Title: Proposed Updates and Expansions of Unihan Numeric Fields
Author: Wang Yifan, Eiso Chan
Date: 2022-10-01

This document describes our updated proposal regarding Unihan numeric fields based on L2/22-111, consolidating subsequent discussions and recommendations including L2/22-121 and L2/22-127. It proposes in total 2 new Unihan properties, 3 updated property descriptions, and 73 additions, 1 change, and 1 deletion in entries of respective fields.

All English number names referred in this document are in the short scale, unless otherwise stated.

Please see the attached text file for the summary of proposed updates in the UAX #38-like format.

1. Changes to properties

With UAX #38 having been revised in the latest Unicode 15.0, we still see the necessity of further clarification and disambiguation on how existing numeric properties should be categorized. Changes are highlighted in underline.

1.1. kAccountingNumeric

Change the current description:

> *The value of the character when used as an accounting numeral to prevent fraud. A numeral such as 十 (ten) is easily transformed into 千 (thousand) by adding a single stroke, so monetary documents often use an accounting form of the numeral, such as 拾 (ten), instead of the more common—and simpler—form. Characters with this property will have a single, well-defined value, which a native reader can reasonably be expected to understand.*
>
> *The three numeric-value fields should have no overlap; that is, characters with a `kAccountingNumeric` value should not have a `kOtherNumeric` or `kPrimaryNumeric` value as well.*

to:

> *The value of the character when used as an accounting numeral to prevent fraud in Chinese (and Sino-Xenic) numeric systems. A numeral such as 十 (ten) is easily transformed into 千 (thousand) by adding a single stroke, so monetary documents often use an accounting form of the numeral, such as 拾 (ten), instead of the more common—and simpler—form. Characters with this property will have a single, well-defined value, which a native reader can reasonably be expected to understand.*
>
> *The three Chinese numeric-value fields should have no overlap; that is, characters with a `kAccountingNumeric` value should not have a `kOtherNumeric` or `kPrimaryNumeric` value as well.*

1.2. kOtherNumeric

Change the current description:

> *One or more values of the character when used as a numeral. Characters with this property are rarely used for writing numbers, or have non-standard or multiple values depending on the region. For example, 𠄘 is a rare character whose meaning, "five," would not be recognized by most native readers. An English-language equivalent is "score," whose numeric value, "twenty," is not recognized by many native readers.*
>
> *The three numeric-value fields should have no overlap; that is, characters with a `kOtherNumeric` value should not have a `kAccountingNumeric` or `kPrimaryNumeric` value as well.*

to:

> *One or more values of the character when used as a numeral in Chinese (and Sino-Xenic) numeric systems. Characters with this property are rarely used, obsolete, domain-specific, non-standard, or non-compositional as a numeral. For example, 幺 "tiny," normally not a numeral, can be used as the phonetic code for "one" in some regions. An English-language equivalent is "score," whose numeric value, "twenty," is not recognized by many native readers.*
>
> *The three Chinese numeric-value fields should have no overlap; that is, characters with a `kOtherNumeric` value should not have a `kAccountingNumeric` or `kPrimaryNumeric` value as well.*

The main reason behind the change is that the Unicode 15.0 description is less compatible with the previous practice in the existing numeric fields. For example, 兆, which represents different numbers in several regional standards, would be unintuitively moved under this field.

1.3. kPrimaryNumeric

Change the current description:

> *The value of the character when used as a numeral. Characters which have this property have numeric values that are common, and always convey the same numeric value. For example, 千 always means "thousand." A native reader is expected to understand the numeric value for these characters.*
>
> *The three numeric-value fields should have no overlap; that is, characters with a `kPrimaryNumeric` value should not have a `kAccountingNumeric` or `kOtherNumeric` value as well.*

to:

> *One or more values of the character when used as a numeral in Chinese (and Sino-Xenic) numeric systems. Characters which have this property have numeric values that are common, or are standardized to convey a fixed numeric value. For example, 千 always means "thousand." A native reader is expected to understand*

the numeric value for these characters.

The three Chinese numeric-value fields should have no overlap; that is, characters with a `kPrimaryNumeric` *value should not have a* `kAccountingNumeric` *or* `kOtherNumeric` *value as well.*

2. New properties

2.1. kVietnameseNumeric

Property	`kVietnameseNumeric`
Status	`Provisional`
Category	`Numeric Values`
Introduced	*– (TBD)*
Delimiter	`space`
Syntax	`[0-9]+`
Description	The value of the character when used as a numeral in Vietnamese with Han script (Hán Nôm). It can be used alongside `kPrimaryNumeric` since the Chinese vocabulary of numbers is also imported in Vietnamese. Nevertheless, in the Vietnamese text, this value should override `kPrimaryNumeric` if the character should have both fields.

2.2. kZhuangNumeric

Property	`kZhuangNumeric`
Status	`Provisional`
Category	`Numeric Values`
Introduced	*– (TBD)*
Delimiter	`space`
Syntax	`[0-9]+`
Description	The value of the character when used as a numeral in Zhuang languages with Han script (Sawndip). It can be used alongside `kPrimaryNumeric` since the Chinese vocabulary of numbers is also imported in Zhuang. Nevertheless, in the Zhuang text, this value should override `kPrimaryNumeric` if the character should have both fields.

3. Data additions

3.1. Deletion from kAccountingNumeric

It contains 1 entry proposed in L2/22-111, Section 1.2.

Glyph	Code	Value	Notes
萬	U+842C	10,000	Move to kPrimaryNumeric as with Simplified U+4E07 万

3.2. Addition to kOtherNumeric

It contains 2 entries proposed in L2/22-111, Section 1, and 5 new additions in this document, all as complements to existing data in certain aspects. Among them, the "phonetic code reading" series is sourced from *Xiandai Hanyu Cidian*, and usage in the Traditional Chinese is to be confirmed.

Glyph	Code	Value	Notes
洞	U+6D1E	0	Complements the phonetic code readings such as U+5E7A 幺 etc.
两	U+4E24	2	Pairing with the existing Traditional U+5169 兩
俩	U+4FE9		Pairing with the existing U+4EE8 仨
倆	U+5006		
拐	U+62D0	7	Complements the phonetic code readings such as U+5E7A 幺 etc.
钩	U+94A9	9	
鈎	U+920E		
皕	U+7695	200	Complements the pre-composite numbers such as U+5EFF 廿 etc.

Additional note:

- U+920E 鈎 is the traditional counterpart of U+94A9 钩 according to the PRC *Table of General Standard Chinese Characters* 通用规范汉字表. We could not confirm usage in other regions.

3.3. Addition to kPrimaryNumeric

It contains 10 entries of larger number names from L2/22-111, Section 2.1, and 1 transferred from kAccountingNumeric. Commas in values are only for presentational purposes.

Glyph	Code	Value	Notes
萬	U+842C	10,000	Moved from kAccountingNumeric

Glyph	Code	Value	Notes
京	U+4EAC	10,000,000,000,000,000	10^{16}; ten quadrillion
垓	U+5793	100,000,000,000,000,000,000	10^{20}; one hundred quintillion
秭	U+79ED	1,000,000,000,000,000,000,000,000 1,000,000,000,000 1,000,000,000	10^{24}; one septillion
秄	U+25771	1,000,000,000,000,000,000,000,000	
穰	U+7A70	10,000,000,000,000,000,000,000,000,000	10^{28}; ten octillion
穣	U+7A63		
溝	U+6E9D	100,000,000,000,000,000,000,000,000,000,000	10^{32}; one hundred nonillion
沟	U+6C9F		
澗	U+6F97	1,000,000,000,000,000,000,000,000,000,000,000,000	10^{36}; one undecillion
涧	U+6DA7		

Additional notes:

- While they are expected to be mostly used in Japan, Simplified/Traditional variants are also supplied.
- Although U+25771 秄 is thought to be a misanalyzed form of U+79ED 秭, it is better included as a regional variant in view of the preference in Japan (coded in JIS X 0213).
- U+79ED 秭 means “billion” (short or long scale) in Vietnamese.

3.4. Changes to kPrimaryNumeric

Commas in values are only for presentational purposes.

Glyph	Code	Value	Notes
兆	U+5146	1,000,000,000,000 100,000	Add a second value 100,000 for PRC and Vietnam conventions

3.5. Addition to kVietnameseNumeric

We suggest assigning Unihan numeric values to the following 42 characters. Commas in values are only for presentational purposes.

Glyph	Code	Quốc ngữ	Value
空	U+7A7A	không	0
沒	U+6C92	một/mốt	1
没	U+6CA1		
蔑	U+8511		
𠬠	U+20B20		
𱍗	U+31357		
𱥺	U+3197A		
台	U+53F0	hai	2
𠄩	U+20129		
𱎔	U+31394		
𠀧	U+20027	ba	3
奔	U+5954	bốn	4
本	U+672C		
𦊚	U+2629A		
𠄼	U+2013C	năm	5
𫡵	U+2B875		
𱎖	U+31396		
𠄻	U+2013B	lăm	
𬂽	U+2C0BD		
𠄶	U+20136	nhăm	
𦒹	U+264B9	sáu	6

Glyph	Code	Quốc ngữ	Value
罢	U+7F62	bảy	7
罷	U+7F77		
𦉱	U+26271		
𬙞	U+2C65E		
𠔭	U+2052D	tám	8
𫤯	U+2B92F		
㐱	U+3431	chín	9
𠃩	U+200E9		
𢒂	U+22482		
𫡦	U+2B866		
迈	U+8FC8	mười/mươi	10
邁	U+9081		
𠦩	U+209A9		
𨑮	U+2846E		
𨒒	U+28492		
𱑕	U+31455		
啉	U+5549	trăm	100
𤾓	U+24F93		
彦	U+5F66	ngàn/nghìn	1,000
𠦳	U+209B3		
𡶨	U+21DA8		

3.6. Addition to kZhuangNumeric

We suggest assigning Unihan numeric values to the following 13 characters from *Sawndip Sawdenj* 古壮字字典.

Glyph	Code	Reading	Value
了	U+4E86	ndeu	1
吊	U+540A		

Glyph	Code	Reading	Value
尞	U+5C1E		
𠮩	U+20BA9		

Glyph	Code	Reading	Value
哅	U+20CA2		
刁	U+2CEB4		
帛	U+3000C		
能	U+80FD	nwngh	
旯	U+2B9C7		

Glyph	Code	Reading	Value
双	U+53CC	song	2
松	U+677E		
侮	U+4FC9	ngux	5
呸	U+3576	haj	

Acknowledgments

Our thanks to Tao Yang, who provides us with a Zhuang Extension G font.

(End of document)