

ISO
International Organization for Standardization
Organisation Internationale de Normalisation
ISO/IEC JTC 1/SC 2/WG 2
Universal Multiple-Octet Coded Character Set
(UCS)

ISO/IEC JTC 1/SC 2/WG 2 N1622
July 3, 1997

Title: Comments on Technical Issues of Mongolian in N 1515 and N1607
Source: Oliver Corff (DE)
Status: Expert contribution
Action: For the consideration of SC 2/WG 2
Distribution: ISO/IEC JTC 1/SC 2/WG 2

1 Document Status and Present Situation

Document N1510 which presented a fairly mature encoding for Mongolian and derived scripts was presented in August 1996. Certain issues relevant to the usage of control codes for selecting appropriate glyph shapes for Mongolian and its derived scripts have stirred controversial opinions among NBs and individual experts. These different opinions crystalized in N1515 (Meeting #32, Singapore, Jan. 1997, source: UNU/IIST and Unicode) and N1607 (Meeting #33, Crete, June/July 1997, source: China).

2 Target Issues

The following items were found to be controversial:

1. *Mongolian Space*: MSP. Mongolian has a characteristic intra-word space which is used between word stems and various types of finals ('floating' vowels and grammatical endings). Document N1515 suggests to use the non-breaking space (NBSP) instead of the MSP, but it is argued in N1607 that the behaviour and functionality of NBSP and MSP is substantially different and that the two characters should not be unified.

2. *Variant Selectors*. Beyond the behaviour determined by 'usual' linguistic constraints (also known as 'context'), a Mongolian character may change its shape in a seemingly 'irregular' manner in foreign loans etc. where usual context rules do not apply. In order to control the character behaviour, the first document (N1510) introduced a set of three variant selectors because probing into the matter had revealed that the generic joiner and non-joiner characters provided by ISO 10646 are not fully functional in the Mongolian environment.

N1515 abolished the three Mongolian variant selectors and introduced a set of *ten* control characters to fulfill the same purpose after it was discovered that there were some rule conflicts in the original interaction of textual characters and control characters. It was also suggested in N1515 to exclude these new control characters from the Mongolian table and relocate them to a common set of control characters for redefinable use by other languages.

Though the word *control character* is used in various places (not only in this document, but also in N1515), the Unicode Consortium disallows the usage of *control code* and *control character* and favours *variant selector* instead. The character name lists in N1510 and N1515 use the word *variant selector* in the tables.

3. *Repertoire Issues*. Document N1510 contains a number of characters which were either questioned as far as their use is concerned (like no. 6, Mongolian Combinatory Symbol [?!]) or which could possibly be unified with existing characters.

3 Suggested Actions

China (in document N1607) has already answered some of the questions raised in N1515 and suggested to hold an experts' meeting in Mongolia in August 1997, benefiting from the circumstance that the International Congress of Mongolian Studies will be held in Ulaanbaatar where many experts will be in Ulaanbaatar anyway. Mongolia responded positively to this suggestion, and all interested parties should be invited if this has not already happened.

At the meeting in August, the following items shall be discussed:

1. *Mongolian Space*. The function of the Mongolian Space should be defined so as to justify the inclusion of this symbol into the Mongolian Character Set.

2. *Variant Selectors*. The Variant Selector issue actually requires several different definitions and decisions:

(a) *Control Rules*. N1515 pointed out correctly that variant selectors of N1510 did not show a very unified set of application rules. Some variant selectors follow the character they control, some precede it, causing those conflicts pointed out by N1515. For the interaction of textual characters and control characters, a Regular Language must be defined which allows the generation of all desired combinations while prohibiting the erroneous generation of faulty character sequences. The Unicode Consortium points out that the variant selectors so far do not have well defined intrinsic semantics. It will be one aspect of defining the rules to clearly define the functionality of each variant selector.

(b) *Set Size*. It is not immediately obvious why a set of variant selectors should include as many as ten elements. In combination with a set of rules the original three selectors should be sufficient.

(c) *Set Location*. Variant opinions exist whether a set of variant selectors should be made specific to one given script or should rather be available to all scripts. The Unicode Consortium prefers to see a solution where these selectors, their intrinsic semantics being defined, pertain specifically to one script, i.e. (in this case) to Mongolian while others prefer to see a general solution where variant selectors are available out of a common pool.

3. *Repertoire Issues*. The characters which are discussed by N1515 and N1607 should be investigated again. Evidence revealing their nature should be brought forward.

3.1 Next WG 2 meeting

The experts' meeting to be held in Mongolia in August 1997 is expected to issue a revised Mongolian Character Set Proposal that can then be submitted to the next WG2 meeting for voting. That meeting will take place in March 1998.