

Digitalization Mining and Protection of Security Language Resources in Big Data Environment

Aruna Bai

Langfang Normal University, Langfang, Hebei, 065000, China

Abstract. The Baoan language is the language of the Chinese security people. It is also used in some parts of Monguor people. As an important carrier for the dissemination and inheritance of minority culture, it is of great practical significance to excavate and protect the resources of the security language. In the context of modern culture, the minority language, represented by the Baoan dialect, is facing a great survival crisis and communication pressure, and even is on the verge of being lost. Faced with the new big data environment, we should give full play to the technical advantages, use the digital way to excavate and protect the security language resources reasonably and scientifically, and ensure the security language to develop in an orderly way. By analyzing the current situation of security language resources and combining with big data environment, this paper puts forward the strategy of digitalization mining and protection of security language resources.

Keywords: Big Data Environment, Security language, Language resources, Digital mining, Protection strategy

In recent years, with the development of computer and Internet technology, the research of minority language has gradually entered the big data era. However, the diversity of sources and recording methods in the database make it difficult for the academic community to use the information. For the construction of security language resources, big data is both an opportunity and a challenge. On the basis of big data environment, we should actively explore the digital mining and protection path of security language resources.

1 BACKGROUND OF DIGITAL MINING AND PROTECTION OF SECURITY LANGUAGE RESOURCES IN BIG DATA ENVIRONMENT

The security guard is a self proclaimed, historically known as Hui people, security guard back and so on. Later, according to the wishes of the people of this nationality, it was officially designated as the Bao'an ethnic group. Bonan people has its own language, no written language. The Altaic languages language belongs to the Mongolian family, and most people are familiar with Chinese. The Bao'an people, who believe in Islam, are largely the same as the Hui, Dongxiang and other ethnic groups in terms of family life, customs and society. Bao'an ethnic group mainly engaged in agricultural production, and engaged in animal husbandry and handicraft industry. The Bao an clan, who had learned to make iron very early on, had superb knife-making skills. Their waist knives, sharp and durable, were exquisite and beautiful, and enjoyed great fame among the local people of all ethnic groups. They were called the Bao an Dao.

Bao an language is the language of the Chinese Bao an ethnic group, which is also used in Monguor people. It belongs to the East Mongolian branch of the Altaic languages Mongolian language family, areas used include the Linxia Hui Autonomous Prefecture Jishishan Bonan, Dongxiang and Salar Autonomous County of Gansu province and Tongren County of Qinghai province. The Baoan language is mainly distributed in the Baoan ethnic group of Jishi Mountain in Gansu province, Dongxiang people, Salar people autonomous county and Linxia County. In places like Xiazhuang and Gastel, near the city of Tongren County in Qinghai province, the language spoken in some parts of Monguor people is basically the same as that spoken in Baoan, with a population of about 9,000 people. Most young adults also speak Chinese. There is no native language, the Chinese language is commonly used. There are two dialects: Dahe family and Tongren family.

Bao'an language is an important communicative tool for the people of Bao'an nationality in China. The security language is closely related to the language of the same language family, and its features are outstanding. Especially compared with Mongolian, the security language has many features in pronunciation, vocabulary and grammar. In terms of vocabulary, in addition to absorbing a considerable number of Chinese loanwords, but also absorbed a lot of loanwords from Tibetan. The Bao'an people, who are mostly fluent in Chinese, have long used the Chinese language. The Bao'an dialect contains a large number of words with the same phonetic origin as Mongols. Because the Bao'an nationality lived together with the Han nationality and the Tibetan (part of Qinghai) in the villages and towns for a long time, the language contains a large number of Chinese loanwords and a certain number of Tibetan loanwords.

In the aspect of grammar, the personal pronoun of Bao'an language is different from the noun, the first and second person are independent of the collar case, the object case and the position case merge and have the wrong root phenomenon. Declarative verbs have definite and non-definite grammatical categories. A quantifier can be placed after the head word when it is used as an attribute. Besides the categories of tense and aspect, the declarative forms of the verbs in the security language also have the categories of definite mood and non-definite mood. The definite mood of the declarative verb expresses the speaker's direct experience, personal experience, subjective decision and will. The indeterminacy of a declarative verb expresses the speaker's non-personal experience, experiences, and situations that are unexpected or against one's will, as well as the tone of the paraphrase. The declarative verbs in Bao'an language include the past tense, the present tense, the future tense, the progressive aspect and the perfect aspect. Each kind of tense and aspect has the difference of definite and non-definite mood, and has the relative different grammar form. According to the investigation, this grammatical category also exists in some languages of the Mongolian and Turkic peoples language families.

2 THE SIGNIFICANCE OF DIGITALIZATION MINING AND PROTECTION OF SECURITY LANGUAGE RESOURCES IN BIG DATA ENVIRONMENT

In 2015, China launched the China Security Language Resource Protection Project, and launched a nationwide survey, preservation, display, development and utilization of security language resources. Today, the protection of security language resources is of great significance in many aspects.

Scientific and effective protection of security language resources, to carry forward the excellent traditional Chinese culture to lay the foundation. At a time when the security language is gradually shrinking, through scientific planning and the use of modern technology, a comprehensive survey will be conducted to preserve the actual language materials of the security language and oral culture, and the endangered security language dialects will be rescued in time, it is of great significance to inherit and carry forward the excellent Chinese traditional culture, and to promote the creative transformation and innovative development of the excellent Chinese traditional culture.

Building a big data platform for security languages to provide services for mastering the national situation of security languages, government decision-making and social applications. The construction of a large database of security languages with an increasing number of resources covering security languages throughout the country, through the survey of security language resources in thousands of locations and the ongoing collection of open security language data, to make it an accurate, authoritative, open and shared Chinese security language resources public service platform, which will help to enhance the level of information and social service ability of Chinese security language, it also provides strong support for a comprehensive understanding of the national situation of the security language and for improving the government's decision-making ability in the security language, it will create a new and wide space for

developing security language data mining, developing security language intelligence and cultural products, and serving local cultural construction.

Building a harmonious language life plays a unique role in building a strong sense of *Volksgemeinschaft*. Security language is an important force for national, social and national stability. On the basis of a comprehensive and systematic investigation and collation of China's security language resources, scientific preservation and protection of dialects and the security languages of various ethnic groups, realizing the promotion of social harmony by living in harmony in the language of security, the promotion of the integration of language, culture and civilization in the language of security, the promotion of the unity and struggle of all nationalities for common prosperity, and the timely prevention and resolution of conflicts in the language of security, to safeguard national security and unity.

We will protect and promote the linguistic diversity of the security forces and contribute Chinese wisdom, Chinese solutions and Chinese experience to building a community of shared future for mankind. The endangered security language has become a global issue. China has successfully implemented the Chinese Security Language Resource Protection Project, leading the world in the protection of the security language, the United Nations Educational, Scientific and Cultural Organization and the international community. The experience, standards and technologies of the language protection project will contribute Chinese wisdom to the protection of language resources and the building of a community of shared future for mankind.

Strengthen, standardize and salvage the investigation and arrangement of the language and culture resources of the minority security languages, classify, arrange and translate all kinds of excellent oral cultural works, and provide core materials for all kinds of development and application, making full use of the language and cultural resources of the security language in the maintenance of the mother tongue and bilingual education, and using the language resources of the security language to provide language services of the security language in the construction of the new countryside, the security language resources are used to serve the administration, judicature, news, tourism and other fields in ethnic areas, and the security language and cultural elements are fully used in the creative development of ethnic culture.

3 DIGITAL MINING AND PROTECTION OF SECURITY LANGUAGE RESOURCES IN BIG DATA ENVIRONMENT

To the security language resources construction talented person and the technique proposed the higher request. For example, how to debug the signal-to-noise ratio, how to avoid phenomena such as flapping and clipping, so as to obtain high-quality and high-fidelity recording materials, and how to build a database of sound security language, how to develop and design homophone checking software, to organize professionals to carry out repeated verification of the survey materials to ensure the accuracy and reliability of the materials; how to further build the database of acoustic parameters, how to mine the data and make use of the security language data to carry out in-depth professional research. Not only do researchers and researchers need to be trained in the appropriate skills, but the paradigm shift of the big data era is forcing researchers to constantly improve their professionalism and research capabilities.

The standardization and standardization of security language resource construction has put forward new requirements. In addition to the need to establish industry uniform recording standards and recording standards, so that the survey data of different personnel and different teams can be developed and utilized in a unified way, we also need to check and collate the survey data of the past and the secondary standards, so as to build a larger database of security language resources. It is necessary for experts in the field to draw up unified specifications and reach a consensus in the academic field. How to realize computer-aided verification and secondary specifications quickly for existing databases is an issue that needs to be deeply discussed.

The preservation and protection of the minority language is a major issue and an effective guarantee for the healthy economic and social development of ethnic minorities. However, the Internet plus action plan, we can take advantage of the internet plus east wind, according to its spirit and concept, analysis of the reasons for the development of the minority language, and how to seize the opportunity of the Internet Plus, it can promote the development of the minority language through the establishment of the Internet + Digital Library of the minority language, the digital service of the literature, and the sharing platform, it is of great significance to promote the prosperity and stability of ethnic minority areas.

At present, with the rapid development of modern information technology, the construction of security language resources is on the fast track, showing a number of new trends, new features. First, it is highly concentrated and pushed forward as a whole. The great attention of the party and the country, the large-scale investment of the government and related units, and the active participation of the majority of outstanding talents will make the construction of security language resources in our country achieve fruitful results. As the related projects become more and more effective, the investigation becomes more and more concentrated, and the research becomes more and more standardized, the construction of security language resources will become relatively concentrated and highly integrated, and form a certain degree of scale effect, and then promote the overall development. The construction of security language resources in China will be more concentrated, more comprehensive and more systematic, so as to promote the unified planning and comprehensive development of security language resources, so as to achieve overall development and form scale effect, in order to better national security language strategy, Social Security language life and security language academic research services.

Looking at the development and Operation Track of the national minority language, we can find that every step and every link in its development and operation can not be separated from CO construction, sharing and cooperation. The construction of security language resources in the new era has put forward new and higher requirements for this. This can not be done by any single unit or team, but needs many domains, many units, many region's experts and scholars and the scientific research personnel to participate together, the group work together, displays the combat superiority. Through cooperation and co-construction, the possibility of extracting the general rules of the security language has been increased, and the evolution of the security language has been promoted. The security language can not only display the Eight Immortals, but also gather the English and the Chinese people together. On the one hand, field survey recording equipment and technology can be shared in order to improve the overall recording quality; by recording their own good at and familiar with the security language or dialect, can improve the overall professional quality. On the other hand, the results of the investigation can also be shared, and the research results can be mutually reinforcing. Once supported by a secure language database, sometimes seemingly random phenomena can be calculated, processed, and analyzed to uncover potentially important information.

The former security language resources construction and protection work, the form is relatively simple, the content is also very limited, the final form only professionals can understand the security language or dialect survey report. In contrast, the current security language resource construction and protection work has undergone a historic change. In form, it is no longer limited to paper-and-pencil records, but has been extended to high-quality, high-fidelity sound and video recording. Content, not limited to dialect word list, word list or national semantic item list of records, but also covers a wealth of folk songs, stories, intangible cultural heritage and so on. The achievements include not only the security language survey report and its audio materials, but also dialects and national languages. In effect, the achievements of the construction of security language resources in the past were mainly used in the research of the minority security language, and the achievements of the construction of the present security language resources provide a

platform and services for exchanges among various ethnic groups, it has greatly enriched the language and cultural life of the security guards.

4 CONCLUSION

Looking to the future, the security language resources construction is full of hope, is a need to work together, long-term adherence to the task. As long as we emancipate our minds and keep pace with the times, set the language resources of the security language in the context of social progress and cultural prosperity, fully mobilize all kinds of resources and widely pool all kinds of strength, we will be able to better develop, protect and use China's security language resources, so that the construction of security language resources can always keep pace with the country's development and progress at the same time.

5 ACKNOWLEDGEMENTS

Langfang Normal University scientific research project "Exploration of minority language resources in colleges and universities"; project number: XBQ202057

REFERENCES

- [1] Xiangyou Liu. Protection and inheritance of endangered minority languages under the background of "Internet +" [J]. Guizhou Ethnic Studies, 2017, 38(12): 217-220.
- [2] Guanghong Qiu, Zixuan Zhu, Rui Liao. Research on the Strategy of Qiang Language Inheritance in the We-Media Era [J]. Drama Home, 2017, (20): 212-213.
- [3] Shanjie Zhuang, Shumei Meng. Digital Mining and Protection of Hezhe Language in Big Data Environment [J]. Frontier Economy and Culture, 2017, (11): 44-45.
- [4] Wen Li. A Brief Discussion on the Active Research on the Digital Construction of Yunnan Minority Historical Archives [J]. Archives Research, 2018, (06): 64-69.
- [5] Jing Xie. Design strategy of digital learning resources with local characteristics: Taking Chaoshan dialect as an example [J]. Digital Media Research, 2019, 36(11): 73-77.