ISO/IEC JTC1/SC2/WG2 N 4379

Date: **2012-10-25**

ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)

Doc. Type: **Input to ISO/IEC 10646:2012**

Title: **Editor Response to Japanese translation issues concerning ISO/IEC 10646: 2012**
Source: **Unicode (authored by Michel Suignard)**
Project: **JTC1 02.10646**
Status: For review by WG2
Date: 2012-10-25
Distribution: WG2
Reference: WG2 N4365

Summary

The following document provides an answer to issues found by Japan when translating ISO/IEC 10646:2012 into Japanese. These issues will be addressed in the next edition of the standard. In the meantime, Japan can use this feedback to complete the translation of the current edition of the standard.

a) In 4.1, NOTE 2, "They can be" appears a mistake and should read as "There can be". It further appears to be better to be rewritten as "There are" to avoid "can".

This is correct. The sentence will read:
There are exceptions for some complex writing systems.

b) In 4.5, NOTE, "A character" should be "A graphic symbol".

This is correct. The sentence will read:
A graphic symbol can be represented by a sequence of one or several coded characters.

c) In 4.44, the number "65536" should be written as "65 536" to follow the directives.

This is correct. The definition will read:
subdivision of the UCS codespace consisting of 65 536 code points

d) In 4.49, the definition for row is now incomplete, because we removed detailed specification formerly placed in clause 6. The suggested definition is: "subdivision of a plane consisting of contiguous 256 code points beginning at a multiple of 256."

This is correct. To also address the concern raised in u) the definition will read:

subdivision of a plane consisting of contiguous 256 code points beginning at a multiple of 256 which can be identified by a number from 00 to FF
Note however that the plane definition could potentially be perceived as having the same issues as the row definition and this is not raised by Japan.

e) In 4.59, the definition contains an undefined term surrogate code unit. The current 10646 has definitions of high-surrogate code unit and low-surrogate code unit, but it has no definition of (unqualified) surrogate code unit. In this particular context, just say "code unit" in place of "surrogate code unit" seems sufficient. Alternatively, we could add a definition of surrogate code unit as "either high-surrogate code unit or low-surrogate code unit".

To avoid renumbering of the terms in clause 4 it is sufficient at this stage to simply replace 'surrogate code unit' with 'code unit'. The definition will read:
code unit in a code unit sequence that is either a high-surrogate code unit that is not immediately followed by a low-surrogate unit, or a low-surrogate code unit that is not immediately preceded by a high-surrogate code unit

f) In 6.3.8, the sentence "Future editions of this International Standard will not allocate any characters to these reserved code points." is wrong, because this subclause is for code points reserved for future standardization. The sentence should read as "Future editions of this International Standard may allocate characters to some of these reserved code points."

This is correct. The sentence will read:
Future editions of this International Standard may allocate characters to some of these reserved code points.

g) In 9.1, there is a word "CC-sequences", which should be "code unit sequences."

This is correct. The second sentence in the third item of the list will read:
This provides compatibility with existing file-handling systems and communications sub-systems which parse code unit sequences for these octet values.

h) In 23.2, the 5th item, "5h field" should be "5th field" ("t" is missing).

This is correct. The fifth item will read:
5th field: Kanji J sources

i) In 23.2, the 5th item, the format for JH source is specified as "(JH-xxxxxx)" but it is wrong. JH source reference is either 6 digits or 7 digits, so the 7 digits version "(JH-xxxxxxx)" should be inserted after the 6 digits version, e.g., "(JA-hhhh), (JH-xxxxxx), (JH-xxxxxxx), (JK-ddddd)". (Note that the last digit for the 7 digits version of JH source reference is always "S", so we can write the format as "(JH-xxxxxxS)" if it is preferred.).

This is correct. A new source will be inserted after (JH-xxxxxx):
(JH-xxxxxxS)

j) In 31.2, EXAMPLE, the character name for 01C9 is written as "LATIN SMALL LETTER IJ", but it should be "LATIN SMALL LETTER LJ" ("LJ" for "IJ").

This is correct. The line will read:
01C9 lj LATIN SMALL LETTER LJ

k) In A.1, the code point range for the collection 105 RTL ALPHABETIC PRESENTATION FORMS uses 2013 EN DASH to indicate a range, whereas other ranges use 002D HYPHEN-MINUS.

This is correct. The collection definition will read:
105 RTL ALPHABETIC PRESENTATION FORMS FB1D-FB4F

l) In A.1, definition for the collection 401 is inappropriate; it says "G=00" to mean "Group 00", but the notion of the group has been removed from 10646. We can simply erase it. The notation "P=" looks strange, because specification that explained such notation (in clause 6) has been removed. It's better to say "Planes 0F and 10".

This is correct. The collection definition will read:
401 PRIVATE USE PLANES-0F-10 Planes 0F and 10

m) In A.1, NOTE 3 (the list of keywords appearing in the collection names), there is an entry "Counting Rod numerals" (with C and R in uppercase), but it was "Counting Rod Numerals" (C, R, and N in uppercase) in the previous edition.

There is no justification to upper case the term 'numerals' in this case. It is also not consistent with the convention used for other items in the list. Therefore it will stay unchanged. However, when reviewing the list, two inconsistencies with this rule were found concerning 'Egyptian Hieroglyphs' and 'Game Tiles'. These entries will read:
Egyptian hieroglyphs 1031
Games tiles 1028 1029

n) In A.1, NOTE 3, the collection numbers for the entry "Game Tiles" is written as "1028, 1029" (with comma), but it should be "1028 1029" (without comma.)

This is correct. The entry will read
Games tiles 1028 1029

o) In A.1, NOTE 3, the entry "Nko" should be spelled as "NKo" (with uppercase K.)

This is correct. The entry will read
NKo 128

p) In A.4.2, NOTE 2, the file name "JIEx.txt" should be "JIExt.txt" (with "t" before a dot.)

This is correct. The sentence will read
The file is named: "JIExt.txt".

q) In A.5.7, it says "Planes 00-10", but it should be "Plane 00", because all code points of BASIC JAPANESE are on the Plane 00.

This is correct. The first plane definition will read:
Plane 00

r) In Annex I, there are I.1.1 and I.1.2 but no I.1. They should be renumbered as I.1 and I.2, respectively.

This is correct. There will be renumbered as I.1 and I.2 in the next edition. This will also make them visible in the table of contents.

s) Throughout A.6.x, there are many occurrences of a phrase "a collection from A.1" (or its plural form). Repeating "from A.1" many times doesn't make sense. It's better to remove the "from" part and simply to say "a collection" or "collections". Alternatively, each "from" part can refer to the exact subclause that gives the definition, as in past editions.

Collection definitions appear in many sub-clauses of annex A, not just A.1. While the phrase is repetitive in some way, it makes sense. No change will be made.

t) In Annex P, NOTE, 1st line, a closing parenthesis should be inserted after "amendments 1 to 5".

This is correct. The sentence will read:
The first edition of this International Standard (ISO/IEC 10646:2003 and amendments 1 to 5) used this annex to provide additional information on all characters.

u) In many places such as 17, 22.2, 26, or A.3, there are texts that assume each row has an identification number in range 00-FF. It was true in the previous editions, but the current edition dropped it. Re-introducing the notion of "row number" seems a better idea than rewriting all related texts.

This is correct in essence. However to avoid the addition of a new term, the definition of the term row can further modified to include the numbering. The definition for row will read:
subdivision of a plane consisting of contiguous 256 code points beginning at a multiple of 256 which can be identified by a number from 00 to FF

-end