

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 `kTangutNumeric` property to UAX #60
Source: Eiso Chan (陈永聪/𐰃𐰆𐰏𐰐)
 Jerry You (游程宇/𐰃𐰆𐰏𐰐)
Status: Individual Contribution
Action: For consideration by UTC & WG2
Date: 2025-02-03 (Beginning of Spring, the first day of 𐰃𐰆𐰏𐰐)

Michel Suignard, the UCS project editor, submitted the proposed draft of UAX #60 ([L2/25-052](#), *Sources and Other Ancillary Data for Non Han Ideographic Scripts*). This UAX will include the properties for Jurchen, Nüshu and Tangut, which is very similar to [UAX #38 Unihan DB](#). There is one numeral related property for Jurchen, which is `kJurchenNumeric`, but there are only two properties for Tangut, that means there is no numeral related property for Tangut currently. We notice there is necessarily to add the new informative `kTangutNumeric` property to make the data more complete.

The suggested property information for UAX #60 is shown below.

Property	<code>kTangutNumeric</code>
Status	Informative
Category	Other Data
Introduced	<i>TBD...</i>
Delimiter	N/A
Syntax	<code>\d+(\.5)?</code>
Default	N/A
Description	Numeric value of the Tangut character. It only applies to a few characters. The main bibliographic information for this property is: Jiǎ Chángyè 贾常业, ed. <i>Xīxiàwén Zìdiǎn</i> 西夏文字典 (<i>Tangut Dictionary</i>). Lanzhou: 甘肃文化出版社 (Gansu Culture Press), 2019.5, ISBN 978-7-5490-1785-0, p. 995

The Tangut encoding sources (see `kTGT_MergedSrc` property) include the entries on different Tangut characters which could be translated into the Chinese basic numerals (see [kPrimaryNumeric](#) property in Unihan DB), but not all of them are allowed to be used as the basic numerals for Tangut language (or written language). We checked the Tangut ancient books and referred the 《西夏文字典》 (*Xīxiàwén Zìdiǎn*, p. 995) which is edited by 贾常业 (Jiǎ

Chángyè) to make the following list. SN #15 isn't mentioned on p. 995, we add it based on the new encoding source H2021.

SN	UCS	Character	Basic Hanzi	Value
1	U+18229	𠄎	一	1
2	U+1736B	𠄎	二	2
3	U+18555	𠄎	三	3
4	U+17943	𠄎	四	4
5	U+173C1	𠄎	五	5
6	U+17901	𠄎	六	6
7	U+174B9	𠄎	七	7
8	U+1824B	𠄎	八	8
9	U+178AD	𠄎	九	9
10	U+17C17	𠄎	十	10
11	U+1829D	𠄎	百	100
12	U+1785E	𠄎	千	1,000
13	U+17551	𠄎	萬/万	10,000
14	U+179B2	𠄎	億/亿	100,000,000
15	U+17E15	𠄎	半	0.5

The basic numeric usage for Tangut language is very close to Chinese and its dialects, but not exactly the same, and there is not the corresponding word meant “zero” in Tangut language. We need to pay attention to the following points when we use the Tangut numerals.

1) There are three parts for the numerals, the first part is SN #1 through #9, which could be recorded as Part A tentatively; the second part is SN #10 through #14, which could be recorded as Part B tentatively; the third part is SN #15, which could be recorded as Part C tentatively. Part B is like the suffix in each corresponding digit (such as U+17C17 𠄎 for tens place, U+1829D 𠄎 for hundreds place, etc.) after Part A when the number is greater than or equal to 10. Part C isn't used with Parts A & B commonly.

2) For part A, the numeral which corresponds to 1 (aka U+18229 𠄎) is only used in ones place, ten thousands place, hundred millions place and so on, and it must be omitted in tens place, hundreds place and other places, such as 1 is 𠄎, 11 is 𠄎𠄎, 151 is 𠄎𠄎𠄎, 910,000 is 𠄎𠄎𠄎𠄎𠄎.

3) As mentioned above, there is not the corresponding word meant “zero”, so 0 must be omitted in each digit, such as 101 is 𠄎𠄎, 2003 is 𠄎𠄎𠄎, 2015 is 𠄎𠄎𠄎𠄎, 9015,004 is 𠄎𠄎𠄎𠄎𠄎𠄎𠄎.

4) For Part C, U+17E15 𠄎 is often used with some units of measurement in date, such as 𠄎𠄎 (half a year), 𠄎𠄎 (half a month), 𠄎𠄎 (half a day) and so on. The word order is different from Chinese.

5) U+17C2D 𠄎 is sometimes used for 10 in days of months as 𠄎𠄎 or 𠄎𠄎, but the proper form is 𠄎𠄎. We only keep U+17C17 𠄎 for 10.

The property data are shown below.

U+1736B	kTangutNumeric	2
U+173C1	kTangutNumeric	5

U+174B9	kTangutNumeric	7
U+17551	kTangutNumeric	10000
U+1785E	kTangutNumeric	1000
U+178AD	kTangutNumeric	9
U+17901	kTangutNumeric	6
U+17943	kTangutNumeric	4
U+179B2	kTangutNumeric	100000000
U+17C17	kTangutNumeric	10
U+17E15	kTangutNumeric	0.5
U+18229	kTangutNumeric	1
U+1824B	kTangutNumeric	8
U+1829D	kTangutNumeric	100
U+18555	kTangutNumeric	3

Bibliography

贾常业: 《西夏文字典》, 兰州: 甘肃文化出版社, 2019.5, ISBN 978-7-5490-1785-0
[CCAMC](#)

Acknowledgement

Andrew West (魏安) reviewed the draft of this proposal and provided some useful comments.

(End of Document)