



Research on the Promotion of Tibetan New Words and Platform Design

Hao Li^{1,2}, Yuan Sun^{1,2(✉)}, Wei Song¹, and Jie Qin^{1,2}

¹ School of Information Engineering,
Minzu University of China, Beijing 100081, China
tracy.yuan.sun@gmail.com

² Minority Languages Branch, National Language Resource and Monitoring
Research Center, Minzu University of China, Beijing 100081, China

Abstract. With the rapid development of the country's technology and culture, more and more Tibetan new words appear, which add vitality to the language. This paper investigates the current usage of Tibetan new words by questionnaire and statistics of relevant websites. We find there exist some problems including time delay, non-standard form of expression and the publishing approach is too single. These problems bring great barriers to communicate and use Tibetan new words. To solve these problems, we design a platform for the promotion of Tibetan new words in the website, which can make the Tibetan new words become more standardized and effective, and we can get statistics on the usage of new words. Meanwhile, it is conducive to the spread of new words in Tibetan, and the platform can also collect effective data to serve the related work of Tibetan new words.

Keywords: Minority language · Tibetan new words · Language promotion · Learning platform

1 Introduction

China has 56 ethnic groups and many different languages. Tibetan is the language used by the Tibetan people. It is used by more than 6 million people in China, and is distributed in Tibetan, Qinghai, Sichuan, Gansu and Yunnan provinces [1]. With the development of society, the frequency of new words in Tibetan has become more and more frequent. However, there are dialects in the five provinces of Tibetan, the lag of standardization work, and the lack of uniform and standardized translation, these brings great challenges to the development of Tibetan.

The development of language has a decisive influence on education. For the Tibetans in different Tibetan areas, the research on their familiarity and habitual expressions of new words is conducive to promote Tibetan new words.

At present, the release way of new words in Tibetan is mainly published by some traditional paper documents and the government, which are relatively scattered and the real-time performance is poor. It takes a good platform to promote the Tibetan new words. At present, the website development technology is already in a mature stage. The educational resources are gradually running in the form of cloud services in cloud

personal learning environments and devices such as smart phones, pads, and e-book bags [2]. Based on this background, the construction of a new words learning and promotion platform for minority languages can be a method and a way to promote new words in minority languages. Starting from the actual research, the author decided to start this work from the Tibetan first, in order to meet the needs of the Tibetan compatriots represented by the majority of young students.

2 Related Works

The study of Tibetan new words began as early as the 1950s. Today, nearly 70 years later, the work related to Tibetan new words has made great achievements, at the same time there are some problems. After entering the Internet era, the traditional way of promoting new words has been transformed from audio media and offline media into Internet and mobile Internet media. This allows the promotion to have new developments in this era.

2.1 Tibetan New Words

The progress of the times and society is always accompanied by the development of language. And the emergence of new words in different languages has a great relationship with the socio-economic, political and cultural. Natural Language Processing is a major branch of artificial intelligence [3]. The application of machine learning in the field of natural language processing has also led to the development of new words [4]. For example, for the study of Chinese and English new words, firstly, it is carried out by an authoritative department such as the Standardization Committee of the International Standardization Committee, at the same time, Xinhua News Agency and the authoritative publications of “China Translation” also regularly compile the translated names of some new English nouns, secondly, the definition and translation of some new words are given by the English-Chinese dictionary and the new words dictionary. All of the above methods have certain lags and are not comprehensive, so the translation and naming of many new words must be done by individual translators. There are two ways of transliteration and paraphrases.

From the 70s to the 80s, the number of Tibetan new words began to expand, which greatly enriched the vocabulary of Tibetan, and the Tibetan new words also existed [5]. Since then, from the end of the 1980s to the beginning of the 21st century, all parties have invested a great deal of energy to solve the problem of standardization of Tibetan new words and achieved positive results. Since 1999, the Office of the Standardization and Validation Committee for Tibetan Words in Qinghai Province has published the “Tibetan Terminology Bulletin” every year. The new words are translated by experts from various universities and related language research centers, and has good normative and authoritative. In the academic research of information on Tibetan new words. In 2011, Jia Jiji of Northwest University for Nationalities compiled more than 5000 new words in the reference books such as the New Word Terminology Dictionary and the Chinese-English New words Dictionary translated by the Chinese National Translation Center. Through the processing of the corpus, a new vocabulary related to

the Tibetan news page is established, which includes 3017 new words, and contains additional information such as part of speech, word length, and word class of the new word [6].

In words of Tibetan new words, the translation of the three major dialect districts in the five provinces is based on their local language habits, such as publishers and information networks.

2.2 New Words Platform

The promotion of new words in English is mainly carried out by various paper and electronic reading materials. At the same time, the dictionary is also an important means of promoting the new words in English. For example, the Webster English Dictionary includes a word or phrase that is widely used within a certain period of time, or is quoted by multiple publications. The editors of the dictionary spend hours reading various paper and electronic versions of the readings every day in order to discover new words and new usages in time. The words that have been marked by the editors will be stored in a computer system. The promotion of Chinese new word words does not seem to have a systematic promotion platform, but it is always spread in China in the form of Weibo, WeChat, Forum, Baidu Encyclopedia and News.

In words of Tibetan, the release of new terminology on the Internet is more limited, most of which are published in the form of pictures in each issue of the Tibetan Terminology Bulletin, they cannot fully display all new words. And there are a few new words published on the Tibetan version of several major websites, such as People's Daily, Xinhuanet Tibetan Channel and China Tibet. The Tibetan Department of Gansu National Normal University has published some new words in Tibetan through the website, but the latest one was published in September 2015, and the website entrance is not easy to find. In addition, the Tibetan-Chinese bilingual website is still a Tibetan learning website that is still running, which is rich in functions, including Tibetan grammar, Tibetan encyclopedia, Tibetan calligraphy and Tibetan vocabulary. But after using, the author found that the site's lexical information is too old. New words such as "WeChat", "Alipay" and "Chinese Dream" have not been found. Even the words "online" and "internet" that are common in Chinese are not included.

3 Preliminaries

In order to better understand the status of Tibetan new words. The author conducted two aspects of research: first, the author conducted a questionnaire of more than 100 Tibetan students in an ethnic university, second, in view of the existing Tibetan new word words, the author counts the word frequency of the mainstream news website in Tibetan in recent years.

The research mainly contains two aspects, one is questionnaires among Tibetan students, and the other is statistical analysis of Tibetan websites. Participants were 104 Tibetan students from an ethnic university. Students are a representative group of Tibetans, they from different regions have different expressions in the use of new words and replace Tibetan expressions with Chinese. Therefore, the investigation of Tibetan

new words in Tibetan students can be used to some extent to understand the use of new words in Tibetan. In addition, in view of the existing Tibetan new words, the author counts the frequency of words in the mainstream news websites of Tibetan in recent years, and analyzes the use of new words.

3.1 Collection of Tibetan New Words

Based on great deal of investigation and material, the author collected a total of 1,841 Tibetan new words from 2011 to 2016, mainly from People’s Daily Online, China Tibetology Network and the National Committee for the Validation of Tibetan New Word Words, as the new words library for Tibetans. The specific composition is shown in Table 1.

Table 1. Number of new words in each year.

Years	2011	2012	2013	2014	2015	2016
Quantity	141	304	819	444	74	59
Total	1841					

3.2 Questionnaire Design

Questionnaire survey is a common method used by researchers to collect data in empirical research. It uses language as the medium and uses strictly designed questions or tables to collect data of research objects [7]. In designing the questionnaire, we followed the relevant theories of questionnaire survey method and consulted senior professors of ethnic linguistics to ensure the rationality and objectivity of questionnaire setting [8, 9].

The rigor of the choice of words. The Tibetan new words involved in the questionnaire mainly comes from the mainstream formal website.

The first part of the questionnaire about the background is mainly to know the influence of the region on the use of Tibetan new words in Tibetan students. The subjective questions are set to know their subjective willingness to some significant problems. The second part is about the familiarity of 70 Tibetan new words in order to know their familiarity with them. The third part of the topic is set to know the current non-standardization of new words.

3.3 Tibetan Corpus from Internet

This study uses the machine method to complete the word frequency statistics of 1841 Tibetan new words. The survey database is a collection of news corpora obtained from four Tibetan websites, including four websites: Qinghai Lake Network, Qinghai Tibetan Broadcasting Network, China Tibet News Network, and Xinhuanet Tibet Channel. The corpus totals 74,284 texts with a size of 568.9. MB, as shown in Table 2.

Table 2. Corpus composition table.

Website name	Number of texts						Size (MB)
	2011	2012	2013	2014	2015	2016	
Qinghai Lake Network	1953	1273	1697	1544	999	451	89.8
China Tibet news network	126	33	3408	31	1564	5439	104
Qinghai Tibetan radio network	3699	1729	4278	12290	17115	13570	351
Xinhuanet Tibet channel	41	188	397	763	1294	402	24.1
Total	74284						568.9

4 Data Analysis

4.1 Questionnaire Data Analysis

Analysis on the Awareness of New Words. The author made statistics on the familiarity degree of 70 new words in Tibetan in the second part of the questionnaire, and divided into five levels according to the degree of familiarity, and the value ranges from 1 to 5 points. The awareness score takes the average of 70 new words scores. The questionnaire survey issued 107 questionnaires and 103 valid questionnaires were retrieved. The Tibetan college students who participated in the questionnaire came from four provinces (Tibet, Gansu, Qinghai and Sichuan) among the five major provinces and regions of Tibetan language in China. As shown in Table 3. It can be seen from the data in the figure that the overall average score of all Tibetan students participating in the questionnaire is 3.64, which indicates that the respondents' overall familiarity with new words in Tibetan is quite good. From a regional point of view, the Tibetan students from the Qinghai region have a relatively high level of cognition, which can reach 3.76. The Tibetan students in Tibetan and Sichuan areas are close to the overall level and are slightly lower. Tibetan students in the Gansu region score relatively lowest.

Table 3. Number of participants from different regions.

	Tibet	Gansu	Qinghai	Sichuan	Total
Number of people	41	13	29	20	103
Average score	3.61	3.48	3.76	3.58	3.61

Analysis of Different Expressions of Tibetan New Words. For a Chinese neologism, a term corresponds to multiple Tibetan expressions, in the third part of the questionnaire, we set 30 questions to investigate the customary expressions of new words. The options include various Tibetan, Chinese and English forms of new words. The statistical results are shown in Table 4. It can be seen that most of the respondents prefer to use the Tibetan form to express new words.

Table 4. The situation of different expressions of new words.

Form	Tibetan	Chinese
Percentage	65.61%	34.39%

Respondents responded to “the use of other ways to express the reasons for the new terminology of Tibetan,” the two highest rates of the reasons for the “cross-regional exchange barriers” and “the lack of unified expression of the language of the nation”, accounting for 35.58%. Both options reflect the existence of non-standardized words in Tibetan new words. Among them, “barriers exist in inter-regional communication” are caused by the non-standard new words of Tibetan, and “the lack of unified expression of national language” is the root cause.

Analysis of Tibetan New Words’ Lagging Condition and the Tibetan University Students’ Expectation. Among respondents, more than 75% of them expected to have the expression of new words in Tibetan within a short time after the emergence of the Chinese new words. However, more than 89% of the respondents believe that there are still more than half a year lags in the expression of words in Tibetan.

Nearly 60% of Tibetan college students in the questionnaire survey learned new words through the Internet, such as chat tools, blogs, and public numbers. This result is in line with the trend that college students can better use the Internet.

Regarding “whether it is expected to have a special platform to learn the words of Tibetan new words”, 89.42% of the fillers chose very much to expect, and 57.6% of fillers expect to obtain the words of Tibetan new words through the website and APP. The specific situation is shown in Table 5.

Table 5. Lag time and publish way of Tibetan new words.

Lag time	Less than half a year	Half a year	1 year	1–2 years	More than 2 years
Percentage	11.54%	31.73%	27.88%	20.19%	8.65%
Tolerance	Synchronization	Short time	Long time	Don’t care	
Percentage	39.05%	38.10%	16.19%	6.67%	

4.2 Internet Data Analysis

Through analysis, among the 1841 Tibetan new words, 18.70% of new words appears in the database text, and 81.30% of new words appear in the text as 0. The specific situation is shown in Table 6 and Fig. 1.

From the statistical analysis of questionnaire results and the 1841 new words comparisons on the four mainstream Tibetan news websites, it can be seen that Tibetan college students have a better overall understanding of the words of Tibetan new words, but there are also some problems. Summarized as follows.

Table 6. Tibetan new words frequency distribution table

Frequency segment	Number of words	Percentage (%)	Examples words
0	1497	81.31	ག་ལྷག་རྒྱུ་མེན་(clenbuterol) ཕྱ་ཚོའི་དངུལ་འབྱོར་(bubble economy)
1-10	184	9.99	རང་འདེམས་ཡུལ་སྐོར་(self-help travel) ནག་དུད་(unregistered citizens)
11-50	78	4.24	བཅོག་སྐྱོན་ཉེང་གཏོང་(energy-saving reduction) ནང་སྐྱེས་སྐུལ་ཤུགས་(endogenous power)
51-100	20	0.92	མཐུན་མོགས་(buddy) ཏིའོ་ཡུལ་སྐོང་ཟུན་(Diaoyu islands)
101-500	37	1.09	ཤུགས་བརྒྱན་པ་(power)
501-1000	8	0.43	ཡུལ་ཚུལ་(tacky)
>=1000	17	1.09	ཡུང་གི་འདུགས་འདུན་(The Chinese dream) ཞིང་(Qin)
total	1841	100	

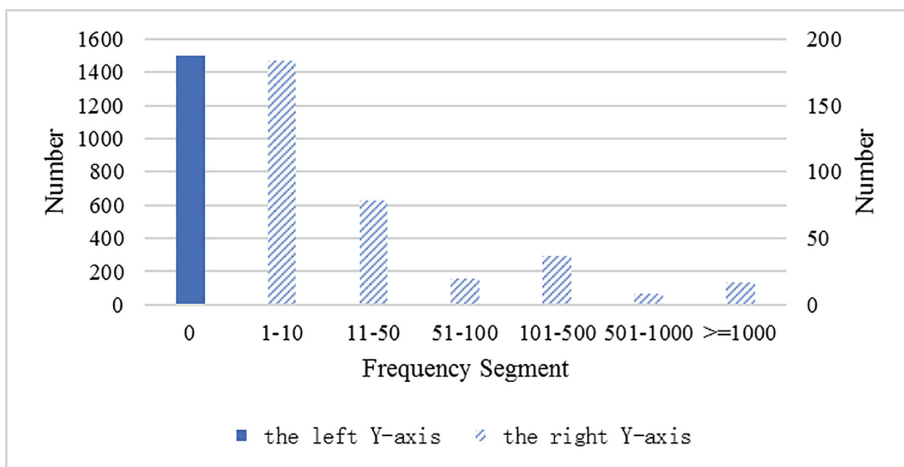


Fig. 1. Tibetan new words frequency distribution chart

First, there is a lag phenomenon in new words in Tibetan. Second, Tibetan students are more inclined to use the Tibetan form to express new words, but there are irregularities in the form of Tibetan expression. Third, the Internet has become a common way for Tibetan college students to obtain new words, and early 60% of Tibetan college students mainly use the Internet to obtain new words. In the survey of expected ways of the new word words promotion platform, the proportion of expressing expectations through the Internet has reached 57.6%.

According to the conclusions of the survey, the author proposes to build a platform for the promotion of Tibetan new words to meet the Tibetan students' learning needs.

5 Design of a Platform for Promoting New Words in Tibetan

5.1 Design Ideas and Principles of the Platform

Design Principles. It is a process of learning new knowledge for Tibetan college students to master new words. In the process, we follow the constructivist theory to build an extension platform that allows Tibetan students to spontaneously learn new words of Tibetan through the promotion platform under the guidance of existing experience and knowledge [10].

At the same time, combining with the theory of communication, the promotion platform is not just promoting the terminology of new words, but also can collect some information feedback from platform learners.

Function of Target System. Tibetan new words promotion platform design considering two parts, first of all, the author analyzes the status quo and main users of the platform for the promotion of new words in Tibetan, second, combining with the theory of constructivism and connectivism knowledge. The overall structure is shown in Fig. 2.

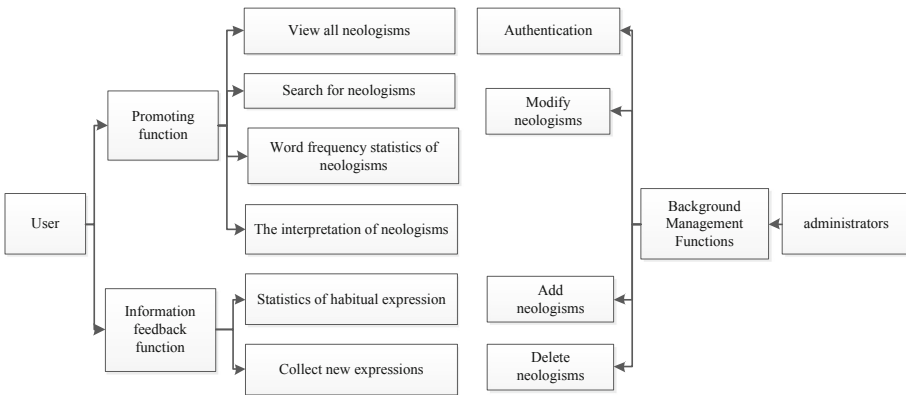


Fig. 2. Promotion platform function char.

- (1) Tibetan new words display function. This function is the core function of the promotion platform and is used to satisfy users' needs for knowing new words in Tibetan. In addition to the basic Tibetan-Chinese counterpart, the platform also has detailed information on the new words, such as explanations, time of occurrence, and so on.

- (2) Search function. This function allows users to query the Tibetan new words based on their own needs. Based on the non-standard phenomenon expressed in words of Tibetan new words, the search function is designed to be an accurate search.
- (3) Information feedback function. For the polysemy term of Tibetan new words, the platform can record the user's choices, at the same time, the user can also supplement the expressions of the new words in Tibetan.
- (4) Network frequency display function. This function provides users with data support for word analysis and research. The platform uses a line chart to display the number of new words in each Tibetan word appearing on the web text in the past five years. At the same time, it provides small function buttons such as screenshots to facilitate users to save data.
- (5) Show all words of Tibetan new words by time. The platform provides the ability to display all new words, and allows users to view the screening criteria at the boundary of the year so that users can understand the Tibetan new words that appear in different years.

5.2 Description of the Website

The page design of the website refers to the style of Baidu Encyclopedia. The home page contains a search box and provides access to popular words, views of all new words, and more. The search results page displays details of new words, including Chinese, Tibetan, time of occurrence, feedback, and Chinese interpretation. In addition, the page also contains the word frequency statistics of the word on the four Tibetan news websites for the past five years.

Promotion Function Module. The homepage display is divided into Chinese and Tibetan versions, which can be achieved by clicking the button in the upper right corner of the page. The page is implemented using PHP code. As shown in Fig. 3. The user clicks the View All New words button to access all new words pages. This page is also divided into Tibetan and Chinese versions. As shown in Fig. 4. After performing the search operation, the user enters the new words detail page, and displays information including Tibetan-Chinese translation, appearance time, information feedback, word interpretation, and word frequency trend information. As shown in Fig. 4.

Information Feedback Function Module. Teaching quality is a key factor influencing student achievement [11]. In the new words details page, the author added an information feedback module. The collection of information including usage habits, the selection of different Tibetan forms, and the use of the frequency in the past five years (2011–2016) on the four news websites (Qinghai Lake Network, Qinghai Tibetan Broadcasting Network, China Tibet News Network, Xinhuanet Tibet Channel). As shown in the Fig. 5.



Fig. 3. Chinese version of the home page.



Fig. 4. All new words display page.



Fig. 5. Search results page.

6 Conclusions

Through a questionnaire survey of Tibetan students in Minzu University of China, and a survey of new words on four Tibetan-language news websites, this paper proposes that there are problems such as irregular expression and time lag in the promotion of Tibetan new words, so this paper build a platform for the promotion of new words in Tibetan, users can view the new words in all directions, and also provide feedback on the platform, which satisfies the Tibetan students' learning needs for Tibetan new words. Meanwhile, it promotes the promotion of Tibetan new words.

The author designed the platform based on the educational theories of constructivism, connectivity, new constructivism, and modern educational technology.

The platform is built on the basis of promoting the new words of Tibetan. In the future operation and improvement, it is necessary to maintain the continuous updating of new words in the database, and supplement the terminology of Tibetan new words from the websites of various reliable sources. First, in the new word interpretation part, we can integrate the materials to add Tibetan interpretation content, so that the platform has better language compatibility. In addition, consider adding a module for video interpretation, using a sound media image to vividly introduce new words to help learners better understand, which is also in line with the theory of situational cognition. Second, based on the completed website, build a new words promotion website for other minority languages. Finally, the platform can be combined with the current hot MOOC online learning to design a video of a new words, which makes the interpretation of the platform's new words more vivid, thus creating a resource-rich online learning platform for new words of ethnic minorities.

Acknowledgments. This work is supported by National Nature Science Foundation (No. 61501529), National Language Committee Project (No. ZDI125-36).

References

1. Qiao, C.: Standardization of Tibetan Neologisms is “an Urgent and Urgent Task” – Interpretation of Comrade Redi’s Speech at the Second Session of the Third Session of the National Committee on Standardization of Tibetan Neologisms, vol. 04. China Tibet Publisher (2014)
2. Zhihe, Y.: Study on the Ontology and Technical Norms of Education Resource Cloud Service-Taking the Construction of Electronic Book Resource Service Technical Specification as an Example. East China Normal University, Shanghai (2012)
3. Wang, S.: Natural language semantic construction based on cloud database. *CMC* **57**(3), 603–619 (2018)
4. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint [arXiv:1408.5882](https://arxiv.org/abs/1408.5882) (2014)
5. Wei, Z.: Five important nodes in the initial stage of Tibetan new word terminology work. *Chin. J. Tibetology* **S1**, 164 (2016)
6. Jia, Y.: Research on New Words in Tibetan News Pages. Northwest University for Nationalities, Lanzhou (2011)
7. Tao, Y.: Precautions in the application of questionnaire method. *China Urban Econ.* **20**, 305 (2011)
8. Hao, X., Yin, Z.: Study on the minority language life in China in the past ten years. *Stud. Ethnic Educ.* **01**(28), 74–80 (2017)
9. Wu, M.: A Survey of Language Use and Language Attitudes of Minority College Students in Beijing. Minzu University of China, Beijing (2007)
10. Zhi, C.: The application of constructivism learning theory in network education. *J. Guangzhou Univ. (Soc. Sci. Ed.)* **11**, 90 (2002)
11. Wang, T., Wu, T., Ashrafzadeh, A.H., He, J.: Crowdsourcing-based framework for teaching quality evaluation and feedback using linguistic 2-tuple. *CMC* **57**(01), 81–96 (2018)