

Minority languages in China and the national preservation project

Edition 12, 2022

Professor Jinfang Li & Professor Yongxian Luo

DOI: 10.37839/MAR2652-550X12.19

1. The languages of China and their speakers

China is a diverse multi-ethnic country with a huge population of 1.4 billion people. It is home to a rich array of ethnic groups who speak a wide range of languages. Around 90 percent of the population speak Chinese, including some 84 million Cantonese speakers which is considered by many a separate language, not to mention several hundred million speakers of Shanghainese, Hokkien, Hakka and other Chinese dialects that are mutually unintelligible.

A total of 55 ethnic minority groups have officially been recognised, apart from Han Chinese. This figure was determined by a team of social scientists and cadres (a worker who implements the goals of the communist party) through a massive state-sponsored enterprise undertaken in the 1950s.

It is useful to note that the concept of 'ethnic group' (*mínzú* in Chinese - 'people-clan') carries both political and cultural nuances. The term *mínzú* is variously translated as 'ethnic minority', 'race', 'nationality', and 'nation'. The frequently-used term '*zhōnghuá mínzú*' means 'the Chinese nation; the Chinese state; the Chinese race', depending on context, and *zhōngguó* '(the) middle kingdom' means 'China, the

Han Chinese, the Chinese empire, the Chinese state, the Chinese nation', which may refer to China proper and/or greater China as broadly conceived. As such it is associated with the concepts of 'regionalism' and 'nationalism'. Before the founding of the Republic, only five ethnic groups were recognised, namely the Han, Mongolia, Hui (Chinese Muslims), Manchura, and Tibetan. Thus Sun Yet-Sen, founder of the Republic, spoke of *wǔzúgònghé* 'the unity of the five nationalities' in his inaugural presidential address on January 1, 1912.

1.1 The classification of the languages of China

Linguistically, China's minority languages may be grouped into five major language families:

- Sino-Tibetan, which includes many Sinitic languages and Tibeto-Burman. Most Chinese scholars identify Kam-Tai, and Hmong-Mien as members of Sino-Tibetan, although many western scholars tend to treat these as separate language families belonging to different stocks
- Altaic: Turkish, Mongolian, and arguably, Korean
- Man-Tungus
- Austronesian; and
- Mon-Khmer.

In addition, a few Indo-European languages are also represented, such as Tajik.

Language classification and ethnic identification do not always go hand in hand. Matters are made even more complicated where languages are in contact, resulting in language mixing, or when a group are bilingual or multi-lingual. For example, what is referred to as Jingpo is in fact made up of a group who speak distinct languages belonging to different branches of Tibeto-Burman. Likewise, many who identify themselves as Tibetan are actually Qiang, Gyarong, Pumi, Namuyi, and Muya speakers whose languages are very different from Tibetan. Despite the large-scale language survey conducted in the 1950s, some minority groups remain

unidentified and their status as a separate group has not been determined.

It is worth noting that the criteria differ for language classification. Quite often, geographic location is a key factor and mutual intelligibility is not. For example, it is generally accepted that there are 8-10 Chinese dialects. However, if mutual intelligibility is taken into account, the numbers will go up significantly because the differences between dialects or even varieties within a dialect group are far more striking than between, for example, Swedish and Norwegian, to the extent that many varieties of Chinese bespeak different languages, for example, Cantonese, and Shanghainese. Similarly, many non-Han ethnic minority languages are characterised by multiple dialects or varieties which exhibit a high degree of diversity. For this reason, western scholars seek to classify some languages of China into tens or several dozens of separate languages.

A recent research volume, *The Languages of China*, has identified 128 languages that are spoken in China in addition to Chinese. Of these, 64 have over 10,000 speakers, 48 with 5,000- 10,000 speakers, and 11 with 100 to 4,900 speakers. In terms of the number of speakers who speak their native language, about half of those who speak minority languages have less than 10,000 speakers, whose languages are considered endangered, or facing endangerment. A recent report indicates that there are currently 25 languages in China which have less than 1,000 speakers. Some of these 25 languages have only about a dozen speakers left and are gravely endangered. Of the 128 languages, 48 have good descriptions which have been published in the *Newly-Discovered Language Series* under the general editorship of Professor Sun Hongkai of Chinese Academy of Social Sciences.

1.2 Government language policies towards minority people

Ever since the time of Qin Shi Huang, the founder of the Qin dynasty (221 to 206 BC) and the first emperor of the unified China, language was seen a symbol of unity and identity of the Chinese nation. Indeed, the existence of a unifying language with

over 5,000 years of written history was key to the continuation of Chinese culture and tradition. The Chinese pride themselves in their language, believing that it has specific features that are superior to other languages in the world (such as tones to distinguish meanings and the use of semantic radicals to arrange words into different categories). Thus, the state's imposition of a national standard language, particularly a standardised written script that serves as a lingua franca, has rarely been questioned.

Article 4 of the Constitution of the Peoples Republic of China stipulates that 'each ethnic group has the freedom of using and developing its own language'. Since 1950, the Central Radio Station (China National Radio, the national radio network) has broadcast programs in several minority languages, including Tibetan, Uygur, Kazak, Mongolian, and Korean. The government also set up *minzu* schools (where bilingual education is available) across the country, particularly in the five autonomous regions. In addition, 13 *minzu* universities were established across the country, including the prestigious Minzu University of China, formerly Central University for Nationalities. A major function of these universities is to train ethnic minority cadres to work in minority areas.

It is worth noting that China has probably the world's most extensive regime of minority affirmative action policies-known as 'preferential policies' (*youthui zhengce* 优惠政策) which are far more extensive than the more widely discussed and debated policies in the United States of America and other liberal democratic countries. These policies provide tangible material and other benefits (such as study programs for ethnic minority students to help them transit to university and the possibility of promotion for government officials who can speak the local ethnic language) to nearly 120 million ethnic minorities in many aspects of daily life. Lower requirements (such as lower entry scores for, or extra entry points) are set for ethnic minority students in university admission. Incentives are offered to government officials working in minority areas who are conversant in the local ethnic language. Such preferential policies appear to work, and do not conform to the hypothesis of Thomas Sowell that affirmative action around the world fails to produce substantial

equity, inhibits economic efficiency and creates inter-ethnic tensions.

Since the beginning of this century, work on the preservation of endangered languages and language resources in China has received increasing attention from the government and society.

Beginning in 2006, an annual Green Paper on the current state of languages in China has been prepared by China's State Language Commission, an administrative department under the Ministry of Education. These Green Papers are published under the title Language Situation in China, and the English translation of the key parts of the reports between 2006 and 2017 are now available. With details on many facets of language policies and language use in China, the reports provide key information for the dynamics of socio-cultural changes in Chinese society today.

2. Studies on endangered languages of China and language preservation activities



Working on written scripts, Gelao, Guizhou Province. Credit: Li Jinfang.

2.1 Language endangerment in China

While many ethnic minority languages are still alive and well, many of them are endangered, or even facing extinction. Language endangerment and language extinction has occurred throughout Chinese history. The Khitan people, boasting a history dating back to the 4th century, dominated much of northern China, Manchuria and the Mongolian Plateau and yet their language is no longer in use. Neither are Tangut, Jurchen, and Manchu.

Although many Chinese dialects and certain ethnic minority languages are not officially considered endangered, the number of people actually speaking their native language is declining rapidly at an alarming rate. Equally concerning is the

loss of resources of oral culture imbedded in language, such as stories and legends, ballads, proverbs, and folk theatres, among others.

2.2 The Language preservation project

In 2015, the State Language Commission and the Ministry of Education launched the Language Preservation Project (often referred to as *yu-bao*), with funding from the central government. Thanks to this support, the Chinese linguistic community was able to expand the scope of researching endangered languages to documenting languages and dialects of China, including regional languages spoken in Hong Kong, Macao and Taiwan. Prior to the Language Preservation Project, the survey and description of non-Han minority languages was largely carried out by researchers from the Institute of Ethnography and Anthropology of the Chinese Academy of Social Sciences.

Phase I of the Language Preservation Project spanned five years (2015-2019), during which several hundred languages and dialects were surveyed across the country, with a total of 1,712 survey points, including more than 200 special sites for languages and dialects that were deemed endangered. These included 123 non-Han ethnic languages, and 106 local Chinese dialects. During the initial stage of the project, over 10 million entries of original corpus data were collected, along with more than five million pieces of audio and video, amounting to a total physical capacity of 100TB of data. With participation from 350 research centers and institutions, the project represents the joint efforts of some 4,500 investigators, and more than 9,000 language consultants. A series of landmark research outputs have been published or are in the process of publication, including the *Collection of Chinese Language and Culture* (in 50 volumes), the *Cassettes of Endangered Languages in China* (in 50 volumes), and the *Collections of Chinese Language Resources* (arranged by province, in 100 volumes).

Phase II of the Language Resources Preservation Project started in April 2021, focusing on the scientific preservation of the resources of the Chinese dialects and

minority languages, as well as the development and utilisation of language resources and promotion of the national common language and writing. Several endangered language survey points have also been added. Data from the 'Phase I' survey are being collated, and digitised. Survey results will continue to be published, as will the Collections of Chinese Dialect and Minority Language Resources. Work is being strengthened on data storage and management, and on display platforms. Construction of language museums are also being supported to provide access to language resources and promote the dissemination and use of these resources to help socio-economic development.

The Language Resources Preservation Project adopts a unified technical standard, with clear guidelines for the classification of the content of survey records. Audio and video recording and other forms of image capture as well as text processing, phonetic transcription, among others, are all carried out in accordance with the current internationally accepted technical standards that are conducive to dissemination and archiving. The basic technical parameters are:

- Audio Recording: channel: mono; sampling rate: 44100Hz; sampling precision: 16bit; audio format: Windows PCM (*.wav)
- Video recording: HD mode, video file parameters are not less than 1920×1080/50i, with file format depending on the type of video camera, e.g. *.m2ts, *.mpg.
- Photos: resolutions not less than 4368×2912 pixels, in *.jpg format.

The contents of field investigation are grouped according to typological considerations. For Chinese dialects, survey points are set up according to the county-level administrative units. In principle, the practice of 'one survey point per county' is implemented. More survey points may be added, or reduced, where circumstances so warrant. Four representative language consultants are selected for each survey site according to gender and age factors, including two men and two women from two different age groups, with two from the younger generation, and two from the older generation.

The survey contents focus on language structure and discourse. Language structure relates to the sound system and grammar of a language. Some 2,000 lexical items are used to examine the sound system and the lexicon, including a 1200-word basic vocabulary to ensure a good description of the sound system. Fifty sentences are employed to investigate basic grammatical features - commonly-used sentences that reflect the key grammatical profile of the language under investigation. Natural discourse data also forms part of the language survey. Discourse structure comes from stories and dialogues. Storytelling includes 'prescribed stories' and 'optional discussion topics'. The romantic folk story 'The Cowherd and the Weaver Girl' is the chosen topic for storytelling, but a wide range of other topics are potentially also used.

For non-Han ethnic languages, survey points are set up in ethnic minority areas where the target languages and dialects are spoken. In general, a minimum of one survey site is set up for each language. A separate survey site may be set up for each particular dialect and/or vernacular where significant differences can be observed. A survey point is selected where the most representative dialect of the language is spoken. A main language consultant is chosen for each of these survey points, a male, aged 55-65 at the time of the survey. 'Oral culture' can be provided by different language consultants.

The investigation includes the following:

- General introduction: an overview of the survey point, a brief account of the language consultants, and a general description of the investigation.
- Phonology: systems of tones (if any), initials or consonants, finals or vowels, and example words for each type.
- Vocabulary: 'Common Words' (or 'General Vocabulary', a total of 1,200 lexical items) and 'Extended Word List' (or 'Extended Vocabulary', a total of 1,800 lexical items).
- Grammar: 100 grammatical structures.
- Discourse: seven topics, the consultant can choose one or more topics to

- speak, with a total length of 20 minutes.
- Oral Culture: Ballads, Stories, Self-Selected Topics, with a total length of 20 minutes. (Attention is paid to the investigation and collection of language-related materials involving various types and levels of intangible cultural heritage.)
 - Local pronunciation of Putonghua, the Common Language or Standard Mandarin. Consultants are requested to talk about one topic for three minutes and then read two short articles in Putonghua, using local pronunciation. This is to check the general ability of ethnic minorities to use the national language.

Apart from the standard requirements for general investigation, the special project also includes the social, historical and cultural background of the language, and endangerment status, among others. Researchers are strongly encouraged to make recordings of as many vocabulary items as possible, along with grammatical example sentences and oral cultural corpus, in order to achieve a comprehensive picture of the language's structure, characteristics and cultural activities (such as ballads, folk stories, folk theatre, and talking and singing performances). This provides solid material for writing language descriptions and preserving language resources.

The initial stage of the endangered language documentation project also incorporates a special language and cultural survey. A total of 102 survey points are designated for languages that exhibit distinctive regional and ethnic cultural features. In general, the scope of the survey is set at the county level. The following sets of items are ear-marked for survey: housing construction; daily tools and utensils; clothing; food; agriculture and handicraft; daily activities, including marriage customs, childbirth and funerals; festivals, and ballad singing performances. The ultimate goal is to present an account of traditional culture. Approximately 1,000 cultural items are to be included in the survey. These are to be photographed, classified, labelled, and annotated. Representative activity items such as festivals, social etiquette, and various types of speaking and singing performances also need to be photographed and video recorded. The collected materials will be

archived after processing. Selected contents will be edited as books, to be published in the *Chinese Language and Culture Collection Series* by the Commercial Press, one of the major publishers in China.

The Language Preservation Project and earlier research efforts have uncovered some of the most typologically complex languages. For example, Lawurong, a Tibetan language spoken by a group of Tibetan people residing in the Jinhuanhe river basin in Aba Tibetan-Qiang Autonomous Prefecture in Sichuan Province, boasts a highly sophisticated phonological system with 44 simple consonants, yielding nearly 400 consonant combinations consisting of double, triple, quadruple and even quintuple clusters. In several recent field trips, Professor Li Daqin and his team from Communication University of China discovered three previously unidentified languages—Suku, Songlin, and Zhahua—all spoken in Zayu County in the Southeastern Part of Tibet on the Tibet-Yunnan border. Further investigation may uncover more languages.

2.3 Other efforts

In recent years, a number of local governments—notably the so-called ‘frontier provinces’ of Guangxi, Guizhou, Yunnan, Tibet, Inner Mongolia, and Hainan Island, as well as several neighbouring provinces such as Sichuan, and Hunan—have also actively engaged in local language preservation work. For example, the Language Research Centre of Guangxi Ethnic Language Commission has set up more than 40 ethnic language survey points in Guangxi alone under the national language preservation plan. In addition, more than 50 ethnic language survey sites have been established, bringing the total number of ethnic language survey sites in Guangxi to 100. This has effectively promoted the preservation of ethnic language resources in Guangxi, home to the Zhuang people, the largest minority group in China.

It is worth noting that several entrepreneurs and celebrities have also joined in the language preservation efforts. A businessman from Guangdong has provided funding for the annual ‘Dialect Film Festival’ to encourage people to shoot movies in their

dialects or native languages. A celebrity anchor from the Hunan Television Station has donated five million RMB to sponsor university teachers and students to survey and preserve the Chinese dialect in Hunan Province, hoping that future generations will be able to hear the language of their ancestors in the museum.

To further promote consensus and efforts to protect language resources and language diversity across countries, the Ministry of Education of China, the People's Government of Hunan Province and UNESCO co-organised the World Congress on the Conservation of Linguistic Resources on September 19-20, 2018 in Changsha, Hunan Province. The theme of the congress was 'The Role of Linguistic Diversity in Building a Community of Shared Future for Human Beings: Preservation, Application and Promotion of Language Resources'. The conference released the Yuelu Declaration, the first important permanent document of UNESCO with the theme of 'protecting linguistic diversity', calling for international communities, countries, districts, governments and NGOs to reach a consensus on protecting and promoting the world's linguistic diversity.



Terrace field of Zhuang in Longsheng, Guangxi Province. Credit: Luo

Yongxian.

3. What's next?

Research on the preservation of language resources has achieved remarkable results over the past 20 years. However, language endangerment and the loss of language and cultural resources have not stopped. As the government and society have a deeper understanding of the need to protect endangered languages and other language resources, work on language preservation will enter a new phase. China's basic language policy remains to 'vigorously promote the national common language and scientifically protect the languages of all ethnic groups.'

There are many challenges to China's language policy and language planning. It is a daunting task to strike a balance between building a modern, outward looking nation and maintaining its distinctive cultural heritage and characteristics. Certain preferential policies may be better implemented in ways that help with language and culture preservation efforts rather than undermining them.

Future research work on the preservation of endangered languages and language resources in China may benefit from the following perspectives:

1) Clear and coherent language policies should be in place for effective research on the preservation of endangered languages and language resources. On the basis of a unifying technical standard, practical measures should be applied to different languages and dialects in light of actual circumstances. With certain languages, revitalisation efforts are more important; for languages that are critically endangered, documentation of its language resources is essential before the language disappears.

2) Participation from the broader community, including gender equity and equal participation of people from minority groups, should be encouraged for language resource preservation. Public awareness should be raised of the significance of this

aspect of cultural heritage. To this end, more publicity is needed to identify those who could participate in this work.

3) Digitisation of language resources should be strengthened to boost research on language preservation. Digitisation is a means to effectively preserve and utilise language resources and several universities in China have established language museums. Collection, storage, and display of digitised language resources across different regions, countries, and worldwide will undoubtedly help with language preservation efforts.

4) The development and utilisation of language resources will help to promote the preservation of languages and the maintenance of endangered languages. In addition to display platforms in museums, mini-application programs, short videos, WeChat public accounts, etc. should be developed. New media technology should be utilized to promote language resources, and to develop small language products, such as short conversations, proverbs, idioms, common sayings, songs, stories, among others. Language resources can also be used in literature and creative art products to create economic value. Language resource development can promote language use, and through language use, language vitality will be maintained.

5) Promotion of bilingualism: As society and the economy progress, it is inevitable that dominant languages will put pressure on ethnic minority languages, particularly when urbanisation is high on government agenda. Policy makers should seriously think about the issue of bilingualism as a way to preserve language vitality. As Li Wei points out in his Preface to Volume 5 of *Language Situation in China*, language planning and policies should embrace 'a mentality of diglossia and bilingualism'. Training diglossic and bilingual people is vital in China's multicultural society.

Authors: Professor Jinfang Li and Professor Yongxian Luo.

Main image: Gelao group singing, Guizhou Province. Credit: Li Jinfang.