F., & Chen, C. L. (2024). Linguistics Typology and Chinese Studies ---- A Research Review of the Past Forty Years. *Chinese Language Learning*