



A Study on the Bibliometric Analysis of Xianggui Southwest Mandarin - Guiliu Branch Based on Co-occurrence Analysis and Log-Likelihood Ratio (LLR) Algorithm

Dong Huang¹, HongLi Deng^{2,3}(✉), JingJing Wang⁴(✉), and BaiXue Ning⁵

¹ Nanning Normal University, Nanning, Guangxi, China

² School of Liberal Arts, Jinan University, Guangzhou, Guangdong, China
361389651@qq.com

³ College of Culture and Communication, Guangxi Science and Technology Normal University, Laibin, Guangxi, China

⁴ College of Vocational and Technical Education, Guangxi Science and Technology Normal University, Laibin, Guangxi, China
wangjj1204@126.com

⁵ Guangxi Science and Technology Normal University, Laibin, Guangxi, China

Abstract. Based on co-occurrence analysis and the Log-likelihood Ratio (LLR) algorithm, this paper analyzes the literature of Xianggui Southwest Mandarin - Guiliu Branch by using CiteSpace bibliometrics to generate a visual knowledge graph. This paper makes a systematic analysis on present research situation. The research results suggest that it is uncommon to see studies on Xianggui Southwest Mandarin - Guiliu Branch and cooperation among authors is still very rare. In addition, their study method is simplified, and the literature studies focus on phonology area distribution and pure phonology.

Keywords: co-occurrence analysis · minimum spanning tree (MST) algorithm · Guiliu Mandarin · knowledge graph

1 Introduction

The term “Xianggui Southwest Mandarin - Guiliu Branch” was first proposed by Huang Xuezhen (1986). She held the belief that the southwest mandarin inside Guangxi Province was phonetically similar to Liuzhou dialect and Guilin dialect, which could be categorized as Xianggui Southwest Mandarin - Guiliu Branch [1]. While, Li Lan (2009) proposed that the “Xianggui Southwest Mandarin - Guiliu Branch” should include “Xiangnan (South Hunan) Small Branch”, “Guibei (North Gui) Small Branch” and “Qiannan Small Branch”. The scope of “Xianggui Southwest Mandarin - Guiliu Branch” proposed by Li (2009) included not only the whole Guangxi Province, but also Xiangnan (South Hunan) area in Hunan Province and Qiannan area in Guizhou Province. This paper adopted Li’s (2009) definition of “Xianggui Southwest Mandarin - Guiliu Branch” [2].

Based on co-occurrence analysis and LLR algorithm, this paper firstly conducts a bibliometric analysis on phonetic literature of “Xianggui Southwest Mandarin - Guiliu Branch” in the database of “Academic Journals” of CNKI, then develops a visual knowledge map, in the end analyzes the “Xianggui Southwest Mandarin - Guiliu Branch” in Xianggui. After analyzing the current situation of the phonetic research of “Xianggui Southwest Mandarin - Guiliu Branch” and the relevant reasearch hotspots in recent years, this paper makes an outlook on the future phonetic research of “Xianggui Southwest Mandarin - Guiliu Branch”, and puts forward some suggestions.

2 Theoretical Basis

This paper is based on citespace’s text mining and visual analysis of the relevant literature of “Xianggui Southwest Mandarin - Guiliu Branch” research. The theoretical basis is mainly embodied in the following four aspects, including co-occurrence analysis of keywords, the minimum spanning tree (MST) algorithm, the pathfinder network and the Log-likelihood Ratio (LLR) algorithm [3].

2.1 Co-occurrence Analysis

Co-occurrence analysis refers to the analysis of the occurrence of words in a literature to determine the relationship among topics in the discipline represented by the document.

At first, co-occurrence analysis identifies the keywords contained in the literature and develop a literature-keyword matrix. Then, terms are collected from the titles, authors, keywords, and abstracts. In the end, node types are selected to conduct co-occurrence analysis on these terms, and the analysis results can be visualized by using a knowledge network approach (Fig. 1).

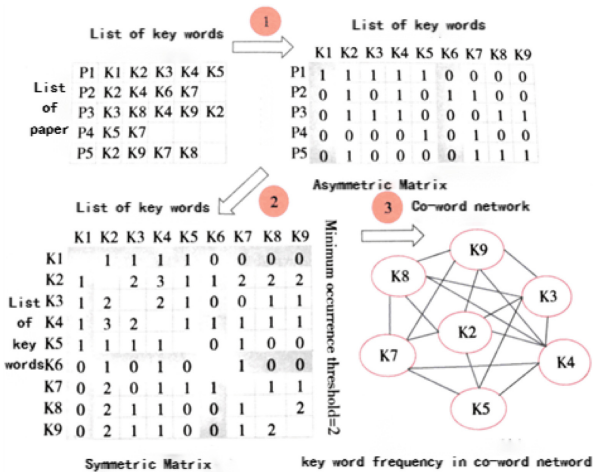


Fig. 1. CiteSpace Text Mining and Visualization in Scientific Literature

2.2 Minimum Spanning Tree (MST) Algorithm and Pathfinder Network

In the Pruning region, the connectivity in the network is trimmed by Minimum spanning tree (MST) algorithm and pathfinder network to improve the readability of the network. The following two network pruning methods are mainly used: Minimum Spanning Tree (MST) and Pathfinder Network (PFNET).

The minimum spanning tree algorithm refers to constructing a minimum spanning tree containing the sum of all vertices and weights from the original graph. The number of network nodes processed by the path-finding network algorithm will not be changed, but the number of connections will be greatly reduced.

The structure of the pathfinder network is mainly determined by the parameters r and q . r measures the length of connected paths of nodes in the network. When a measurement space is given, the triangular inequality relation is defined as: $w_{ij} \leq (2 - r)w_{ik} + rw_{kj}$. It meets the maximum number of connections for the triangle inequality. q can be set to any integer between $[2, N-1]$, where N denotes the number of nodes in the network. When r approaches 1, then $q = N-1$, which shows the network pathfinder algorithm reaches the maximum clipping capacity.

2.3 Log-Likelihood Ratio (LLR)

The Log-likelihood Ratio (LLR) algorithm is used to calculate the similarity between events by counting two different events and is mainly used for keyword cluster analysis in literature databases. This paper uses the Log-likelihood Ratio (LLR) algorithm to obtain the clustering map of the paper before interpretation [4].

3 Data Sources and Research Methods

CNKI is used as data source in this study. The search topic is (Southwest Mandarin + Guiliu dialect + Guiliu mandarin + Guiliu dialect + Guilin dialect + Liuzhou dialect + Yongzhou dialect + Guiyang dialect + Jiangyong branch + Chenzhou City)* (phonetic + phonemes + rhymes + tones + vowels + consonants). The search time ranges from 1985 to 2022. After excluding literature that do not meet the requirements, a total of 190 relevant literature is obtained. Therefore, the research on the phonology of Xianggui Southwest Mandarin - Guiliu Branch can be visually analyzed.

This study adopts a bibliometric analysis method, and the tool used is Citespace v6.1.R3. This software is a visualization software developed by Prof. Chaomei Chen, which can integrate and measure literature of a specific field, explores the key paths of the evolution of a subject area, and predicts the new trends of the development of related disciplines. Based on this, this study analyzes Xianggui Southwest Mandarin - Guiliu Branch to sort out the current status and future trends of research in this area.

3.1 Distribution of Posting Time

According to Fig. 2, we can roughly divide the research on Xianggui Southwest Mandarin - Guiliu Branch into three stages:

1. 1985–2002. It is the initial stage of the research, with few papers published.

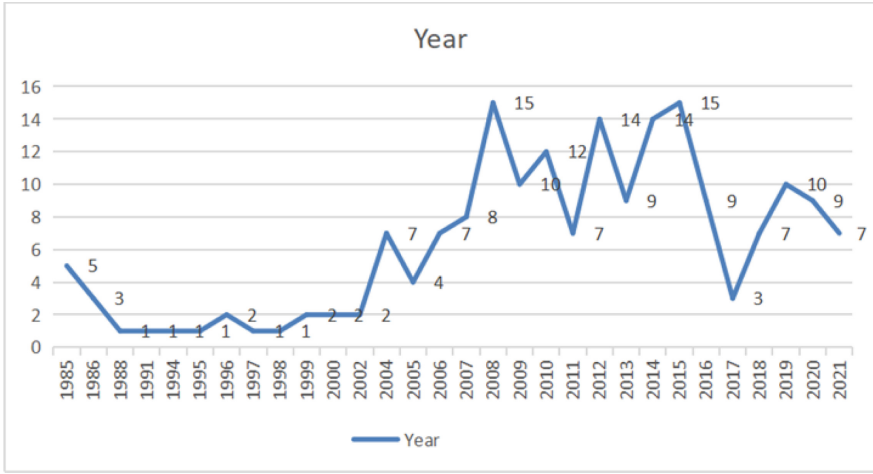


Fig. 2. Distribution of the annual number of publications(owner-draw)

- 2. 2002–2015. It is the stable growth stage, showing a significant growth trend, with 124 publications, 104 more than the first stage.
- 3. 2015–2021. It is the saturation state of publication, and the number of publications decreases year by year, and the downward trend is very obvious.

It should be noted here that this study was conducted in the second half of 2022, and the papers in 2022 have not yet been published on CNKI, so the data are not available for 2022.

3.2 Analysis of Major Publication Vehicles

According to Fig. 3, we can see that 10 of the top 15 journals with more articles are from Guangxi Province, and only 2 from Hunan Province. The author believes that this is related to the geographical scope under the jurisdiction of Xianggui Southwest Mandarin - Guilin Branch. There are only 14 counties and cities in Hunan Province, but there are 42 counties and cities in Guangxi Province. Therefore, there are more literature studying Guangxi Xianggui Southwest Mandarin - Guilin Branch, 36% of which is published in local journals in Guangxi Province. The journal with the largest circulation is Journal of Guilin Normal College of Higher Education, with 18 articles. It has something to do with the fact that the journal opened the columns of “Research on Guibei (North Gui) dialect”, “Research on Guibei (North Gui) dialect and surrounding dialects” and “Research on Guangxi dialect” in 2005, 2006 and 2007 respectively. Because of these column such as “Research on ping dialect of north Guangxi”, “Research on ping dialect of north Guangxi and Surrounding Dialects” and “Research on Guangxi Dialect” in 2005, 2006 and 2007 respectively, it make it a main publication carrier of Guangxi Mandarin.

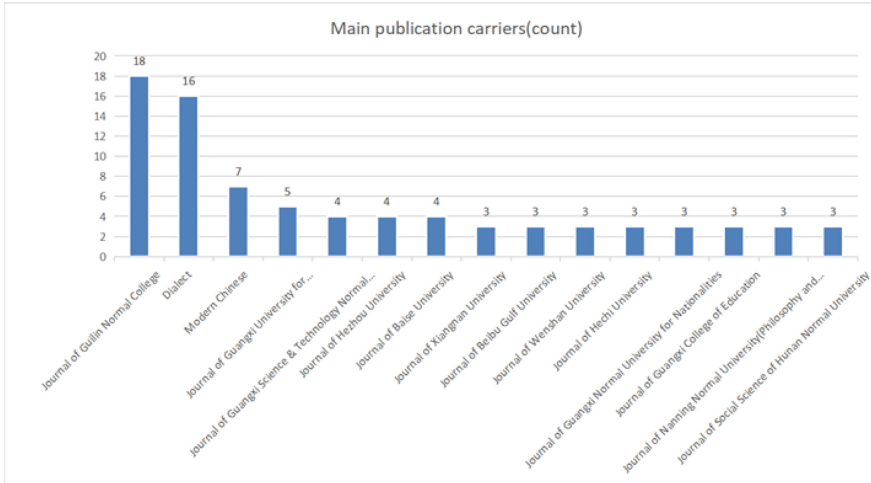


Fig. 3. Volume of articles issued by major carriers (owner-draw)

3.3 Distribution of Major Research Institutions and Cooperation

According to Fig. 4, the top 14 research institutions with more publications are recorded in this paper, among which are 8 research institutions in Guangxi Province and only 1 research institution in Hunan Province. It still has something to do with the fact that there are more phonetic literature studies in Guangxi Mandarin, and 58% of them are Guangxi research institutions. Among them, Guangxi University is the research institution with the largest circulation - 28 articles, which has something to do the establishment of the degree conferring institution of the relevant academic majors in this institution earlier. Guangxi University has been conferred the master’s degree of Chinese Linguistics and Literature as early as 1980s. Among the articles published in Guangxi University on Xianggui Southwest Mandarin - Guiliu Branch, most of them are master’s theses, with a total of 12 papers, accounting for about 43%. Hunan Normal University is the main research position of small pieces in Shonan, with a total of 9 papers.

From the Fig. 5 research institution count chart, it is easy to tell that there remain very rare cooperation among research institutions on Xianggui Southwest Mandarin - Guiliu Branch, and there is only cooperation between Hunan Normal University and Yulin Normal College, with 2 articles. Hence it is can easily be concluded as follows:

1. Research institutions in Xianggui Southwest Mandarin - Guiliu Branch are mainly universities in Guangxi;
2. There are few cooperation among universities at present.

3.4 Core Author Analysis

The concept - Core Author is proposed by scientist Derek John DeSola Price. It refers to the highly productive authors of papers in the field with certain academic influence (Fanxiu Kong and Yao Sun 2020). This paper uses the number of publications and



Fig. 4. Main research institutions (owner-draw)

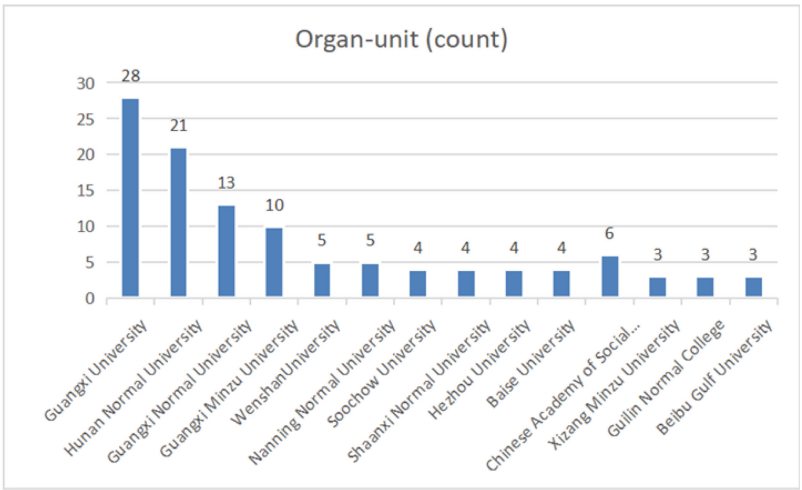


Fig. 5. Research institution count chart (owner-draw)

citations to measure the comprehensive index of authors, so as to screen out the core authors of Xianggui Southwest Mandarin - Guilin Branch.

According to Price’s Law, the minimum number of papers published by core authors is m_p , which can equate as $m_p = 0.749\sqrt{np_{max}}$, where m_p is the minimum number of papers published by core authors in the statistical time period, and np_{max} is the maximum number of papers published by core authors in the statistical time period (Ding Xuedong 1992). This paper calculates the number of papers published by the core authors on Xianggui Southwest Mandarin - Guiliu Branch. And the result is $m_p \approx 3.08$. According to rounding principle, the authors who publish three or more papers are selected as core author candidates.

The core authors, according to the comprehensive index, are Li Lan, Bao Houxing, Liu Cunhan, and 16 others. The first one is Li Lan, whose “Classification/Distribution of Southwest Mandarin (Draft)” is a collection of textual descriptions of the southwest mandarin Distribution Map of the New Atlas of Chinese Language. It is a more comprehensive re-division of southwest mandarin based on the previous studies, and it is of great value to the study of southwest mandarin. From Fig. 5, it can be seen that there are very few cooperation among the these authors, for instance, collaborations among Tang Qiyuan, Qiu Hao Yang and Luo Ting; collaborations between Gong Na and Luo Xinru; and collaborations between Liu Cunhan and Bai Yun. In conclusion, Xianggui Southwest Mandarin - Guiliu Branch of phonetic research is dominated by individual research and lacks cooperation and interaction.

In this paper, keywords in the range of 1985 to 2022 are extracted according to the Log-likelihood Ratio (LLR), after that, keyword clusters can be obtained. Each cluster consists of multiple closely related words, and the smaller the cluster number, the more keywords are contained in that cluster. The figure shows the keywords cluster, and the average value of knowledge mapping module (Q value) in the table is 0.804, and the average value of network homogeneity (S value) is 0.9567. Generally, when Q value > 0.3, it indicates that the cluster structure is significant. When S value > 0.5, it indicates that the cluster is reasonable. Therefore this cluster is significant. The cluster numbers

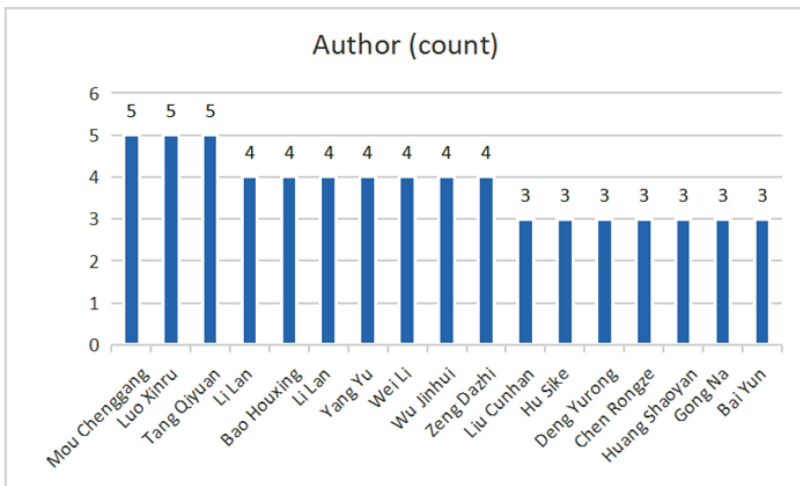


Fig. 6. Author count chart (owner-draw)

0, 1, 2, and 5 mainly examine the dialect partitioning, and 3, 4, 7, and 8 focus on phonological ontology research. Combining with Fig. 6, the author categorizes the key words with similar meaning again, and comes up with two important areas of internal partitioning and phonological ontology research of Xianggui Southwest Mandarin - Guiliu Branch (Table 1 and Fig. 7).

3.5 Exploration on Southwest Mandarin Classification/Distribution

(1) Exploration on Xiangnan (South Hunan) Small Area

Zhou Zhenhe and You Rujie (1985) classified twelve counties, including Chen County, Guiyang, Yongxing, Xintian, Ningyuan, Yizhang, Linwu, Jiahe, Lanshan, Jianghua, Jiangyong, and Dao County, as a mixed area of Southwest Mandarin and Xiang dialect. The Classification/Distribution of Southwest Mandarin (Draft) (1986) and The Classification/Distribution of Xiang Dialects first classified the above-mentioned counties and cities as the “Xiangnan (South Hunan) Branch” of southwest mandarin, meanwhile The Classification/Distribution of Xiang Dialects mentions that the “Xiangnan (South Hunan) Branch” is a place where both local dialect and official dialect coexist [5]. Chen Hui and Bao Houxing (2007) referred to the “Xiangnan (South Hunan) Branch” as the “Yongchen Branch”. Li Lan (2009), for the first time, classified the “Xiangnan (South Hunan) Branch” as the “Xiangnan (South Hunan) Small Area” in the “Guiliu Branch”. Therefore, there is less controversy about the internal division of the Xiangnan (South Hunan) Small Area, and names are different, but the reference is the same.

(2) Exploration on Guibei (North Gui) Small Area

Li Lan (2009) separated the Cantonese, Hakka, Guinan (South Gui), and Min dialect of Guangxi from the Southwest Mandarin according to (“whether the ancient checked tone systematically preserves the final stop consonant”, and any dialect that systematically preserves the final stop consonant can be defined as a non-official dialect. According to (“the present-day pronunciation of the ancient full turbid vowels” of Guibei (North Gui) Ping language, Xiang language, and various other local dialects are separated from the Southwest Mandarin, and only the Southwest Mandarin is “clearing, with the Pingsheng releasing air and the Zesheng not releasing air.” Among the dialects that qualify as the Southwest Mandarin, the first thing to observe is “whether the checked tone is now pronounced as rising tone”, and those that meets this requirement are basically can be defined as the Southwest Mandarin. For dialects with checked tone, it needs to observe whether the forth tone is close to Guilin dialect, and those with intonation close to Guilin dialect can be classified as Southwest Mandarin even if there is a checked tone. (Most of the dialects of Wuming, Xia Guojie, and Fucheng have Yangping tone and a few have checked tone with [-p-t-k] rhyme endings. However, Li (2009) did not simply classify the Wuming dialect as a Southwest Mandarin, but classified the Xia Guojie dialect with the same characteristics as a southwest mandarin. Huang Xuezhen (1986) and Liu Cunhan (2012) classified Wuming Mandarin as a branch of Southwestern Mandarin, and Liu (2012) also divided Li’s (2009) Guibei (North Gui) Small Area into Guibei (North Gui) Small Area and Guinan (South Gui) Small Area [6]. Although Yang (2005) and Li (2010)

Table 1. Keyword clustering table (owner-draw)

Cluster Number	Number of clustered keywords	Profile value	Year Median	Cluster Labeling (LLR)	Other major keywords in the clustering tag (LLR)
0	28	0.993	2005	Chinese dialect	Guangxi, vocabulary distance, day mother character
1	25	0.955	2011	Guangxi dialect	Language distribution, interference, Chinese southwest mandarin
2	24	0.946	2008	bureaucratic language	Compare, Xingan, Xiang
3	24	0.957	2009	Tone	Liuzhou dialect, comparative analysis, ancient intonation
4	24	0.917	2006	Mandarin	Liuzhou language, characteristics, teaching
5	21	0.945	2002	Dialect Division	Tone types, dialect islands, and subdivision methods
6	20	0.949	2008	Southwestern Mandarin	Internal, language attitudes, speech patterns
7	20	0.944	2013	vernacular	Labial fricative, sound change, Yongju
8	16	0.948	2010	Historical sound changes	Present-day sounds, phonetic sounds, ancient sounds

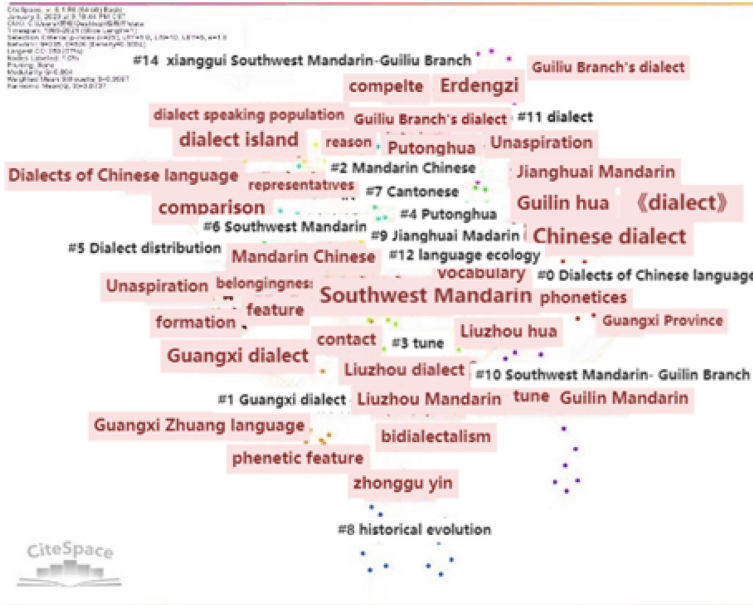


Fig. 7. Knowledge mapping mode (owner-draw)

had different views on the nature of the Wuming dialect inscriptions, they all agreed that Wuming dialect belongs to the Southwest Mandarin [7].

It can be seen that the attribution of Wuming, Fucheng and other southwest mandarin is related to the fact whether Wuming, Fucheng, Xiaguojie and other areas in southern Gui belongs to the northern Gui Small Area. It is of great value to the internal distribution of the Northern Gui Small Area to clarify their territorial attribution.

3.6 Phonological Studies

(1) Study on Tones

Most of the vocal tones in Xianggui Southwest Mandarin - Guiliu Branch have 4 tones, and the most special ones are the Southwest Mandrin of SouthGui and Xiangzhou, where most of the checked tones are classified as yang tones, but some of them are pronounced as checked tones with [-p-t-k] final stop rhyme.

There are currently two views on the nature of the phonetic characters in the southern Guinan region:

A. It is derived from the original structural system. Yang Yuguo (2005) compared Wuming Mandarin, Nanning baihua, and Wuming Zhuang language. They believed that among the 10 characters listed and compared, none of the words in Wuming Zhuang language had the same rhyme ending as Wuming Mandarin, and the Nanning baihua had different rhyme endings except “shi” sound. In addition, the character that retains the ending sound T is generally regarded as the Chinese characters of “shan xian shen zhen”, and the character that retains the ending sound K is generally regared as “tong

jiang zeng geng”, so he believes that the checked tone character of Wuming Mandarin is derived from the original structural system.

B. It arises from linguistic contact and is not a manifestation of antiquity. Li Lianjin (2010), Cao Ping (2010), Zheng Shilei (2012), Lu Miaoyan (2016), etc., Li (2010) and Cao (2010) believed that the checked tones with ending sound t and k in Wuming Mandarin are a kind of pragmatic sound produced by drawing on the surrounding plain pronunciation in order to move closer to the standard of zheng yin in official rhyme books, and are not the inheritance of ancient Chinese invoicing [8].

(2) Study on Initials and Finals

At present, the vocal rhyme descriptions and studies of Xianggui Southwest Mandarin - Guiliu Branch of official dialect generally appear in single-point phonological systems or phonological comparisons, such as Yang Huandian (1964), Liu Cunhan, Li Rong (1995), Deng Haifeng (2004), Zeng Xianfei (2004), Hu Sike (2009), Yang Yu (2008, 2009), Yang Yuguo (2005), Zhang Yibing, Bai Yun (2009), Li Lianjin (2010), Tang Qiyuan, Lu Dapeng (2019) et al., Huang Zhaoyan (2008), Zheng Shilei (2012), Cao Ping (2010), Lu Miaoyan (2013), etc [9]. According to the previous studies, the vocal-rhyme characteristics of Guangxi mandarin can be roughly grasped as follows:

① Except for the four counties in North Guangxi, the small piece of northern Gui has all the ancient turbid sibilants and sibilant consonants cleared, and the “ping” and “ze” tone don’t release air. Some of the dialect points in the small piece of southern Hunan retain turbid consonants.

② The “ni” and “lai” tone is divided into two, the Guilin Mandarin in the small area of southern Hunan and Northern Guizhou, and the “ni” and “lai” tone in the whole irrigation official dialect are not distinguished, while the Mandarin in other areas separate n and l.

③ There is no lingual post-vocalic vowel.

④ The Mandarin in South Gui has a border fricative [ʈ].

⑤ The sharp group sound is mainly divided, only Guilin, Leye and other local official languages are not divided.

⑥ Only the prelingual unrounded labial vowel [ɸ].

⑦ Guilin and Liuzhou have nasalized vowels in their official languages, while Wuming and Xiangzhou do not.

⑧ South Gui Mandarin has [-m, -p, -t, -k] rhyme endings, while other small areas’ do not [10].

The articles in this category record and describe the phonetic system of some dialect points of the Mandarin of the Guangxi - Guiliu Branch, providing important reference materials for further investigation and analysis by later generations.

4 Questions and Reflections

4.1 Research Power

According to the analysis, the research of Xianggui-Guiliu Branch still mainly relies on the university institutions, but there is no strong cooperation among the various research universities. Except that, most researches appear in Guangxi Province and

Hunan Province, which are not evenly distributed. In the author's opinion, the cooperation among universities should be strengthened and resources should be shared to realize cross-regional research on the Xianggui Southwest Mandarin - Guiliu Branch and to promote the development of research.

4.2 Research Methodology

The research method of Xianggui Southwest Mandarin - Guiliu Branch of official phonetic is relatively single, basically the traditional phonetic method to describe the single point phonetic system or use different dialect areas for the contact comparison, rarely can use the method of phonetic experiments. Even if phonological experiments are used, they are only limited to single-word tone experiments. Therefore, the author believes that we should continue to innovate, apply more different phonetic experimental instruments, and we can study regional dialects better through information technology such as speech recognition, and provide a more accurate depiction of the official phonetic system of Xianggui Southwest Mandarin - Guiliu Branch.

5 Conclusion

This paper analyzes the existing literature on Xianggui Southwest Mandarin - Guiliu Branch phonology in a more systematic way based on bibliometric tools such as Citespace and Excel, which helps to clearly show the overall situation and future development trend of Xianggui Southwest Mandarin - Guiliu Branch phonology research. However, there still remain some limitations such as, the normalization of data processing and the setting of related parameters. Thresholds may have more or less influence on the analysis results, and more data analysis functions in Citespace software are not all used, which are to be further developed in the future.

Funding. 1. The key research base of humanities and social sciences of ordinary colleges and universities in Guangdong Province (Chinese Dialect Research Center of Jinan University).

2. Guangxi Philosophy and Sociology Science Planning Research Project "Guangxi Southwest of Mandarin Phonetic Acoustics and Perception Experiment Research" (20FYY018).

3. Research and practice of the integrated teaching mode of "exploration-imitation-practice-evaluation" of language expression courses from the Perspective of New Liberal Arts (2022JGA374) Phased results.

4. The Basic Ability Improvement Project for Young and Middle-aged Teachers in Guangxi Universities -National Language Promotion Base of Guangxi Science & Technology Normal University (YWJD2014502).

5. 2021 at the scientific research platform "Guizhong Culture Research Center" of Guangxi Normal University of Science and Technology (GXKSKYPT2021001).

References

1. LiHuang Xuezheng. Subdivisions of Southwestern Official Dialects (Draft) [J]. *Dialects*, 1986(04):262-272.
2. Lan. Partitioning of Southwestern Official Dialects (Draft)[J]. *Dialects*,2009,31(01):72-87.

3. Chen, Y.,Chen, Chaomei,Hu, C.G.,and Wang, X.W.. 2014. Principles and Applications of Citation Space Analysis - A Practical Guide to CiteSpace.
4. Li J,Chen Chaomei.CiteSpace Technology Text Mining and Visualization (3rd edition) [M]. Capital University of Economics and Business Press.2022.
5. Bao Houxing,Yan Sen. Zoning of Hunan dialects[J]. Dialects, 1986(04):273-276.
6. Yang Huandian, Liang Zhensi, Li Puying, Liu Cunhan. The Chinese dialects of Guangxi (draft) [J]. Dialects, 1985(03):181-190.
7. Yang Yuguo. A survey of accented characters in Guangxi Wuming official dialect [C]//. Chinese Sociolinguistics (2005, No. 2). ,2005:173–183.
8. Li Lianjin. The nature and causes of the accent in Guangxi Wuming official dialect[J]. Ethnic Languages, 2010(02):62-69.
9. Duan Hao. Research on the official language of Nalu Village, Luoxiu Town, Xiangzhou County, Guangxi[D]. Guangxi Normal University,2019.
10. Tang Qiyuan, Qiu Hao Yang. A repertoire of homophones in Zhongping official dialect of Xiangzhou, Guangxi[J]. Journal of Beibuwan University,2019,34(11):68–77.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

