

# **A Quantitative Study on the Linguistic Features of the Creation Epic Hnewo Teyy**

**Yuan Li, Xiaojin Zhang\***

*School of Foreign Languages, North Minzu University, Yinchuan, China*

*\*Corresponding Author.*

**Abstract:** Research on the Yi epic Hnewo teyy has primarily employed qualitative methods, focusing on fields such as translation studies, philosophy, folklore, and poetics. This paper introduces quantitative linguistic methods to examine the linguistic features of the Yi creation epic Hnewo teyy. By analyzing the word frequency distribution characteristics of Hnewo teyy, Zipf's Law is utilized to validate the naturalness of this Yi creation epic. Through clustering experiments comparing Hnewo teyy with modern poetry, regulated poetry, and the folk song "Hua'er," it is found that the narrative style of the epic is more similar to modern poetry and shows a tendency toward ballad-like characteristics. This study addresses methodological gaps in the research of Yi epics, contributing to a more systematic, objective, and scientific approach to the study of ethnic epics.

**Keywords:** Hnewo Teyy; Naturalness; Ballad-Like Characteristics; Quantitative Linguistics

## **1. Introduction**

The Yi people, with their rich history, are an indispensable part of the Chinese nation. As China's sixth largest ethnic minority, they possess the valuable Yi language and script, belonging to the Tibeto-Burman branch of the Sino-Tibetan language family. Based on differences in pronunciation, vocabulary, and grammar, the Yi language is divided into six dialects: northern, eastern, southern, southeastern, western, and central, including 5 sub-dialects and 25 vernaculars. This language is a crucial element of Yi culture, reflecting their production, life, and customs. Through their long historical development, the Yi have developed a distinct cultural identity, with Hnewo teyy being one of its fruitful cultural expressions, holding significant research value.

Hnewo teyy has profoundly influenced Yi culture. As one of the creation epics, it is widely circulated among the Yi and deeply impacts their thoughts, beliefs, literature, arts, and social life. It is not only a cultural treasure of the Yi but also an integral part of Chinese culture. In February 2022, Hnewo teyy was included in Sichuan's list of provincial intangible cultural heritage representative projects, marking another recognition of ethnic minority literature and culture in China. With the growing influence of the Chinese nation, the translations of Hnewo teyy have gained increasing academic attention. Thus, this paper uses the 2016 Chinese translation by Feng Yuanwei as a research corpus.

Recent studies have explored Hnewo teyy from various perspectives, revealing its roots in the spiritual pursuits of the Yi people and their deep contemplation of social life. The epic embodies harmonies consistent with Confucian traditional culture, forming a philosophical foundation for the Yi's future and serving as a cultural link within the Chinese national family. It is regarded as a "genealogy" and a cultural treasure, encompassing history, ethnicity, culture, religion, and customs, with significant potential for international dissemination. The epic metaphorically expresses the formation of a community of blood ties among multiple ethnic groups, reflecting the consciousness of the Chinese national community. Moreover, its numerical symbols not only describe and drive the narrative but also convey textual meanings and cultural attributes. The depiction of industrious female figures holds significant importance for real-life women, and its exploration of water and fire cultures reflects a dialectical philosophy of mutual generation and restriction, embodying a life aesthetics consciousness of virtue and beauty.

The epic's cultural roots allow for perpetual reinterpretations across different historical

contexts, requiring historical interpretation by readers. Its multifaceted functions span literature, folklore, religion, and law, while its artistic authenticity and vividness offer recognition, education, and artistic enjoyment. These studies have made significant contributions to understanding and preserving the Yi epic, yet there remains a need for further linguistic analysis of Hnewo teyy to explore its intrinsic patterns.

Traditionally, studies on ethnic epics have predominantly used qualitative methods, lacking the support of scientific linguistic data. Modern linguistic research emphasizes real language data examination, aligning well with the methods of Quantitative Linguistics. This discipline, both young and ancient, was notably marked by the publication of George Kingsley Zipf's "The Psycho-Biology of Language: An Introduction to Dynamic Philology" in 1935. Quantitative Linguistics uses real language materials and precise mathematical methods to explore language structures and developmental patterns, explaining linguistic phenomena with precision, authenticity, and dynamism, which have gained favor internationally.[1-4]

Research has revealed various findings, such as the inverse relationship between the Gini coefficient and lexical richness, and the descriptive and active characteristics of Romanian poetry. However, focus has been primarily on German and Eastern European languages, with limited quantitative studies on Chinese.[2] Domestic scholars, led by Liu Haitao, have made significant strides over the past fifteen years, utilizing Zipf's Law as a cornerstone. Studies have examined word frequency distributions in modern Chinese poetry, linking them to classical Chinese poetry, and explored noun predominance in language, narrative vs. dialogic styles in novels, and the rich vocabulary and storytelling in the Gesar epic "Huo Ling." [5-7]

Quantitative comparisons of Northwestern Chinese poetry "Hua'er," Chinese, and English poetry have shown "Hua'er" to closely align with Chinese folk songs, highlighting fidelity and variation in translation. Research on Miao ancient songs confirmed their high conformity to Zipf's Law, with narrative styles akin to modern near-form poetry. Studies on Mongolian verbs revealed a long-tail distribution, while analysis of student

compositions identified new parameters for written expression development. Further investigations into word length and polysemy within collaborative linguistics have validated theoretical hypotheses and mathematical models.[9]

The interdisciplinary paradigm of qualitative and quantitative research has been underscored by the word frequency statistics of adverbial folk songs. Studies on "Hua'er" using quantitative linguistic methods and Zipf's Law have validated the intrinsic mechanisms of these texts. Comparisons between human and machine translations have highlighted differences in translation style and quality. Research within the context of the Chinese national community has focused on lexical, syntactic, and semantic features of folk songs from the Yellow River Basin, revealing mixed word order types and thematic consistency linked to the national community. These results collectively affirm the feasibility and scientific validity of this research approach.

As mentioned above, the Northwestern folk song "Hua'er" and the Miao ancient songs highly conform to Zipf's Law. Does the orally transmitted Yi epic Hnewo teyy also conform to Zipf's Law? What are the stylistic connections between Yi oral poetry and different styles of regulated verse and modern poetry? "Hua'er" is a folk ballad, and the Yi epic Hnewo teyy is also an orally transmitted "poem." Do they share similarities in stylistic features? To answer these questions, we will adopt a combination of qualitative and quantitative methods to explore the linguistic features of the Hnewo teyy text, while comparing them with regulated verse, modern poetry, and "Hua'er," providing a new perspective for the study of ethnic oral epics.

## **2. Corpus and Methods**

This study constructs a corpus of Hnewo teyy, using the 2016 Chinese translation by Feng Yuanwei as the source material. The corpus is processed and analyzed using tools such as AntConc, Segment, WordSmith, and Altman Fitter. Hnewo teyy consists of 14 chapters, and this research extracts 10 chapters as the text to explore the word frequency distribution and use the Altman Fitter to test whether the high-frequency words in Hnewo teyy conform to Zipf's Law. Additionally, the author selects 10 regulated verses from the Ancient Poetry

Website, 10 modern poems from the Chinese Poetry Database, and 10 "Hua'er" songs from "The Complete Collection of Chinese Hua'er." Using verbs and nouns as clustering indicators, the author conducts a systematic clustering experiment with Hnewo teyy and regulated and modern poems using SPSS 27.0 software. A clustering experiment is conducted with the

epic Hnewo teyy and "Hua'er," using nouns, pronouns, and verbs as clustering indicators, with data standardized using a proportional method. Finally, the quantitative characteristics of the Hnewo teyy text are described and theoretically interpreted. Basic information on the corpus used in this study is shown in Table 1.

Table 1. Corpus Information

Epic	Regulated Poetry	Modern Poetry	Hua'er
Epic1	Dù Fǔ - Chūn Yè Xī Yǔ1	Xú Zhīmó - Bàn Yè Shēn Xiàng Pí1	Hua'er1
Epic2	Dù Fǔ - Bā Zhèn Tú2	Hú Shì - Xī Wàng2	Hua'er2
Epic3	Dù Fǔ - Jué Jù3	Lǐ Jīnfā - Xīn Yuàn3	Hua'er3
Epic4	Lǐ Bái - Dú Zuò Jīng Tíng Shān4	Ài Qīng - TàiYángYú 4	Hua'er4
Epic5	Lǐ Bái - Yuè Xià Dú ZhuóQí Yī5	Guāngzhōng - Jì Gěi Huà Jiā5	Hua'er5
Epic6	Mèng Jiāo - Quàn Xué6	Dài Wàngshū - Yǔ Xiàng6	Hua'er6
Epic7	Mèng Jiāo - Gǔ Yuè Bié7	Lín Huīyīn - Rén Rán7	Hua'er7
Epic8	Wáng Wéi - Shān Jū Qiū Míng8	Hǎi Zi - Chóng Jiàn Jiā Yuán8	Hua'er8
Epic9	Qí Jǐ - Zǎo Méi9	Duō Duō - Néng Gòu9	Hua'er9
Epic10	Lǐ Hè - Chóng Yì Lǐ Zhì Yǔ10	Xī Chuān - Shàng Dì De Cūn Zhuāng10	Hua'er10

### 3. Results and Discussion

#### 3.1 The Naturalness of the Epic Hnewo Teyy Text

Within the realm of Quantitative Linguistics, two core metrics often utilized for analyzing the linguistic characteristics of literary works are frequency and probability. Frequency refers to the number of times a specific linguistic unit appears in a given text, while probability is calculated as the ratio of the occurrence of this linguistic unit to the total occurrences of all linguistic units in the text. Frequency provides a direct reflection of how often a linguistic unit appears in the text, whereas probability indicates the likelihood of its occurrence. These metrics are closely related to Zipf's Law, which is a significant principle describing the word frequency distribution in natural language, indicating that a word's frequency is inversely proportional to its rank in the frequency table. In other words, words that rank higher appear more frequently, while those ranked lower appear less frequently. The mathematical expression of this law is  $y = ax^{-b}$ , where  $y$  represents the frequency of a word,  $x$  denotes the word's position in the sequence, and  $a$  and  $b$  are parameters to be determined.

Table 2 presents the frequency and probability of the top 30 words in the Hnewo teyy samples, arranged in descending order of frequency

(frequency equals the word frequency count divided by the total word count). Figure 1 illustrates the frequencies of words in the 10 samples of Hnewo teyy arranged in descending order, along with a fitted curve based on the formula of Zipf's Law.

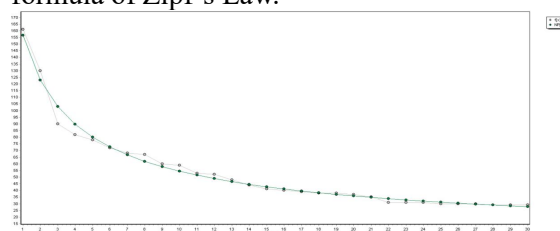


Figure 1. Zipf's Law Fit for High-Frequency Words in Hnewo Teyy

Table 2. Frequency Table of Vocabulary Words in the Top 30 Order of Hnewo Teyy

Rank	Frequency Count	Frequency%	Chinese pinyin
1	161	2.1	de
2	130	1.69	zài
3	90	1.17	le
4	82	1.02	lái
5	78	0.94	zhù
6	72	0.89	a
7	68	0.87	qù
8	67	0.78	yě
9	60	0.78	dào
10	59	0.77	shì
11	53	0.69	difāng
12	52	0.68	bù
13	48	0.63	wǒ
14	44	0.57	zuò

15	41	0.53	zī
16	40	0.52	yòu
17	39	0.51	dì
18	38	0.49	shàng
19	38	0.48	zhè
20	37	0.46	zhàn
21	35	0.46	shuō
22	31	0.40	yào
23	31	0.40	bèi
24	31	0.40	búyuàn
25	30	0.39	hòu
26	30	0.39	yào
27	30	0.39	zhǐ
28	29	0.38	zhōng
29	29	0.38	búshì
30	29	0.35	dà

From Table 2, several observations can be made: Firstly, among the top 30 most frequent words in Hnewo teyy, monosyllabic words account for the majority, with a frequency as high as 90%, while disyllabic words constitute only 10%. This aligns with Zipf's "principle of least effort." The three disyllabic words are "dì fāng" (place), "bù shì" (is not), and "bù yuàn" (unwilling). Secondly, the most frequent word across the 14 stories is "de" (of), with a probability (Pr) of approximately 0.021, which is less than 0.1, consistent with Zipf's later findings. Thirdly, among the high-frequency words, function words dominate, monosyllabic words are most prevalent, disyllabic words are few, and adjectives are least represented. This reflects "Lege Ertai's" nature of being recited rather than sung, where emphasis is often placed on phonetics and rhythm. Such artistic forms demand control over word count and tonal harmony.

Moreover, Table 2 and Figure 1 show that the frequency and rank order of words in the Hnewo teyy samples conform to Zipf's Law, confirming the first hypothesis of this study. The goodness-of-fit coefficient ( $R^2 = 0.9501$ ) indicates a strong fit between the word frequency distribution and Zipf's Law, suggesting that the internal linguistic mechanisms of the epic samples possess "self-organizing" properties. From a quantitative perspective, the distribution of words in Hnewo teyy is neither overly concentrated nor excessively dispersed, indicating its unique formulaic characteristics.

Hnewo teyy, as an oral art form, showcases the Yi people's distinctive artistic charm and

cultural traditions through its rich expression and deep cultural heritage. During the process of transmission, Yi singers often improvise lyrics to reflect the current times, singing scenes, and emotional shifts, an essential aspect of Yi oral art. Meanwhile, certain forms of singing, such as "Ā Dū Gāo Qiāng," maintain stable emotional expressions, adhering to traditional expressive habits, thus embodying the combination of stability and improvisation in oral art.

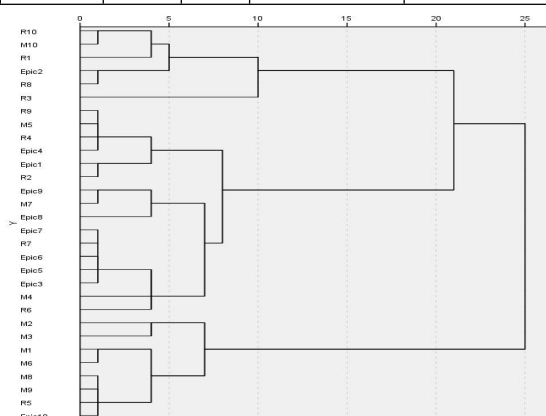
### 3.2. Stylistic Features of Hnewo Teyy

Hnewo teyy, along with modern and regulated verse, belongs to the realm of poetry. This study aims to explore the relationships among them by conducting a clustering experiment using the frequency of nouns and verbs as clustering indicators. Table 3 presents the word frequencies of nouns and verbs in the epic Hnewo teyy, regulated verse, and modern poetry. Figure 2 illustrates the results of the clustering analysis based on the frequency of nouns and verbs in the texts. To facilitate chart creation, we represent the ten regulated verses from Corpus Table 1 with "R," labeled as R1 to R10, and the ten modern poems with "M," labeled as M1 to M10.

**Table 3. Noun and Verb Frequency Table of Hnewo Teyy, Regulated Verse, and Modern Poetry**

Title & Number	Nouns	Verbs	Standardized Verbs	Standardized Verbs
Epic1	42	45	15	15
Epic2	22	15	25	20
Epic3	151	164	15	20
Epic4	66	42	20	15
Epic5	146	197	15	20
Epic6	143	171	15	20
Epic7	58	85	15	20
Epic8	51	71	10	15
Epic9	25	61	10	20
Epic10	154	236	20	30
R1	10	5	30	15
R2	3	3	15	15
R3	5	2	30	10
R4	10	8	20	15
R5	3	5	20	30
R6	5	8	15	25
R7	15	18	15	20
R8	8	5	25	20
R9	7	7	20	15
R10	12	8	25	15

M1	26	32	20	25
M2	7	13	15	30
M3	10	23	15	35
M4	20	22	20	20
M5	46	33	20	15
M6	38	34	20	25
M7	16	25	10	20
M8	15	21	20	30
M9	20	33	20	30
M10	23	12	25	15



**Figure 2. Clustering Analysis Results Based on Noun and Verb Frequencies in Texts**

Firstly, as seen in Figure 2, the clustering experiment using nouns and verbs as indicators completely distinguishes regulated verse from modern poetry, reaffirming the assertion that linguistic structural features can differentiate various literary forms (Huang Wei, Liu Haitao 2009). Secondly, the clustering results indicate that, compared to regulated verse, the Yi epic Hnewo teyy is more similar to modern poetry. Sections like Zhige Alon, Aoshu Bu, and Xuezi Twelve Branches from the epic cluster together with modern poem No. 4 "The Sun" by Ai Qing, modern poem No. 6 "Rain Lane" by Dai Wangshu, and modern poem No. 7 "Still" by Lin Huiyin. The clustering of multiple epics and modern poems within short distances suggests a high similarity in certain features. For instance, "Calling the Sun and Moon" clusters with modern poem No. 7 at the first level, indicating strong similarities. "Shooting the Sun and Moon" and "Summoning the Sun and Moon" form a group with modern poem No. 7 at a short distance, showing similarity. Simultaneously, some epics cluster closely with regulated verses, indicating certain similarities. For example, "The Evolution of Heaven and Earth," a chapter from the Yi creation epic Hnewo teyy, clusters with regulated verse No. 2, Du Fu's "The Eightfold

Formation." "The Genealogy of the Monkey Clan," also part of Hnewo teyy, clusters with regulated verse No. 4, Li Bai's "Sitting Alone on Jingting Mountain," and regulated verse No. 2, "The Eightfold Formation." This section involves Yi family history and legends, similar in theme or style to regulated verse, thus clustering together.

Overall, the Yi epic Hnewo teyy shows a slightly closer proximity to modern poetry than to regulated verse in the clustering diagram. This may be due to greater similarities in the frequency of verb and noun usage and literary characteristics between Hnewo teyy and modern poetry. As a creation epic of the Yi people, Hnewo teyy is rich and grand, primarily transmitted orally, with many various Yi and Chinese handwritten versions. This oral tradition shares certain similarities with the free form and modernity of modern poetry. Additionally, the narrative style of Hnewo teyy is closer to that of modern poetry, which also tends toward direct and clear storytelling.

Despite these similarities with both modern and regulated verse, Hnewo teyy maintains its uniqueness. In the clustering diagram, it does not fully converge with either modern or regulated verse, maintaining a certain distance, which reflects its distinctive cultural background and literary characteristics. The translation and dissemination of Hnewo teyy involve an effort to balance heterogeneity and cross-cultural communication, reflecting a process of understanding, interpreting, and integrating different cultures. This unique cultural and literary value allows Hnewo teyy to retain a degree of independence in its clustering with modern and regulated verse.

In summary, the Yi epic Hnewo teyy clusters more closely with modern poetry while maintaining its distinctiveness.

### 3.3 The Folk Song Characteristics of Hnewo Teyy

The aforementioned author noted that Hnewo teyy exhibits features of oral singing art. To further validate this conclusion, the author conducted a clustering experiment using the frequencies of nouns, verbs, and pronouns in the epic and "Hua'er" as clustering indicators. "Hua'er" is a form of folk song from the northwest region of China. According to Zhang Xiaojin and Liu Haitao (2017), clustering results using nouns, adverbs, and pronouns as

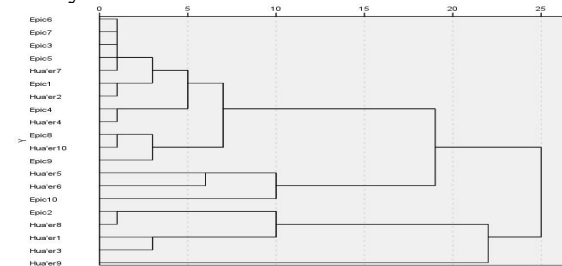
indicators show that the "May Fourth" new poetry has the folk and colloquial characteristics of ballads. Therefore, this study will employ a similar research method to gain deeper insights.

Table 4 presents the frequencies of nouns, verbs, and pronouns in the Yi epic Hnewo teyy and "Hua'er," and Figure 1 shows the results obtained from systematic clustering using SPSS 27.0 based on the data from Table 4.

The arrangement of texts from Hnewo teyy and "Hua'er" demonstrates a clustering trend, indicating that the clustering method based on pronouns, nouns, and verbs can effectively group texts with similar linguistic features. Although the first ten chapters of the epic Hnewo teyy do not cluster closely with the randomly selected ten northwest folk songs "Hua'er," this may be due to Hnewo teyy primarily narrating the Yi people's creation stories, heroic legends, and significant historical events through a grand narrative structure. Its content often involves profound themes such as the origin of the universe, the birth of humanity, and the formation of the nation. In contrast, the northwest ballads "Hua'er" are more about personal emotions, covering topics like love, labor, natural scenery, and everyday anecdotes. These themes are

usually closer to daily life and express inner emotions and reflections on life through vivid language and imagery.[10]

However, some sections in Figure 3 show smaller gaps between the epic and "Hua'er" texts, indicating certain similarities in textual expression when using verbs, nouns, and pronouns as features. For example, Epic5 clusters with "Hua'er"7, Epic1 with "Hua'er"2, Epic4 with "Hua'er"4, and Epic4 with "Hua'er"8. Both rely on oral transmission, spreading and preserving through singing. This mode of transmission endows them with rhythmical characteristics that are easy to memorize and sing. Moreover, both employ a folkloric style of expression, using simple and direct language, vivid metaphors, and straightforward emotional expressions, which easily resonate with the audience.



**Figure 3. Clustering Analysis Results Based on the Frequencies of Nouns, Verbs, and Pronouns in Texts**

**Table 4. Frequency of Nouns, Verbs, and Pronouns in Hnewo teyy and Hua'er**

Title & Number	Nouns	Verbs	Pronouns	Standardized Nouns	Standardized Verbs	Standardized Pronouns
Epic1	42	45	0	15	15	0
Epic2	22	15	1	25	20	1
Epic3	151	164	3	15	20	1
Epic4	66	42	0	20	15	0
Epic5	146	197	11	15	20	2
Epic6	143	171	6	15	20	1
Epic7	58	85	3	15	20	1
Epic8	51	71	2	10	15	1
Epic9	25	61	0	10	20	0
Epic10	154	236	29	20	30	8
Hua'er1	4	1	0	30	15	0
Hua'er2	4	7	1	15	15	3
Hua'er3	6	5	0	30	10	0
Hua'er4	5	5	1	20	15	3
Hua'er5	8	3	0	20	30	0
Hua'er6	7	5	0	15	25	0
Hua'er7	9	11	2	15	20	4
Hua'er8	10	8	0	25	20	0
Hua'er9	8	7	1	20	5	2
Hua'er10	3	7	0	8	17	0

Based on this, it can be concluded that the Yi creation epic Hnewo teyy exhibits certain folk-like oral characteristics. Compared to the northwest folk song "Hua'er," both possess unique content.

#### 4. Conclusion

Within the theoretical framework of Quantitative Linguistics, this paper systematically explores the textual features of the first ten chapters of Hnewo teyy using quantitative linguistic methods. The following conclusions are drawn: In terms of word frequency distribution, the text of the Yi creation epic Hnewo teyy strongly adheres to Zipf's law, once again confirming the inherent mechanistic laws of human language. From the perspective of the probabilistic distribution of text parts of speech, clustering analysis reveals that the narrative of Hnewo teyy is closer to modern poetry, yet it maintains independence between modern and regulated verse. Additionally, it is found that Hnewo teyy exhibits folk-like oral characteristics, suggesting a close connection between epic poetry as a distinct form and folk ballads. This supports the conclusion that poetry largely originates from folk songs and, over time, differentiates from ballads to become an independent literary form.

This study demonstrates that using quantitative linguistic methods to research epics is feasible, scientific, and objective. It also provides a reference perspective for future studies of national epics. The use of quantitative indicators in this paper requires further enrichment, and the author will continue to explore the possibility of using more quantitative linguistic indicators to study epic issues in the future.

#### References

- [1] Xiaojin Zhang, Haitao Liu. Quantitative Characteristics of Chinese Folk Song "Hua'er". *Journal of Ningxia University (Humanities and Social Sciences Edition)*, 2017, 39(05): 76-80+91.
- [2] Haitao Liu, Wei Huang. Current Status, Theory, and Methods of Quantitative Linguistics. *Journal of Zhejiang University (Humanities and Social Sciences Edition)*, 2012(02).
- [3] Xiaxing Pan. *Quantitative Study of Modern Chinese Poetry*. Zhejiang University, 2015.
- [4] Aiyun Wei. *Quantitative Study of the Zhuang Language Lexical System*. Zhejiang University, 2019. DOI:10.27461/d.cnki.gzjdx.2019.002582.
- [5] Duo La, Zhaxi Jia. Analysis of Lexical Quantification and Epic Elements: Using Corpus Methods to Deconstruct the Epic of King Gesar "Holing". *Journal of Tibet University (Social Science Edition)*, 2014, 29(03): 103-110. DOI:10.16249/j.cnki.1005-5738.2014.03.003.
- [6] Yan Hua Bao, Wuyun Ga. Quantitative Analysis of the Distribution Features of Modern Mongolian Verb Meanings. *Journal of Minzu University of China (Natural Sciences Edition)*, 2024, 33(02): 54-61.
- [7] Shuwen Li, Hongbing Xing. Study on the Development of Lexical Richness in Written Expression of Primary and Secondary School Students. *Applied Linguistics*, 2020(1): 121-131.
- [8] Huiyan Xia, Fenglan Sun. Quantitative Linguistic Study on the Relationship Between English Lexical Length and Lexical Meaning. *Foreign Languages and Their Teaching*, 2011(3): 44-49.
- [9] Xiaojin Zhang, Haitao Liu. Quantitative Study of the Linguistic Features of Folk Songs in the Yellow River Basin from the Perspective of the Chinese National Community. *Journal of Northern Nationalities University*, 2024, (02): 111-119. DOI:10.20076/j.cnki.64-1065/G4.2024.02.013.
- [10] Bender M. 2008. "Tribes of Snow": Animals and Plants in the Nuosu "Book of Origins". *Asian Ethnology University: Social Science Edition*, (4):76-77.