



A Case Study of Linguistic Research Methods in the Age of Computing

TieJun Shao^{1,3,4}(✉), JianShe Zhou^{1,3}, and WenYan Zhang²

¹ Capital Normal University, Beijing, China

2180101027@cnu.edu.cn

² China Fire and Rescue Institute, Beijing, China

³ Research Center for Language Intelligence of China, Beijing, China

⁴ North China Institute of Science and Technology, Beijing, China

Abstract. With the development of technologies such as computers, Artificial Intelligence, and Big Data new research methods have emerged in the traditional social sciences. With the help of new research methods, the efficiency and accuracy of quantitative analysis can be improved, which helps to improve research efficiency and demonstrate research trends. This paper uses the data analysis software CiteSpace as an example to analyze the current research situation of domestic Tibetan-Burman Language research, and graphically display the analysis results.

Keywords: Artificial intelligence · Bibliometric analysis · CiteSpace

1 Introduction

With the development of technologies such as computers, Artificial Intelligence, and Big Data, new research methods have emerged in the traditional social sciences. With the help of new research methods, the efficiency and accuracy of quantitative analysis can be improved, which is helpful to improve research efficiency and display research trends. Linguistics is the subject of human language research. The scope of exploration includes the nature, function, structure, application and historical development of language, as well as other language-related issues. The object of linguistic research is objectively existing language facts. Whether it is modern language or ancient language, it is an objective language phenomenon.

In China, due to historical, geographic and other factors, ethnic minority languages are an integral part of linguistic research. Tibeto-Burman language is the language family with the widest distribution, the largest internal differences in the Sino-Tibetan language family. The Tibetan-Burman languages in China are mainly distributed in Tibet, Gansu, Yunnan, Sichuan, Guizhou, Guangxi, Hunan, Hubei and other provinces and autonomous regions. Tibeto-Burman language is an important content in linguistic research, and it has special value for the construction of linguistic theory and national culture research. The comparative study of Tibeto-Burman language in China started relatively late, but the study of Tibeto-Burman language has developed from the original research interest of a

few scholars to an important branch of current linguistics. Through the efforts of several generations, Tibetan-Burman language research has achieved considerable results. In the face of vigorous Tibetan-Burman language research, we must not only think about what results have been obtained in Tibetan-Burman language research, which aspects have been sufficiently studied, and which aspects are still available. Very weak, how can we make greater breakthroughs in Tibetan-Burman language research? Therefore, this article uses the Tibetan-Burman language research from 1992 to 2019 as the data source, and uses the visualization function of the Citespace software to quantitatively analyze the development status and trends of Tibetan-Burman language research.

1.1 Research Method

Bibliometric analysis is to use mathematical and statistical methods to quantitatively analyze a given knowledge carrier. Through a quantitative analysis of literature in a certain field, basic information such as research overview, research hotspots, and research distribution in the field can be obtained. This information has a good reference value for researchers to sort out the research context of the field and discover the research innovation points in the field. CiteSpace is a visual bibliometric analysis software developed by the team of information visualization expert Chen Chaomei. This software can perform keyword clustering analysis of scientific literature, author cooperation network analysis, etc., and display the analysis results in visual forms such as timeline and knowledge structure map. At present, few scholars use professional tool software to carry out bibliometric analysis of Tibetan-Burman language research. This article combines the needs of Tibetan-Burman language research and the characteristics of CiteSpace's visualization tools, and uses CiteSpace to do quantitative analysis and visual display of Tibetan-Burman literature.

1.2 Research Data Sources

The object of quantitative analysis in this article is related papers on Tibetan-Burman language research. The research data comes from CNKI database of CNKI. The search condition is "topic = Tibetan-Burman", the source category is "core journals", and the retrieval time is April 2020. On the 27th, a total of 425 documents with the theme of "Tibetan-Burman" were retrieved, and the time span was from 1992 to 2019. Since this article conducts statistical analysis on Tibetan-Burman language research on a yearly basis, 2020 has not yet ended, so the time span ends in 2019 to facilitate statistics throughout the year. In order to ensure the accuracy of the retrieved literature data, the retrieval results were manually screened before analysis, and related seminars, linguistic conferences, person news, seminar reviews, and published information were deleted to obtain a total of 376 documents that were effectively.

2 Analysis of Annual Paper Volume

This article counts the papers on the Tibetan-Burman language in the core journals of CNKI from 1992 to 2019, and analyzes the annual number of papers in this field.

The annual volume of papers is one of the important indicators to measure the development of a certain research field. To a certain extent, it can reflect the improvement of knowledge in the field, the research progress of researchers, and the popularity of field research. Figure 1 shows the distribution of the number of articles issued in the Tibetan-Burman language research field from 1992 to 2019. From the figure, it can be seen that the core amount of articles issued in the Tibetan-Burman language research field in China is a research trend that tends to be stable while changing. The average annual number of articles published is 13.5.

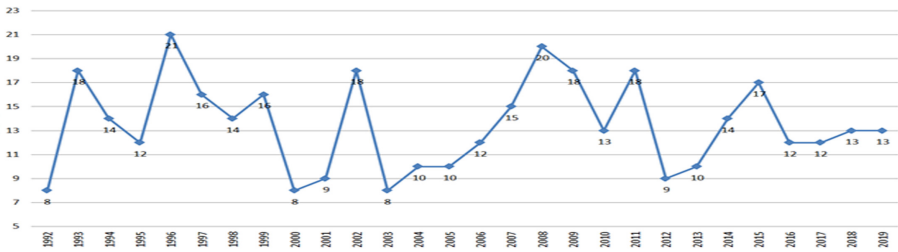


Fig. 1. Annual paper volume from 1992–2019.

From 1992 to 2015, the number of publications in Tibeto-Burman language research was wavy, and there were several peaks and valleys in the number of publications. Among them, the number of publications increased year by year from 2005 to 2008, indicating that the study of Tibeto-Burman language in China during this period is relatively hot, new research hotspots have emerged, knowledge has been improved more, and more research results have been published. Language contact, cross-language comparison, and linguistic issues related to the Tibetan-Yi Corridor are all research hotspots in this period; the research content involves specific issues such as language family issues, phonetics, vocabulary, and grammar, such as “Some Thoughts on Grammar Comparison”, “Language Contact and Cultural Diffusion on the Silk Road”, etc., the research field covers a wide range and rich results.

After 2008, Tibetan-Burman language research has experienced a fluctuating decline. From 2016 to 2019, the number of publications has been steadily changing, but the annual publication volume of 2016–2019 is lower than the average annual publication volume of the 1992–2019 interval. The changes in the number of articles published every year indicate that there is still a large research space in the Tibetan-Burman language research field on the basis of existing subjects, and there is the potential for new research hotspots.

3 Source Journal Analysis

Professional academic journals are the main carrier for the release of subject research results, an important platform for subject exchanges, and the main medium for researchers to obtain research materials and understand academic frontiers. The ability of research results to be published in professional journals also recognizes the academic level of scholars. The chart of journals containing Tibetan-Burman research papers from

1992 to 2019 and the number of collections. Through statistics and analysis of domestic journals containing Tibetan-Burman research papers, it is not difficult to find that the journal with the largest number of Tibetan-Burman research papers is “National Language and Language Studies. During the statistical period, “Ethnic Language” published a total of 144 papers on Tibetan and Burmese, and “Language Studies” published a total of 50 papers, accounting for 38.3% and 13.3% of the number of statistical papers. The total issue of the two journals The number of publications accounted for 51.6% of the number of statistical papers, that is, more than 50% of the Tibetan-Burman language research papers were published in “Ethnic Languages” and “Language Studies” during the statistical period.

“National Language” was founded in 1979 and is sponsored by the Institute of Ethnology and Anthropology of the Chinese Academy of Social Sciences. The journal focuses on the facts of various Chinese languages as the main research object. It is based on Sino-Tibetan, Altaic, South Asian and Austronesian languages. The publication is the main research object. “Language Studies” was founded in 1981 and sponsored by the Institute of Chinese Languages of Huazhong University of Science and Technology. The journal focuses on various languages in China as its research objects. The content includes the exploration of linguistic theoretical methods and the evaluation of new theories, Chinese and minority languages And the description of dialects, the historical comparative study of relative languages, etc.

Due to the different positioning and focus of the journal, the focus of the Tibetan-Burman essays published in National Language and Language Studies is also different. The Tibetan-Burman essays published in “Ethnic Languages” pay more attention to specific ethnic languages, language homologous relations and other issues, placing the study of Tibetan-Burman languages within the framework of minority languages, such as “Liangshan Yi Tone and Expressions”, “Innovation and Division of Rong Language Branch”, “Phonetic Features and Variations of Zhuang Language in Yunnan Ninglang”, “Range and Range of Word Choices in Comparison of Cognate Words Standards—Take the formulation of the comparison table of Tibetan-Burman cognate words as an example” and “The Distribution and Sources of Tibetan-Burmese Words”, etc.; Tibetan-Burman papers published in “Language Studies” more Pay attention to the language itself, and conduct research from the perspectives of language types, phonetics, and grammar, such as “The Territorial Structure of Tibeto-Burman Language in Southeast Tibet”, “The Yinsheng Vowel System of Miao-Yao Language”, “Bisu Language Status and Significance in Historical Comparison”, “Types and Characteristics of Affricates in Ethnic Languages of Southern China”, etc.

Domestic research on Tibeto-Burman language is not only limited to linguistic studies, but also involves the study of ethnic minorities in the southwest. Therefore, some relevant research papers on Tibeto-Burman language are also published in journals such as Guizhou Ethnic Studies and Ideological Front, such as Huang Lihong Published in “Guizhou Ethnic Studies”, “The Historical Textual Research of my country’s Tibetan-Burman Ethnic Minorities Migrating Southward”, etc. It can be seen that academic papers on Tibetan-Burman language studies are mainly published in academic journals related to linguistics and ethnology.

4 Keyword Analysis

4.1 Keyword Co-occurrence Analysis

Keywords are used to express the main content of the article. The keywords can reflect the main research results of the article. CiteSpace is used to analyze the keywords of the literature. The distribution of high-frequency keywords in the research field can be seen through intuitive visualization, so as to understand the corresponding Research hotspots.

The high-frequency keywords mainly include: Tibeto-Burman, Sino-Tibetan, Tibeto-Burman, Sino-Tibetan, Yi branch, Dong-Taiwan, historical comparison, prefix, vowels, etc. Since this article is a literature search on the subject of “Tibetan-Burman”, it is not meaningful to analyze the high-frequency words Tibetan-Burman separately, so it mainly analyzes other high-frequency words. Table 1 shows the top 15 keywords of frequency.

Table 1. Words with high-frequency.

| No | Frequency | Word |
|----|-----------|------------------------------|
| 1 | 127 | Tibetan Burmese Language |
| 2 | 46 | Sino Tibetan language family |
| 3 | 32 | Tibeto-Burman Branch |
| 4 | 28 | Dong Tai language |
| 5 | 24 | Yi language branch |
| 6 | 23 | Chinese Tibetan |
| 7 | 19 | Historical comparison |
| 8 | 16 | Construction |
| 9 | 14 | Prefix |
| 10 | 14 | Initial consonant |
| 11 | 14 | chinese |
| 12 | 14 | Personal pronouns |
| 13 | 13 | Three body pronouns |
| 14 | 12 | Miao-Yao |
| 15 | 11 | Voice |

From the above frequency table, it is not difficult to see the relationship between the Tibetan-Burman, Chinese and Sino-Tibetan language families, as well as the specific research content involved in the Tibetan-Burman language research. The high-frequency keywords obtained through CiteSpace analysis belong to the scope of Tibetan-Burman language research, and can also directly reflect the characteristics and key issues of the current Tibetan-Burman language research, such as the Dong-Tai language, Yi branch, structure, prefix shown in Table 1. Initials, phonetics, etc. These are the hotspots of

current research, and it also reflects that the current Tibetan-Burman language research focuses on phonetics, which is consistent with the characteristics of most languages of the Tibetan-Burman language that only have phonetics without text. On the other hand, it can be seen that the current Tibetan-Burman grammar and vocabulary research is relatively small, and there is still much room for research in this area in the future.

Co-occurring words that appeared earlier include Sino-Tibetan, Sino-Tibetan language family, historical comparison, construction, affix (1993), Burmese branch, Dong-Taiwan, phonetic (1997), vocabulary comparison, cognate characters, Ancient Sounds (1998), Thai, Siamese (1999); Co-occurring words that appeared later include Bai (2010), historical typology, typology (2011), copula, judgment word (2013) etc. Through the temporal changes of keyword co-occurrence, we can summarize the brief evolutionary context of the study of Tibeto-Burman language in my country. In the early stage of the study, the study focused on specific language from the perspectives of word formation and phonetics. Based on these studies, it began to appear in the later period. Linguistics and typology study the integrity of Tibeto-Burman language. It reflects the evolution of Tibetan-Burman language research from concrete to abstract, from individuality to generality, from linguistic data investigation to the establishment of disciplinary research framework.

Summarizing the evolution of Tibetan-Burman language research is to start with a specific research problem of a single language in the early stage of the research, such as word prefixes and suffix issues, and then use historical comparison to analyze the relationship between languages in the Tibetan-Burman language family from the perspective of morphology. The evolution relationship is analyzed. With the increase of research materials and the advancement of research theories, typological theories are used to conduct more in-depth research on Tibetan-Burman language.

4.2 Keyword Cluster Analysis

Using CiteSpace to perform clustering analysis on co-occurring keywords can get research hotspots in the research field. This paper uses the LSI clustering algorithm provided by Cite Space to cluster high-frequency co-occurring words in Tibetan and Burmese, a total of 7 cluster labels are obtained, namely #0 Tujia language, #1 morphology (grammar), #2 Tibetan, #3 root, #4 tone system, #5 series and #6 Dong Dai language family. These key The clustering of words can reflect the research hotspots of Tibeto-Burman in China.

#0Tujia Language Cluster. The main content includes Tujia language, pure four-class rhyme, Yi language, ethnic linguistics, historical typology, Daofu language, Muya language, comparative research, ethnic minorities, folk literature, language investigation, open syllables, Continuous tone sandhi, inflection, Li Fanggui, ethnology, language branch, Bai nationality, etc. This clustering shows that the current research on Tujia language is mainly focused on phonetics, and compared with other languages, the characteristics of the phonetics are obtained to explore the family of Tujia language.

#1 Morphology (Grammar). The main content of clustering includes morphology (grammar), person, pronunciation, polyphony, imperative, vowel, pre-addition, attribute,

Burmese, ancient Chinese, consistent relationship, post-addition, rhyme ending, Tones, Burmese branch, etc. This clustering fully shows how the morphology changes and its effects.

#2 The Main Content of Tibetan. Clustering includes Tibetan, sayings, Siamese, Miao-Yao, Zhuang-Dai, Dong-Taiwan, Thai, Dong, Dong-Shui, ancient sounds, phonetics, cognate characters, Alternate relations, Shui language, dialectology, Jingpo language, Chinese dialects, Sino-Tibetan, primitive Sino-Tibetan, Tibeto-Burman language family, trishen pronouns, vowels, phonetics, personal pronouns, etc. There are many keywords in this cluster, which reflects the depth of Tibetan language research so far, and also reflects the important role and significance of Tibetan language in Tibetan-Burman language research.

#3 The Main Content of Root Clustering. Includes roots, associated markers, morphemes, cognates, Shang and Zhou dynasties, causal complex sentences, underlying words, Zhou Dynasty, hypothetical sentences, common words, correspondence, verb overlap, methodological issues, and imitation of ancient, Tibetan language, Chinese studies, verb overlap, etc. This clustering reflects scholars' tracing of roots, the relationship between roots and cognates, and the important role of roots in language research.

#4 Tone System Clustering. Mainly includes tone system, grammatical category, typology, Tibeto-Burman language, Dong Dai language family, balance state, Zhuang Dong language family, language typology, causative usage, directional verbs, grammar, syntactic format, language Groups, linguistic commonalities, Yin tones, Jianyu style, phase space, intransitiveness, tones, order, Qiang branch, suffix, language system, continuum, modern Chinese, etc. Tone is an important feature of Chinese-Tibetan languages in phonetics. This clustering reflects the extensiveness of tonal research and the involvement of new theories and new perspectives, such as equilibrium and phase space.

#5 Cohort Clustering. Mainly includes copulas, judgment words, historical comparison, interpretive sentences, demonstrative pronouns, conceptual structure, verbal characteristics, judgment sentences, difference, construction, first person, Tibetan branch, Sino-Tibetan Department etc. This clustering can see the role and significance of copulas in the field of research.

#6 The Main Content of Dong Dai Language. Cluster includes Dong Dai language family, equilibrium, cognate words, Zhuang-Dong language family, three-shen pronouns, tones, roots, Bai language, etc.

It can be seen that cluster analysis can provide intuitive and visual research focus and direction. Facilitating the extraction and processing of key information is conducive to our grasp of research directions and research hotspots.

4.3 Keyword Evolution Trend Analysis

CiteSpcae provides a knowledge graph displayed in a time zone view (Timezone View). The time zone view displays the changes of keywords in the field literature from the time

dimension. At the same time, you can intuitively understand the number of keywords in the current year through the distribution of nodes and lines in the corresponding year, that is, more lines indicate high There are more frequent keywords. It can be seen from the figure that there were no high-frequency keywords in 2003–2006 and 2014–2019. Combined with the analysis of the number of publications in 2.1 years, the annual publication volume of these two time periods was continuously lower than the annual average publication volume, that is, the study of Tibetan-Burman There are no major research hotspots in the field, and the output of research results is small.

5 Conclusion

This paper uses the bibliometric software CiteSpace to conduct a simple quantitative analysis of the domestic Tibetan-Burmese research field. The analysis includes the author, the publishing organization, and the co-occurrence of keywords. The conclusions are drawn as follow:

- (1) New research methods such as computers, artificial intelligence, and big data can improve the efficiency and accuracy of quantitative analysis, help improve research efficiency and show research trends.
- (2) The cooperation mode is single. Most of the authors in the field of Tibetan-Burman language research in China are independent researches and articles, and the existing cooperative relationships are mostly teachers or colleagues. Although this kind of cooperation has cultivated new forces for Tibetan-Burman language research, it is at the same time Lack of research and cooperation within the field and across fields, and there are certain limitations in research methods and research theories, which is not conducive to the creation of academic innovations;
- (3) There is currently a lack of new research hotspots in the Tibetan-Burman language field, and there have been few Tibetan-Burman research papers in recent years;
- (4) (Research methods lack innovation. The existing Tibetan-Burman language research is mostly based on traditional linguistic research methods, and there is no combination of information and intelligent research methods in the Internet age.

Acknowledgement. This research was financially supported by “Research on key technologies and model verification of prose genre oriented text understanding (ZDI135-101)”, “Research and Application of Key Technologies of Intelligent Auxiliary Reading System (ZDI135-79)”, Capacity Building for Sci-Tech Innovation-Fundamental Scientific Research Foundation (20530290082).

References

1. Yang, H., Shao, X.X., Wu, M.: A review on ecosystem health research: a visualization based on CiteSpace. *Sustainability* **11**, 4908 (2019)
2. Zheng, K.Y., Wang, X.Q.: Publications on the association between cognitive function and pain from 2000 to 2018: a bibliometric analysis using CiteSpace. *Med. Sci. Monitor.* **25**, 8940–8951 (2019)

3. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Int. Med.* **151**, 264–269 (2009)
4. Yao, Q., et al.: Scientometric trends and knowledge maps of global health systems research. *Health Res. Policy Syst.* **12** (2014)
5. Archambault, E., Campbell, D., Gingras, Y., Lariviere, V.: Comparing of science bibliometric statistics obtained from the web and scopus. *J. Am. Soc. Inf. Sci. Technol.* **60**, 1320–1326 (2009)
6. Hou, J.H., Yang, X.C., Chen, C.M.: Emerging trends and new developments in information science: a document co-citation analysis (2009–2016). *Scientometrics* **115**, 869–892 (2018)
7. Martin, D.J., Simon, J., Daan, S., Zhan, C.J., Margot, W.: Sustainable-smart-resilient-low carbon-eco-knowledge cities; making sense of a multitude of concepts promoting sustainable urbanization. *J. Clean. Prod.* **109**, 25–38 (2015)
8. Cabeza, R., Luis, J., Sanchez, C.S., Fuentes, G.F.J.: Past themes and tracking research trends in entrepreneurship: a Co-word, Cites and usage count analysis. *Sustainability* **11**, 3121 (2019)
9. Small, H., Griffith, B.C.: The structure of scientific literatures I: identifying and graphing specialties. *Sci. Stud.* **4**, 17–40 (1974)
10. Cambrosio, A., Limoges, C., Courtial, J.P., Laville, F.: Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis. *Scientometrics* **27**, 119–143 (1993)
11. Ying, D., Chowdhury, G.G., Foo, S.: Bibliometric cartography of information retrieval research by using co-word analysis. *Inform. Process. Manag.* **37**, 817–842 (2001)
12. Zong, Q., Shen, H., Yuan, Q., Hu, X., Hou, Z., Deng, S.: Doctoral dissertations of Library and Information Science in China: a co-word analysis. *Scientometrics* **94**, 781–799 (2013)
13. Diego, C.G. Marta, O.U.C. Eva-María, M.V.: Knowledge areas, themes and future research on open data: a co-word analysis. *Gov. Inform. Q.* **36**, 77–87 (2019)
14. Zhang, Q., Ma, F.C.: On paradigm of research knowledge management: a bibliometric analysis. *J. Manag. Sci. China* **10**, 65–74 (2007)
15. An, X.Y., Wu, Q.Q.: Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics* **88**, 133–144 (2011)
16. Cho, C.H., Patten, D.M.: The role of environmental disclosures as tools of legitimacy: a research note. *Account. Org. Soc.* **32**, 639–647 (2007)
17. Tang, Y.H., Wu, S.M., Miao, X., Pollard, S.J.T., Hrudehy, S.E.: Resilience to evolving drinking water contamination risks: a human error prevention perspective. *J. Clean. Prod.* **57**, 228–237 (2013)
18. Boston-Fleischhauer, C.: Enhancing healthcare process design with human factors engineering and reliability science, part 2: applying the knowledge to clinical documentation systems. *J. Nurs. Adm.* **38**, 84–89 (2018)
19. Davis, M.V., Macdonald, P.D.M., Cline, J.S., Baker, E.L.: Evaluation of public health response to hurricanes finds north carolina better prepared for public health emergencies. *Public Health Rep.* **122**, 17–26 (2007)
20. Gao, W.Y., Zhang, H., Luo, Y.: Preventing false negative results in detecting new coronavirus nucleic acid. *Int. J. Lab. Med.* **41**, 641–643 (2020)