1,* **Fei Li**

Computer Bibliometric Analysis of Hakka Dialect Phonetics Based on Citespace and VOSviewer

Abstract: - The findings of research on Hakka dialect phonetics has mainly been published in the form of journal papers, dissertations and monographs. With "China National Knowledge Infrastructure (CNKI)" and "Duxiu Academic Search" as data sources, in this work, relevant research reports on the phonetics of Hakka dialect in the 30-year period from 1993 to 2023 were collected, which included 186 journal papers, 79 dissertations and 36 academic monograph. Firstly, bibliometric analyses were used by Citespace and VOSviewer on computer; then, EXCEL was employed for collecting and sorting out the general situation of the thesis and evaluating the combination of qualitative and quantitative methods. The goal was to investigate the general research on Hakka dialect phonetics in the past 30 years as comprehensively as possible, reveal the development trend of Hakka dialect phonetics objectively and scientifically to provide theoretical basis for the development of new research directions in the future, and provide data support for the development and utilization of Hakka dialect language conservation project resources.

Keywords: Hakka Dialect Phonetic; Computer Bibliometric Analysis; Journal Thesis; Dissertation; Academic Monograph

I. INTRODUCTION

"Hakka" is relative to "indigenous". In order to escape the war, Hakka ancestors have carried out five large-scale southward migration activities and gradually formed "Hakka people" and "Hakka dialect" during southward migration [1]. Hakka dialects are distributed in more than 200 counties and cities in eight provinces and regions, including Guangdong, Guangxi, Fujian, Taiwan, Hainan, Jiangxi, Jiangxi, Hunan, Sichuan and Chongqing. Among them, the eastern and central regions of Guangdong Province, the western part of Fujian Province, and the southern regions of Jiangxi Province have the most concentrated populations [2].

Regarding regional divisions of Hakka dialect, Xiong [3] divided Hakka dialects in Guangdong and Taiwan into four areas: Guangdong and Taiwan (Jiaying, Xinghua, Xinhui, and Shaonan), Guangdong, Huizhou, and northern Guangdong and those in Jiangxi and Hunan into Ninglong, Yugui, and Tonggu. Then Liu [4], based on Xiong [3], divided the Xinghua small piece in Meizhou Hakka dialect into Xingning-Wuhua and Dabu-Fengshun. In addition, Xie [5] of the Language Atlas of China divided the Hakka dialectics of Jiangxi, Hunan, Fujian, Guangdong, Taiwan, Hong Kong and other provinces into eight areas: Yue-Tai (Meihui and Longhua), Hailu, Yuebei, Yuexi, Tingzhou, Ninglong, Yuxin and Tonggui.

Hakka dialect investigation in the real sense began with the Grottoes written by Huang Zhao in the early years of Jiaqing, which has a history of more than 200 years. With the global introduction of modern language theories and methods in 1920s, Chinese dialect science has been transformed into an independent discipline. The first Hakka dialect seminar in line with Hakka dialect investigation was held in Longyan, Fujian province in 1993, and the first Hakka dialect seminar collection was published in the following year under the title of "Collection of the first Hakka dialect academic seminar". So far, the seminar has been held for 15 sessions and has a history of 30 years. According to Peng [6], since 1993, the number of Hakka dialect research journals has continuously increased.

According to Li [7], Hakka dialect research results have mainly focused on phonetics. For this reason, this research work took 1993 as the node and collected phonetic research literature of Hakka dialect from 1993 to 2023.

In this work, the adopted documents focused on Hakka dialect phonetics and base on computer bibliometric analysis by use of Citespace and Vosviewer.

The bibliometric analysis method adopted in this study mainly uses computer software CiteSpace and Vosviewer to carry out data mining, metrological analysis and visual map rendering. In order to analysis the research of Hakka dialect phonetics. Bibliometric analysis tool is CiteSpace developed by Professor Chen Chaomei of Drexel University. The principle of the software is based on the co-citation analysis theory and path-finding network algorithm to carry out quantitative analysis on the literature data of a certain field. Then explore the critical path and knowledge inflection point of the evolution of the subject field.[8]

1 School of Sports Media, Guangzhou Sport University, Guangzhou, China
*Corresponding author: Fei Li

II. BIBLIOMETRIC ANALYSIS OF JOURNALS

A. Data Sources and Study Methods

In this section, "academic journals" and "academic collected papers" databases in "CNKI" were adopted to review papers on Hakka dialect phonetics. Search was performed in September 2023 and covered the period from 1993 to 2023. After cleaning the retrieved documents by weight, error and manual selection, 186 journal papers were considered as research data source.

Output the relevant data into Excel data sheets and Refworks literature data sheets. Then, the effective data is standardized, including the standardization of synonyms and abbreviations, the merger of institutional names, the keyword extraction of literature lacking keywords, the deletion of non-standard keywords, etc., so as to ensure the objectivity and accuracy of software operation results.. Then, the annual publication volume was calculated through Excel and chart, and then, co-occurrence network, co-occurrence network of research institutions, co-occurrence of high-frequency keywords, clustering network and keyword emergence trend charts were obtained. Finally, the resulting charts and networks were interpreted through both qualitative and quantitative methods.

B. Annual Distribution of Published Documents

The development trend and dynamics of the field was predicted by drawing distribution curves to obtain historical and comprehensive statistics of literature distribution, which is of great significance for stage evaluation [9]. Annual number of Hakka dialect research publications is illustrated in Figure 1, the horizontal coordinate represents the year and the vertical coordinate represents the number of articles published. Before 2007, the number of articles was only in single digits and was declined after reaching 10 in 2007. Since 2009, the number of articles has increased continuously and reached 13 in 2014. The 21st century was an important turning point for research on Hakka dialect phonetics. Previous research works on Hakka dialect phonetics has mainly focused on the recording and accumulation of phonetic materials. Since 21st century, however, research on Hakka dialect phonetics began to explore the internal laws of phonetics while recording, developed into interdisciplinary and social applications, and achieved a number of excellent research results. However, the number of Hakka dialect users in Lingnan population is less than that of Cantonese and Min dialects, and the stability of Hakka dialect research is far less than that of Cantonese and Min dialects. Anyway, the number of Hakka dialect research papers is increasing every year showing that Hakka dialect research teams and language community literature also pay increasing attention to Hakka dialect research works.
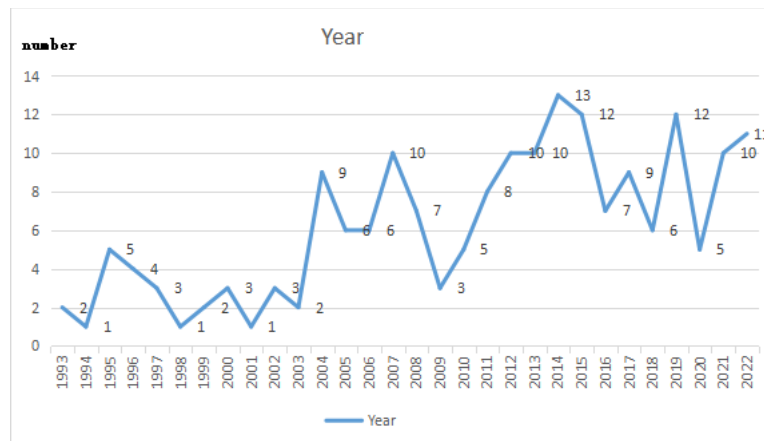

Figure 1: Distribution of Annual Publications on Hakka Dialectic

C. Distribution of Core Authors

We also evaluated the core authors of Hakka dialect research according to Price's law and comprehensive index method. We first used Price's Law to select the core authors. Regarding the academic level of core authors, we comprehensively evaluated the number of publications and citations. The number of publications represented the importance of the authors to the journals and the amount of citation denoted the academic influence of the authors [10].

First, Price's Law was applied for the selection of the core authors. The calculation equation was $m_p = 0.749\sqrt{np_{max}}$ where m_p is the number of papers published at least as core authors during the statistical period and np_{max} is the number of papers published by the most published authors during the statistical period. In this way, the number of papers published by the core authors on Hakka dialect phonetics was at least m_p which was

calculated to be 5.65 Then, based on the principle of integration, the authors of the above papers were selected as core author candidates.

Then, core author candidates were evaluated based on composite index method. Different researchers had different views on the weight values of the two indicators and citation. For example, Shuping [9] considered the above weight values as 0.5 and 0.5, respectively. This paper held that importance and academic influence had similar importance degrees; therefore, the weight values of both the number of articles and citations of core authors were set at 0.5. Comprehensive index value was calculated by the equation: $z_i = \frac{x_i}{\bar{x}} \times 100 \times 0.5 + \frac{y_i}{\bar{y}} \times 100 \times 0.5$ z_i is the composite index of the i candidate, $\bar{x}$ is the average number of papers published by core author candidates, and $\bar{y}$ is the average number of citations published by core author candidates, x_i is the cumulative number of posts for the i th candidate, and y_i is cumulative citation amount for the i th candidate.

As was seen from Figure 2, phonetics study on Hakka dialect has formed a number of small cooperative group. Most of the core authors of cooperative study were teachers, students or colleagues, such as Xie and Huang, Wen and Hou, Zhuang and Huang. Considering comprehensive index table, it was seen that the top core authors had less interaction and cooperation. Based on the statistics of the native location and age of authors, the core authors were mostly Hakka researchers, followed by native Minnan speakers. The emergence of young and middle-aged researchers showed that the echelon construction of Hakka dialect phonetics research has a long way to go.
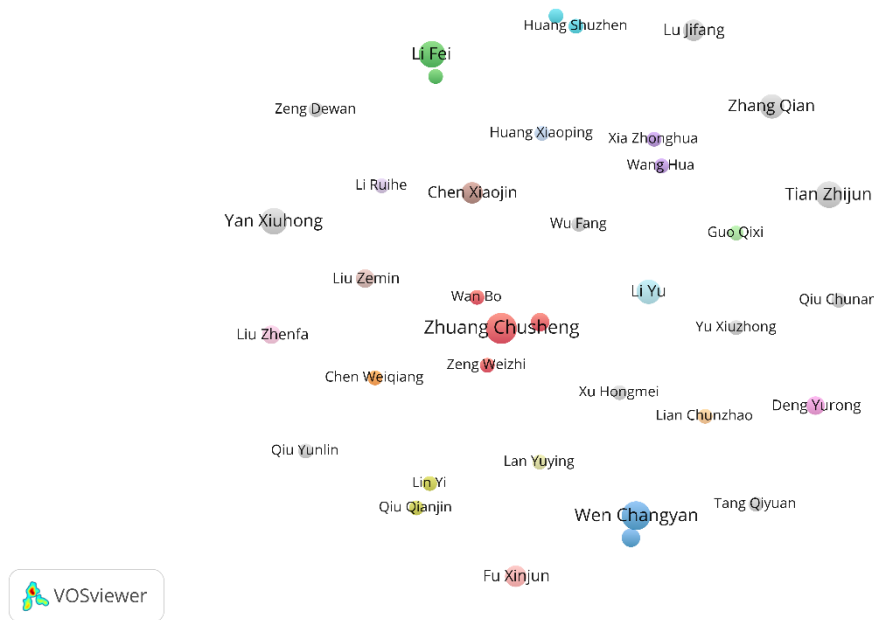

Figure 2: Co-occurrence Network of Journal Paper Authors

D. Distribution of Journal Paper Publishing Institutions

In this work, research institution co-occurrence network was generated with Citespace. As was seen from Figure 3, high-yield institutions of Hakka dialect phonetics research were mainly distributed in Guangdong, Guangxi and Jiangxi, which was completely consistent with main Hakka dialect distribution areas. Top ranking units: Jiaying College, as the gathering place of Hakka people and the highest institution in Meizhou, had high-yielding authors such as Wen and Hou; Jinan University had the only provincial dialect research base —— Chinese Dialect Research Center, a team of teachers and students; Sun Yat-sen University was a senior university in Guangdong Province, represented by the research team of Zhuang. In addition, Professor Qiyuan of Guangxi University and Professors Xiaoyan and Baiyun of Guangxi Normal University have made great contributions to phonetic research on Hakka dialect. From the perspective of institutional cooperation, Hakka Research Institute / School of Literature of Jiaying University, Chinese Dialect Research Center of Jinan University, and Department of Chinese Literature of Sun Yat-sen University as cores formed a relatively obvious cooperative network group, which promoted research development and cultivation of Hakka dialect.
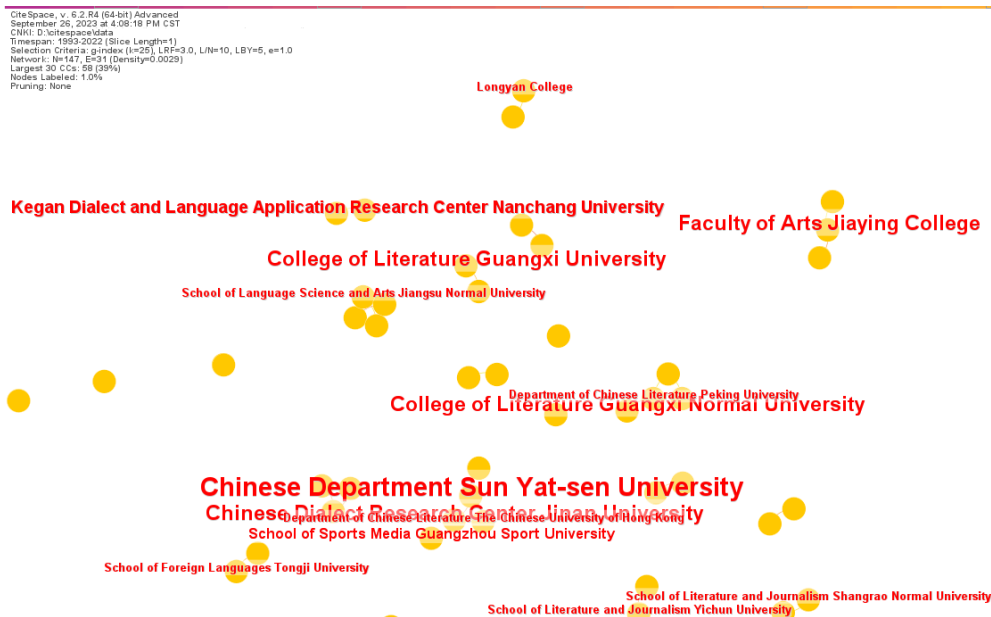
Figure 3: Co-present Network of Research Institutions

E. Main Carrier Analysis of Journal Papers

We comprehensively investigated publishing carriers and distribution of core journals and the number of journal publications on the subject field.

From the perspective of core journal areas of Hakka dialect research, there were 27 journals of Dialect, 12 Journal of Jiaying College, 12 journals of Longyan College (including Longyan Teachers College), and 6 journals of Language Research Collection. Based on data analysis, Dialect and Language Research Collection were professional journals of domestic language disciplines, which were both source journals of Chinese Social Science Citation Index (CSSCI) (2021-2022) and Overview of Chinese Core Journal of Peking University. Journal of Jiaying College was the journal of Hakka dialect research and Longyan Journal was located in Longyan, the place where Hakka people are gathered. In terms of overall distribution, the majority of top ranking journals were university journals in Guangdong, Jiangxi and Fujian provinces and the comprehensive journals of Hakka dialect papers showed certain regional features, mainly in universities in Guangdong, Jiangxi and Guangxi provinces where Hakka people gathered.

Although interdisciplinary collaborative research is the trend of dialect research, it has not formed core journal groups in other related disciplines from the perspective of publication carrier distribution. Academic circle should vigorously promote interdisciplinary research on Hakka dialect and analyze and discuss "Hakka dialect" in a more diversified disciplinary background.

F. Statistical Analysis of Keywords in Journal Papers

As the theme of article extraction, keywords can reflect core article contents and help dig deep into research connotation in this field. Research hotspots and predicted evolution trends can be analyzed through occurrence frequency, cluster analysis and co-occurrence map of keywords.

1) Keyword Co-occurrence Map Analysis: Keywords express the topic of the paper, which is the core and essence of the research. As shown in Figure 4, in the keyword co-occurrence network drawn by VOSviewer, Hakka dialect phonetics research field mainly included homophoncharacters, phonetic characteristics, rhyme characteristics and rhyme, which reflected the hot spots of phonetic research on Hakka dialect in the past 30 years and constructed knowledge network in this field. As was seen from the keyword co-occurrence map, the whole high-frequency keyword network was relatively close, which to some extent showed that research on Hakka dialect phonetics was developed around these cores.
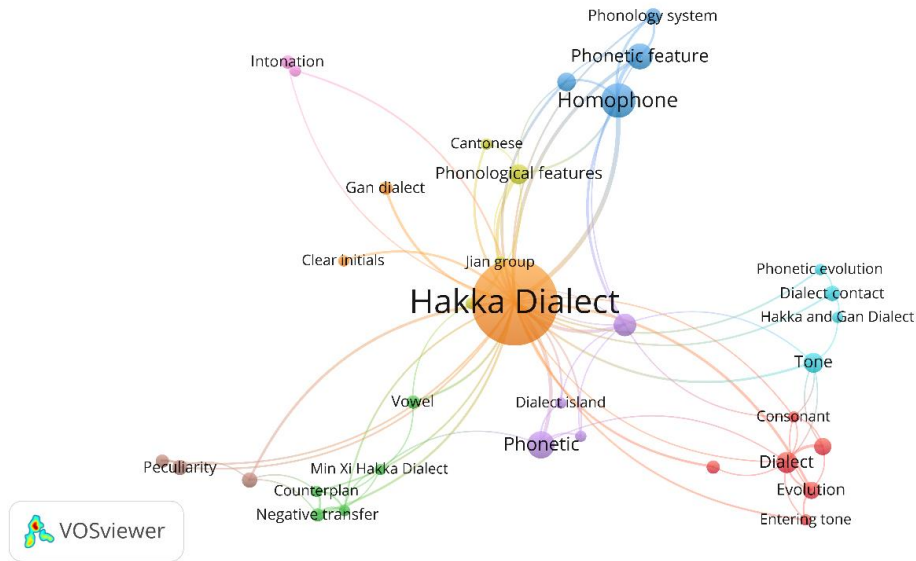
Figure 4: Keyword Co-occurrence Network (including clustering)

2) Keyword Cluster Analysis: Keyword cluster analysis was applied to classify similar keywords based on keyword co-occurrence. The principle was to extract keywords within the retrieval time range on the basis of log-likelihood algorithm (LLR). After operation, keywords were clustered and each cluster was consisted of multiple closely related words.

3) Keyword Emergence Trend Analysis: Keyword outburst could show the evolutionary dynamics of hot research spots and predict its development trend. The outburst trend chart includes Keywords, Year and Strength, where Strength is prominence intensity; the greater the prominence, the more obvious the research frontier; Begin is the starting year of a keyword research hotspot; End is the ending year. Also, Begin and End were the starting and ending years of a keyword research hotspot, respectively. The highlighted year of the keyword was the hot spot tracked in the research field in a relatively short period of time.

Figure 5: Trend Chart of Emerging Keywords

As was seen from the time of sudden termination of the keywords shown in Figure 5, since 1995, Hakka dialect researchers have tried to investigate language ontology "clear voice" and "full voiced voice". From 1996 to 2004, research on "Hakka people" and "partial characters" of Hakka people was performed; that is, investigation on dialect culture and dialect application began to attract Hakka dialect phonetics researchers. In the following several years, ontology research remained the predominant role. In addition, research on "Hakka folk songs", "Banshan Hakka" and "dialect island" also gained popularity. In recent years, "speech evolution" and "dialect contact" have become the hot topics in common academic attention. This showed that in recent Hakka dialect research works, the above-mentioned keywords were relatively new fields in China and need to be further explored.

III. GENERAL OVERVIEW OF MASTER'S THESIS IN HAKKA DIALECT PHONETICS RESEARCH

Another important sign reflecting academic attention to Hakka dialect phonetics, especially the research progress of Hakka dialect phonetics, was that many universities across the country adopted Hakka dialect phonetics as research topic, where in the full-text database of "Bo and Master" dissertations, a total of 8 related doctoral dissertations and 71 related master's dissertations were published by September 2023.

Regarding the number of published papers, Guangxi Normal University issued the highest number of papers, with a total of 15, followed by Jinan University, with a total of 11 and Guangxi University with a total of 10. Guangxi is one of the main places where Hakka dialects are distributed. Both Guangxi Normal University and Guangxi University in Guangxi Province have a large number of theses on Hakka dialect research, most of which are on Hakka dialect research in Guangxi. Jinan University, based on the Chinese Dialect Research Center, has trained a number of graduate students in Hakka dialect research at home and abroad, whose research scope is mainly in Guangdong Province and overseas Hakka dialects. This is due to the key project of the National Social Science Fund "Construction and Research of Guangdong and Fujian Dialects Geographic Information System" undertaken by Gan Yu 'en and the major project of the National Social Science Fund "Research of Chinese Dialects and Culture in Overseas Chinese communities" undertaken by Chen Xiaojin. In addition, some other universities such as Fujian Normal University, Guangdong Normal University of Technology, Jiangxi Normal University and other universities also had many Hakka dialect phonetics research dissertations. The distribution of relevant papers in different universities is shown in Figure 6.
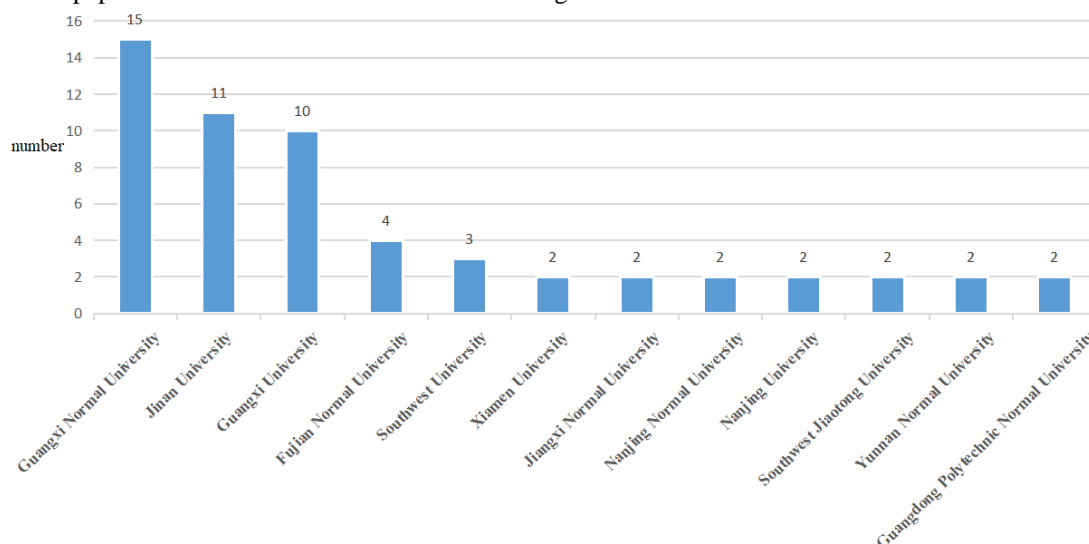

Figure 6: Distribution of Relevant Papers in Different Universities

In terms of thesis level, doctoral dissertation should put forward innovative theories, methods or applications, with higher far-reaching value and significance than master's thesis. Corresponding training of doctoral students should be a purposeful export to the academic field, while master's students only need to develop their ability to analyze problems. Will retrieved 8 doctoral thesis from East China Normal University, Jinan University, Fujian Normal University, Shanxi Normal University, Beijing Language and Culture University, Shanghai Normal University, Hunan Normal University, and Graduate School of Chinese Academy of Social Sciences. It can be seen that these 8 universities have cultivated a group of late talents in the field of Hakka dialect phonetic research.

Based on the findings of this work, the above 79 dissertations could be divided into two major types. The first category included comprehensive studies of Hakka dialect, which were mainly found in the dissertations of Guangxi Normal University and generally selected a certain Hakka dialect site in Guangxi to describe phonetic

vocabulary and grammar. For example, Chen Ling, Study on Hakka, Jinbao Township, Yangshuo County, Guilin, Yuan Xin, Study on Hakka, Tongan Town, Pingle County, Guangxi. Two or more Hakka dialect points were selected to compare their phonetic vocabulary and grammar. For example, Liu Xiaoyi "Comparative Study of Guigang Hakka Dialect", Li Ling "Contact Study and Historical Level of Hakka Dialect around Pitou Town, Lianping County". The second category was Hakka dialect voice (rhyme) pure study, which in turn could be divided into two categories. Class 1 included Hakka dialect voice comprehensive study which also included a dialect point voice research and two or more dialect point voice comparative studies such as Jun-Feng Wu "Bobai county Hakka speech dialect geography study", Liu Tao "Meizhou guest words rhyme comparative study". Class 2 covered Hakka dialect phonetic research, mainly focusing on Hakka dialect vowel, tone and pitch, such as Lv Xiaojia "Hakka dialect of elastic vowel sound output research", Dan-Dan Shao" based on EGG Mei country, Fuzhou, Changsha dialect tone experimental research", Ma Jieqiong" Meizhou hakka pitch research".

These dissertations not only recorded the phonetic features of Hakka dialect, but also tried phonological features by means of geographical linguistics and experimental phonetics.

IV. ACADEMIC COMPILATION OF HAKKA DIALECT PHONETICS

Academic compilation is a centralized, systematic, comprehensive and in-depth discussion of a certain subject, field or topic. Compared with natural science, it is more extensive, complex and social. Hence, in the fields of humanities and social sciences, Academic compilation is undoubtedly a more advantageous type of literature than journal papers. We retrieved research works on Hakka dialect phonetics from the "Book" column of "Reading Search" through manual screening and excluded Hakka dialect vocabulary, grammar, culture or other research results to reach a total of 36 books. They included 24 comprehensive studies, 5 collections of seminars, 3 phonetic records and 4 monographs on phonetic studies of Hakka dialect. The distribution of annual published number is shown in Figure 7, the horizontal coordinate represents the year and the vertical coordinate represents the number of academic compilation.
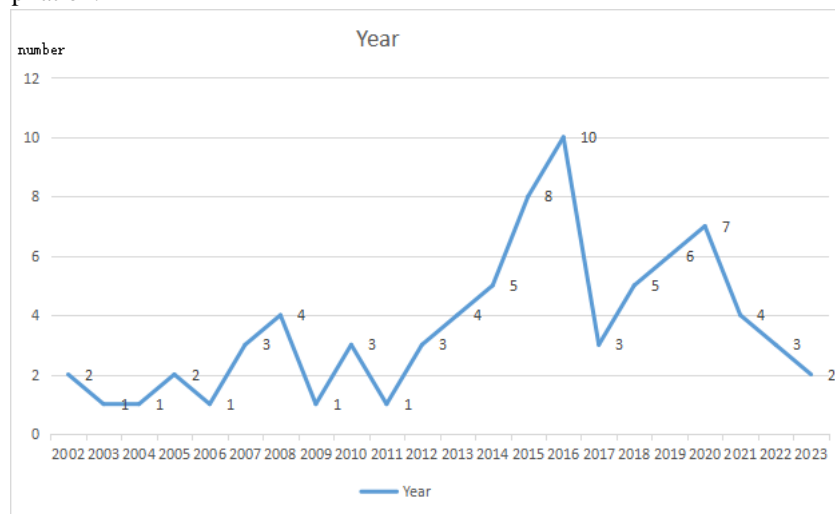

Figure 7: Annual Academic Compilation Distribution Chart

Echoing Hakka dialect journal papers, since the Survey report of Hakka-Gan dialect, it showed a relatively stable growth trend year by year, with four peaks in 2008, 2010, 2016, and 2020. From the perspective of types, in the 1990s and before, Hakka dialects were mainly recorded and described. In the 21st century, monographs on the characteristics and origins of Hakka dialects appeared from the perspective of phonetics. Later, with the emergence of interdisciplinary research, the introduction of geographical linguistics and experimental phonetics contributed to the study of Hakka dialects.

Among Hakka dialect phonetics research theses, 36 monographs focused on comprehensive study on Hakka dialect, in which phonetics research accounted for a certain proportion. There were only 4 monographs on research on pure Hakka dialect phonetics among which, Xie [11] discusses the phonetic problems of Hakka dialect from the phonology, studies and probes into some important characteristics of Hakka dialect consonants, vowels and tones, and probes into the special pronunciation of "milk, mouth, ze" in Hakka dialect, and also discusses the relationship between Hakka dialect and Gan dialect Zhao [12] selects 8 representative Hakka dialect points in Leizhou Peninsula, makes a detailed survey of their phonetic features, and makes a detailed report and distribution map of Hakka population in Leizhou Peninsula.

Qiu [13] taking "Minnan Hakka Dialect" as the research object, this paper studies the time structure of the Hakka vowels of Guanbei and Meilin, the influence of the vowel end on the vowel sound, and the synchronal variation of the vowel sound through the acoustic analysis of experimental speech. It also uses the SS ANOVA to fit the trajectory of the vowel formant, revealing the dynamic characteristics of the formant sound, and carries out acoustic analysis of the meson sound. It can be said that the first systematic and comprehensive study of Hakka dialect vowels from the point of view of experimental phonetics is made by constructing the sound change chain of the muon. Li [14] studied geographic linguistics mainly in Meizhou Hakka dialect voice as research object, with dialect map features and dialect speech features. She used dialect maps to display dialect phonetic features and regional characteristics, explore the commonalities and differences within regions, try to find the evolution mechanism of regional phonetic features in the distribution of regional dialects, and realize the dynamic display of Hakka dialect evolution on the map. At the same time, Hakka dialect materials from Guangdong-Hong Kong-Macao Greater Bay Area are introduced. The study on the geographical variation of Hakka dialect under the influence of strong Cantonese dialect is a representative work that first introduces geographical linguistics into the study of Chinese dialects.

Among the four monographs on research on pure Hakka dialect, three were developed based on doctoral thesis, which showed that doctoral dissertation played a pivotal role in Hakka dialect phonetic research.

V. CONCLUSION

This research showed that the achievements of research on Hakka dialect phonetics in the past 30 years were mainly reflected in the following aspects:

1. In terms of research type, we changed our research type from single description record to comparative research, and carried out thematic discussion based on comparative research. It was found that research means was more and more in-depth.

2. The scope of research was becoming more and more detailed, starting from big to small.

3. Research methods ranged from single description to multidisciplinary joint research.

According to current research situation, there are still some problems:

1. Researchers pay more attention to common phonology of Hakka dialect and ignore the regional characteristics of Hakka dialect distribution.

2. Based on descriptive analysis and comparison of ancient and modern documents, speech development clues are rarely speculated from phonetic correspondence and development history of speech is seldom explained.

3. Taking purely phonological level as observation point and considering phonological phenomenon from phonological level, cross-level research based on unique syllable structure of Chinese is still insufficient.

4. Multidisciplinary research methods are not enough. Combination of theory and method should be taken seriously and the interdisciplinary and comprehensive research of disciplines should be strengthened.

From the overall data statistics and analysis, it was seen that Hakka dialect phonetics research recorded many phonetic materials, with few core journal articles, strong literature research and regional features and regional publications lacked core publications. Research content lacked single, interdisciplinary or interdisciplinary investigations, lack of computer-based software quantitative analysis. Previous Hakka dialect seminars have resulted in the publication of many quality papers but not included in CNKI literature database. Research contents, methods and scopes should be more diversified and broadened.

ACKNOWLEDGEMENTS

REFERENCES

[1] Yuan Jiahua, et al, Chinese dialect summary, 2nd ed., China, Writing Reform Publishing House, 1983.06:145.

[2] Zhan Bohui, Zhang Zhenxing, On the Grand Dictionary of Chinese Dialect Studies, Guangzhou: Guangdong Education Press, 2017(05):359.

[3] Xiong Zhenghui, Division of the Cantonese dialect, Dialect, 1987, (3):162.

[4] Liu Tao, A comparative study on the rhyme of Meizhou guest, Master dissertation of Jinan University, 2003:4.

[5] Xie Liuwen, Huang Xuezhen, Division of Hakka dialect (draft), Dialect, 2007 (03): 238-249.

[6] Peng Zhifeng, Li Fei, Review of Hakka dialect research in 40 years —— Measurement analysis based on CNKI literature data, Southern Linguistics, 2022 (02).

[7] Li Fei, Overview of Hakka dialect research trend based on CNKI computer bibliometric —— focuses on phonetics analysis, Southern Linguistics, 2017(13).

[8] Chenyue,Chen Chaomei,Hu Zhigang, Wang Xianwen, Principles and applications of Analyzing a Citation Space, Science Press,2014.

[9] Qiu Junping, Yang Siluo, and Song Yanhui, Visual analysis of the research status of knowledge exchange. Library Journal of China, .2012(2): 78-89.

[10] Zong Shuping.2016. Core author evaluation based on Price's Law and comprehensive Index method —— takes Chinese Science and Technology Journals Research as an example. Research in Chinese Science and technology journals.(12).

[11] Xie Liuwen, Research on Hakka dialect pronunciation. Beijing: China Social Sciences Press, 2003, 12.

[12] Zhao Yue, The Hakka dialect pronunciation of Leizhou Peninsula, Beijing: China Social Sciences Press, 2015.12.

[13] Qiu Chun'an, Vowel experiment in Zhangzhou Guanpo and Meilin Hakka, Beijing: Intellectual Property Press, 2018(01).

[14] Li Fei, Research on the pronunciation of The Hakka dialect in Meizhou from the perspective of geolinguistics, Guangzhou: Guangzhou Sun Yat-sen University Press, 2022(01).