

Universal Multiple–Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная Организация по Стандартизации

Doc Type: Working Group Document
Title: Final proposal to add 22 characters for Tai Don writing system
Source: Kushim JIANG (姜兆勤)
Status: Individual Contribution
Action: For consideration by UTC
Date: 2022–06–30

This proposal is a revised version of:

- [L2/20–207](#): “Preliminary proposal to encode the Tai Khao script in UCS”
- [L2/22–098](#): “Updated proposal to encode the Tai Don script in UCS”

It incorporates comments made by the UTC Script Ad Hoc Committee and other experts in:

- [L2/21–016](#): “Recommendations to UTC #166 January 2021 on Script Proposals”
- [L2/22–068](#): “Recommendations to UTC #171 April 2022 on Script Proposals”
- [L2/20–208](#): “A response to Kushim JIANG, ‘[Preliminary proposal to encode the Tai Khao script in UCS](#)’”
- [L2/22–099](#): “Comments on ‘[Updated proposal to encode the Tai Don script](#)’”
- Eiso CHAN: “Comments on the Tai Don or Tai Khao encoding”

1 Introduction

There are more than 1.26 million Dai people in China (2010) ([HE, 2014](#)), mainly gathered in the border areas of Xishuangbanna Dai Autonomous Prefecture and Dehong Dai Jingpo Autonomous Prefecture in southwestern Yunnan Province. The Dai ethnicity has its own languages and scripts, and is a multi–dialect and multi–script ethnic ethnicity. The Dai languages are in the Zhuang–Tai languages of the Zhuang–Dong languages of the Sino–Tibetan languages and are divided into four dialects: Xishuangbanna dialect, Dehong dialect, Jinping dialect and Hongjin dialect ([ZHOU & LUO, 2001](#)). And there are several kinds of writing systems used to record the Dai languages in China ([HE & XIONG, 1999](#)).

The Jinping Dai writing system is used to record the Jinping dialect of the Dai languages, and is mainly found in the Dai area of Jinping Miao and Yao Dai Autonomous Prefecture in Yunnan Province, with a population of 20689 (Bureau of Statistics in Jinping, Mar 2019). As the local Dai people call themselves /tăi⁴³ dôn⁵⁵/ or /tăi⁴³ xau³³/, both meaning “White Dai”, the writing system is also known as Tai Don writing system or Tai Khao writing system. The writing system is also used in Vietnam, mainly in Lai Châu Province, Sơn La Province and Điện Biên Province (LAN et al, 2015); in Laos, mainly in Bolikhamsai Province (HAO et al, 2016); and in Thailand.

With the implementation of the Belt and Road strategy, especially the launch of Chinese Language Resources Protection Project by the Ministry of Education and the State Language Commission in 2015, a strong impetus is given to the rescue and protection of minority languages. The Jinping dialect of the Dai languages was included in the 2015 language preservation project and the project was completed. This is the first time that this dialect has been approved as a special research project. In May 2019, with the support of the Jinping County Party Committee and the county government, the Jinping County Cultural Centre hosted the first Jinping Dai writing system standardization seminar, discussing the normalization, standardization and informatization of the Jinping Dai writing system. Experts from the Office of the Steering Committee for Ethnic Minority Languages in Yunnan Province, Yunnan Minzu University and other Dai experts and scholars from Honghe Prefecture reached a consensus on how to solve the problems of irregularities in Jinping Dai writing system and inconsistencies between the written and spoken languages, and initial results were achieved in the standardization process.

2 Script Details

2.1 Script Name

NGÔ & BRASE, 2006 and BRASE, 2007 attempt to construct the Unified Tai character set for use in Vietnam, which leads to the corresponding script name being identified as the Tai Viet script. Specifically, the current character subset includes a complete set of alphabets for the Tai Dam writing system, with other three pairs of consonant letters used in the Tai Don writing system. The authors of the two aforementioned documents consider the current subset to be phonetically adequate for the Tai Don language, but still formally incomplete. This argument suggests that the content of the Jinping Dai writing system (Tai Don writing system) should be integrated into the Tai Viet script class, and that the relevant characters should be added to the Tai Viet block.

According to the results of the intercharacter relationship analysis, it is reasonable to group the characters used in the Tai Don writing system with those used in the Tai Dam writing system. However, the demands

for use from China are not compatible with the name Tai Viet. Therefore, Tai Don–Tai Dam script or Tai Dam–Tai Don script is a reasonable alternative name, which should also lead to a change in the block name.

2.2 Encoding Model

The existing encoding model is a graphetic model based on visual order, which is compatible with the soon-to-be-completed Tai Don writing system. Therefore, the character will be added either as a base character or as a combining mark and no rearrangement will occur during shaping.

2.3 Character Name

The character name consists of the writing system name with the Latinization form corresponding to the character. There is currently no Latinization scheme for Tai Don language in use. This proposal gives a Latinization scheme modified from the Latinization scheme of New Tai Lue script showing below.

/b/	/ʔ/	/d/	/ts/	/k/	/x/	/kʰ/	/h/	/t/	/tʰ/	/p/
B	Q	D	TS	K	X	KH	H	T	TH	P
/pʰ/	/tʰ/	/f/	/s/	/n/	/m/	/l/	/ŋ/	/v/	/j/	/ɲ/
PH	TSH	F	S	N	M	L	NG	V	Y	NY
/ǎ/	/a/	/i/	/e/	/ɛ/	/ɔ/	/o/	/u/	/ə/	/u/	
A	AA	I	EE	EH	O	OO	U	E	UE	

2.4 Consonants

See [Figure 1](#) and [Figure 2](#) for a list of 44 consonant letters (numbered 1 to 44) used in Tai Dam writing system and Tai Don writing system. The images are from *Bảng chữ Thái* quoted by [CHAN, 2022](#). In two tables, the fourth column contains the consonants in the Tai Dam writing system, with the corresponding codepoints marked yellow on the left. From the 37th letter to the 42nd letter are consonants for Tai Don language, so the codepoints on the left are enclosed with square brackets.

The fifth to ninth columns give the consonants in Tai Don writing systems used in Mường Lay (Điện Biên), Phong Thổ (Lai Châu), Phú Yên (Sơn La), Mộc Châu (Sơn La) and Jinping (Yunnan) respectively. For Jinping writing system, the unencoded characters are marked in red and the corresponding codepoints are marked in yellow on the right. See [Table 1](#) for unencoded consonants in Jinping Dai writing system.

Số thứ tự	Tên chữ	Thống nhất	Thái đen	Mường Lay	Phong Thổ	Phù Yên	Mộc Châu	Jin Ping	
1	Bơ	AA9A	✓	✓	✓	✓	✓	✓	AA9A
2	Bồ	AA9B	ư	ư	ư	ư	ư	ư	AA9B
3	Pơ	AA9C	✓	✓	✓	✓	✓	✓	AA9C
4	Pồ	AA9D	ư	ư	ư	ư	ư	ư	
5	Đơ	AAA4	ơ	ơ	ư	ơ	ơ	ư	
6	Đồ	AAA5	ơ	ơ	ơ	ơ	ơ	✓	AAA5
7	Sơ	AA8E	x	ư	ư	x	x	ư	
8	Sồ	AA8F	ư	ư	ư	ư	ư	ư	
9	Phơ	AA9E	ư	ư	ư	ư	ư	ư	AA9E
10	Phồ	AAA1	ư.ư	ư	ư	ư	ư	ư	AAA1
11	no	AA98	ư	ư	ư	ư	ư	ư	AA98
12	áo	AA99	ư	ư	ư	ư	ư	ư	AA99
13	mo	AAA2	ư	ư	ư	ư	ư	ư	
14	mồ	AAA3	ư	ư	ư	ư	ư	ư	AA9D
15	ngơ	AA88	ư	ư	ư	ư	ư	ư	
16	ngồ	AA89	ơ	ơ	ơ	ơ	ơ	ơ	AA89
17	ươ	AAAA	ư	ư	ư	ư	ư	ư	AAAA
18	ưồ	AAAB	ơ	ơ	ơ	ơ	ơ	ơ	AAAB
19	to	AA94	ư	ư	ư	ư	ư	ư	AA94
20	tồ	AA95	ư	ư	ư	ư	ư	ư	AA95
21	Kơ	AA80	ư	ư	ư	ư	ư	ư	AA80
22	Kồ	AA81	ư	ư	ư	x	ư	ư	
23	Hơ	AAAC	ư	ư	ư	ư	ư	ư	AAAC
24	Hồ	AAAD	ư	ư	ư	ư	ư	ư	AAAD

Figure 1 Consonants in Tai Don–Tai Dam script (*Bảng chữ Thái*)

Số thứ tự	Tên chữ	Thống nhất	Thái đen	Mường Lay	Phong Thổ	Phù Yên	Mộc Châu	Jin Ping	
25	khô	AA84	ㄴ	ㄴ	ㅈ	ㅈ	ㅈ	ㅈ	AA84
26	khô	AA85	ㄴ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA85
27	đo	AA92	ㄴ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	
28	đo	AA93	ㄴ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA93
29	lo	AAA8	ㄴ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AAA8
30	lô	AAA9	ㄴ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AAA9
31	o	(AAAE)	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AAAE
32	ố	(AAAF)	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AAAF
33	tho	AA96	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA96
34	tho	AA97	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA97
35	nhô	AA98	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	
36	nhô	AA91	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	
37	kh'	[AA82]	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA82
38	kh'	[AA83]	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA83
39	pho	[AAA0]	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AAA0
40	phô	[AA9F]	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	
41	êo	[AA8C]	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA8C
42	êo	[AA8D]	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA8D
43	êho	AA8A	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA8A
44	êhố	AA8B	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	AA8B
45	quing	AADC	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	
46	kôn	AADB	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	ㅈ	

Figure 2 Consonants in Tai Don–Tai Dam script (*Bảng chữ Thái*)

The unification process, which is also a graphetic unification process, is based on [BRASE, 2008](#), and the result in this proposal is broadly similar to the reference. The differences come from the standardized glyphs from Jinping, stated in [DAO et al, 2019](#), selected in the standardization meeting held in Jinping County, influencing the unification decision. All these differences are:

- The characters 𑜋 and 𑜌 are uncertain whether to unify to 𑜊 and 𑜍 in [BRASE, 2008](#), and disunified to 𑜊 and 𑜍 in this proposal (see [Figure 3](#)). 𑜊 comes from the x shape, 𑜍 comes from the bent 𑜌 shape, but 𑜋 and 𑜌 come from turned 𑜊 shape. So, their cognition is different from each other.

TAI VIET LETTER LOW SO	𑜊	𑜋	/s1/	𑜌 Glyph 16	/s1/	𑜍 Glyph 16	/s1/	𑜎 Glyph 16	/s1/	𑜏 Glyph 16	/s1/
TAI VIET LETTER HIGH SO	𑜐	𑜑	/s4/	𑜒 Glyph 17	/s4/	𑜓 Glyph 17	/s4/	𑜔 Glyph 17	/s4/	𑜕 Glyph 17	/s4/

Figure 3 Consonants in Tai Don–Tai Dam script

- The characters 𑜖 is unified to 𑜗 in [BRASE, 2008](#), but disunified to 𑜗 in this proposal (see [Figure 4](#)). 𑜗 comes from the gyrating 𑜘 shape, but 𑜖 comes from 𑜙 shape, which is disunified from 𑜚 in [BRASE, 2008](#), so their cognition is different from each other.

TAI VIET LETTER LOW NGO	𑜛	𑜜	/ŋ1/	𑜝^{**}	/ŋ1/	𑜞^{**}	/ŋ1/	𑜟^{**}	/ŋ1/	𑜠^{**}	/ŋ1/
-------------------------	------------	------------	------	-----------------	------	-----------------	------	-----------------	------	-----------------	------

TAI VIET LETTER LO DO	𑜣	𑜤	/d1/	—	—	—	—	—	—	—	—
New char based on GLYPH 25 (*TAI DON LETTER LOW DO)	—	—	—	𑜥	/d1/	𑜦	/d1/	𑜧	/d1/	𑜨	/d1/

Figure 4 Consonants in Tai Don–Tai Dam script

- The characters 𑜩 is disunified to 𑜪 in [BRASE, 2008](#), but unified to 𑜪 in this proposal (see [Figure 5](#)). [BRASE, 2008](#) shows 𑜫 shape for the character, which is actually different from the standardized shape determined by Jinping, so it is reasonable to unify them.

New char based on GLYPH 18 (*TAI DON LETTER HIGH NYO)	—	—	—	ᵛ	/n ^h 4/	—	—	—	—	ᵛ	/k ^h 4/
--	---	---	---	---	--------------------	---	---	---	---	---	--------------------

38	ꨀ	[AA83]	—	ᵛ	ᵛ	—	—	ᵛ	AA83
----	---	--------	---	---	---	---	---	---	------

Figure 5 Consonants in Tai Don–Tai Dam script

- The characters ᵛ is unified to ᵛ in [BRASE, 2008](#), but disunified to ᵛ in this proposal (see [Figure 6](#)). ᵛ comes from a tailed ᵛ, while ᵛ comes from a tailed ᵛ, which is disunified to ᵛ in [BRASE, 2008](#), so their cognition is different from each other.

TAI VIET LETTER HIGH PHO	ᵛ	ᵛ	/p ^h 4/	ᵛ	/p ^h 4/	ᵛ	/p ^h 4/	ᵛ	/p ^h 4/	ᵛ**	/p ^h 4/
--------------------------	---	---	--------------------	---	--------------------	---	--------------------	---	--------------------	-----	--------------------

40	ꨀ	[AA9F]	ᵛ	ᵛ	—	ᵛ	ᵛ	ᵛ
----	---	--------	---	---	---	---	---	---

Figure 6 Consonants in Tai Don–Tai Dam script

Table 1 Unencoded consonant characters in Jinping Dai writing system

Codepoint	Glyph	Character Name	Value
AAC3	ᵛ	TAI DON LETTER HIGH KO	/k ₂ /
AAC4	ᵛ	TAI DON LETTER LOW NGO	/ŋ ₁ /
AAC5	ᵛ	TAI DON LETTER LOW SO	/s ₁ /
AAC6	ᵛ	TAI DON LETTER HIGH SO	/s ₂ /
AAC7	ᵛ	TAI DON LETTER LOW YO	/j ₁ /
AAC8	ᵛ	TAI DON LETTER HIGH PO	/p ₂ /
AAC9	ᵛ	TAI DON LETTER HIGH PHO	/p ^h ₂ /
AACA	ᵛ	TAI DON LETTER LOW MO	/m ₁ /
AACB	ᵛ	TAI DON LETTER LOW DO	/d ₁ /
AACC	ᵛ	TAI DON LETTER LOW NYO	/n ₁ /
AACD	ᵛ	TAI DON LETTER HIGH NYO	/n ₂ /

2.5 Vowels

See [Figure 7](#) for a list of 18 finals (numbered 1 to 18) used in Tai Dam writing system and Tai Don writing system. The image is also from *Bảng chữ Thái*. The fourth column contains the consonants in the Tai Dam writing system, with the corresponding codepoints marked yellow on the left. The fifth to eighth columns give the consonants in Tai Don writing systems used in Phong Thổ (Lai Châu), Mộc Châu (Sơn La), Yên Châu (Sơn La) and Jinping (Yunnan) respectively. For Jinping writing system, the unencoded characters are marked in red and the corresponding codepoints are marked in yellow on the right. The unification process is identical to that of [BRASE, 2008](#), that is, there are four unencoded characters. See [Table 2](#) for unencoded vowels in Jinping Dai writing system.

Table 2 Unencoded consonant characters in Jinping Dai writing system

Codepoint	Glyph	Character Name	Value
AACE	⦿	TAI DON MAAI KAA	/a/
AACF	⦿	TAI DON MAAI TSAM	/ɔ/
AADO	✂	TAI DON MAAI KAT	/at/
AADI	⦿	TAI DON MAAI KU	/u/

2.6 Tones

Jinping Dai writing system use U+AABF ⦿ for tone 2 and tone 5, and U+AAC1 ⦿ for tone 3 and tone 6. [DAO et al, 2019](#) also mentions that, the /-kʰ/ with *rusheng* tone becomes /-ʔ/ and it is necessary to introduce text element to represent the glottal stopper, so ⦿ should be encoded as a Tai Don mark. See [Table 3](#) for the unencoded tone mark and [Figure 8\(a\)](#), [Figure 9\(d\)](#) for the evidence.

Table 3 Unencoded consonant characters in Jinping Dai writing system

Codepoint	Glyph	Character Name	Value
AAD2	⦿	TAI DON MAAI QAA	ʔ


Số thứ tự	Tên may	Chữ Thống nhất	Vùng Thái đen	Lai Châu Phong thổ	Mộc Châu	Yên Châu	Jin Ping
1	khít	AAB7	✓	huy	✓	✓	○ AAB7
2	ca	AAB1	2	2 huy	2	2	○ AAB1
3	cay	AABC	✓	✓	✓	✓	✓ AABC
4	ke	AAB5	✓	✓	✓	✓	✓ AAB5
5	củi	AABB	9	9	9 (ư)	9 (ư)	9 AABB
6	cắm	AABE	✓	✓	✓	✓	○ AABE
7	caw	AAB9 AAB9	✓	✓	✓	✓	6-6 AAB9 AAB9
8	căn	AABD	✓	✓	✓	✓	✓ AABD
9	cua	AABA	✓	✓ (ư)	✓	✓	✓ AABA
10	cổng	AAB6	✓	✓ (ư)	✓	✓	✓ AAB6
11	cu	AAB4	✓	✓	✓	✓	○ AAB4
12	của	AAB9	✓	✓ (ư)	✓	✓	6 AAB9
13	cổ	✓	✓	✓	✓	✓	
14	củ	AAB3	✓	✓	✓	✓	○ AAB3
15	kia	AAB8	✓	✓ (ư)	✓	✓	○ AAB8
16	Ky	AAB2	✓	✓	✓	✓	○ AAB2
17	cảng	AAB0	✓	✓	✓	✓	○ AAB0
18	cát			✓			✓
19	quy	✓	✓	✓	✓	✓	
20	que	✓	✓	✓	✓	✓	
21	quô	✓	✓	✓	✓	✓	
22	quya	✓	✓	✓	✓	✓	

Figure 7 Vowels in Tai Don–Tai Dam script (*Bảng chữ Thái*)

3 Character Properties

AAC3;TAI	DON	LETTER	HIGH	KO;Lo;0;L;;;;N;;;;;
AAC4;TAI	DON	LETTER	LOW	NGO;Lo;0;L;;;;N;;;;;
AAC5;TAI	DON	LETTER	LOW	SO;Lo;0;L;;;;N;;;;;
AAC6;TAI	DON	LETTER	HIGH	SO;Lo;0;L;;;;N;;;;;
AAC7;TAI	DON	LETTER	LOW	YO;Lo;0;L;;;;N;;;;;
AAC8;TAI	DON	LETTER	HIGH	PO;Lo;0;L;;;;N;;;;;
AAC9;TAI	DON	LETTER	HIGH	PHO;Lo;0;L;;;;N;;;;;
AACA;TAI	DON	LETTER	LOW	MO;Lo;0;L;;;;N;;;;;
AACB;TAI	DON	LETTER	LOW	DO;Lo;0;L;;;;N;;;;;
AACC;TAI	DON	LETTER	LOW	NYO;Lo;0;L;;;;N;;;;;
AACD;TAI	DON	LETTER	HIGH	NYO;Lo;0;L;;;;N;;;;;
AACE;TAI	DON	MAAI	KAA;Mn;220;NSM;;;;N;;;;;	
AACF;TAI	DON	MAAI	TSAM;Mn;220;NSM;;;;N;;;;;	
AAD0;TAI	DON	MAAI	KAT;Lo;0;L;;;;N;;;;;	
AAD1;TAI	DON	MAAI	KU;Mn;230;NSM;;;;N;;;;;	
AAD2;TAI	DON	MAAI	QAQ;Mn;230;NSM;;;;N;;;;;	
AAD3;TAI	DON	SYMBOL	BON;Lo;0;L;;;;N;;;;;	
AAD4;TAI	DON	SYMBOL	MAA;Lo;0;L;;;;N;;;;;	
AAD7;TAI	DON	SYMBOL	MAAI	MEHT;Po;0;L;;;;N;;;;;
AAD8;TAI	DON	SYMBOL	MAAI	SAM-1;Po;0;L;;;;N;;;;;
AAD9;TAI	DON	SYMBOL	MAAI	SAM-2;Po;0;L;;;;N;;;;;
AADA;TAI	DON	SYMBOL	MAAI	SAM-3;Po;0;L;;;;N;;;;;

Reference

- Jim BRASE. 2007. *Tai Viet Script for Unicode*. [L2/07-039R](#).
- Jim BRASE. 2008. *Writing Tai Don: Additional characters needed for the Tai Viet script*. [L2/08-217](#).
- EISO CHAN 陈永聪. 2022. *Comments on the Tai Don or Tai Khao encoding*. Not posted yet.
- DAO Jie 刀洁, ZHOU Yan 周焱, WANG Yuqing 王玉清. 2019. "The standardization of the phonetics of Jinping Dai writing system in Yunnan 云南金平傣文语音的规范化问题". In *Language Plannology Studies* 语言规划学研究, No. 2: pp. 57-68.
- HAO Yong 郝勇, HUANG Yong 黄勇, QIN Hailun 覃海伦. 2016. *An Introduction to Lao* 老挝概论. World Publishing Guangdong Co., Ltd 世界图书出版广东有限公司.
- HE Lifeng 和丽峰, XIONG Yuyou 熊玉有. 1999. *Outline of the Scripts of Yunnan's Ethnic Minorities* 云南少数民族文字概要. Yunnan Ethnic Publishing House 云南民族出版社.
- HE Shaolin 何少林. 2014. *Dictionary of Chinese Ethnic Minorities: Dai* 中国少数民族大辞典·傣族卷. Yunnan Ethnic Publishing House 云南民族出版社.
- LAN Qiang 兰强, XU Fangyu 徐方宇, LI Huajie 李华杰. 2015. *An Introduction to Vietnam* 越南概论. World Publishing Guangdong Co., Ltd 世界图书出版广东有限公司.
- LUO Wenbin 罗文斌 (chief). 2017.  金平傣文教程. Cultural Center of Jinping Miao, Yao and Dai Autonomous County 金平苗族瑶族傣族自治县文化馆.
- Ngô Trung Việt, Jim BRASE. 2006. *Unified Tai Script for Unicode*. [L2/06-041](#).
- ZHOU Yaowen 周耀文, LUO Meizhen 罗美珍. 2001. *Research on Dai Languages and Dialects* 傣语方言研究. The Ethnic Publishing House 民族出版社.

Appendix A Normalization of the Phonology of Jinping Dai writing system

DAO *et al*, 2019 describes the standardized Jinping Dai orthography, shown below.

A.1 Normalization of initials

There are 44 initials in Jinping Dai writing system, traditionally known as ᐃᐃᐃ ᐃᐃᐃ [to³³ mɛ⁵²] (“mother character”), divided into the high level group (ᐃᐃᐃ ᐃᐃᐃ [to³³ bau³³], “light character”) and the low level group (ᐃᐃᐃ ᐃᐃᐃ [to³³ nak⁵⁵], “heavy character”). There are 22 phonemes, traditionally arranged in the following order:

ᐃ	[ʔo ³³]	ᐃ AAAE	ᐃ	[p ^h o ³³]	ᐃ AAAO
ᐃ	[ʔo ⁴³]	ᐃ AAAF	ᐃ	[p ^h o ⁴³]	ᐃ AAC9
ᐃ	[ko ⁴³]	ᐃ AA8O	ᐃ	[mo ³³]	ᐃ AACA
ᐃ	[ko ⁴³]	ᐃ AAC3	ᐃ	[mo ⁴³]	ᐃ AA9D
ᐃ	[xo ³³]	ᐃ AA84	ᐃ	[fo ³³]	ᐃ AA9E
ᐃ	[xo ⁴³]	ᐃ AA85	ᐃ	[fo ⁴³]	ᐃ AAAI
ᐃ	[ɲo ³³]	ᐃ AAC4	ᐃ	[vo ³³]	ᐃ AAAA
ᐃ	[ɲo ⁴³]	ᐃ AA89	ᐃ	[vo ⁴³]	ᐃ AAAB
ᐃ	[tso ³³]	ᐃ AA8A	ᐃ	[lo ³³]	ᐃ AAA8
ᐃ	[tso ⁴³]	ᐃ AA8B	ᐃ	[lo ⁴³]	ᐃ AAA9
ᐃ	[so ³³]	ᐃ AAC5	ᐃ	[ho ³³]	ᐃ AAAC
ᐃ	[so ⁴³]	ᐃ AAC6	ᐃ	[ho ⁴³]	ᐃ AAAD
ᐃ	[jo ³³]	ᐃ AAC7	ᐃ	[do ³³]	ᐃ AACB
ᐃ	[jo ⁴³]	ᐃ AAA5	ᐃ	[do ⁴³]	ᐃ AA93
ᐃ	[to ³³]	ᐃ AA94	ᐃ	[bo ³³]	ᐃ AA9A
ᐃ	[to ⁴³]	ᐃ AA95	ᐃ	[bo ⁴³]	ᐃ AA9B
ᐃ	[t ^h o ³³]	ᐃ AA96	ᐃ	[k ^h o ³³]	ᐃ AA82
ᐃ	[t ^h o ⁴³]	ᐃ AA97	ᐃ	[k ^h o ⁴³]	ᐃ AA83
ᐃ	[no ³³]	ᐃ AA98	ᐃ	[ts ^h o ³³]	ᐃ AA8C
ᐃ	[no ⁴³]	ᐃ AA99	ᐃ	[ts ^h o ⁴³]	ᐃ AA8D
ᐃ	[po ³³]	ᐃ AA9C	ᐃ	[no ³³]	ᐃ AACC
ᐃ	[po ⁴³]	ᐃ AAC8	ᐃ	[no ⁴³]	ᐃ AACD

With the introduction of Chinese loanwords, round-labialized phonemes that had disappeared have reappeared, and a number of phonemes have been added that were not present in the original Jinping Dai language, including:

$n\text{ɔ}$	$[k^{w\text{ɔ}^{43}}]$	n_{AA80}	ɔ_{AAAB}	$\check{v}\text{ɔ}$	$[ts^{w\text{ɔ}^{33}}]$	\check{v}_{AA8A}	ɔ_{AAAB}
$n\text{ɔ}$	$[k^{w\text{ɔ}^{43}}]$	n_{AAC3}	ɔ_{AAAB}	$\eta\text{ɔ}$	$[ts^{w\text{ɔ}^{43}}]$	η_{AA8B}	ɔ_{AAAB}
$\text{ɛ}\text{ɔ}$	$[x^{w\text{ɔ}^{33}}]$	ɛ_{AA84}	ɔ_{AAAB}	$\text{ɛ}\text{ɔ}$	$[k^{hw\text{ɔ}^{33}}]$	ɛ_{AA82}	ɔ_{AAAB}
$\eta\text{ɔ}$	$[x^{w\text{ɔ}^{43}}]$	η_{AA85}	ɔ_{AAAB}	$w\text{ɔ}$	$[k^{hw\text{ɔ}^{43}}]$	w_{AA83}	ɔ_{AAAB}

A.2 Normalization of finals

There are 83 finals in Jinping Dai writing system, including 18 $n\text{ɔ}w$ $“\text{ɔ}”$ $[to^{33} ma:i^{33}]$ (“symbols”) and 65 $n\text{ɔ}w$ $“\text{ɔ}”$ $[to^{33} p^hwi^{55}]$ (“matching characters”), they are (Take n as the position of the consonant):

Monophthong (9 in 11 shapes)

$n\text{ɔ}$	$[a]$	n_{AA80}	ɔ_{AABI}
ɔ	$[a^{(25)}]$	ɔ_{AAAC}	ɔ_{AAAE}
\check{n}	$[i]$	n_{AA80}	ɔ_{AAB2}
\check{n}	$[e]$	n_{AA80}	ɔ_{AAB8}
$“n$	$[\varepsilon]$	$“_{AAB5}$	n_{AA80}
$n\text{ɔ}$	$[\text{ɔ}]$	n_{AA80}	ɔ_{AAAE}
$n\text{ɔ}$	$[\text{ɔ}^{(43)}]$	n_{AAC8}	ɔ_{AACF}
$n\text{ɔ}$	$[o]$	n_{AA80}	ɔ_{AABA}
\check{n}	$[u]$	n_{AA80}	ɔ_{AAB4}
\check{n}	$[u]$	n_{AA80}	ɔ_{AAB3}
\check{n}	$[\text{ə}]$	n_{AA80}	ɔ_{AAB7}

Composite Phoneme (9 in 9 shapes)

$\text{ɛ}n$	$[ai]$	ɛ_{AABC}	n_{AA80}
$\text{ɛ}n$	$[au]$	ɛ_{AABB}	n_{AA80}
$\text{ɛ}n$	$[au]$	ɛ_{AAB9}	n_{AA80} ɛ_{AAB9}
$\check{n}\text{ɔ}$	$[a\eta]$	n_{AA80}	ɔ_{AABO} ɔ_{AA89}
$n\text{ɔ}$	$[an]$	n_{AA80}	ɔ_{AABD}
$\check{n}w$	$[am]$	n_{AA80}	ɔ_{AABE} w_{AAAB}
$\check{n}n$	$[ak]$	n_{AA80}	ɔ_{AABO} n_{AA80}

$n\text{ɔ}$	$[at]$	n_{AA80}	ɔ_{AADO}
$\check{n}v$	$[ap]$	n_{AA80}	ɔ_{AABE} v_{AA9A}

Matching Final (65 in 67 shapes)



















































$n\text{ɔ}$	$[a:i]$	n_{AA80}	ɔ_{AABI} \check{v}_{AAA5}
$n\text{ɔ}$	$[ɔi]$	n_{AA80}	ɔ_{AAAE} \check{v}_{AAA5}
$n\text{ɔ}$	$[oi]$	n_{AA80}	ɔ_{AABA} \check{v}_{AAA5}
$n\text{ɔ}$	$[ui]$	n_{AA80}	ɔ_{AAB4} \check{v}_{AAA5}
$\check{n}v$	$[ui]$	n_{AA80}	ɔ_{AAB3} \check{v}_{AAA5}
$\check{n}v$	$[əi]$	n_{AA80}	ɔ_{AAB7} \check{v}_{AAA5}
$n\text{ɔ}$	$[a:u]$	n_{AA80}	ɔ_{AABI} ɔ_{AAAB}
$\hat{n}\text{ɔ}$	$[iu]$	n_{AA80}	ɔ_{AAB2} ɔ_{AAAB}
$\check{n}\text{ɔ}$	$[eu]$	n_{AA80}	ɔ_{AAB8} ɔ_{AAAB}
$“n\text{ɔ}$	$[\varepsilon u]$	$“_{AAB5}$	n_{AA80} ɔ_{AAAB}
$\check{n}\text{ɔ}$	$[\text{ə}u]$	n_{AA80}	ɔ_{AAB7} ɔ_{AAAB}
$n\text{ɔ}$	$[a:\eta]$	n_{AA80}	ɔ_{AABI} ɔ_{AA89}
$n\text{ɔ}$	$[a:n]$	n_{AA80}	ɔ_{AABI} ɔ_{AA99}
$n\text{ɔ}$	$[a:m]$	n_{AA80}	ɔ_{AABI} w_{AAAB}
$n\text{ɔ}$	$[a:k]$	n_{AA80}	ɔ_{AABI} n_{AA80}
$n\text{ɔ}$	$[a:t]$	n_{AA80}	ɔ_{AABI} ɔ_{AACB}
$n\text{ɔ}$	$[a:p]$	n_{AA80}	ɔ_{AABI} v_{AA9A}

$\hat{n}\acute{6}$	[iŋ]	n_{AA80}	$\hat{\circ}_{AAB2}$	$\acute{6}_{AA89}$	$n\acute{w}\acute{1}$	[on]	n_{AA80}	\acute{w}_{AABA}	$\acute{1}_{AA99}$
$\hat{n}\acute{1}$	[in]	n_{AA80}	$\hat{\circ}_{AAB2}$	$\acute{1}_{AA99}$	$n\acute{w}\acute{w}$	[om]	n_{AA80}	\acute{w}_{AABA}	\acute{w}_{AAAB}
$\hat{n}\acute{w}$	[im]	n_{AA80}	$\hat{\circ}_{AAB2}$	\acute{w}_{AAAB}	$n\acute{w}n$	[ok]	n_{AA80}	\acute{w}_{AABA}	n_{AA80}
$\hat{n}n$	[ik]	n_{AA80}	$\hat{\circ}_{AAB2}$	n_{AA80}	$n\acute{w}\acute{6}$	[ot]	n_{AA80}	\acute{w}_{AABA}	$\acute{6}_{AACB}$
$\hat{n}\acute{6}$	[it]	n_{AA80}	$\hat{\circ}_{AAB2}$	$\acute{6}_{AACB}$	$n\acute{w}\acute{1}$	[op]	n_{AA80}	\acute{w}_{AABA}	$\acute{1}_{AA9A}$
$\hat{n}\acute{1}$	[ip]	n_{AA80}	$\hat{\circ}_{AAB2}$	$\acute{1}_{AA9A}$	$\acute{n}\acute{6}$	[uŋ]	n_{AA80}	$\acute{\circ}_{AAB4}$	$\acute{6}_{AA89}$
$\acute{n}\acute{6}$	[eŋ]	n_{AA80}	$\acute{\circ}_{AAB8}$	$\acute{6}_{AA89}$	$\acute{1}\acute{n}\acute{6}$	[uŋ]	$\acute{1}_{AAB6}$	$\acute{n}\acute{1}_{AAC5}$	$\acute{6}_{AA89}$
$\acute{n}\acute{1}$	[en]	n_{AA80}	$\acute{\circ}_{AAB8}$	$\acute{1}_{AA99}$	$\acute{n}\acute{1}$	[un]	n_{AA80}	$\acute{\circ}_{AAB4}$	$\acute{1}_{AA99}$
$\acute{n}\acute{w}$	[em]	n_{AA80}	$\acute{\circ}_{AAB8}$	\acute{w}_{AAAB}	$\acute{n}\acute{w}$	[um]	n_{AA80}	$\acute{\circ}_{AAB4}$	\acute{w}_{AAAB}
$\acute{n}n$	[ek]	n_{AA80}	$\acute{\circ}_{AAB8}$	n_{AA80}	$\acute{1}\acute{w}$	[um]	$\acute{1}_{AA8A}$	$\acute{\circ}_{AADI}$	\acute{w}_{AAAB}
$\acute{n}\acute{6}$	[et]	n_{AA80}	$\acute{\circ}_{AAB8}$	$\acute{6}_{AACB}$	$\acute{n}n$	[uk]	n_{AA80}	$\acute{\circ}_{AAB4}$	n_{AA80}
$\acute{n}\acute{1}$	[ep]	n_{AA80}	$\acute{\circ}_{AAB8}$	$\acute{1}_{AA9A}$	$\acute{n}\acute{6}$	[ut]	n_{AA80}	$\acute{\circ}_{AAB4}$	$\acute{6}_{AACB}$
$\acute{n}\acute{6}$	[eŋ]	\acute{n}_{AAB5}	n_{AA80}	$\acute{6}_{AA89}$	$\acute{n}\acute{1}$	[up]	n_{AA80}	$\acute{\circ}_{AAB4}$	$\acute{1}_{AA9A}$
$\acute{n}\acute{1}$	[en]	\acute{n}_{AAB5}	n_{AA80}	$\acute{1}_{AA99}$	$\acute{n}\acute{6}$	[uŋ]	n_{AA80}	$\acute{\circ}_{AAB3}$	$\acute{6}_{AA89}$
$\acute{n}\acute{w}$	[em]	\acute{n}_{AAB5}	n_{AA80}	\acute{w}_{AAAB}	$\acute{n}\acute{1}$	[un]	n_{AA80}	$\acute{\circ}_{AAB3}$	$\acute{1}_{AA99}$
$\acute{n}n$	[ek]	\acute{n}_{AAB5}	n_{AA80}	n_{AA80}	$\acute{n}\acute{w}$	[um]	n_{AA80}	$\acute{\circ}_{AAB3}$	\acute{w}_{AAAB}
$\acute{n}\acute{6}$	[et]	\acute{n}_{AAB5}	n_{AA80}	$\acute{6}_{AACB}$	$\acute{n}n$	[uk]	n_{AA80}	$\acute{\circ}_{AAB3}$	n_{AA80}
$\acute{n}\acute{1}$	[ep]	\acute{n}_{AAB5}	n_{AA80}	$\acute{1}_{AA9A}$	$\acute{n}\acute{6}$	[ut]	n_{AA80}	$\acute{\circ}_{AAB3}$	$\acute{6}_{AACB}$
$n\acute{6}$	[oŋ]	n_{AA80}	$\acute{\circ}_{AAAE}$	$\acute{6}_{AA89}$	$\acute{n}\acute{1}$	[up]	n_{AA80}	$\acute{\circ}_{AAB3}$	$\acute{1}_{AA9A}$
$n\acute{1}$	[on]	n_{AA80}	$\acute{\circ}_{AAAE}$	$\acute{1}_{AA99}$	$\acute{n}\acute{6}$	[əŋ]	n_{AA80}	$\acute{\circ}_{AAB7}$	$\acute{6}_{AA89}$
$n\acute{w}$	[om]	n_{AA80}	$\acute{\circ}_{AAAE}$	\acute{w}_{AAAB}	$\acute{n}\acute{1}$	[ən]	n_{AA80}	$\acute{\circ}_{AAB7}$	$\acute{1}_{AA99}$
$n\acute{n}$	[ok]	n_{AA80}	$\acute{\circ}_{AAAE}$	n_{AA80}	$\acute{n}\acute{w}$	[əm]	n_{AA80}	$\acute{\circ}_{AAB7}$	\acute{w}_{AAAB}
$n\acute{6}$	[ot]	n_{AA80}	$\acute{\circ}_{AAAE}$	$\acute{6}_{AACB}$	$\acute{n}n$	[ək]	n_{AA80}	$\acute{\circ}_{AAB7}$	n_{AA80}
$n\acute{1}$	[op]	n_{AA80}	$\acute{\circ}_{AAAE}$	$\acute{1}_{AA9A}$	$\acute{n}\acute{6}$	[ət]	n_{AA80}	$\acute{\circ}_{AAB7}$	$\acute{6}_{AACB}$
$n\acute{w}$	[oŋ]	n_{AA80}	\acute{w}_{AABA}	$\acute{6}_{AA89}$	$\acute{n}\acute{1}$	[əp]	n_{AA80}	$\acute{\circ}_{AAB7}$	$\acute{1}_{AA9A}$

A.3 Normalization of tones

For a considerable period of time after its creation, there was no symbol for the tones in Jinping Dai writing system. According to the sources, it was only after the 1940s that tonal symbols were introduced for precise pronunciation, and they were derived from the Kinh language. There are eight tones in Jinping Dai writing system, divided into six relaxed tones and two abrupt tones, borrowing two tone symbols of the Kinh language (in Latin script), [ma:i³¹ sak⁵⁵] and [ma:i³¹ həi²⁵]. The final [-k⁷] variant have a change in pronunciation, but the tone value remains the same. This situation cannot be represented in writing, so a





symbol should be introduced to indicate a syllable with a variation in the final consonant, which is read as [ma:i³³ ʔaʔ⁵⁵]. The tones are:

	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]
	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]
	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]
	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]
	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]
	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]
	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]
	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]
	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]
	[ka ³³]		[ka ⁵⁵]		[ka ²⁵]		[ka ⁴³]		[ka ⁵²]





A.4 Special symbols

There are 12 special symbols, neither initials nor finals. 6 fixed characters and 6 end-of-sentence symbols are included.

A fixed character is a specific symbol used to represent a particular word of a fixed written form that constitutes a monogram for independent use. Such fixed characters are few and far between, appearing in traditional texts only in the words “one”, “person”, “moon” and “horse”. The computer font adds two similar monograms, “child” and “late”, as required. [Note: The typeface designer joining in [Dao et al, 2019](#) failed to handle GPOS, so these two monograms will not be submitted.] These monograms can also be written in a non-independent way. The fixed characters are:

	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]
	[nuw ⁵²]		[k ⁴³]		[b ³³]		[ma ³¹]		[lu ⁴³]		[la ²⁵]

Jinping Dai writing system does not generally use punctuation, but ends each sentence with an end-of-sentence symbol. There is no fixed character to be used, and most of them vary from person to person. There are six end-of-sentence symbols. The first five of these are collectively known as [ma:i³³ sam³¹], meaning “ending symbols”, and are often used in long stories or poems. The last one is called [ma:i³³ met⁵⁵], which means “completion sign” and is used after a sentence to indicate that a sentence has been completed. The end-of-sentence symbols are:

	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]
	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]
	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]
	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]
	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]
	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]
	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]
	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]
	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]
	[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]		[ma:i ³³ sam ³¹]

Appendix B Multi-column Tai Don–Tai Dam character set

A reference forms table for two columns is given in this appendix. The first column shows the reference forms for Vietnamese use, in which *Tai Heritage Pro* was chosen as the typeface. The second column is the reference forms for Chinese use, which contains characters from the Jinping Dai writing system.

The codepoints marked in brown are those for which the allocation of characters is requested in the proposal, and the codepoints marked in grey are reserved.

Table 5 Multi-column Tai Don–Tai Dam character set

AA80	n	n	AA90	၁၇	AAA0	၁	၁
AA81	၁		AA91	၂	AAA1	၂	၂
AA82	၃	၃	AA92	၃	AAA2	၃	
AA83	၄	၄	AA93	၄	AAA3	၄	
AA84	၆	၆	AA94	၆	AAA4	၆	
AA85	၇	၇	AA95	၇	AAA5	၇	၇
AA86	၈		AA96	၈	AAA6	၈	
AA87	၉		AA97	၉	AAA7	၉	
AA88	၁၀		AA98	၁၀	AAA8	၁၀	၁၀
AA89	၁၁	၁၁	AA99	၁၁	AAA9	၁၁	၁၁
AA8A	၁၂	၁၂	AA9A	၁၂	AAAA	၁၂	၁၂
AA8B	၁၃	၁၃	AA9B	၁၃	AAAB	၁၃	၁၃
AA8C	၁၄	၁၄	AA9C	၁၄	AAAC	၁၄	၁၄
AA8D	၁၅	၁၅	AA9D	၁၅	AAAD	၁၅	၁၅
AA8E	၁၆		AA9E	၁၆	AAAE	၁၆	၁၆
AA8F	၁၇		AA9F	၁၇	AAAF	၁၇	၁၇

AAB0	✓ ○	✓ ○	AAC0	e ✓ ○	○	AAD0	✓ ○
AAB1	7	9	AAC1	○	○	AAD1	○
AAB2	^ ○	^ ○	AAC2	7		AAD2	~ ○
AAB3	✓ ○	✓ ○	AAC3		7	AAD3	✓ ✓
AAB4	○ ;	○ ;	AAC4		916	AAD4	W ✓
AAB5	11	11	AAC5		7	AAD5	
AAB6	✓	✓	AAC6		7	AAD6	
AAB7	○	○	AAC7		917	AAD7	⊗
AAB8	○	○	AAC8		7	AAD8	🐟
AAB9	1	6	AAC9		7	AAD9	🐟
AABA	6	W	AACA		9W	AADA	1611
AABB	9	9	AACB		16	AADB	7
AABC	6	6	AACC		9W3	AADC	W
AABD	14	14	AACD		W3	AADD	6
AABE	○	○	AACE		○	AADE	1611
AABF	○	○	AACF		○	AADF	6

Appendix C Figures

(续表)

变异前	变异后	汉义
nɔk ⁴³	nɔ ⁴³	外面
lok ⁴³	lo ⁴³	烫一下
luk ⁴³	lu ⁴³	儿女
lək ⁴³	lə ⁴³	挑选

(三) 声调的规范化

金平傣文在创制后相当一段历史时期内，都没有标记声调的符号。根据有关资料分析，在 20 世纪 40 年代后，为了精确表达读音才有了声调符号，其声调符号来自京语（越南语）。如前所述，金平傣文使用的最大领域在越南，由于长期的语言接触导致语言间的相互借用，这是语言演变的直接表现。京语有 6 个声调，5 个声调符号。金平傣文有 8 个声调，其中分

为 6 个舒声调和 2 个入声调，但只借用了京语其中的两个声调符号 [mai³¹ sak⁵⁵] “ ” 和 [mai³¹ hɔi²⁵] “ ”，即有效解决了声调的区分问题。这两个声调只用于舒声字里，第 1 调和第 4 调以及入声字都不用声调调号，只有第 2 调和第 5 调用声调符号 [mai³¹ sak⁵⁵] “ ”，第 3 调和第 6 调用声调符号 [mai³¹ hɔi²⁵] “ ”。详见表 7。

(a)

表 7 金平傣文声调

调类		调次	调值	声调符号	例词
阴调类	阴平	1	33		𐄢 [ka ³³] 乌鸦
	阴去	2	55	’	𐄢’ [ka ⁵⁵] 跑了
	阴上	3	25	’	𐄢’ [ka ²⁵] 秧苗
阳调类	阳平	4	43		𐄢 [ka ⁴³] 枷锁
	阳去	5	52	’	𐄢’ [ka ⁵²] 价格
	阳上	6	31	’	𐄢’ [ka ³¹] 生意
入声类	阴入	7	55		𐄢 [lak ⁵⁵] 聪明；𐄢 [laɽ ⁵⁵] 陌生
	阳入	8	43		𐄢 [lak ⁴³] 小偷

上述尾音 -k 变异的字，音变调不变。但文字上无法表示出来，我们采取在辅音上增加符号“~”来表示尾音变异的音节，此符号暂读为 [ma: i³³ ʔaɽ⁵⁵]。例如“水果” 𐄢 [ma: k⁵⁵] 变异为 𐄢 [maɽ⁵⁵]，

“陌生” 𐄢 [la: k⁵⁵] 变异为 𐄢 [laɽ⁵⁵]。
从表 7 来看，金平傣文第 2 调和第 7 调的调值相同，都是 55 调；第 4 调和第 8 调的调值相同，都是 43 调。因此，可以归纳为 6 个声调调位，即/33/、

/55/、/25/、/43/、/52/、/31/。其中调位/55/包含了舒声调第2调的调值[55]和入声调第7调的调值[55]；调位/43/包含了舒声调第4调的调值[43]和入声调第8调的调值[43]。

四、特殊符号

金平傣文有一些符号较为特殊，既不是声母，也不是韵母，而是一种固定的特殊符号，专门用来书写某个词或表示句末停顿的符号。这样的符号不多，可分两类：一种是固定字符；另一种是句末符号。

(一) 固定字符

固定字符是指用一个特定的符号来表示某一个词的固定书写形式，构成独立使用的独体字。这种固定字符不多，传统文字中只出现在“一、人、月亮、马”4个词中。即

(b) 𑜀𑜂𑜆 [nɯŋ⁵²] 一 𑜁𑜪 [kɯŋ⁴³] 人
 𑜂𑜆𑜂𑜫 [bən³³] 月亮 𑜂𑜆𑜂𑜫 [ma³¹] 马

计算机字体根据需要增加了2个类似的独体字。即

𑜂𑜆𑜂𑜫 [lu⁴³] 孩子 𑜂𑜆𑜂𑜫 [la²⁵] 迟

上述的独体字也可以用声韵调拼合的方式书写，如： 𑜀𑜂𑜆 (一)、 𑜁𑜪 (人)、 𑜂𑜆𑜂𑜫 (月亮)、 𑜂𑜆𑜂𑜫 (马)、 𑜂𑜆𑜂𑜫 (孩子)、 𑜂𑜆𑜂𑜫 (迟)。使用时可根据需要选择使用独特字或声韵调拼合的字。

(二) 句末符号

金平傣文一般不使用标点符号，每一句话结束时就在后面标注一个句末符号。但并没有固定用哪一个字符，大多因人而异。有的字符与其说是字，不如说是图画，有的字形类似象形文字。句末符号有如下6个。即



其中，前5个均统称为[ma: i³³ sam³¹]，有“结束符号”之意，通常用在长篇的故事或诗歌里。最后1个称为[ma: i³³ mɛt⁵⁵]，有“完成符号”之意，用在句子之后，表示一句话的完成。

五、结语

综上所述，金平傣文是一种源于印度梵文字母系统的拼音文字，也是跨境自称“傣端”或“傣皓”的傣族白傣支系使用的民族文字。该文字由于创制的年代久远，与现行口语在声母、韵母、声调等方面都出现了一定的差异。加之长期以来均通过手抄方式流传于民间，字体形式因人而异，既没有规范的字体，也没有计算机字体，这对跨境傣族语言文化的传承和发展带来了不利的影响。文字是一个民族进步的标志性特征之一，它对推动民族文化的发展起到积极的作用。尤其是当今数字化、信息化的时代，文字所发挥的作用更加具有不可替代性。为了能更好保护与传承优秀的民族文化，本文通过对金平傣文的实地调查，并根据现行口语的语音特征，对其语音的规范化问题进行了探讨，并提出金平傣文语音规范化的基本原则和标准。具体归纳总结表8。

1. 金平傣文规范字体

以金平县融媒体中心罗文斌同志所设计的金平傣文计算机字体小睿体作为金平傣文数字化、信息化的规范字体。该字体包含了在传统文字符号基础上所设计的字体和为完整记录现行语言而新增的文字符号字体。

2. 金平傣文标准音

以金平县勐拉镇旧勐村的白傣语语音作为金平傣文的标准音。该村落的历史文化、地理位置、语音特点，符合标准选择的一般原则。

3. 金平傣文规范声母

以22个辅音音位44个傣文字母作为金平傣文的规范声母。列表8。

表 8 规范声母

高音组			低音组		
ᵛ			ᵝ		
n	ᵛ	ᵛᵛ	n	ᵝ	ᵛ
ᵛ	ᵛᵛ	ᵛᵛᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛ
m	ᵛ	ᵛᵛ	m	ᵛ	ᵛᵛ
ᵛ	ᵛᵛ	ᵛᵛᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛ
ᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛ	ᵛ	ᵛᵛ
ᵛᵛ	ᵛᵛ	ᵛᵛᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛᵛ
ᵛᵛ	ᵛᵛ	ᵛᵛᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛᵛ

4. 金平傣文规范韵母

以 83 个韵母作为金平傣文的规范韵母。包含 9 个单元音、14 个双元音、60 个带辅音韵尾的韵母（简称“辅音尾韵母”）。列表 9。

表 9 规范韵母

单元音		ᵛ	ᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛ
双元音	ᵛᵛ									
	ᵛᵛ	ᵛᵛᵛ				ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ
	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ					ᵛᵛᵛ
辅音尾韵母	ᵛᵛ	ᵛᵛ	ᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ
	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛᵛ
	ᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛ
	ᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛ
	ᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛ
	ᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛᵛ	ᵛᵛᵛᵛᵛᵛᵛᵛ

5. 金平傣文规范声调

符号作为规范调符。列表 10。




以六个调位作为金平傣文的 6 个声调，3 个声调

表 10 规范声调





调位	一/33/	二/55/	三/25/	四/43/	五/52/	六/31/
调值	33	55	25	43	52	31
调符		ˊ ~	ˊ		ˊ	ˊ
例词	ᵛᵛᵛ	ᵛᵛᵛᵛ ᵛᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛ	ᵛᵛᵛᵛ	ᵛᵛᵛᵛ

Figure 8 Jinping Dai orthography shown in [Dao et al, 2019](#)

元音字母表

前标元音: (4个)		
	/ai/ /oŋ/ /aa/ /ε/	
上标元音: (7个)		
	/i/ /an/ /ɜ/ /e/ /u/ /anm/ /enm/	
后标元音: (5个)		
	/a/ /o/ /o/ /ant/	
下标元音: (2个)		
	/u/ /aa/	
前后标元音: (1个)		
	/ao/	






说明: 在金平傣文元音字母中前标元音“ [ong]、 [on]、 [an] “”不常使用, 在使用时必须与其它字母搭配使用。




二、固定字符

			
/neun/	/kun/	/ma/	/vbun/
一	人	马	月




在金平傣文中有四个字符是固定的, 相当于中文中的独体字可以单独使用。对等于中文的“一、人、马、月(亮)”。当然我们也可以单独拼读出以上的四个独体字, 拼读出的字与上面的发音相同用法一样没有其它的用意。两者都能使用。







三、俚语、敬语

		
/f'aa/	/lu/	/kun/
民族	人(族)	人(族)
俚语	敬语	敬语

白傣源起中国, 历史上被称作“白夷、摆夷”一支, 在勐梭王国解体后白傣先民顺红河流域迁徙进入越南。后因与安南交趾的京族发生矛盾回迁至现在的居住地。至今在越南的白傣仍然被称为“Tai Don”, 系来自中国的外迁民族。按照“以夷制夷”, 稳固边疆的政策, 1640年勐梭寨土司刁氏接受明朝崇祯帝的任命和册封。成为白傣有史以来最早的世袭土司之一, 管辖地域被称作“泰族十二州”。中法战争结束后, 中国的勐梭、勐莱、勐蚌(约1600平方公里)被划界给法属印度支那联邦, 后归为越南管辖。勐立、勐梭被称作白傣居住地最大的两个勐, 在当时所有进贡给朝廷的贡品及粮草均要从勐梭接转至勐拉后上贡给安府(今建水)。白傣作为上层阶级的民族, 视其它民族为下等民族, 故“”一词也有歧视之意。俚语。如今更多的白傣民族已经意识到了这个词语的使用, 为维护地方民族的团结, 提倡民族平等, 该词已不再被视为傣语中民族的称谓, 而是使用敬语“”、“”来作为对其他民族的称谓。








第四课 声调、符号

金平傣文共有六个声调值, 两个声调符号, 需要说明的是, 声调随低、高辅音字母的使用而产生不同的调值。在日常的傣语发音中通常分为低音组轻声、一声、二声和高音组轻声、一声、二声。以 、、 作为韵尾的复合元音在拼读时不再标注声调符号, 发音时根据辅音字母所在组决定平声调值。

					
an'at n'it	an'at n'it	an'at n'it	an'at n'it	an'at n'it	an'at n'it
33调	35调	25调	43调	52调	31调
平声	一声	二声	平声	一声	二声
乌鸦	逃跑	秧苗	卡住	玩	生意

一、常用符号

						
may'sannm/	may'sannm/	may'sannm/	may'sannm/	may'met/	may'sannm/	may'taanm/
省略号	省略号	省略号	省略号	句号	省略号	重音号

金平傣文傣文符号形式多样, 归类整理后共有以上几种, 符号多见于文章、歌曲的开头及结尾。其用意多为省略及沿长。符号、、多用于文章的标题修饰或内容的省略, 符号、则用于记录歌曲的演唱延音的标注, 如: oy.....oy.....。符号则是白傣文的标点符号, 代表着一句话的表述完成。即是逗号也是句号, 同时也代表叹号, 符号成则为重音符号通常在无法标注声调的词语中使用, 代表声音的轻重。

第十三课 颜色

			
u'it	u'it	u'it	u'it
/rit/	/rit/	/rit/	/rit/
红	橙	黄	绿
			
u'it	u'it	u'it	u'it
/xew don/	/pe/	/lao/	/ranm/
青	蓝	紫	黑
			
u'it	u'it		
/ron/	/hmuun/		
白	灰		

练习题:
根据所学颜色写出下面的句子

红色的国旗
绿色的叶子
黑色的衣服
紫色的茄子
黄色的小桶
蓝色的天空
白色的衬衣

Figure 9 Jinping Dai orthography shown in Luo, 2017

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P&P) from

<http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Final proposal to add 22 characters for Tai Don writing system		
2. Requester's name:	Kushim JIANG		
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution		
4. Submission date:	July 3, 2022		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:			yes
(or) More information will be provided later:			

B. Technical – General

1. Choose one of the following:			
a. This proposal is for a new script (set of characters):			
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:	yes		
Name of the existing block:	Tai Viet (block name requested to update)		
2. Number of characters in proposal:	22		
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	<input checked="" type="checkbox"/>	B.1-Specialized (small collection)	<input type="checkbox"/>
C-Major extinct	<input type="checkbox"/>	D-Attested extinct	<input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic	<input type="checkbox"/>	E-Minor extinct	<input type="checkbox"/>
G-Obscure or questionable usage symbols	<input type="checkbox"/>		
4. Is a repertoire including character names provided?			
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	yes		
b. Are the character shapes attached in a legible form suitable for review?	yes		
5. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	Kushim JIANG		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	Kushim JIANG, kushim_jiang@outlook.com		
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	yes		
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	yes		
7. Special encoding issues:			
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?			yes
A sorting order for the writing system is provided.			
8. Additional Information:			
<p>Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behavior information such as line breaks, widths etc., Combining behavior, Spacing behavior, Directional behavior, Default Collation behavior, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.</p>			

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	no
If YES explain	
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	no
If YES, with whom?	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	yes
Reference: Luo Wenbin. 2017. <i>Tai Don script Tutorial</i> . Binge Design Studio. [Luo, 2017]	
4. The context of use for the proposed characters (type of use; common or rare)	common
Reference:	
5. Are the proposed characters in current use by the user community?	yes
If YES, where? Reference: Jinping County, Yunnan Province, PRC	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	yes
If YES, is a rationale provided?	
If YES, reference: The existing block is in BMP.	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	no
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	no
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	no
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	yes
If YES, is a rationale for such use provided?	
If YES, reference: Combining characters are an inherent part of the writing system.	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference: Appendix A of this proposal.	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	no
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility characters?	no
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	