



Universal Multiple Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: Working Group Document

Title: Proposal to disunify certain fencing CJK punctuation marks from similar-looking Math fences

Source: Kent Karlsson

Status: Action item from Ad-hoc on Mathematical Symbols

Action: For consideration by JTC1/SC2/WG2 and UTC

Related: N2336

Math and CJK fencing punctuation characters are not treated consistently in the UCS. This is a proposal to treat them in a consistent manner and in that way to avoid typographic issues with using CKJ punctuation for math expressions.

Already disunified in practice

The following math and CJK fence symbols are already disunified, and exist in forms suitable for math and for CJK punctuation respectively:

- () (ASCII/math (0028, 0029) vs. full width-ASCII/CJK (FF08, FF09))
- [] (ASCII/math (005B, 005D) vs. full width-ASCII/CJK (FF3B, FF3D))
- { } (ASCII/math (007B, 007D) vs. full width-ASCII/CJK (FF5B, FF5D))

"Disunified" but disunification broken

The UCS originally provided for a disunification of the angle brackets as used in math and the single quotation marks as used in CJK.

- ⟨ ⟩ (CJK: big ⟨ ⟩ (3008, 3009); math: tuple fences (2329, 232A))

However, Unicode introduced a canonical mapping from 2329 to 3008 and from 232A to 3009. Although these canonical mappings should be seen as mistakes, at this time it is too late to change them. The canonical mappings for existing characters have been frozen for at least two years. So they are now in effect unified. For a disunification, new single angle brackets/braces need therefore to be added.

Currently unified

The following CJK fences in the CJK Symbols and punctuation block are also currently intended for math usage.

- 《 》 (300A, 300B) (CJK: big « »; math: Z notation/chevron brackets)
- [[]] (301A, 301B) (CJK; math/cs: abstract syntax fences)
- () (3014, 3015) Tortoise shell brackets
- () (3018, 3019) White tortoise shell brackets

The following is in the (new; amd.1) Miscellaneous mathematical symbols block are also used for CJK:

- () (2985, 2986)

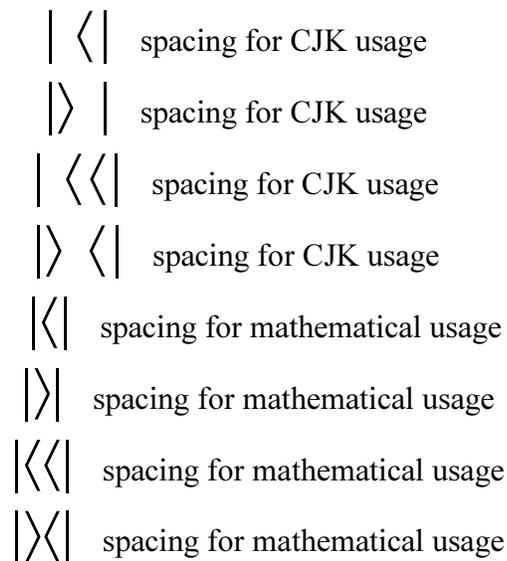
The CJK glyph may have a double parenthesis appearance.

CJK only

The rest of the fences from CJK punctuation block are currently assumed to be used in CJK context only.

Math only

- () (amd.1: 2997, 2998)
- { } (amd.1: 2983, 2984)
- () (amd.1: 2987, 2988)
- ‹ › (amd.1: 2989, 298A)
- and several more math fences in amd.1



where | marks the edge of the preceding or following character.

Figure 1.: Spacing and layout of brackets

Motivations for disunification of CJK punctuation from math fencing punctuation

1. Math fencing punctuation has typographic properties that are different from CJK fencing punctuation. In many cases it is sufficient to just typeset the glyphs for the characters as most glyphs are typeset (in particular: no automatic deletion of inherent whitespace; there aren't ever excessive amounts of it). In some cases, the fencing brackets size up with the size of the math expression fenced. This does not apply to CJK punctuation.
2. On the other hand the CJK fencing punctuation not only have inherent whitespace, but sometimes this inherent whitespace should be removed during typesetting for aesthetic reasons. (See figure 1)
3. The angle/curvature, baseline position, and sizing differ between the math use and the CJK use, even when the inherent whitespace is removed from the CJK fences. Thus, just changing "beginning of line" to "beginning of line or math context" does still not result in the proper glyphs to be used, unless the glyph is also changed for in a math context.
4. Further, without markup it is not always easily determinable whether it is a math use or a CJK use in a text with CJK text and math expressions mixed.
5. The "inherent whitespace removal" for CJK is often done through code position testing, not through "virtual ink testing". Therefore all fonts should uniformly have the inherent whitespace for the CJK fences. This is not expected for any of the math fences.

Therefore it is not helpful to have some CJK fences unified with some (similar-looking) math fences.

Proposal

Let the fences in the CJK punctuation block all be wide with CJK specific whitespace.

Add in a "math symbols" block:

```
2B00 MATHEMATICAL LEFT WHITE SQUARE BRACKET
2B01 MATHEMATICAL RIGHT WHITE SQUARE BRACKET
2B02 MATHEMATICAL LEFT ANGLE BRACKET
2B03 MATHEMATICAL RIGHT ANGLE BRACKET
2B04 MATHEMATICAL LEFT DOUBLE ANGLE BRACKET
2B05 MATHEMATICAL RIGHT DOUBLE ANGLE BRACKET
```

Rename these two characters in amd1

```
2985 MATHEMATICAL WHITE LEFT PARENTHESIS
2986 MATHEMATICAL WHITE RIGHT PARENTHESIS
```

Add in a "CJK block"

```
33DE, WHITE LEFT PARENTHESIS
33DF, WHITE RIGHT PARENTHESIS
```

Code positions to be determined by WG2.

Postpone for further study

- , MATHEMATICAL LEFT TORTOISE SHELL BRACKET
- , MATHEMATICAL RIGHT TORTOISE SHELL BRACKET
- , MATHEMATICAL LEFT DOUBLE TORTOISE SHELL BRACKET
- , MATHEMATICAL RIGHT DOUBLE TORTOISE SHELL BRACKET

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

Please fill Sections A, B and C below. Section D will be filled by SC 2/WG 2. For instructions and guidance for filling in the form please see the document "Principles and Procedures for Allocation of New Characters and Scripts"

(<http://www.dkuug.dk/JTC1/SC2/WG2/prot>) 1.1. A. Administrative

1. Title: **Disunify braces/brackets for math, computing science, and Z notation from similar-looking CJK braces/brackets**

2. Requester's name: **Kent Karlsson**

3. Requester type (Member body/Liaison/Individual contribution): **Individual contribution**

4. Submission date: **2001-04-04**

5. Requester's reference (if applicable):

6. (Choose one of the following:)
This is a complete proposal

1.2. B. Technical - General

1.b. The proposal is for addition of character(s) to an existing block: **X**
Name of the existing block: **Miscellaneous mathematical symbols**

2. Number of characters in proposal: **8**

3. Proposed category (see section II, Character Categories): **A (contemporary)**

4. Proposed Level of Implementation (see clause 15, ISO/IEC 10646-1): **1**

Is a rationale provided for the choice? **Yes**

If Yes, reference: **(simple graphical characters, no combining or other implementation difficulties)**

5. Is a repertoire including character names provided?: **Yes** a. If YES, are the names in accordance with the 'character naming guidelines' in Annex K of ISO/IEC 10646-1? **Yes**

b. Are the character shapes attached in a reviewable form? **(disunification proposal, see description below)**

6. Who will provide the appropriate computerized font (ordered preference: True Type, PostScript or 96x96 bit-mapped format) for publishing the standard?

If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

7. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

8. 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): **No.**

1.3. 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.)? 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? **No.**

Reference:

4. The context of use for the proposed characters (type of use, common or rare) **Common in math, computing science, and Z notation**

Reference:

5. Are the proposed characters in current use by the user community? **Yes**

If YES, where? Reference:

6. After giving due considerations to the principles in N 1352 must the proposed characters be entirely in the BMP? **Yes**

If YES, is a rationale provided? **Yes**

If YES, reference: **(Co-location with similar (and less used) characters in the misc. math. symbols block.)**

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **Nearly (see detailed proposal below).**

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? **No, though there are characters that look similar, but have other typographic properties.**

If YES, is a rationale for its inclusion provided?

If YES, reference:

9. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? **Yes.**

If YES, is a rationale for its inclusion provided? **Yes.**

If YES, reference: **Though similar in appearance to some CJK punctuation, the use context, typographic appearance, and typographic spacing properties are different.**

10. Does the proposal include use of combining characters and/or use of composite sequences (see clause 4.11 and 4.13 in ISO/IEC 10646-1)? **No.**

If YES, is a rationale for such use provided?

If YES, reference:

Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

If YES, reference: **N/A**

11. 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)

1.4. D. SC 2/WG 2 Administrative (To be completed by SC 2/WG 2)

1. Relevant SC 2/WG 2 document numbers:

2. Status (list of meeting number and corresponding action or disposition):

3. Additional contact to user communities, liaison organizations etc:

4. Assigned category and assigned priority/time frame:

E. References

ISO/IEC WD2.6 13568, Formal Specification 'Z Notation Syntax, type and semantics.

Donald E. Knuth, The TeX, Book

Leslie Lamport, LaTeX

Mathematical and CS text books and literature