

The Literature and Problems in the Research of Chinese Dialects

Bohan Li

Dec 18, 2021

LING195 Independent Paper

ABSTRACT: The study of Chinese speech forms has always been a fascinating and confounding topic. The rich diversity poses many challenges while examining the concepts of language and dialect, and it furthermore inspires various approaches to account for the complexity. The controversy in regards to the categorization and identification of dialects within the Chinese language family arises from many different aspects. While no universal criterion or standard exists, the paper will introduce several major theories and literature that should be studied before conducting research in the field. As a brief literature review and general introduction, the paper intends to summarize the state of research in Chinese languages and dialects and inspires myself to discover potential research topics. It may also serve as a guideline for those interested in this discipline yet have minimal background in traditional dialectology, language variations, or sociolinguistics.

The paper has three major components. Starting with the discussion about the intricate relations between languages and dialects, it will highlight some important milestones in the development of Chinese dialectology and then introduce existing resources for Chinese dialects. In the last section, we will also be discussing the problem of inadequate documentation and organization of dialect resources and why and how it is beneficial to resolve the

problem.

KEYWORDS: Dialectology, Chinese Dialects

1 Introduction

The study of Chinese languages and dialects has a very long history. The rich geographic and social variations make the topic extra meaningful yet challenging. To non-linguists, a dialect is usually a substandard and sometimes rustic form of a language that is less prestigious (Chambers & Trudgill 1998). It is often considered the opposite concept to "standard language" or "common speech", despite the fact that "standard language" is, in fact, merely a form of language that is widely recognized and used. Today, as more and more discussion has been carried out by scholars, it is often agreed that dialects are a very important component of any language or culture, natural or social science, and humanity. In the past several decades, dialectology has gained more attention and well-designed systematic research. The uniqueness of Chinese dialects and languages has inspired many linguists around the world, and it is important to identify what the field has achieved so far and what remains to be uncovered to facilitate further studies. The paper intends to address literature from Western and Chinese sources in the space available.

The paper has three major components, excluding the two sections of the introduction and the conclusion. The first section mainly sets the prerequisites in general dialectology and sociolinguistics for further discussions explicitly in Chinese dialects. The second section will briefly go over the history of Chinese dialectology with a few remarkable works, and talk about the general taxonomy of Chinese dialects, followed by the third section that summarizes the existing dialect resources. The last section will also illustrate a number of problems encountered during the research process and share some personal thoughts and insights.

2 Dialectology: The Basis

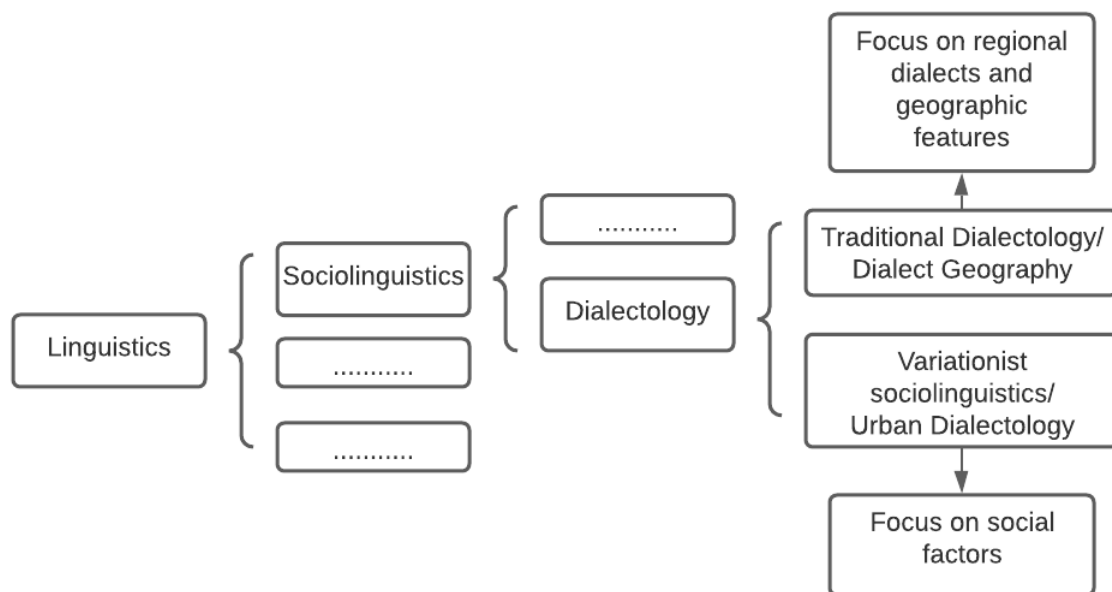
The origin of the word "dialect" can be traced back to Classical Greek. It first had the meaning of a discourse or a way of speaking, then gradually shifted to represent a regional variety of a language (Hickey 2018). The term commonly refers to a geographically delimited form of a language. A dialect, by definition, cannot exist on its own, as it must belong to a language. Therefore, one of the oldest and most critical questions that bothers linguists is how to distinguish between a language and a dialect.

Before getting lost in this question, it is important to obtain a bigger picture of the field and think about several prerequisites. Specifically,

- What disciplines are involved in this question?
- Why is it one of the most confounding questions?
- What is the common ground that needs to be established before approaching Chinese dialects?

2.1 Dialectology and Sociolinguistics

In a broader sense, dialectology is the study of linguistic dialects. Dialectology is the primary discipline that is associated with dialects. Nevertheless, the term bears different meanings as the field develops. The graph below illustrates where dialectology stands within linguistics nowadays.



Graph 1: Dialectology and Related Fields

It is now commonly accepted that dialectology is one of the sub-fields of sociolinguistics. This systematic study officially began in Europe after 1850 (Chambers 2015), but it has much earlier roots, even before the establishment of sociolinguistics.

The rise of sociolinguistics involved three renowned linguists: Labov, Chambers, and Trudgill. Labov was the founder of sociolinguistics. His seminal book, *Sociolinguistic Patterns* (1972), laid a solid foundation for this new field. The core of sociolinguistics is language variation. A simple interpretation of language variation, as Walker (2021) summarized, is "different ways of saying the same thing," where "different ways" refer to the various forms and aspects of language, and "same things" refer to the same meanings that are conveyed by the forms. Linguists generally divide language variation into two overarching areas: diachronic variation and synchronic variation. The former covers the differences found in a language throughout history, while the latter studies the many varieties of a language at a single point in time. Labov promoted synchronic variation as a major concentration of sociolinguistics to explain the quandaries of language change, social change, social groupings, and dialectology (Nerbonne et al., 2021). Chambers and Trudgill's important book *Di-*

alectology (1998) continued the discussion and struggled about how to incorporate Labov's work into the study of dialectology, and how the terminologies should be redefined. Eventually, the new structural relations were established. As Chambers and Trudgill introduced in their book, traditional dialectology, which is sometimes interchangeably referred to as dialect geography, was majorly concerned with spatial features or regional variations. They considered that the new meaning of dialectology should be "the study of language variety by any methodology," and traditional dialectology should be taken as one of the methodologies. From then on, dialectology transitioned to being the study of one component of linguistic variation that focuses more on the geographical causes and differences than on the variation found within speakers or speaker communities (Krug et al. 2013). Correspondingly, urban dialectology (also called social dialectology or variationist sociolinguistics) is the sub-field that emphasizes the importance of social factors, such as gender or social class.

The information provided above may appear to be off-topic from the main theme of the study. However, in order to limit down the topic of investigation, it is necessary to first gain a more comprehensive image. For the sake of this literature review, I will limit my attention to synchronic regional language variations that fall within the purview of traditional dialectology. According to the research I reviewed, the terms can be used in a casual manner. As a result, it is vital to repeat the meanings of terms when conducting research.

2.2 Differentiate a Language and a Dialect: The Ultimate Rabbit Hole

Referring back to Chambers and Trudgill's book *Dialectology* (1998), one way to look at the difference between a language and a dialect is that "a language is a collection of mutually intelligible dialects". The concept of "mutual intelligibility" is a well-known criterion for determining the link between speech forms in this context. I couldn't find the term's origin, but it's a pretty popular notion that should be covered in any freshman linguistics textbook. The book *Language Files* is an excellent resource (Department of Linguistics, Ohio State

University, 2011). In general, the phrase refers to the ability of speakers of various speech types to comprehend and communicate with one another without the need for formal instruction. And, rather than being two languages, the two speech forms are dialects of the same language. While this concept is useful, it is frequently insufficient to distinguish between a language and a dialect for a variety of reasons, as discussed in Chambers and Trudgill's book (1998):

- To begin with, linguistic characteristics aren't the only criteria in determining what constitutes a language and what constitutes a dialect. Most of the time, geopolitical factors play a larger impact. Serbo-Croatian languages are a well-known example. Despite the fact that speakers from various speech communities can communicate with each other with little difficulty, these speech forms are not considered dialects but languages. The vast diversity of the Chinese language is a well-known counterexample. The majority of northern and southern dialects are mutually unintelligible. It is not rare in southern China for people from two neighboring regions to be completely unable to understand each other's speech. This is an old Chinese proverb that eloquently illustrates such an occurrence: local tongues vary over every 10 miles, and local cultures vary over every 100 miles.
- Mutual intelligibility in both directions may not always be equal. In Scandinavian languages, for example, it is frequently stated that Danes understand Norwegians better than vice versa. In addition, one's ability to speak or understand a dialect may be influenced by one's amount of exposure to other languages, education, and social class.
- Another possibility is that dialects are scattered over a continuous continuum with no defined borders. Chambers and Trudgill introduced a particular puzzle: "a traveler walking in a straight line from village to village notices successive small changes, but seldom if ever, observe large differences." It's difficult to count how many dialects can be recognized on a continuum because these minor changes are typically subtle

and vary in different dimensions. Furthermore, because the difference is cumulative, the speech forms from the villages located at the ends of the straight line can differ significantly, yet they are still linked within the chain and share many characteristics with other speech forms in the continuum. Without more constraints, it's difficult to tell whether they're dialects or different languages.

Another drawback of using intelligibility as a criterion is that it can be difficult to quantify. Extra-linguistic elements can have a significant impact on intelligence tests, which are often instinctive and impressionistic judgements (Gooskens 2018). Nonetheless, some researchers, such as Tang and Heuven (2009), created experimental designs that use linguistic distance measurements to predict mutual intelligibility, and furthermore tend to draw clear boundaries between dialects.

In general, separating a dialect from a language, as well as a dialect from other dialects, is a challenging task. It is well acknowledged that "language" is not an absolute concept, particularly in the Chinese language. Chambers and Trudgill attempted to define these words precisely. A neutral term, "variety," is introduced to represent a particular kind of language as a single entity. Specifically, "accent" refers to a variety with particular phonetic and phonological characteristics. And "dialect" refers to a variety with particular grammatical, lexical, and phonological differences. The above illustration and summary are intended to draw attention to the universal problem of dialect classification and language-dialect identification. While surveying the literature of Chinese dialects, these background details will be taken into consideration.

Furthermore, from a variationist's standpoint, we can look at this issue from a different angle. Dialect identification and classification consider a language and its variants as a whole, which contradicts the diversity of dialect features to some extent. It is sometimes desirable to focus on each variety or dimension of variation while standing on a generally known common ground of classification rather than giving a uniform answer for all types.

3 Chinese Dialectology

The scope of Chinese dialectology is relatively large due to the diversity of Chinese dialects. Despite the fact that there are several issues with identifying dialects and no universal standard, it remains an essential research area and a basic topic in dialectology. This chapter will cover the history of Chinese dialectology, the most generally used taxonomy, as well as other prevalent study methods and goals.

3.1 Dialect or Topolect

Dialectology seems to be plagued by terminological conundrums at all times. Victor Mair, a Sinologist, distinguished between "dialect" as a Western term and "dialect (fangyan)" as a Chinese term (Mair 1991). The English word "dialect" refers to "one of two or more mutually intelligible varieties of a given language distinguished by vocabulary, idiom, and pronunciation." The two names may be interchangeable in some cases, such as in distinct varieties of Mandarin Chinese, but not in others, such as Mandarin and Cantonese. Mair felt that the translation of "fangyan" as "dialect" is erroneous because of the latter example. To deal with the problem, he coined the term "topolect" to describe Chinese dialects (fangyan). The name "topolect" is derived from the words "topo" (location/fang) and "lect" (language variety/yan). In certain publications, the term "regionalect" is used to describe mutually intelligible varieties, whereas "topolect" is used to describe speech that is difficult to categorize as a dialect or language (Tang 2017).

Despite the fact that the concept of topolect is often mentioned in literature to emphasize the uniqueness of Chinese varieties, the word is rarely used. By convention, researchers will still call Chinese varieties, whether mutually intelligible or not, Chinese dialects. However, it's vital to remember that the nature of Chinese dialect differs significantly from the concept of dialect in the Western world, leading me to believe that the methodology or mindset adopted by Chinese linguists from Western resources may not be appropriate. The field's

progress hinges on determining the differences between Chinese dialectology and dialect studies in other nations, and then modifying research models and methods accordingly.

3.2 A Brief History

Tang’s article (2017) in the Handbook of Dialectology is a wonderful place to start because it provides a brief timeline and emphasizes various noteworthy works, both historical and contemporary. This section summarizes several notable landmarks that she listed out in her work. *Fangyan*, the first composition on Chinese dialects, dates back to the Han Dynasty. Yang Xiong (B.C. 53–A.D. 18), the author, was a poet, philosopher, and linguist. *Fangyan* is the first ancient Chinese dialect dictionary. The root of the term ”fangyan” is the title of the book (dialect). Yang Xiong spent 27 years in collecting and editing *Fangyan*, through the tradition of annually sending imperial emissaries all over the country and surveying the regional dialect vocabulary from folk songs and popular sayings. The diversity of dialects had become noticeable by his time. He accumulated roughly 9000 dialectal characteristics in the end. Modern Chinese dialect research burgeoned in the early 1900s. “Geyao Yanjiuhui (folk song seminar)” from Peking University made a significant contribution to the early stage of Chinese dialectology. They investigated folk songs or ballads that were usually sung in dialects, similar to Yang Xiong’s work. Several professional institutes (usually schools or government affiliations) specializing on one or more groups of dialects were founded and began to survey and collect dialect data from the 1920s to the 1950s. Surprisingly, the Putonghua (common speech, standard Chinese) initiative in 1956 marked a key turning point in Chinese dialectology. The movement’s original goal was to promote the use of Putonghua over regional dialects. However, it inspired a thorough examination of Chinese dialects. To collect dialect data, a variety of fieldwork and instruction manuals were used. Facilitated by the boom, a remarkable work *The Language Atlas of China* was first published in 1987 and updated in 2012. This collaborative work from the Australian Academy of the Humanities and the Chinese Academy of Social Sciences is a collection of dialects maps in all

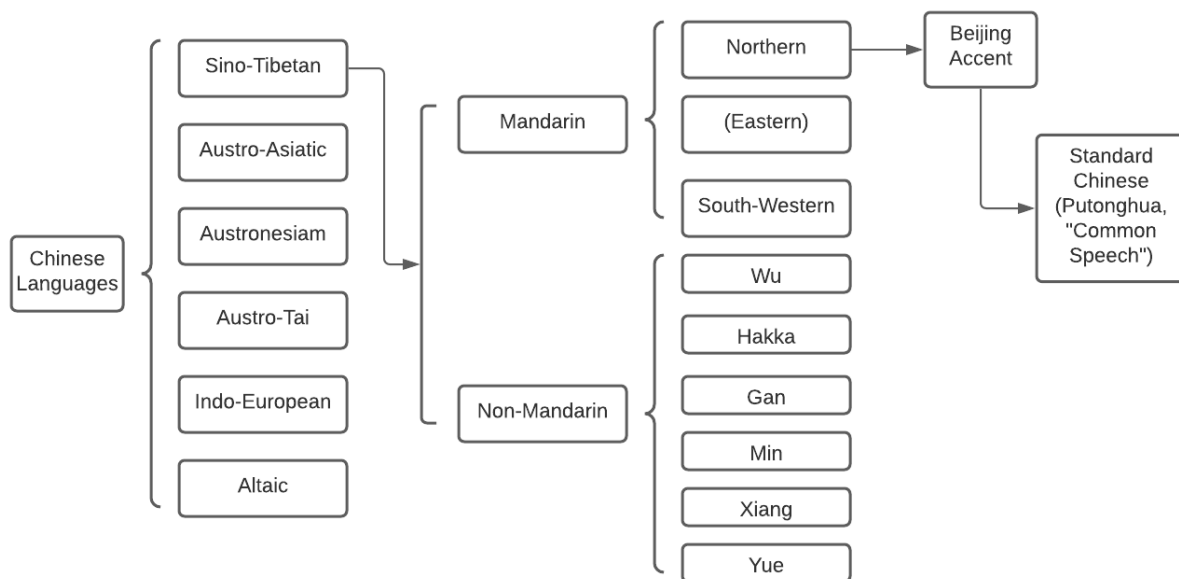
regions in the country. *The Linguistic Atlas of Chinese Dialects* (Cao 2008) was published in three volumes in 2008, and it contains information on the various regional variants of Chinese. The former covers both ethnic minority languages and Chinese dialects, whereas the latter includes just Chinese dialects. The latter are concerned with creating maps of various features of dialects (Iwata 2010).

3.3 Classification and Taxonomy

The classification of Chinese dialects may differ depending on the criteria and dimensions that have been taken into consideration throughout the classification process. The most important factors, in particular, are the phonetic features of the language and the historical relationship to Middle Chinese (Dong 2014). A good description of the changes in classification may be found in Tang (2017). In general, researchers proposed several layers of classification: supergroup; group; subgroup; cluster; and local dialect. The process of group categorization has gone through several stages. Early classifications of Chinese dialects were primarily concerned with geographical distribution and phonological change, rather than with any other characteristics. Later, another method transferred the emphasis from the phonological and phonetic development of phonemes to the development of phonemes by Wang Li (cf. Wang 1936). Following a series of major phonetic features in the 1930s, five groups were established based on those features in the 1937 (cf. Wang 1996); eight major groups were proposed based on Middle Chinese phonological features (Li 1937); and seven major groups were adapted based on Middle Chinese evolutionary features in 1982 (Ting 1982). The evolutionary characteristics of Middle Chinese served as an important trait to be included in the criteria, and helped to draw a clear distinction between Mandarin Chinese and non-Mandarin Chinese. Subsequently, the dialects were subdivided into three primary groups: The Northern group, Central group, and South group (Norman 1988). In the surveyed literature, dialectologists tend to split Chinese dialects into seven groups (Mandarin, Gan, Hakka, Min, Wu, Xiang, and Yue) or ten groupings (Mandarin, Min, Wu, Yue, Jin, Gan,

Hakka, Xiang, Huizhou, Pinghua). Tang’s work was the primary source for many of the references listed above. A dearth of readily available materials limited my ability to trace the origins of a few of the pieces. According to Tang’s study, Yan’s book *Introduction to Chinese Dialectology* (Yan 2006) appears to have a basic introduction to how the classification criteria developed over the course of the previous century. As a result, it may also be a worthwhile resource to investigate.

Generated based on Tang (2017) and Bauer (2006)’s articles, the diagram below depicts the common taxonomy of Chinese dialects. Mandarin, which is frequently referred to as the Northern dialect, is distinguished from the other six groups of dialects that make up the South dialects. Overall, dialects are mutually intelligible to a certain extent within Mandarin branches, although Mandarin and southern dialects are not usually mutually intelligible within Mandarin branches.



Graph 2: Taxonomy of Chinese Varieties

Apart from Mandarin and Non-Mandarin varieties, it can be observed that there are varieties from other language families. In fact, China is the home of six major language families: Altaic, Austro-Asiatic, Austronesian, Austro-Tai, Indo-European, and Sino-Tibetan (Bauer 2012). In a narrow sense, it is common to consider Chinese dialects to be those belonging

to the Sino-Tibetan Family, which are often referred to as the Han Chinese spoken by the Han ethnic groups, as opposed to other dialects. Because China is a multi-ethnic country, approximately 10% of the population, namely the ethnic minorities, speak dialects of the remaining five language families. Despite the fact that it is clearly in opposition to the linguistics notion of dialects, both Han varieties and minority languages are considered to be Chinese dialects for the geopolitical reasons that we discussed earlier in this article.

3.4 Several Valuable References

During my research for this topic, I came across three introductory books that I believe are excellent resources for clarifying the directions that have been taken and the types of research that have been conducted in both Chinese and English literature.

Yan's *Introduction to Chinese Dialectology* is the first book in the series (2006). Yan's research interests are in the fields of anthropological-linguistics and sociolinguistics. Her studies of Chinese dialectology began as early as the 1970s, according to her. The book is divided into four major chapters, which are introduced in the following order: the background of Chinese dialectology, the modern period of dialectology, and the major dialect groups. I was unable to gain access to the publication. Bauer's book review (2012), on the other hand, provided some interesting insights into this book. Many of the summaries and introductions to notable works are consistent with other works of literature that I have read. It appears to be a dependable source of information.

The second book is Kurpaska's *Chinese Language(s), a Prism of The Great Dictionary of Modern Chinese Dialects* (2010). Kurpaska is a dialectologist who also happens to be a Sinologist. Despite the fact that the book is written in English, Kurpaska conducted extensive research on Chinese resources while studying at Peking University, as being mentored by renowned Chinese linguists throughout the time. Kurpaska addresses a number of important issues in Chinese dialectology in this book, including but not limited to the general situation of language varieties in China, historical phonetic features, the discussion of Putonghua as

the common speech, the classification of dialects, and guidelines for conducting field research in China. One of the most important compositions in modern Chinese dialectology, as well as in Chinese linguistics in general, is *The Great Dictionary of Modern Chinese Dialects*, which is another major component of her work analysis. In addition, the dictionary is closely related to the *Language Atlas of China* (Li 1989), which I mentioned previously.

Introduction to Chinese Dialectology also happens to be the title of the third book, which was written by You around 2000 and republished in 2018 (You 2018). The idea for this Chinese composition came from the perspective of a Chinese dialectologist. Overall, You began with an overview of the research methods that have been used and progressed to discussions of dialect geography, historical Chinese dialects, and dialect variations in relation to regional and historical factors as well as socio-economic factors. Then he went into detail about the interaction between dialects. The discussion on written dialect scripts is a particularly interesting aspect. In spite of the fact that it is widely known that Chinese dialects all share the same writing system, certain dialects incorporate the orthography in a distinctive manner. Not all dialectal words and sentences can be represented by standard scripts. This section of his work provides a brief overview of how dialects can be written with existing scripts or other transcripts. There haven't been many surveys conducted on this specific topic, which makes his discussion all the more valuable.

Surprisingly, the field of dialectology has benefited greatly from the contributions of Japanese scholars. Iwata's (2010) work provided an overview of how Japanese researchers conduct projects on Chinese dialects and the current trend in the field. I didn't look into this paper because I didn't have enough time due to time constraints. The fact that significant work is being done outside of Chinese and English literature should not be overlooked. In particular, Shi's article (Shi 1996) summarizes the important work conducted by Japanese academics.

3.5 Research on One or Few Varieties

The fact that many researches tend to focus on one or a few varieties or variations has come to my attention while browsing through the research in Chinese literature. As previously stated, researchers proposed several layers of dialect classification, including supergroup, group, subgroup, cluster, and local dialect. Every level of classification may contain enormous phenomena that are worth investigating. The narrow scopes of work make it impossible to summarize what has been done in general, especially when the study object is specific. As an illustration, I'll use two articles that I came across as examples.

Inoue's research on "Continuum of Fujian Language Boundary Perception: Dialect Division and Dialect Image (2018)" concentrates on the varieties within the Fujian Region, a province famous for its multi-dialect situation. The paper conducted a survey of individuals from the region, who were divided by location and age group, and used the methodologies of perceptual dialectology in the process. Yuan et al. (2019) have narrowed the scope of their investigation to a specific phonetics property in the Wu dialect, which is spoken along the border between Jiangsu and Anhui provinces. In the paper, the authors looked at the acoustic characteristics of the voice using waveforms and analyzed the circumstances surrounding the emergence of a special bilabial trill. In general, there are a plethora of different directions that prospective studies could take. Despite the fact that the field is currently in a complicated and perplexing state, it is critical to establish realistic expectations and to clarify on what common ground the paper will be based. It is also necessary to obtain a more comprehensive picture of the field in order to accurately assess the current state of a specific sub-field or topic.

4 Existing Resources and Problems

There are a wide variety of different formats available for dialect resources. Traditional dialectology oftenly documents the characteristics and features of varieties in a written format

and conducts research on a small number of recordings culled from field studies that have been carefully surveyed. New methods have been developed in conjunction with the advancement of language technologies.

During the course of my research, it has come to the realization that a lack of proper documentation, organization, and archive of dialect resources could be a significant problem in the field.

4.1 Traditional Resources

Early missionary documents produced by European and American missionaries who came to China to spread their religious beliefs may have served as the first source of modern Chinese dialect documentation. During their interactions with the locals, they have created a large number of observation logs. Generally speaking, according to Kloter (2012), there were three major sources for this type of documentation. The first portion was created by missionaries from the Society of Jesus in the late 16th century. Secondly, some Min resources was gathered by European missionaries and early Southern Fujian immigrants residing in the Philippines, dating back to the early seventeenth century. Protestant missionaries from the mid-nineteenth century provided a third source of information that gathered resources in Southern dialects.

Other useful resources include dialect dictionaries and dialect atlases, among other things (Tang 2017). This section will not include any further discussion of the two major works, *Language Atlas of China* (Li 1989) and *The Great Dictionary of Modern Chinese Dialects* (Li et al. 2002), because we have already discussed them in the previous section.

4.2 Corpus and Audio Data

As we previously discussed, because Han dialects all use the same writing system, corpus data is rather rare and insufficient. Ethnic minority dialects (minority languages) on the other hand, frequently have distinctive writing scripts. Zhang (2011) provided a few examples of

this type of corpus, including the following:

- Xinjiang Normal University Uyghur corpus: 2 million words
- The Tibetan corpus of 5 million Tibetan characters from the Institute of Ethnic Studies of the Chinese Academy of Social Sciences
- The Mongolian corpus of Inner Mongolia University: segmented and part-of-speech tagged

Audio data is one of the most important resources, as written data does not appear to capture all of the distinctive characteristics of dialects, according to some. The ability to reuse audio data for the purpose of studying dialectology is another significant advantage. The following is a list of several frequently-mentioned audio files and dialect sound databases, in no particular order:

- Dialect Sound Database for Modern Chinese (Hou 2003) in the format of tape
- RASC863 - G2 - 863 Database: Founded by the national 863 High-Tech Program, this database is a annotated accent speech corpus that contains recordings of six representative regional accents. It was produced by Laboratory of Phonetics and Speech Science, Institute of Linguistics, CASS in 2004
- Chinese Language Resources Preservation Project (<https://zhongguoyuyan.cn/>) It is a large-scale national project on language and culture, which is led by the Ministry of Education and the State Language and Script Working Committee and initiated by the Ministry of Finance of China. The project includes a survey of Chinese dialects, a survey of minority languages, a documentary collection, and the construction of a collection and display platform. The specific work is organized by the China Language Resources Conservation Research Center (located at Beijing Language and Culture University) and has been implemented since 2015. Till today, the project collected 8324 speech samples from 1284 research locations in 34 provinces.

4.3 Problems Detected in the Research

Several issues have arisen as a result of the surveying and accessing of existing data that have been encountered. Only a few dialectal speech databases have been created and made publicly available, which is the first major problem. As a researcher, it is unclear how to gain access to the data from the majority of projects, which were founded by national high-tech programs and put to use by universities and research institutes. To make matters more complicated, the sources of the databases mentioned above are only mentioned in other citations or on the websites of several research institutions, making it difficult to properly attribute them. Neither the databases nor the projects that I was looking for had any information or instructions on them. There is a second problem in that there does not appear to be a comprehensive archive that collects existing resources, making it even more difficult to survey what has been done and comprehend the diversity of these works. Many databases and corpora had main web pages that pointed to an online archive called the Chinese Linguistic Data Consortium. Despite this, the website appears to be inaccessible.

There were also issues with projects that were not led by national projects. To give an example, the Dialect Preservation Project, which was founded by IflyTech, a major speech technology company, developed a web-based application that allows users to freely record themselves speaking in a dialect. This method of crowd-sourcing was successful in amassing 1.5 million contributions from 700,000 participants. While the project encourages users to help preserve Chinese dialects by providing sample speech, it is unclear whether the data collected will be made available for use by the general public in dialectology research.

4.4 Further Impact caused by the Unavailability

In fact, this appears to be a common problem encountered when surveying or conducting research in Chinese academia outside of universities, research institutes, or large corporations, regardless of the source of the data. The scarcity of open-domain resources will impede the development of certain language technologies, such as accented speech recognition. Further-

more, it will be counterproductive to the preservation of endangered languages because there are so few resources available for research.

Regarding dialects, Chinese language technologies encompass a wide range of topics. Due to space and time constraints, this survey was unable to conduct a thorough investigation into this topic. In any case, the problem remains the same, as the documentation and organization of dialectal language technologies are equally chaotic and perplexing, which makes for an excellent prospective research topic to be resolved.

On the other hand, researchers have addressed the issue of endangered dialects and made recommendations for how to better preserve dialects and varieties that are on their way out. In his article "On the Endangered Chinese Dialects," Cao (2011), for example, identified several major causes of endangered dialects, including the following:

- The most visible endangered dialects are those that are spoken by a small number of people and are surrounded by other dialects that are dominant in the area.
- Those dialects that are less prestigious will gradually fade away in an area where multiple dialects coexist, if the distribution of dialects is not equal.
- As a result of the strong influence of Putonghua in mono-dialect areas, some small towns and rural areas have adopted Putonghua as their daily mode of communication, which is a major contributor to dialect extinction.

Cao has also argued that despite the fact that language variations and change are unavoidable, linguists should take a proactive role in addressing them. As a reminder, when we say "intervention," we are not referring to the simplistic notions of "rescue" or "protection," but rather to a rational perspective on the change of dialects in line with the trend. This means that we should approach dialect change from a logical standpoint and guide the process of dialect change in order to minimize the negative consequences of dialect extinction. The most critical solutions are to begin language preservation projects by collecting and publishing Chinese dialect corpora and audio/video database, among other things.

5 Conclusion

Following the clarification of several terminologies and key concepts within general dialectology and sociolinguistics, we move on to the discipline of Chinese dialectology, where we continue to identify several issues that must be addressed in order to facilitate the field's development.

In the first section, we first lay the groundwork for dialectology, and then we address the problem of distinguishing between languages and dialects. What mutual intelligibility is, how it is associated with the identification of languages versus dialects, and what some of the issues are associated with using it as a classification criterion are discussed in this episode.

Now that the common ground has been established, we can move on to learning about Chinese dialectology. In this section, we first introduce the concept of topolect, which emphasizes the distinctiveness of Chinese varieties, and then provide a brief timeline of the development of modern Chinese dialectology. The taxonomy presented in this section serves as the foundation for any type of dialectal research and forms the heart of this section. Also included in the package is a list of introductory books and articles that will enable further understanding of the current state of this discipline.

In the third section, we examine the existing resources associated with Chinese dialects and recognize the problem of a scarcity of readily available information on these dialects. Despite the fact that this paper is merely a review of the literature, we believe that this is a problem that should be investigated further in the future because it is necessary to resolve it in order to first promote the development of language technology and second, to preserve endangered dialects.

For a conclusion, it should be noted that the survey of research into Chinese dialects is far from comprehensive. However, it is critical to have an increasing number of literature reviews in order to assist future researchers in identifying their research goals and to contribute to the continued growth of this fascinating field.

References

- Hickey, Raymond. 2018. Dialectology, Philology, and Historical Linguistics. *The handbook of dialectology*:23-38.
- Chambers, J. K. 1980. *Dialectology*. Cambridge Textbooks in Linguistics. Cambridge [Eng.]; Cambridge University Press.
- Chambers, J. K. 2015. *International Encyclopedia of the Social and Behavioral Sciences*. Second edition. London: Elsevier: 348-352
- Labov, William. 1973 *Sociolinguistic Patterns*. University of Pennsylvania Press.
- Walker, James. 2021. Language Variations. *Similar languages, varieties, and dialects: a computational perspective*. Studies in Natural Language Processing. Cambridge: University Press: 3-16
- Nerbonne et al. 2021. Dialectology for Computational Linguistics *Similar languages, varieties, and dialects: a computational perspective*. Studies in Natural Language Processing. Cambridge: University Press: 3-16
- Krug, Manfred, and Julia Schlüter. 2013. *Research Methods in Language Variation and Change*. West Nyack: Cambridge University Press.
- Gooskens, Charlotte. 2018. Dialectology Intelligibility. *The handbook of dialectology*:204-218.
- Bergmann, Anouschka, et al. 2007. *Language files: materials for an introduction to language and linguistics*. Tenth edition. Columbus: Ohio State University Press.
- Tang, Chaoju, and Vincent J. van Heuven. 2009. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119: 709–732.
- Tang, Chaoju. 2017. Dialects of Chinese. *The handbook of dialectology*:548-557.
- Mair, Victor. H. 1991. What Is a Chinese “Dialect/Topolect”? Reflections on Some Key Sino-English Linguistic Terms. *Sino-Platonic Papers*, 29, 1-31.
- Li, Rong. 2008 Theo Baumann, and Mei W. Lee. 1988. *Language atlas of China: parts I and II*. Hong Kong: Longman Group (Far East) Limited.

- Li, Rong, et al. 2002. *Great Dictionary of Modern Chinese Dialects. 2002(Chinese)*. Institute of Linguistics, Chinese Academy of Social Sciences.
- Iwata, Ray. 2011. Chinese geolinguistics: history, current trend and theoretical issues. Universitat de Barcelona.
- Dong, Hongyuan. 2014. *A history of the Chinese language*. London ; Routledge.
- Wang, Li. 1936. *Zhongguo Yinuunxue*. [Phonology of Chinese]. Beijing: Shangwu Yinshuguan. [Tlle Commercial Press].
- Wang, William S-Y. 1996. Linguistic Diversity and Language Relationships.
- Li, Pang-kuei. 1937. *Languages and Dialects of China*. The Chinese Yearbook Shanghai: Commercial Press (Reprinted in *Journal of Chinese Linguistics* 1: 1-13.
- Ting, Pang-Hsin. 1982. Hanyu Fangyan qufen de diaojian. [Phonological features for classification of the Chinese dialects] *Tsing Hua Journal of Chinese Studies* (14): 257-273.
- Norman, Jerry R. 1988. *Chinese*. Cambridge University Press.
- Hou, Jinyi. 2003. *Xiandai hanyu fangyan yiku*. (CD-Rom version, in Chinese) Shanghai: Shanghai Education Press.
- Yan, Margaret Mian. 2006. *Introduction to Chinese dialectology*. Vol. 22. LINCOM Studies in Asian Linguistics. Muenchen, [Germany]: LINCOM Europa.
- You, Rujie. 2018. *Han Yu Fang Yan Xue Dao Lun*. Shanghai Jiao Yu Chu Ban She.
- Bauer, Robert S. 2012. "REVIEW: INTRODUCTION TO CHINESE DIALECTOLOGY BY MARGARET MIAN YAN." *Journal of Chinese Linguistics*, vol. 40, no. 2, Journal of Chinese Linguistics Publication Office.
- Inoue, Fumio. (2018). Continuum of fujian language boundary perception: Dialect division and dialect image. *Dialectologia*. 147-180.
- Henning Kloter. 2011. Mandarin and Chinese Dialects in Western Sources : The Earliest Documents. *Fudan journal of the humanities and social sciences* 4: 59–83.

袁丹 et al. 2019. BILABIAL TRILL INDUCED BY FRICATIVE HIGH ROUNDED VOWEL: THE EMERGING OF "T_B" IN THE WU DIALECT OF SU-WAN BOUNDARY. *Journal of Chinese linguistics* 47. Journal of Chinese Linguistics Publication Office.

张华平.2005. "中文信息处理技术发展简史 A." *Volume*, DOI: <http://www.nlp.org.cn>.

曹志耘.2001.关于濒危汉语方言问题. *语言教学与研究*(01),8-12.
doi:CNKI:SUN:YYJX.0.2001-01-001.

中国语言资源保护工程采录展示平台. 2021.
<https://zhongguoyuyan.cn/index.html?lang=cn>. Accessed December 19.

方言保护计划. 2021. <https://fangyan.xunfei.cn/#/>. Accessed December 19.