

L2/00-0032

1 February 2000

Compatibility Ideographs in the Unicode Standard

John H. Jenkins
Apple Computer, Inc.
jenkins@apple.com

Unicode 1.0 included a Compatibility Ideographs block for ideographs which were required by Unicode member companies but which were not a part of national standards or were duplicates. This block has been essentially unchanged since then, except that in Unicode 3.0 a number of these “compatibility” ideographs are now treated as part of the CJK Unified Ideographs set.

In the fall of 1999, the UTC was asked to consider a proposal to add 56 new compatibility ideographs to this block for the sake of round-trip compatibility with JIS X 0213. Given the importance of the new JIS standard, the relatively small number of characters involved, and the fact that they would all fit within the bounds of the currently unused portions of the Compatibility Ideographs block, the UTC approved the idea.

At the 14th IRG meeting in Singapore, December 1999, Taiwan observed that the same need potentially exists for other important East Asian character encoding standards, most notably CNS 11643-1992. A total of some 700 compatibility ideographs added to Unicode would provide complete round-trip compatibility with CNS 11643-1992; if they were to be added as compatibility ideographs, moreover, there would be no confusion as to their status or intended use. Taiwan further, recognizing that there is no room for additional compatibility ideographs in the BMP, recommended that a 2048-character block in Plane 2 be earmarked for compatibility ideographs. A survey of members at the meeting revealed that this is nearly twice as big as current needs might possibly require.

Discussion at the IRG meeting showed a general agreement with Taiwan’s overall position. Hong Kong, at least, also has a set of ideographs they might desire to have added as compatibility ideographs; other members were less certain as to whether or not they would have submissions.

The main issues that came up in the discussion were whether or not submissions for compatibility ideographs should be unified against each other (e.g., if Taiwan and Hong Kong both submit the same compatibility ideograph, only one code point should be allocated for it, not two), and, more seriously, how best to *restrict* the submission process so as to avoid overwhelming it with every glyphic variant of

every ideographs in existence.

Another concern raised was that currently the IRG has no charter for dealing with compatibility ideographs; that is solely the provenance of WG2. It was determined, however, that it would be advisable to solicit WG2's direction in the matter.

So far as the UTC and L2 are concerned, we should be willing to provide input to WG2 on the matter. Some issues for discussion—

Does the UTC or L2 want to make a submission for compatibility ideographs (e.g., from EACC)?

Do we support the addition of a relatively small number of ideographs for full round-trip compatibility with CNS 11643-1992?

Do we support the idea of setting aside 2048 code points on Plane 2 for compatibility ideographs?

Do we support the idea of instructing the IRG to unify proposals for new compatibility ideographs against one another?

How do we feel that the potential problem of large submissions of compatibility ideographs can be solved?