Explanations

This file provides explanations for the questions raised in this paper, which are listed below.

1) There is a lack of explanation about the classes of characters to be collected. What is the reason for deciding on 47 classes?

2) Is it appropriate to collect characters written within a square frame?

3) Are the positions and sizes of the characters that make up syllables appropriate?

Detailed Explanations

---

1. There is a lack of explanation about the classes of characters to be collected. What is the reason for deciding on 47 classes?

It is written that the Tibetan language can be expressed using 47 types of characters (vowels, consonants, numbers, symbols), and that syllables are constructed by combining these "characters" (Figure 1). However, the literature [9] states that it includes 238 character classes, and MRG-OHTC [13] states that it includes 910 character classes. These are unlikely to be the number of vowels or consonants, so they can be interpreted as corresponding to the "syllables" in these papers. In other words, there is confusion over the meaning of "character" here. The meaning of the term needs to be clarified so that there are no contradictions.

Furthermore, I think that there is insufficient explanation of the appropriateness of the number of classes (47). Reference [13] states that there are 30 consonant classes and 4 vowel classes, which is consistent with this paper, but is this number of 30+4 standardized? If there is a standard that defines the number of classes, please let us know. Also, this paper adds 10 number classes and 3 symbol classes, but are there any other classes that should be added? Please show us why the author decided that 47 classes is appropriate.

[Response]

(1) Meaning of "character"

The meaning of Tibetan characters does not have a single standardized definition. In some contexts, a character may refer to an alphabet, a numeral, or a composite character made up of multiple elements. In standard files [1][2], characters are categorized into basic characters and extended characters. Extended characters are combinations of several components and typically represent parts of a syllable rather than a complete syllable. For instance, characters ཀྱ, སྐྱ and ཀ are extended characters, which consist

of more than one component such as top vowels, superscripts, base characters, hyphens, and/or bottom vowels (as shown in Figure 1, components 0,2,3,4,5). In References [9] and MRG-OHTC[13] of the paper, both basic and extended characters are included in the datasets. Since extended characters can be derived from basic characters and the challenges and costs associated with offline data collection are considerable, in this work, we chose to focus on basic characters, as defined in the Unicode standard [3].

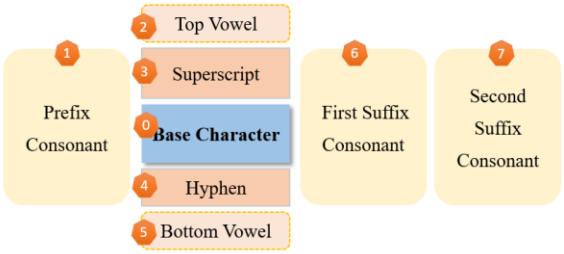


Figure 1 Tibetan Syllable Structure

(2) Explanation of the number of classes (47)

According to Unicode standard [3], Tibetan characters range from 0F00 to 0FFF, totaling 211 available characters. As described in Reference [4], these characters are categorized into Tibetan characters, Sanskrit transliteration Tibetan characters, and other non-text characters, as shown in Figure 2. In this study, we focus only on Tibetan characters, and the reasons for selecting the 47 classes are as follows.

(a) Consonant character. The core set of the Tibetan text consists of 30 consonant characters, as shown in Figure 3. Our dataset includes all of these.

(b) Vowel character. Tibetan characters typically include 9 vowel characters, as shown in Figure 4. Four of them (i.e., ི, ུ, ེ, ོ) are commonly used in Tibetan text. The vowel 'ཻ', a reversed vowel character, and is rarely used in modern Tibetan and is mostly found in ancient Tibetan documents. We analyzed a Tibetan text with a size of around 120MB, which crawls from the website and contains more than 41.67 million characters, and found that 'ཻ' only appears with a ratio of 0.000535%. Additionally, it is a reversed form of 'ི' and can be generated from 'ི'. Another vowel 'ཪ' usually serves as a consonant and has been included in the consonant set. The remaining three vowels are borrowed from Indian and other languages, which are seldom used in Tibetan. In the above Tibetan text with around 120MB, they are appear only 7 times, i.e., a frequency of 1.68e-7. Therefore, we have selected the 4 most commonly used vowels for inclusion in the dataset. While we did not find a standardized definition for the exact number of vowels, many references directly consider Tibetan to have 4 vowels, such as [5] and Reference [13].

(c) Numeral. Tibetan numerals consist of 20 symbols, which include the digits (0-9) and 10 special numeral symbols (half-numbers), as shown in Figure 5. However, the half-number symbols have largely been replaced by the dot (.) in modern Tibetan text. In our analysis of the 200MB text, we found that they do not appear. Therefore, the dataset only includes the 10 numerals (0-9).

(d) Punctuation. Tibetan punctuation includes 3 categories: intersyllabic mark ('·'), intersentence marks (e.g., single vertical stroke '།', double vertical stroke '॥', and quadro-vertical stroke for topic end mark '꠸') and other marks. The quadro-vertical stroke can be composed of two double vertical stroke, and thus it can be generated or recognized based on this. Other punctuation marks are primarily sourced from religion document which usually appear in ancient Tibetan text, and rarely used in modern Tibetan, accounting for only 0.0019% of the characters. Thus, we selected 3 typical punctuation marks, i.e., '·', '།', '॥', in our dataset.

In summary, the characters in our dataset consist of 30 consonants, 4 vowels, 10 numerals and 3 punctuation marks, resulting in 47 classes. Since offline collection of handwritten characters is time-consuming and costly, we selected these character set because it covers the majority of common Tibetan characters in modern Tibetan. To assess this further, we also analyzed around 120MB Tibetan text, and found that these 47 classes account for over 99.9975% of the characters in the text. Furthermore, our future work will include to expand the character dataset to increase more characters.

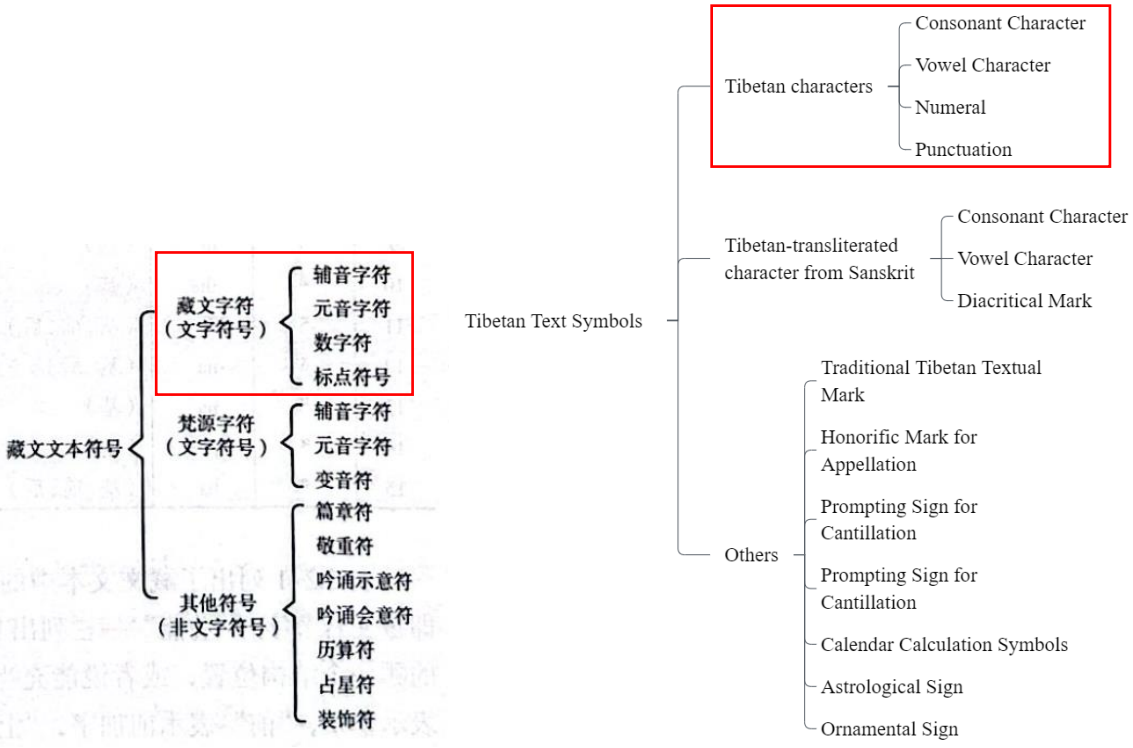


Figure 2 The categories of the Tibetan text characters (Left) and the corresponding English translation (Right)

表 3.2-1 基本字符：藏文 30 个辅音字符

序	符号	转写	功 能	序	符号	转写	功 能
1	ཀ	ka	（基）	16	མ	ma	（基，前，后）
2	ཁ	kha	（基）	17	ཙ	tsa	（基）
3	ག	ga	（基，前，后）	18	ཚ	tsha	（基）
4	ང	nga	（基，后）	19	ཛ	dza	（基）
5	ཅ	ca	（基）	20	ཝ	wa	（基，下）
6	ཆ	cha	（基）	21	ཞ	zha	（基）
7	ཇ	ja	（基）	22	ཟ	za	（基）
8	ཉ	nya	（基）	23	འ	va	（基，前，后）
9	ཏ	ta	（基）	24	ཡ	ya	（基，下）
10	ཐ	tha	（基）	25	ར	ra	（基，上，下，后）
11	ད	da	（基，前，后）	26	ལ	la	（基，上，下，后）
12	ན	na	（基，后）	27	ཤ	sha	（基）
13	པ	pa	（基）	28	ས	sa	（基，上，后）
14	ཕ	pha	（基）	29	ཧ	ha	（基）
15	བ	ba	（基，前，后）	30	ཨ	a	（基）

ID	Character	Transliteration	Function	ID	Character	Transliteration	Function
1	ཀ	ka	base	16	མ	ma	base, prefix, suffix
2	ཁ	kha	base	17	ཙ	tsa	base, prefix, suffix, post-suffix
3	ག	ga	base, prefix, suffix	18	ཚ	tsha	base, prefix, suffix, post-suffix
4	ང	nga	base, suffix	19	ཛ	dza	base, prefix, suffix
5	ཅ	ca	base	20	ཝ	wa	base, prefix, suffix, post-suffix
6	ཆ	cha	base	21	ཞ	zha	base, subscript
7	ཇ	ja	base	22	ཟ	za	base, subscript
8	ཉ	nya	base	23	འ	va	base, subscript
9	ཏ	ta	base	24	ཡ	ya	base, subscript
10	ཐ	tha	base, prefix, suffix	25	ར	ra	base, subscript
11	ད	da	base, prefix, suffix	26	ལ	la	base, subscript
12	ན	na	base, suffix	27	ཤ	sha	base, subscript
13	པ	pa	base	28	ས	sa	base, subscript
14	ཕ	pha	base	29	ཧ	ha	base, subscript
15	བ	ba	base	30	ཨ	a	base, subscript

Figure 3 Consonant characters (Left) and the corresponding English translation (Right)

符号 Character	转写 Translit eration	来源 Source	符号 Character	转写 Translit eration	来源 Source	符号 Character	转写 Translit eration	来源 Source
ཨ	a	本体 inherent	ཇ	e	本体 inherent	ཱ	ā	借用 borrowed
ི	i	本体 inherent	ོ	o	本体 inherent	ཱི	ī	借用 borrowed
ུ	u	本体 inherent	ེ	·i	本体 inherent	ཱུ	ū	借用 borrowed

Figure 4 Vowel characters and the corresponding English translation

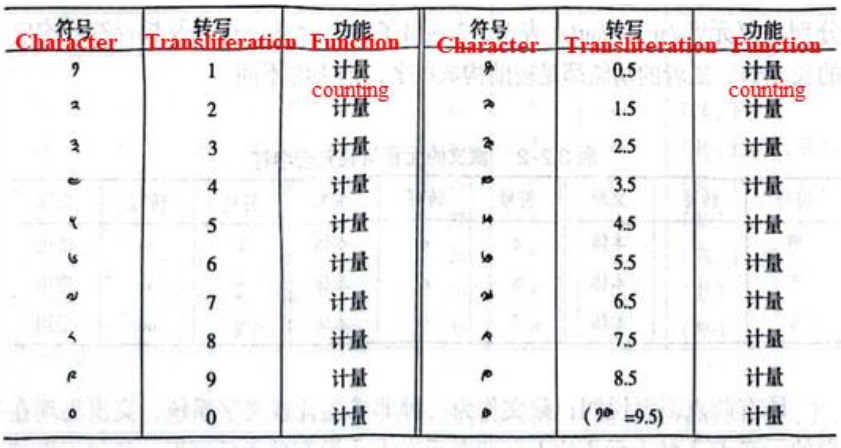


Figure 5 Numeral symbols and the corresponding English translation

Reference

[1] State Administration of Quality and Technical Supervision. GB 22323-2008 information technology, Tibetan coded character set (Basic set and Extended set A). Beijing: China Standards Press; 2008.

[2] State Administration of Quality and Technical Supervision. GB/T 25913-2010 information technology, Tibetan coded character set (Extended set B). Beijing: China Standards Press; 2010.

[3] The Unicode Standard, Version 16.0, https://www.unicode.org/charts/PDF/U0F00.pdf

[4] 江荻, 龙从军. 藏文字符研究: 字母, 读音, 编码, 字频, 排序, 图形, 拉丁字母转写规则研究. 社会科学文献出版社,北京,中国, 2010.

Di Jiang, Congjun Long, "On Characters of Tibetan Writing System." On Characters of Tibetan Writing System: Alphabetic Characters, Pronunciations, ISO Codes, Frequencies, Sorting Orders, Picture Symbols, and Transliterations", Social Sciences Academic Press, Beijing, China, 2010. (in Chinese).

"On Characters of Tibetan Writing System: Alphabetic characters, pronunciations, ISO codes, sorting sequences, picture symbols, and transliterations". Paths Publishing Group; Translation edition, 2017.

[5] 龙从军, 刘汇丹, 安波, 等. 藏文编码字符集标准应用中的问题及对策. 信息技术与标准化, 2016 (1): 46-51.

C. Long, H. Liu, B. An, et al,. "Problems and Improvement Measures of Tibetan Character Coded Set in Application", Standardization Research, pp46-51, 2016

2. Is it appropriate to collect characters written within a square frame?

According to Figure 1, the components (what are referred to as characters in this paper) that make up syllables appear to be arranged in a two-dimensional pattern, and each component has a different position and size. So, was it appropriate to have the participants write the characters within square

frames in the data collection in this study? Referring to Figure 1, there seem to be both vertically and horizontally elongated characters, so the character patterns written within the square frames are not considered to be natural shapes.

There are two possible answers to this question: a) the use of square frames is appropriate, and b) the use of square frames is not optimal, but this method was adopted for this project. Which is the author's opinion? It is necessary to clarify the author's position. If it is a), please explain why you think there is no problem. If it is b), please explain why you adopted this method.

[Response]

This approach may not be optimal in every case, but it is chosen for several practical reasons.

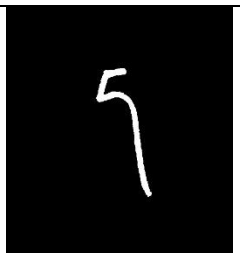
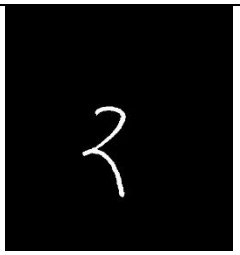
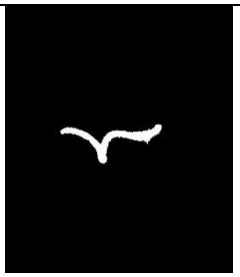
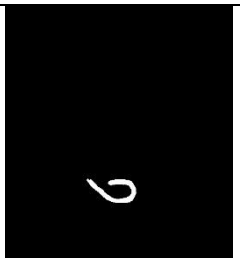
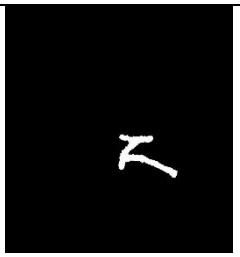
(a) In our data collection process, we split an A4 sheet into a 10*10 grid, with each grid cell measuring 2cm by 2.5cm. This grid size is large enough to accommodate typical handwritten characters. Table 1 provides several character samples processed during data processing, including both vertically (‘ঁ’, ‘ঃ’) and horizontally (‘ঁ’, ‘ঁ’, ‘ঁ’) elongated characters. The second column of Table 1 demonstrates that the size of one grid cell is sufficient for characters of various shapes. Therefore, the use of square frames does not negatively impact the natural appearance of the characters.

(b) From a practical perspective, the grid design is crucial for automating data processing. As described in Section 2.2, this design provides clear boundaries for each handwritten character, which aids in the automatic detection and segmentation of the characters. Once a character is detected, it can be extracted using character edge detection and then removing blank margins, as shown in the third column of Table 1. Thus, the use of the square frames significantly improves the efficiency of dataset construction by enabling automated processing, reducing manual labor and time.

In summary, while the use of square frames may not be optimal for capturing every possible variation of character shapes, it has little impact on most handwritten characters and can reduce the time and manual cost for dataset construction. Therefore, we consider this design to be well-suited to the needs of our study.

Table 6 Character samples during data processing. For clear comparison, the character in the third column is highlighted with a red border, which does not appear in the final dataset.

Character	After grid split	After character extraction

ཅ		
ཇ		
ཉ		
ཏ		
ཐ		
ད		

3. Are the positions and sizes of the characters that make up syllables appropriate?

In this paper, the collected data is combined to synthesize syllable patterns. It is stated that the position, size, and tilt of the components are randomly varied when doing so (Section 3.1). First of all, it is not clear how the position and size of each component within the syllable are determined. Also, it is difficult to believe that a natural pattern can be created by randomly varying the values. The information provided in Section 3.1 alone is insufficient to synthesize the syllable patterns shown in Figure 1.

[Response]:

The synthetic data mentioned in Section 3.1 are pseudo syllables, which are used as augmented training data aimed at improving model performance. The random selection of characters is intended to simulate the variability of handwritten styles and increase the data diversity, thus improving the model's ability to generalize across different styles and handwriting conditions.

Regarding the natural syllable pattern, we will use a dictionary-based approach to construct it. This dictionary consists of over 20,000 syllables, each with a corresponding set of character components and position IDs within the syllable structure. The position ID specifies the location and alignment of each component, as shown in Figure 1. This method will allow us to generate more realistic syllables. As illustrated in Table 2, the position ID for each component is mapped to the corresponding character and its placement within the syllable. Currently, we have constructed a dictionary consisting of 22,976 syllables and their corresponding components, with the position ID still under construction.

Table 7 Examples of syllable and the corresponding component characters

Syllable	Characters	Position ID
འགྲོན་	འ ར ོ ན ་	1 0 2 6
སྒོར་མ	ས ྒ ོ ར ་ མ	0 4 2 6 7
བསྟོད་ཚིག	བ ས ་ ད ོ ་ ཚ ི ག	1 3 0 2 6 7