

Title: Comments on JCS Proposals
Source: Unicode Technical Committee
Author: Lisa Moore, Chair, UTC
Distribution: Kohji Shibano, Chairman, JCS Committee
 Takayuki Sato, Japanese SC2
 Zhang Zhoucai, Rapporteur, IRG
 Mike Ksar, Convenor, JTC1 SC2/WG2
 Arnold Winkler, Co-chair, UTC
Action: For Review and Response by JCS
Date: November 23, 1999

The members of Unicode Technical Committee (UTC) wish to thank Kohji Shibano, Chairman, JIS Coded Character Set (JCS) Committee, Japan Standards Association for forwarding the JCS proposals for our consideration. These proposals were reviewed at UTC #81, the week of October 26-29, 1999. The UTC took a number of actions with regards to the proposed JCS characters and had a number of questions for which we would welcome answers.

The UTC had major concerns with aspects of these proposals and the recently balloted standard JIS X 0213:

Y JIS X 0213 gives character mappings to unassigned Unicode characters. These mappings are invalid and use of them is not conformant to the Unicode Standard or to ISO/IEC 10646. As is made apparent in the detailed results which follow, the UTC has accepted only a few of the JCS characters at their proposed code positions.

Y The UTC strongly discourages encoding further precomposed characters which can be represented with combining characters already in the standard. A new normalization form, canonical composition, was defined in the Unicode Standard, Version 3, based on the Unicode Version 3 Character Database. Many companies and organizations (including the W3C) are adopting this new normalization form, and it is expected that most programs will use normalized data. For stability, the normalized form of new precomposed characters will be the decomposition to a base character plus combining characters. Thus there is little value in adding new precomposed characters. For more information, see Unicode Technical Report #15 (<http://www.unicode.org/unicode/reports/tr15/>).

The detailed results of the UTC discussions follow, organized by proposal.

1) *Fifty Six Kanji Compatibility Ideographs*. Because the UTC had originally proposed the addition of the fifty six Kanji compatibility characters during the development of the URO, the UTC now supports the addition of the fifty six Kanji characters and will relay this position to the IRG. We also support the proposed code positions given in your proposal (FA30..FA67).

We request that you provide us with the compatibility mappings, as these mappings are required for the Unicode Standard.

2) *Seven Hiragana Characters*. The UTC accepted the two small Hiragana characters at the proposed code sequences:

| | |
|--------------------------|------|
| HIRAGANA LETTER SMALL KA | 3095 |
| HIRAGANA LETTER SMALL KE | 3096 |

Such acceptance is provisional, since the Unicode consortium and ISO/IEC SC2/WG2 maintain synchronization between the Unicode Standard and ISO 10646. This requires that both organizations agree to the characters before they will be added to the respective standards.

The five extended Hiragana characters were not accepted because they are already represented in the Unicode Standard by the following character code sequences:

| | | |
|--|------|------|
| HIRAGANA LETTER KA WITH SEMI-VOICED SOUND MARK | 304B | 309A |
| HIRAGANA LETTER KI WITH SEMI-VOICED SOUND MARK | 304D | 309A |
| HIRAGANA LETTER KU WITH SEMI-VOICED SOUND MARK | 304F | 309A |
| HIRAGANA LETTER KE WITH SEMI-VOICED SOUND MARK | 3051 | 309A |
| HIRAGANA LETTER KO WITH SEMI-VOICED SOUND MARK | 3053 | 309A |

3) *Twenty Five Katakana Characters*. The UTC accepted the sixteen small Katakana characters at the following code sequences:

| | | |
|--------------------------|------|------|
| KATAKANA LETTER SMALL KU | 31F0 | |
| KATAKANA LETTER SMALL SI | | 31F1 |
| KATAKANA LETTER SMALL SU | 31F2 | |
| KATAKANA LETTER SMALL TO | 31F3 | |
| KATAKANA LETTER SMALL NU | 31F4 | |
| KATAKANA LETTER SMALL HA | 31F5 | |
| KATAKANA LETTER SMALL HI | 31F6 | |
| KATAKANA LETTER SMALL HU | 31F7 | |
| KATAKANA LETTER SMALL HE | 31F8 | |
| KATAKANA LETTER SMALL HO | 31F9 | |
| KATAKANA LETTER SMALL MU | 31FA | |
| KATAKANA LETTER SMALL RA | 31FB | |
| KATAKANA LETTER SMALL RI | 31FC | |
| KATAKANA LETTER SMALL RU | 31FD | |
| KATAKANA LETTER SMALL RE | 31FE | |
| KATAKANA LETTER SMALL RO | 31FF | |

Such acceptance is provisional, since the Unicode consortium and ISO/IEC SC2/WG2 maintain synchronization between the Unicode Standard and ISO 10646. This requires that both organizations agree to the characters before they will be added to the respective standards.

Note: The code position allocations are not the same as those in the JCS proposal. The extended small Katakana character was not accepted because it will be represented in the Unicode Standard by the following character code sequences:

| | | |
|--------------------------|------|------|
| KATAKANA LETTER SMALL PU | 31F7 | 309A |
|--------------------------|------|------|

The eight extended Katakana characters were not accepted because they are already represented in the Unicode Standard by the following character code sequences:

| | |
|--|-----------|
| KATAKANA LETTER KA WITH SEMI-VOICED SOUND MARK | 30AB 309A |
| KATAKANA LETTER KI WITH SEMI-VOICED SOUND MARK | 30AD 309A |
| KATAKANA LETTER KU WITH SEMI-VOICED SOUND MARK | 30AF 309A |
| KATAKANA LETTER KE WITH SEMI-VOICED SOUND MARK | 30B1 309A |
| KATAKANA LETTER KO WITH SEMI-VOICED SOUND MARK | 30B3 309A |
| KATAKANA LETTER SE WITH SEMI-VOICED SOUND MARK | 30BB 309A |
| KATAKANA LETTER TU WITH SEMI-VOICED SOUND MARK | 30C4 309A |
| KATAKANA LETTER TO WITH SEMI-VOICED SOUND MARK | 30C8 309A |

4) *Forty Enclosed Numbers*. The UTC will discuss in the future a general mechanism for applying a mark to a sequence of characters. This general mechanism will address the JCS proposal for additional circled numbers. This topic will be an agenda item to be covered at a future UTC meeting.

5) *Sixteen Publishing Characters*. The UTC accepted the following four characters at the proposed code sequences:

| | |
|----------------------|------|
| DOUBLE QUESTION MARK | 2047 |
| WHITE SHOGI PIECE | 2616 |
| BLACK SHOGI PIECE | 2617 |
| RETURN SIGN | 2618 |

Such acceptance is provisional, since the Unicode consortium and ISO/IEC SC2/WG2 maintain synchronization between the Unicode Standard and ISO 10646. This requires that both organizations agree to the characters before they will be added to the respective standards.

The remaining twelve characters were not accepted due to insufficient information on their usage. Please provide to the UTC examples of usage in documents (not just in code charts), and explain if any of these twelve characters are used for emphasis or as combining characters.

6) *Twenty Seven Dentist Characters*. The UTC will consider the ten double circled numbers as part of the general mechanism to be defined in the future. See 4) above. The remaining seventeen dentist symbols were not accepted due to insufficient evidence of usage. Please provide documents with examples of usage, and explain if any of these characters are combining, or if any extend across other symbols to delineate quadrants of the jaw.

7) *Fourteen Linguistic Education Characters*. The nine precomposed Latin characters were not accepted because they are already represented in the Unicode Standard by the following character code sequences:

| | |
|--|-----------|
| LATIN SMALL LETTER AE WITH ACUTE | 00E6 0301 |
| LATIN SMALL LETTER OPEN O WITH GRAVE | 0254 0300 |
| LATIN SMALL LETTER OPEN O WITH ACUTE | 0254 0301 |
| LATIN SMALL LETTER TURNED V WITH GRAVE | 028C 0300 |
| LATIN SMALL LETTER TURNED V WITH ACUTE | 028C 0301 |
| LATIN SMALL LETTER SCHWA WITH GRAVE | 0259 0300 |

| | |
|--|-----------|
| LATIN SMALL LETTER SCHWA WITH ACUTE | 0259 0301 |
| LATIN SMALL LETTER HOOKED SCHWA WITH GRAVE | 025A 0300 |
| LATIN SMALL LETTER HOOKED SCHWA WITH ACUTE | 025A 0301 |

The two spacing modifier letters were not accepted because they are already represented in the Unicode Standard (see *The Unicode Standard, Version 2*, page 6-13) by the following character code sequences:

| | |
|----------------|-----------|
| RISING SYMBOL | 02E9 02E5 |
| FALLING SYMBOL | 02E5 02E9 |

The two arrow characters (RISING ARROW and FALLING ARROW) will be added to the Math and Technical Symbols proposal for future encoding. Code positions were not assigned.

8) *313 New Kanji Characters*. The UTC took no action on these proposed ideographic characters due to a number of serious concerns:

Ÿ Many of the proposed radicals are already encoded, such as AB99 (encoded at 2ECC), AB6C (encoded at 2EC0), AB6D (encoded at 2EBF), and ABBE (encoded at 2EDE).

Ÿ There are glyph variants of unified characters

Ÿ It is unclear if these 313 new ideographs already included in Extension B for encoding in Plane 2.

Ÿ If these characters are not in Extension B, then they must be proposed to the IRG for resolution