

Universal Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Proposal on adding one provisional `kTayNumeric` property to Unihan DB
Source: Eiso Chan (陈永聪, Culture and Art Publishing House)
Status: Individual Contribution
Action: For consideration by UTC
Date: 2025-02-03 (Beginning of Spring, the first day of 乙巳蛇年)

Tày people (Người Tày/𑜋𑜂𑜆𑜐) are one of the ethnic groups in Vietnam, and they use Tày Nôm characters (Chữ Nôm Tày) to record their languages, which is similar to Zhuang characters. The Vietnam NB and VNPF have submitted many to IRG and many of them have been encoded. Numbers for Tày language(s) can be written in Tày Nôm characters. I suggest adding a new provisional `kTayNumeric` property to include the values of the following 7 characters¹ based on [Từ Điển Chữ Nôm Tày](#) (A Dictionary of Nôm Tày).

UCS	Character	Reading	Value
U+5200	刀	đeo, đéo ²	1
U+6761	条	đeo	1
U+80FD ³	能	nâng	1
U+2B871	𑜂𑜆𑜐	soong ⁴	2
U+31EC7	𑜂𑜆𑜐	sí ⁵	4
U+2B872	𑜂𑜆𑜐	hả	5
U+2C954	𑜂𑜆𑜐	chất	7

The proposed `kTayNumeric` property in this document is similar to [kVietnameseNumeric](#) and [kZhuangNumeric](#) properties proposed in [L2/22-111](#). The suggested property information for UAX #38 is shown below.

Property	<code>kTayNumeric</code>
Status	Provisional

¹ Other numbers are the same as common Hanzi or Vietnamese Nôm characters.

² This entry and the next two ones are only used for the single digits.

³ There is the same value for `kZhuangNumeric` property under U+80FD.

⁴ This word is spelled as “sloong” in Reference 2. The etymology of this word for old Chinese is 雙.

⁵ This word is spelled as “sí” in Reference 2.

Category	Numeric Values
Introduced	TBD...
Delimiter	space
Syntax	\d+
Description	The value of the ideograph when used as a numeral in Tày languages with the Han script (Chữ Nôm Tày). It can be used alongside kPrimaryNumeric or kAccountingNumeric since the Chinese vocabulary of numbers is also imported in Tày; it can be also used alongside kVietnameseNumeric since the Vietnamese vocabulary of numbers is also imported. Nevertheless, in Tày text, this value should override kPrimaryNumeric , kAccountingNumeric and kVietnameseNumeric if the ideograph should have any of these properties.

As the above description shows, Chinese basic numerals, Chinese accounting numerals (參捌拾) and Vietnamese numerals (文卅匹) are also often used in Tày text, but the Sino-Vietnamese readings for Chinese basic numerals, Chinese accounting numerals and Vietnamese numerals are just used for some literary words like 八音 (bát âm), 九重 (cửu túng) and 參春 (sam xuân). [“âm” and “xuân” are also the Sino-Vietnamese readings.] There is no need to include these parts for `kTayNumeric` property.

The following table shows the same usages and values like `kPrimaryNumeric` and `kAccountingNumeric` properties for Tày language(s).

UCS	Character	Reading	Value
U+4E00	一	ất, nhất ¹	1
U+4E8C	二	nhĩ ²	2
U+4E09	三	sam, tam ³	3
U+53C3	參	sam	3
U+56DB	四	sí, tứ ⁴	4
U+4E94	五	ngụ, ngũ ⁵	5
U+516D	六	xôc, lục ⁶	6
U+4E03	七	chất ⁷	7
U+516B	八	pét, bát ⁸	8
U+634C	捌	pét	8
U+4E5D	九	cầu, cửu ⁹	9
U+5341	十	síp ¹⁰	10

¹ Reference 2 also includes “ê”. “nhất” is borrowed from Sino-Vietnamese reading.

² It also reads as the same as 𠄎. “nhĩ” is derived from Sino-Vietnamese reading “nhị”.

³ This word is spelled as “slam” in Reference 2. “tam” is borrowed from Sino-Vietnamese reading.

⁴ The reading is the same as 肆, and this word is spelled as “slí” in Reference 2. “tứ” is borrowed from Sino-Vietnamese reading.

⁵ It also reads as the same as 𠄎. “ngụ” is derived from Sino-Vietnamese reading “ngũ”, and “ngũ” is borrowed from Sino-Vietnamese reading.

⁶ This word is also spelled as “hôc” in Reference 2. “lục” is borrowed from Sino-Vietnamese reading.

⁷ The reading is the same as 𠄎, and this word is spelled as “chêt” in Reference 2.

⁸ This word is spelled as “pet” in Reference 2. “bát” is borrowed from Sino-Vietnamese reading.

⁹ “cửu” is borrowed from Sino-Vietnamese reading.

¹⁰ This word is spelled as “slíp” in Reference 2.

UCS	Character	Reading	Value
U+62FE	拾	sip	10
U+767E	百	pác, bách ¹	100
U+5343	千	xiên, thiên ²	1,000
U+842C	萬	vản, vạn ³	10,000
U+4E07	万	vản, vạn	10,000

Notice 1: The Tày word of “eleven” reads as “sip ất” (written as 十一).

Notice 2: The Tày word of “twelve” reads as “sip nhĩ” (written as 十二) and “sip soong” (written as 十𠄎 or 十二)⁴; the Tày word of “twenty” reads as “nhĩ sip” (written as 二十).

Notice 3: The Tày word of “fifteen” reads as “sip hả” (written as 十五 or 十五); the Tày word of “fifty” reads as “hả sip” (written as 五十). “ngụ” is just used for some literary words like 五音 (ngụ âm), 五福 (ngụ phúc) and 五色 (ngụ sắc).

The following table shows the same usages and values like `kVietnameseNumeric` property for Tày language(s).

UCS	Character	Reading	Value
U+20B20	𠀎	một	1
U+20129	𠀎	hai	2
U+20027	𠀎	ba	3

Bibliography

- [1] Hoàng Triều Ân; Dương Nhật Thanh, Hoàng Tuấn Nam: *Từ Điển Chữ Nôm Tày*, Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội, 2003
- [2] 韦树关, 颜海云, 黎莎: 《国外壮侗语族语言词汇集》, 广州: 世界图书出版广东有限公司, 2019.11, ISBN 978-7-5192-5877-1
- [3] Eugene Chan: *Numeral Systems of the World's Languages*, 2025.1.25
- [4] 喻翠容, 罗美珍: 《傣仂汉词典》 (《中国少数民族语言词典丛书》), 北京: 民族出版社, 2004.1, ISBN 7-105-05834-X
- [5] 孟尊贤, 方伯龙: 《汉傣词典》, 昆明: 云南民族出版社, 1991.6, ISBN 7-5367-0239-6/G·53

Acknowledgement

Lee Collins (康立論), Ngô Trung Việt (吳中越), Dr. Ken Lunde (小林劍) and Yifan Wang (王一凡) reviewed the drafts of this proposal and provided some useful comments.

(End of Document)

¹ This word is spelled as “pac” in Reference 2. “bách” is borrowed from Sino-Vietnamese reading.

² “thiên” is borrowed from Sino-Vietnamese reading.

³ This word is spelled as “fan” in Reference 2. “vản” is derived from Sino-Vietnamese reading “vạn”, and “vạn” is borrowed from Sino-Vietnamese reading.

⁴ The etymology of the first two syllables of the Chinese geographic name Xishuangbanna (西双版纳) in Yunnan Province (云南省) is the same as the Tày word “sip soong”. The Tai Lue word is “ᨧᩢ᩠ᨦᩉ᩠ᩅ”, the Tai Le word is “ᨧᩢ᩠ᨦᩉ᩠ᩅ” and the Thai word is “สิบสอง”.