International Conference on Education, Language, Literacies, and Literary Studies (ICONELS)
Volume 02 Nomor 02
http://dx.doi.org/10.63011/iconels.v2i1.97

Documentation of Low-Resource Languages in Southern China in the Digital Era：Interdisciplinary Fieldwork, Practice, and Values

Zhiwei Xu[1]; Yancheng He[2]

[1] Kyung Hee University, Korea, xzw0704@gmail.com
[2] Guangxi Normal University, China, yanchenghe@126.com

Corresponding author:
E-mail: yanchenghe@126.com

Abstract

In the digital era, documenting low-resource languages in Southern China, especially minority and endangered languages in Southwest China, is of great significance. These languages, embodying distinctive cultural heritages and cognitive systems, are under threat from language shift and modernization. Rooted in documentary linguistics—a newly born and still growing field dedicated to comprehensively documenting languages through digital media technologies for dynamic recording, long-term preservation, and easy accessible dissemination—for about two decades, we have conducted a couple of projects in documenting several languages or dialects of the Tai-Kadai family, utilizing interdisciplinary fieldwork to preserve linguistic data and cultural heritages embedded in these languages, thus obtaining some achievements in language preservation and revitalization, enriching global linguistic diversity and cultural heritage.

This comprehensive interdisciplinary approach, aligned with documentary linguistics' methodologies of participatory recording and cross-disciplinary data integration, not only comprehensively documents language data but also integrates diverse cultural knowledge, bridging the gap between linguistics and other disciplines. It provides a new model for low-resource language documentation, highlighting the value of preserving linguistic and cultural diversity, and offers essential data for language preservation and research, cultural inheritance, and community development.

Keywords: *Low-resource languages, Language documentation, Documentary linguistics, Interdisciplinary approach, Tai-Kadai languages*

Introduction

Low-resource languages "are those spoken by smaller populations and are often characterized by a lack of comprehensive linguistic resources such as written documentation, digital tools, or academic research. Compared to widely spoken ones, these languages typically have fewer speakers and are often overshadowed by technological and educational advancements."(Violette Spector, 2025)

According to UNESCO's document *Language Vitality and Endangerment* (2003), "A language is in danger when its speakers cease to use it, use it in an increasingly reduced

number of communicative domains, and cease to pass it on from one generation to the next. That is, there are no new speakers, adults or children. (p.2)."

"In the digital and AI-driven world, low-resource languages face the risk of extinction. The lack of written records or digital content for these languages makes their preservation a pressing challenge, threatening communication and the cultural identity embedded in these linguistic treasures." (Violette Spector, 2025)

With technological advancements, language preservation methods have evolved from initial written documentation to audio and video recording, leading to a continuous improvement in data accuracy. High-quality audio and visual materials have significantly enhanced data reliability, thereby facilitating the emergence of a new interdisciplinary field—documentary linguistics. The past two decades have seen rapid progresses and great achievements of language documentation in China, especially of low-resource or endangered languages in the southwestern regions. With the focus on the current status of low-resource languages in southern China, particularly minority and endangered languages in the southwest region, this paper introduces recent accomplishments achieved through the comprehensive interdisciplinary methodology of documentary linguistics, with particular reference to the documentation of several languages or dialects from the Tai-Kadai family conducted by the authors.

Language Endangerment in the Digital Era

The rapid advancement of technology and the accelerated pace of globalization has given rise to profound transformations in people's lifestyles, communication patterns, and cultural outlooks. These changes not only pose threats to low-resource languages but also bring about new opportunities. During the modernization process, urbanization and industrialization have substantially enhanced population mobility and inter-ethnic integration. However, this process has also led to a noticeable decline in cultural diversity. Many young individuals migrate from their hometowns to cities in search of better employment and education opportunities. To integrate more effectively into new social environments, they are often required to adopt a dominant or standardized language. As a result, speakers of endangered languages—who are usually geographically dispersed—face increasing difficulties in passing their native languages down to future generations. Consequently, the number of speakers of these languages has sharply declined. Most remaining speakers are middle-aged or elderly individuals, while younger generations show limited interest or engagement. This generational gap places many of these languages at serious risk of disappearance. Moreover, although technological progress has broadened the scope and accessibility of language communication, digital and mass media platforms predominantly favor mainstream languages and dominant cultures. For minority and indigenous languages—particularly those with small speaker populations or no standardized writing system—this trend represents a shrinking linguistic ecological space, leading to their progressive marginalization.

The disappearance of endangered languages not only signifies the loss of a linguistic system, but also means the erosion of an entire community and its rich cultural heritage. As a fundamental carrier and integral component of human culture, a language serves as a repository for expressing the social identity, cultural traditions, historical narratives, and unique knowledge of the ethnic group to which its speakers belong. Enrique Uribe-Jongbloed (2007) points out that, "Language serves as a medium for culture's traditions and lore, in

the same way that artifacts or buildings resemble its beliefs and architectural prowess. On the other hand, the loss of such a precious part of a culture might have detrimental effects to the concerned group as a whole (p. 69).” The disappearance of endangered languages will inevitably result in profound and irreplaceable losses to the continuity and sustainable development of human civilization.

Therefore, the issue of “endangered languages” (Constance, 1991, p. 159), first introduced by *Science* magazine, has since garnered increasing attention and recognition from UNESCO as well as cultural and academic communities. Fan Junjun (2022) argues that addressing the issue of language endangerment primarily involves two strategies: the first is language protection, which seeks to sustain and revitalize endangered languages through active use; the second is language preservation, which entails systematically documenting these languages to ensure reliable data are available for future research, revitalization efforts, and cultural inheritance (p. 3).

China has implemented a series of targeted measures to address the issue of endangered languages. For example, in ethnic minority regions, numerous primary and secondary schools have progressively introduced bilingual education programs, enabling students to acquire proficiency in the national common language while systematically learning their native ethnic language and writing system. The China Language Resources Protection Project, launched by the State Language Commission in 2015, employs modern technologies to systematically collect and document linguistic data related to Chinese dialects, minority languages, and oral linguistic traditions. The first phase of the project completed fieldwork and protection efforts at over 1,700 survey sites across all provinces, covering more than 120 distinct dialects. This initiative has provided substantial support for cultural continuity and academic research in China, standing as a significant national effort in the field of language preservation.

Method

The Current Situation of Endangered Languages

According to the document of UNESCO, an endangered language refers to a language with a significant reduction in the number of speakers, a lack of new users, and the potential to completely disappear in the near future. Based on inter-generational transmission, UNESCO further classifies endangered languages into four levels: Vulnerable, Definitely Endangered, Severely Endangered, and Critically Endangered.* These levels not only reflect the differences in the degree of language endangerment, but also reveal the severity of the language survival situation. Currently, according to the website, Ethnologue—Languages of the World, there are 7,159 languages in use worldwide, but approximately 45% of them are in an endangered state, with typically no more than 1,000 speakers.

The emergence of endangered languages is mainly attributed to the following reasons:

（1）Cultural Assimilation: With the advancement of globalization and urbanization, mainstream cultures and languages increasingly take the dominant position, leading to some minority language speakers gradually abandon their mother tongues in favor of mainstream languages that offer greater social advantages. Many indigenous communities, upon exposure to modern civilization, feel compelled to relinquish their native languages to better

integrate into mainstream society. This process of cultural assimilation not only alters language use patterns but also profoundly impacts their cultural identity, lifestyle, and value systems.

(2) Economic Pressure: Communities that speak endangered languages often face significant economic challenges. Mastering mainstream or international language is widely perceived as essential for securing better employment opportunities and achieving higher social status. Consequently, younger generations frequently migrate to urban areas in pursuit of improved living conditions, where they adopt mainstream languages. As a result, opportunities to use their mother tongues diminish over time, leading to gradual language attrition and a weakened sense of identity and belonging among younger speakers.

(3) Educational policies: Many educational systems do not support or promote the use of native languages, causing the younger generation to only learn and use mainstream languages at school. National education policies often prioritize official languages while neglecting the preservation and transmission of native languages. This policy orientation not only deprives students of the opportunity to learn and engage with their native languages but also accelerates the decline and eventual disappearance of endangered languages.

(4) Influence of Mainstream Media: Online streaming media such as Weibo and short-video platforms have strengthened the dominance of mainstream languages, expanding their dissemination advantages while shrinking the linguistic space available to minority languages. As a result, minority languages with already limited speaker populations become increasingly marginalized. Furthermore, media content often prioritizes trending topics and attention, leading to superficial and one-sided portrayals of endangered languages that ignore their linguistic diversity and cultural complexity. For instance, media narratives may reinforce certain stereotypes by associating endangered languages with labels such as “backwardness” or “primitiveness,” thereby intensifying social stigma and discrimination against their speakers.

(5) Social Attitudes: Negative perceptions of native languages or the belief that they lack practical utility among social groups lead to a decline in the number of native language speakers. Prejudice and social discrimination foster feelings of inferiority among minority language speakers, discouraging them from using their mother tongues in public and further reducing the frequency of native language use. According to Marianne Mithun (2007), “the traditional language is barely thought about at all, regarded simply as a utilitarian tool for conveying information. In some it is even viewed with contempt or shame, considered a sign of backwardness” (p. 42).

(6) Population Migration: Population mobility alters original language environments, dispersing previously cohesive language communities and disrupting the inter-generational transmission of languages. With the acceleration of urbanization, large numbers of rural residents move to cities, significantly reducing the occasions in which their mother tongues are used. As the next generation grows up in urban environments, the continuity of native language transmission interrupts. Population migration not only influences the sociolinguistic landscape but also intensifies the challenges faced by endangered languages in maintaining their vitality.

Every language constitutes a vital component of human cultural diversity, and its extinction entails irreversible losses to the collective heritage of humanity. Many endangered languages encapsulate unique systems of ecological knowledge, traditional practices, and oral narratives. Once these invaluable cultural assets vanish, they cannot be recovered. Furthermore, linguistic diversity holds profound significance for various academic disciplines, including linguistic research, cognitive science, and sociolinguistics. Each language encodes distinct structures for conveying thought, emotion, and interpersonal relationships. The disappearance of language is a loss of human cognition and communication patterns.

Strategies and Measures for Addressing the Issue of Endangered Languages

To protect endangered languages, organizations and researchers have undertaken proactive initiatives, including documenting and analyzing endangered languages, supporting community-based language revitalization programs, and incorporating endangered language instruction into formal education systems. Specifically, the following strategies are suggested to deal with the issue of endangered languages:

(1) Legal Guarantees and Policy Support. The state and local governments are encouraged to attach greater importance to the protection of endangered languages, formulate and implement relevant laws and regulations, and provide a solid legal foundation for the protection of endangered languages. Through legislation, the responsibilities and tasks of relevant departments such as culture, education, and ethnic affairs in the protection of endangered languages can be clearly defined to ensure the effective implementation of various measures. At the same time, dedicated funding can be allocated to support projects focused on the documentation and revitalization of endangered languages. This can help encourage universities, research institutions, and members of the broader community to get involved in related research and practical efforts. Furthermore, governments may consider offering tax incentives, financial subsidies, and other supportive policies to inspire businesses and social organizations to contribute to the protection and transmission of these valuable linguistic heritages.

(2) Enhance the self-rescue awareness of endangered languages. The communities that speak endangered languages are not only the inheritors of the languages but also the core force in language protection. Therefore, it is of great significance to enhance their awareness and initiative in language protection. Systematic education and training help them fully recognize the cultural value and inheritance responsibility of their own language, enhance their crisis awareness in language protection, and master basic language recording and preservation skills, so as to actively engage in the rescue and revival of endangered languages. At the same time, the government can play a supportive role in encouraging the use and sharing of endangered languages within schools and local communities. Incorporating endangered languages into the school curriculum and encouraging community members to learn and use their native languages in daily life are important paths to achieve sustainable language transmission. In addition, traditional and new media platforms can be fully utilized to carry out various forms of publicity and promotion activities to enhance the awareness and interest of people of different ages and cultural backgrounds in endangered languages, and attract more social forces to participate in the protection and dissemination of endangered languages.

(3) International Cooperation. In September 2018, UNESCO and the Chinese government jointly held the first World Conference on Language Resources Protection in Changsha, Hunan Province. The conference adopted the *Draft Yuelu Declaration*, which was revised by UNESCO and officially released in 2019. The declaration advocates promoting exchanges and mutual learning among civilizations, reducing cultural conflicts, and providing a “diversity in harmony” solution for global governance through language resource protection, and promoting the building of a community with a shared future for mankind. The *Yuelu Declaration* marks a new stage in global language protection, moving from scattered actions to systematic collaboration. Specifically, all countries are supposed to share the common goal of protecting endangered languages, which is to maintain global linguistic diversity. On this basis, they are encouraged to formulate targeted protection strategies based on their own language ecology and social realities. Different countries may have different focuses in their protection approaches. For instance, some countries may prioritize enhancing the vitality of endangered languages through the education system and cultural promotion, while

others may emphasize legal guarantees and systematic planning of language policies. On this basis, countries may leverage international conferences, academic symposiums and other platforms to strengthen information exchange, resource sharing and experience sharing in the field of endangered language protection, and form a collaborative international cooperation mechanism. For example, an international network for the protection of endangered languages could be established to address transnational language crises and common challenges. In addition, they could actively draw on international successful experiences, such as Brazil's "Indigenous Language Literacy Program" and the European Endangered Languages Documentation Project (ELDP), to promote the digital preservation and systematic documentation of endangered languages, and enhance the scientific and professional levels of global language protection.

Language documentation and digital preservation. According to David Nathan (2008), before the emergence of documentary linguistics, audio played a relatively minor role in the epistemology of linguistics. As a result, linguistic data were predominantly derived from written materials, leading to the tragic loss of a substantial amount of linguistic information. Documentary linguistics, however, introduced a new paradigm by emphasizing the systematic collection and preservation of primary linguistic data. With the continuous advancement of technology, digital technology is playing an increasingly important role in the protection of endangered languages. Researchers can use high-fidelity audio and video equipment to systematically record the phonetic materials of endangered languages and their actual usage scenarios, ensuring clear sound quality, stable video, and controllable environmental noise, thereby guaranteeing the authenticity and integrity of the collected data. At the same time, the development of cloud computing and cloud storage technology provides efficient, secure, and sustainable storage solutions for the digital preservation of endangered languages, laying a solid technical foundation for the "infinite preservation" (Junjun, 2022, p. 46) of endangered language materials.

Fieldwork and practice of documentation of some Tai-Kadai languages

We have hosted or conducted four projects of documentation of some Tai-Kadai languages: (1) Project of "Documentation of Two Gelao Varieties: Zou Lei and A Hou, Southwest China," conducted from 2006 to 2010, funded by Endanger Languages Documentation Projects (ELDP). (2) Project of "Documentation and protection study of the oral texts of Rongshui Dong (Kam) in Guangxi," conducted from 2012 to 2017, funded by The National Social Science Foundation Project of China. (3) Project of "Documentation of languages and cultures---the Dong (Kam) language of Sanjiang," conducted from 2017 to 2019, funded by Special Foundation of Language Resources Protection Project of China. (4) Project of "Documentation and study of endangered Hongfeng Gelao," conducted from 2019 to 2024, funded by The National Social Science Foundation Project of China.

These projects reveal that two Gelao varieties spoken in Guizhou Province are severely endangered, with only a small number of elderly speakers remaining in each case. Additionally, two Dong (Kam) dialects from Guangxi—specifically from Rongshui County and Sanjiang County—are still widely used and well preserved. All four of these languages belong to the Tai-Kadai language family. They can be classified as low-resource languages due to the limited availability of linguistic data, including text corpora, annotated datasets, and other linguistic resources. Next we will focus on the documentation of Sanjiang Dong.

The practice of "Documentation of languages and cultures—the Dong (Kam) language of Sanjiang" was funded by the Special Project of the China Language Resources Protection

Project and was initiated in 2017, concluding in 2019. The project was structured into four distinct phases. In the first phase, Professor He Yancheng developed a comprehensive investigation plan based on the "Manual for Collection and Investigation" provided by the China Language Resources Protection Project, and prepared technical equipment such as cameras, video recorders, and laptops.

In the second phase of the project, to collect comprehensive and representative primary linguistic data, we conducted three fieldwork campaigns. The first fieldwork took place in July 2017 and had three primary objectives: first, to visit a range of villages, become familiar with the local geography and cultural preservation status, and identify suitable fieldwork sites; second, to build rapport with local communities, deepen understanding of Dong culture, and select appropriate language informants; third, to assess the surrounding environments of the villages in order to address logistical concerns such as food and accommodation. During this investigation, the team visited more than twenty villages and towns, including Dudong, Linxi, Doujiang, Bajiang, Tongle, Buyang, and Meilin. Ultimately, Gaoding Village in Dudong Town was selected as the main investigation site due to its well-preserved traditional Dong village characteristics, and Mr. Mo Renzheng was identified as the key language informant. Mr. Mo, aged 60 at the time, is a native speaker of Dong and not only a local resident of a Dudong Town, but also a former trainee at the county health school. He possesses extensive knowledge of traditional Dong herbal medicine and is proficient in playing traditional Dong musical instruments, including the Pipa, Dong flute, and Lusheng.

Overall, the primary objective of the first field investigation was to establish a solid foundation for the subsequent phase and ensure the smooth implementation of the language documentation practice. The second fieldwork was launched at the end of July 2018. This investigation was more targeted, with the main tasks being to record cultural items, write down item contents, and visit various villages and towns to use audio-visual equipment to document cultural elements with Dong ethnic characteristics, such as wind and rain bridges, drum towers, farm tools, and costumes. The third field investigation was carried out in July 2019. As the project had entered the stage of checking for omissions and filling in the gaps, the main purpose of this investigation was to supplement and re-shoot the omitted or substandard recorded contents discovered during the data sorting process.

Results

The third phase of the project commenced in 2019, during which the team transcribed and categorized the collected vocabulary entries and discourse materials. At the same time, Mr. Mo Renzheng was invited to participate in language recording sessions, contributing valuable audio data to the project. The final stage of the project involved the publication of the book *Archives of Language and Cultures in China—Sanjiang Dong(Kam) Language*. This work encompasses nine major categories of Dong cultural entries, including architecture, daily utensils, clothing, food, agriculture and handicrafts, daily activities, marriage, childbirth and funeral customs, festivals, as well as oral and performance traditions. Among these, the musical instruments, festivals, architectural styles, and cuisines of the Sanjiang Dong community are especially distinctive.

The primary musical instruments of the Sanjiang Dong people include the Dong Pipa, Dong Flute, and Lusheng, which serve as the core carriers of Dong musical culture. The Dong Pipa is available in three sizes—large, medium, and small—with the body typically crafted from a

single piece of wood such as paulownia, tung, or fir. The large and medium-sized Pipas produce a deep and mellow tone, commonly used to accompany long narrative songs. In contrast, the small Pipa emits a clear and crisp sound, making it suitable for lyrical melodies and frequently played by young men and women during evening performances of short lyrical songs.

The Dong Flute is a traditional bamboo instrument unique to the Dong people. It features a reed at the mouthpiece and produces a clear and melodious tone. Primarily used as an accompaniment for vocal music, it is often played in conjunction with singing to create "Flute Songs." Typically, men perform on the Dong Flute while women sing. Due to its portability and bright tonal quality, the Dong Flute has become a major instrument in Dong youth social interactions. Young Dong men often play the flute at night to meet their romantic partners, using melodies to convey their emotions.

The Dong Lusheng, a traditional reed-pipe wind instrument, is a distinctive traditional aerophone of the Dong people. A Lusheng ensemble typically comprises five types of Lusheng and a resonating tube known as the Mangtong. These five types include the large Lusheng, second large Lusheng, sixth Lusheng, fifth Lusheng, and small lusheng. The large Lusheng produces a solemn and dignified tone, while the small Lusheng offers a bright and clear timbre, often used to guide and lead the ensemble. As a symbol of joy and celebration, the Lusheng is prominently played in festivals and ritual ceremonies.

Sanjiang Dong Autonomous County is widely recognized as the "County of a Hundred Festivals," renowned for its diverse and culturally rich festival traditions that vividly reflect Dong ethnic customs. Among these, two of the most distinctive celebrations are "Yueye" and the "Dong Ethnic March 3rd Festival."

"Yueye," which means "visiting as a guest" in the Dong language, is a traditional inter-village communal gathering. It is typically held between January and March, during the agricultural off-season. The event features a variety of cultural performances, including blowing the Lusheng, singing Dong folk songs, and staging Dong operas. These activities serve to strengthen social bonds between villages and provide a significant occasion for young people to interact and develop romantic relationships.

The "Dong Ethnic March 3rd Festival" is celebrated on the third day of the third lunar month. One of its central activities is the "firecracker-snatching" competition. The firecracker, an iron ring wrapped in red cloth or silk, is launched into the air using a gunpowder-propelled device. Participants then compete to catch the ring as it descends. The individual who successfully captures the firecracker is believed to be blessed with promising future and good fortune.

The architectural structures of the Dong people in Sanjiang are entirely constructed from wood, with their most distinctive feature being the exclusive use of mortise-and-tenon joinery. Without employing a single nail or rivet, the precise interlocking of tenons and mortises ensures structural integrity and durability. Among these wooden structures, the drum tower and the wind and rain bridge stand out as exemplary representatives.

The drum tower is typically the tallest structure in a Dong village, serving as a central communal space for ceremonies, meetings, gatherings, welcoming guests, festivals, performances, and recreational activities. Its upper section features a multi-eaved pagoda-style roof available in four-, six-, or eight-cornered designs, with the number of eaves always being odd. The main framework generally consists of four principal load-bearing columns

and twelve eave columns, symbolizing the four seasons and twelve months of the year. Some drum towers adopt a unique single-column design, where a single fir tree with a diameter exceeding 80 centimeters serves as the central pillar throughout the entire structure, symbolizing the unity and cohesion of the village, much like a giant tree rooted firmly in the earth.

The wind and rain bridge represents another hallmark of featured architecture in Sanjiang. It is composed of stone piers, a wooden superstructure, a covered corridor, and decorative pavilions. The piers are constructed from stone, while the bridge body is entirely made of wood. The corridor forms a long walkway, providing shelter from wind and rain. The pavilion roofs and upturned eaves are adorned with carvings of symbolic motifs, including flying birds, gourds, which represent freedom, prosperity, and agricultural abundance.

Within the dietary culture of the Dong people in Sanjiang, pickled food holds the most distinctive position. A local saying goes, "The Dong people cannot live without pickled food," reflecting its central role in daily life. The tradition of pickled food preparation dates back to the hunting era. Due to the hot and humid climate of the mountainous regions, pickling was developed as a method to extend the shelf life of food and manage seasonal surpluses.

Pickled vegetable products include pickled cabbage, pickled ginger, pickled radish, pickled bamboo shoots, and pickled peppers. These are typically prepared by salting and dehydration, then wrapped in glutinous rice or chili powder before being stored in sealed earthenware jars. The main types of pickled meat products are cured pork, cured duck, cured goose, and pickled fish. After removing the internal organs, the meat is usually smoked for three to five days, then wrapped in a mixture of salt and glutinous rice before being sealed in wooden barrels. This preservation method allows the food to remain edible for several years. Pickled meat aged over ten years is considered a rare delicacy and is often reserved for serving to special guests. Pickled food is an essential component of major social events such as weddings and funerals. Pickled fish symbolizes "abundance every year," while pickled duck conveys the wish for "family prosperity."

In the language documentation practice of Sanjiang, we systematically applied the methodologies of documentary linguistics to comprehensively record naturalistic linguistic data. As for the outputs, we constructed an annotated corpus with time-aligned multimedia and transcriptions, as well as a lexical database enriched with audiovisual exemplars. This effectively made up for the limitations of traditional descriptive linguistics in terms of the singularity of its recording methods and the limited scope of its service objects. This practice not only verified the guiding value of methodologies of documentary linguistics but also provided solid data support for the inheritance of language and culture as well as related disciplinary research.

Discussion

Documentary Linguistics and Language Documentation

The Concept of Documentary Linguistics

The term "Documentary Linguistics" was first proposed by Nikolaus Himmelmann in 1998 (P. 161). Linguists recognized that thousands of language—especially indigenous and minority languages—were disappearing without systematic records. As a result, documentary linguistics emerged in the late 20th century as a response to the rapid loss of linguistic

diversity. Documentary linguistics is a subfield of linguistics and an emerging interdisciplinary discipline that differs from traditional linguistics. Given its integration of multiple fields—including descriptive linguistics, cognitive linguistics, anthropology, and folklore. Essentially, unlike descriptive linguistics, which focused on grammar rules, documentary linguistics prioritizes rich contextualization. It documents language within its socio-cultural ecosystem, including gestures, discourse, patterns, and cultural practices. Thus, its application targets are not limited to linguists but also to anthropologists, folklorists, literary figures, musicians (Huang, 2022, p. 3).

With the support of modern science and technology, documentary linguistics aims to create comprehensive and everlasting records of languages through systematic approaches, which include: (1) Multimedia Arching. Systematically recording audio, video, and annotated texts to capture contextualized speech, such as rituals, narratives, and daily interactions. (2) Collaborative Methodologies. Partnering with native speakers as co-researchers to ensure the authenticity of linguistic data and promote ethical, community-centered research practices. (3) Accessibility. Developing open-access databases, such as the Endangered Languages Archive, to provide researchers, educators, and language communities with reliable access to linguistic resources. Specifically, documentary linguistics intend to achieve three core objectives: (1) Preservation. To systematically document and preserve endangered languages and dialects for future generations, researchers, and public. (2) To investigate language within its social and cultural contexts, including its use in various cultural practices and social interactions. (3) Revitalization. To support language revitalization initiatives by providing well-documented linguistic materials that communities can use to reclaim and promote their linguistic heritage.

Language Documentation in the vain of Documentary Linguistics

Documentary linguistics prioritizes the systematic creation of multifunctional, long-lasting records of linguistic practices, emphasizing three foundational principles: primary data authenticity, ethnographic embedding, representative. Firstly, as Nancy Dorian highlights the “record’s accuracy” (p. 179), it requires researchers to record naturalistic data in unedited audio and video formats during the process of language documentation practice. Secondly, on the basis of authenticity, the data should ensure comprehensiveness and diversity.Therefore, in addition to documenting basic phonetics, vocabulary and grammar, researchers also need to record language within socio-cultural contexts, including material and spiritual culture.

In the domain of material culture, researchers should document key aspects such as the styles, patterns, and craftsmanship of traditional clothing; the design principles and functional applications of tools used in production and daily life; the architectural styles, structural features, and historical backgrounds of buildings; as well as the evolution of transportation systems and their influence on community dynamics.

Regarding spiritual culture, scholars are expected to carefully record cultural expressions closely related to language, including the specific details of production and living practices, the forms and social functions of song and dance entertainment, the ritual processes and cultural connotations of religious customs, the origins and evolution of customs and their diverse manifestations in modern society. In addition, researchers should particularly focus

on the systematic collection of long oral texts, such as folk tales, poems, proverbs and other oral traditional resources. Finally, researchers must also observe and document how language is used across various communicative contexts, including informal family conversations, marketplace transactions, and formal recitations or prayers during religious ceremonies.

Building upon the aforementioned three principles, language documentation should be systematically conducted through a four-stage process:

Phase 1: Pre-fieldwork Planning. This stage serves as the foundational step for the entire research endeavor. Researchers must make thorough preparations in two key aspects: community consultation and equipment selection. Specifically, they should first clarify the research objectives and thematic focus, which will guide the selection of appropriate field sites. To ensure representativeness and uniqueness, researchers are advised to gather comprehensive information regarding the geography, history, culture, and local customs of potential locations. Additionally, practical factors such as accessibility, safety, and the willingness of local communities to participate should be carefully evaluated. In terms of equipment selection, researchers should prepare essential tools for capturing high-quality primary data, including digital voice recorders, lavalier microphones, camcorders, and laptops. It is also crucial to test all equipment before fieldwork to ensure proper functionality and to prevent technical failures during data collection.

Phase 2: Data Collection. Researchers need to formally enter the investigation site, establish trust and friendly relations with local participants.They should gain a comprehensive understanding of the community's demographic composition, historical background, geographical context, and cultural customs. This process would facilitate smoother integration into the local environment and supports more effective research implementation. In this phase, researchers are expected to collect unelicited discourse through naturalistic methods such as direct observation, in-depth interviews, and active participation in traditional cultural events. This includes capturing authentic linguistic data such as spontaneous conversations, oral histories, and ceremonial speeches.

Phase 3: Annotation & Analysis. Researchers are advised to edit the collected audio and video files with ELAN, FLEx, and Praat. They need to annotate and transcript each word of the recorded raw data with the Unicode International Phonetic Alphabet and analyze vocabulary and long texts with the Toolbox software.

Phase 4: Archiving & Access. Researchers are required to deposit data into sustainable archives to ensure the long-term preservation and accessible retrieval of the corpus. To achieve comprehensive documentation, an XML-based tagging system should be developed according to the corpus's attributes and functional purposes, ensuring all linguistic features are systematically encoded. The corpus should then be disseminated in widely used formats such as Word, MP4, and PDF. In addition, a tiered access framework must be implemented, clearly defining access and usage rights—such as distinguishing between publicly available and community-restricted materials—to ensure transparency in public access while safeguarding intellectual property rights.

Progress and Achievements of Documentary Linguistics Research in China

Progress

Documentary linguistics research in China has developed over many years, with its core focusing on the protection, documentation and cultural inheritance of language resources. It has presented a diverse pattern of "regional pioneering exploration—national project promotion— in-depth research by academic institutions—continuous support from funds."

Taiwan started earlier than the mainland in the field of language digital archives, the Academia Sinica's development of the Digital Archive of Formosan Languages, the Digital Archieve of Min and Hakka Languages, and the establishment of the Multimedia Database of Formosan Languages have laid the technical and methodological foundation for the construction of language digital archives on the mainland.

The Chinese Academy of Social Sciences has advanced the development of documentary linguistics through the support of key disciplinary programs and extensive fieldwork. From 2014 to 2019, the Institute of Ethnology and Anthropology designated "Documentary Linguistics and Descriptive Linguistics" as a key discipline, with the Southern Language Research Office responsible for the sub-discipline of "Documentary Linguistics." During this period, the office carried out comprehensive documentation of the Miao language in Dananshan, Bijie, Guizhou, and the Qiang language in Zawo, Heishui, Sichuan. These efforts effectively facilitated the transition of domestic research in documentary linguistics from theoretical exploration to practical application.

Since 2010, the National Social Science Foundation has incorporated support for documentary linguistics into a regularized system, promoting the development of language documentation, protection, and theoretical exploration through diverse project types such as general projects, key projects, and major tender projects. General projects include "Research on the Construction of a Vocal Database of Minority Languages in Yunnan" and "Research on the Construction of a Vocal Database of the Xibe Oral Dictionary," focusing on the construction of specific language resource databases. Key projects include "Research on the Construction of a Language Materials Platform for Uyghur, Kazakh, and Kyrgyz Languages" and "Research on the Theoretical System of Vocal Archives for Endangered Languages," concentrating on the construction of corpus for the same language family and theoretical system construction. Major tender projects include "Research on Digital Methods for the Protection and Inheritance of China's Vocal Languages and Oral Culture" and "Research on the Construction of a Digital Museum for Endangered Languages in China," driving technological innovation in language documentation and cultural revitalization.

The "China Language Resources Protection Project," a language and culture project initiated by the Ministry of Education and the State Language Commission and launched in 2015, represents a core national-level initiative in language preservation. It is the first large-scale series in China to systematically record dialects and minority languages and their cultural vocabularies. It encompasses 1,712 survey sites nationwide, including 1,134 Chinese dialect sites and 324 minority language sites. The survey covers three key dimensions: basic linguistic information, phonetic structure, and cultural context. In addition, the "China Language and Culture Collection" project was advanced to complement the protection effort, establishing standardized protocols for each phase of language documentation. By 2018, a total of 100 survey sites had been established, where multimedia technologies were systematically applied to preserve 9 categories of cultural elements, such as architecture,

clothing, and dietary customs. This initiative constitutes a significant practice of documentary linguistic and cultural heritage.

Values of Documentation of Low-Resources Languages

Documentary Linguistics, as a discipline focusing on endangered and low-resource languages, operates under a sense of urgency by systematically documenting phonetics, vocabulary, grammar, pragmatics, and cultural contexts. In doing so, it preserves irreplaceable "living archives" that reflect the diversity of human civilization. Its significance extends beyond the academic boundaries of linguistics, deeply embedded in a multidimensional value system encompassing cultural preservation and scholarly innovation. From a cultural perspective, documentation linguistics serves as a guardian of the diversity of human civilization. Language functions not only as a medium of communication but also as a "living fossil" of cultural heritage. Each language encapsulates a unique worldview held by its respective ethnic group, often preserving the original wisdom of humanity in understanding and adapting to the natural and social environment—ranging from ecological knowledge and interpersonal ethics to alternative cognitive frameworks that may transcend modern scientific paradigms. The extinction of a language entails the irreversible loss of associated cultural knowledge, thereby diminishing humanity's understanding of itself. Through methods such as audio recording, transcription, and annotation, documentation linguistics transforms languages into durable digital and textual records, effectively creating a "rescue backup" for the human cultural civilization and ensuring that every culture remains preserved within the collective memory of humankind.

From an academic dimension, documentary linguistics serves as a vital source of innovation for linguistic theory. The advancement of linguistics depends on comparative analysis across diverse languages. While mainstream languages have been extensively studied, their long-standing integration with dominant cultural norms has led to a high degree of standardization, limiting their capacity to reflect the full range of human linguistic potential. In contrast, low-resource languages, with endangered languages in particular—often shaped by prolonged isolation or minimal external influence—frequently retain unconventional linguistic features. These may include atypical syntactic structures, intricate morphological systems, or alternative modes of expression that challenge conventional linguistic models. Such features offer critical insights for various subfields of linguistics: typology can expand its understanding of linguistic universals and variation, morphology can develop new analytical frameworks from complex word-formation patterns, and historical linguistics can reconstruct previously unknown stages of language evolution based on their relatively undisturbed, pre-modern states. In essence, each documented endangered language enriches the "theoretical toolkit" of linguistics, enabling the discipline to shift from explaining known phenomena to exploring uncharted linguistic territories.

From a social perspective, documentary linguistics plays a crucial role in empowering the cultural rights and identities of minority communities. Language extinction is, at its core, an erosion of cultural identity. Many small ethnic groups, due to historical colonization, cultural assimilation pressures, or the forces of modernization, are compelled to abandon their native languages in favor of dominant languages. This shift often severs the emotional and generational ties between younger members and their cultural heritage, perpetuating a destructive cycle of "language extinction — cultural marginalization — accelerated language

loss." According to Leanne Hinton (2010), "it has become a given that linguistic research must also serve the interests of the community whose language is being documented" (p. 35). Documentary linguistics addresses this challenge through a participatory research approach: it not only documents linguistic data but also emphasizes collaboration with community members. By involving native speakers in documenting the content priorities—such as myths, oral traditions, and daily conversations rather than solely grammatical structures—and returning the achievements (e.g., dictionaries, audio recordings, and cultural guides) to the community, which will serve as a "pedagogical material" (Cablitz, 2011, p. 446) at a later stage, this method provides practical tools for language revitalization, more importantly, it fosters a renewed recognition of the value of the mother tongue to community members. More importantly, it fosters a renewed recognition of the value of the mother tongue. As younger generations engage with their ancestral language through these resources, they can reestablish a meaningful connection with their cultural heritage, thereby revitalizing cultural pride and identity.

In an era marked by rapid language extinction, the value of documentation linguistics has transcended the mere preservation of linguistic data. It represents a comprehensive effort to safeguard the diversity of human civilization, to rescue cultural knowledge, to advance interdisciplinary research, and to uphold the cultural rights of delating "all" linguistic communities. When we record an endangered language, document its grammar, and preserve its narratives, we are not only preserving the past but also safeguarding the future's capacity for cultural and intellectual diversity. The richness of human civilization is rooted in linguistic diversity, and such diversity demands our collective commitment to its protection.

Conclusion

During the past two decades, the emergence of documentary linguistics has brought about great achievements in the documentation and preservation of low-resource or endangered languages in China. The interdisciplinary fieldwork, practice, and values that the projects of documentation of low-resource or endangered languages in southern China, particularly those from the Tai-Kadai family, have gained and displayed provide good cases of language documentation and preservation in the digital era. More cases of similar vain should be done as soon as possible as an active response to low-resource or endangered languages in this new era.

References

Cablitz, Gabriele. (2011). Documenting Cultural Knowledge in Dictionaries of Endangered Languages. *International Journal of Lexicography*, 24(4), 446-462. https://doi.org/10.1093/ijl/ecr017

Constance, Holden. (1991). Endangered Languages. *Science*, 251(4990), 159.

Dorian, Nancy. (2010). Language and responsibility. *Language&Communication,* 30, 179-185.

Ethnologue—Languages of the World. (2025). How Many Languages Are There In the World? [Data set] . Summer Insitute of Linguistics Internation. https://www.ethnologue.com/insights/how-many-languages/

Fan, Junjun. (2022). The Building of Digital Museum for Endangered Language. *Museum Management*, 4, 45-55.

Himmelmann, Nikolaus. (1998). Documentary and Descriptive Linguistics. *Linguistics*, 36, 161-195.

Hinton, Leanne. (2010). Language Revitalization in North America and the New Direction of Linguistics. *Transforming Anthropology*, 18(1), 35-41. https://doi.org/10.1111/j.1548-7466.2010.01068.x.

Huang, Chenglong. (2024). Linguistic Fieldwork and Research in the Era of Digital Intelligence—Documentary Linguistics: The State of the Art. *Journal of International Chinese Teaching*, 1, 3-12. https://doi.org/10.3969/j.issn.2095-798X.2024.01.002

Mithun, Marianne. (2007). What is a language? Documentation for diverse and evolving audiences. *Language Typology and Universals*, 60(1), 42-55. https://doi.org/10.1524/stuf.2007.60.1.42

Nathan, David. (2008). Minding Our Words: Audio Responsibilities in Endangered Languages Documentation and Archiving. *Taiwan Journal of Linguistics*, 6(2), 59-78. https://doi.org/10.6519/TJL.2008.6(2).3

Spector, Violette. (2025). AI in Language Preservation: Safeguarding Low-resource and Indigenous Languages. https://www.welocalize.com/insights/ai-in-language-preservation-safeguarding-low-resource-and-indigenous-languages/

Uribe-Jongbloed, Enrique. Endangered languages: Heritage of humanity in dire need of protection. *Folios*, 65-70. https://doi.org/10.17227/01234870.26folios65.70