On the "Hangzhou-style" Numerals in ISO/IEC 10646-1

Thomas EMERSON (湯姆·愛摩森)
Basis Technology Corporation
One Kendall Square, Building 200
Cambridge, Massachusetts 02139 USA
tree@basistech.com

March 5, 2000

1 Introduction

The BMP of ISO/IEC 10646-1:1993 contains a range of values in the CJK Symbols and Punctuation block (U+3002 – U+3029) named the "Hangzhou-style numerals." ISO/IEC 10646-1:2000 adds three additional "Hangzhou numeral" code points (U+3038 – U+303A) in Ideographic Extension Block A.

The addition of these new code-points led to several discussions on the Unicode[1] and Chinese[2] mail lists about the use of the name "Hangzhou" in relation to these numbers. This paper explores the origin and encoding of these numbers and presents evidence that they should called the "Suzhou" numerals.

This paper is organized as follows: Section 2 describes the encoding of these characters in various in ISO/IEC 10646 and several locale-specific encodings. Section 3 explores the origin of the name used with these. Section 4 concludes.

2 Encoding Methods

This section outlines the encoding of the Hangzhou numerals in a number of national and international coded character set standards.

Taiwan's CNS 11643-1992 and Big Five character sets include code points for all twelve Hangzhou numerals in ISO/IEC 10646-1:2000. The characters for numeral ten and numeral thirty, 十 and 卅 respectively, are duplicated at different code points in both encodings.

These two characters are also encoded in China's GB 2312-80 at CAAE and D8A6. The remaining nine characters are not encoded. GB 13000.1-93, essentially a translation of ISO 10646-1:1993 includes those code points in the CJK Symbols and Punctuation Block as well as the code points for 十 and 卅 in the CJK Ideographic Area.[3]

Similarly, Japan's JIS X 0208:1997 standard encodes 十 and 卅 at 29-29 and 50-33, respectively. Like GB2312-80, the remaining Hangzhou numerals do not appear in that standard.

---

[1] unicode@unicode.org
[2] chinese@kenyon.edu
[3] I have to date been unable to investigate the presence of these characters in GBK/Code Page 936.

Because of the source separation rule these two ideographs were not unified in ISO/IEC 10646-1:1993, round-trip compatibility between ISO/IEC 10646-1:1993 and both Big Five and CNS 11643-1992 is impossible.[4]

It should also be noted that these ideographs, being in the CJK Ideographic Block, do not have the numeric property assigned to them. (This is true, too, of the "regular" Han numerals, and is intentional.)

The addition of the three new Hangzhou numeric ideographs in Ideographic Extension Block A of ISO 10646-1:2000 facilitates the round trip conversion of these code points with Big Five and CNS 11643-1992. They also share the numeric property with the pre-existing Hangzhou numerals.

The following table presents the twelve Hangzhou numerals in ISO/IEC 10646-1:2000 and their mappings in CNS 11643-1992 and Big Five.

Character	UCS-2	CNS 11643-1992	Big Five	Unicode Name
ISO/IEC 10646-1:1993 (CJK Symbols and Punctuation Block)				
〡	3021	12435	A2C3	HANGZHOU NUMERAL ONE
〢	3022	12436	A2C4	HANGZHOU NUMERAL TWO
〣	3023	12437	A2C5	HANGZHOU NUMERAL THREE
〤	3024	12438	A2C6	HANGZHOU NUMERAL FOUR
〥	3025	12439	A2C7	HANGZHOU NUMERAL FIVE
〦	3026	1243A	A2C8	HANGZHOU NUMERAL SIX
〧	3027	1243B	A2C9	HANGZHOU NUMERAL SEVEN
〨	3028	1243C	A2CA	HANGZHOU NUMERAL EIGHT
〩	3029	1243D	A2CB	HANGZHOU NUMERAL NINE
ISO/IEC 10646-1:2000 (Ideographic Extension Block A)				
〸	3038	1243E	A2CC	HANGZHOU NUMERAL TEN
〹	3039	1243F	A2CD	HANGZHOU NUMERAL TWENTY
〺	303A	12440	A2CF	HANGZHOU NUMERAL THIRTY

The following table contains the different mappings for the ideographs 〸 and 〺:

[4] The comments in the Unicode Consortium's (informative) mapping tables between Unicode and CNS 11643-1992 and Big Five mention this. There are other code-points standing in the way of round-tripping between Unicode and these national encodings.

	Hanzi/Kanji					Hangzhou Numerals		
	JIS X 0208	**GB 2312**	**Big Five**	**CNS**	**ISO 10646-1 (1993)**	**Big Five**	**CNS**	**ISO 10646-1 (2000)**
十	29-29	CAAE	A451	14432	5341	A2CC	1243E	3038
卅	50-33	D8A6	A4CA	1452B	5345	A2CF	12440	303A

3 What's in a name?

Where exactly does the name for these characters come from? I have been unable to find any documentary evidence for the use of the name"Hangzhou". The majority of references I have been able to find call them the "Suzhou numerals." This section summarizes these findings..[5]

The Soviet Academy of Sciences' Большой Китайско-Русский Словарь (*Bolshoy Kitaysko-Russkiy Slovar*, Encyclopedic Chinese-Russian Dictionary) contains two entries for these numerals. In Volume 1 these numbers are named the "Chinese Common ('Suzhou', 'Commercial') Numerals", with the *hanzi* 蘇州碼子 *Sūzhōu mǎzi* (Suzhou numerals) given. The Dictionary presents horizontal variants to 丨, 𠄌, and 川, though it is unclear whether these are identical to 一, 二, 三 or should be considered separate ideographs. To numerals one through nine is added 〇, representing zero. The table also includes glyphs for 100 and 1000, but not twenty, thirty, or 10,000.

In his dictionary, DeFrancis (1996) glosses the entry for *Sūzhōu mǎzi* (using the simplified form 苏州码子) as, "Suzhou numerals (used by shopkeepers to mark prices)."

The entry for 蘇州碼子 in 汉语大词典 *Hànyǔ dà Cídiǎn* (Big Chinese Dictionary) Volume 9 states that these numerals were originally used by the people of Suzhou to write numerals, and were later spread to the rest of the country. It lists the numbers one through ten in the forms described above. The entry also gives an alternate name, 草码 *cǎomǎ* (lit. "grass numbers.") DeFrancis also contains an entry for *cǎomǎ*, which he glosses as, "ancient numerical symbols."

The 中文大辞典 *Zhōngwén dà Cídiǎn* (Encyclopedic Dictionary of the Chinese Language) Volume 8 states that the people of Suzhou first used these numerals. The entry shows the numerals one through nine.

In his book *The Universal History of Numbers: From Prehistory to the Invention of the Computer*, Georges Ifrah names these ideographs *gán mà zí*, or "secret marks" [his translation].[6] They are described as the "form used by traders to display the prices of goods." The ideographs for the numerals one thru ten, 100, 1,000, and 10,000 are presented. The glyphs used for 10,000 and 100,000 are 千 and 万, which are identical to those used when using "normal" *hanzi*

[5] This section uses Hanyu Pinyin to romanize Mandarin, and the LSHK Romanization for Cantonese.

[6] Ifrah does not show the *hanzi* used to write this name, and I'm unable to determine the correct translation for these. One *hanzi* read as *gǎn*, 秆, has the meaning "stem, stalk" which may relate this name with *cǎomǎ*, though this is pure conjecture at the present time.

numerals. The glyph for 100 is presented in two forms: one that looks like the Greek capital letter theta and the other like the Arabic numeral three.

Ifrah also indicates that these numbers are used in Japan, where he calls them “commercial forms.” While the ideographs for 十 and 卅 are found JIS X 0208, the remaining numerals are not encoded, and I have not been able to find further evidence that these numbers are in current use in Japan.

One Cantonese dictionary refers to these numbers as 商碼 *soeng1 maa5* “Shop Units” (according to Michael Bauer: I have not yet been able to see this for myself.) These forms are in regular use in Hong Kong, according to Ottfried Cheong.

4 Conclusion

Based on the evidence presented above, it would appear that the so-called “Hangzhou” numerals have been misnamed in ISO 10646.1. They should properly be called “Souzhou” numerals.

Acknowledgements

A number of people provided assistance to me while researching the information contained in this paper. Thanks to Dirk Meyer for going through his extensive library and scanning relevant material for me (and to Ken Lunde for passing me off to Dirk); Dmitri Zdorov for his help with the Russian sources; Ying Su for her help on the Chinese sources; and Jack Halpern for his expertise in lexicography and Japanese. I would also like to thank several members of the Unicode mailing list who responded to my initial query regarding the “Hangzhou” numerals: Patrick Andries, Michael Bauer, Otfried Cheong, Michael Everson, Scott Horne, and Frank Tang. Finally I want to thank Heather Emerson for proof-reading the final copy.

References

Академия Наук СССР, Институт Востоковедения. Большой Китайско-Русский Словарь, Том 1. Издательство "Наука", Москва 1983.

Академия Наук СССР, Институт Востоковедения. Большой Китайско-Русский Словарь, Том 3. Издательство "Наука", Москва 1983

Georges Ifrah. *The Universal History of Numbers: From Prehistory to the Invention of the Computer*. John Wiley & Sons, Inc., 2000.

John DeFrancis, ed. *ABC Chinese-English Dictionary*. University of Hawaii Press. 1996.

Ken Lunde. *CJKV Information Processing*. O’Reilly and Associates. 1999.

The Unicode Consortium. *The Unicode Standard, Version 2.0*. Addison-Wesley. 1996.