Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Background

This document provides additional evidence for four (4) UNC Version 3.0 "D" set ideographs, specifically #025, #069, #072, and #111. The evidence for #025 suggests that unification should not take place, and that this ideograph be moved to the "M" set. The evidence for the other three ideographs supports unification, and simply serves to reinforce others' recommendation to unify them.

Additional UNC #025 Evidence

UNC #025 is an interesting case, because U+43EE (Extension A) established a precedent that suggests that unification with U+5354 should not take place. The first line of scans in the table below are from *The Modern Reader's Japanese-English Character Dictionary*, Second Revised Edition (Charles E. Tuttle Company, 1974), whose mappings are reflected in the "kNelson" field of the Unihan Database. The second line of scans are from its revision, entitled *The New Nelson Japanese-English Character Dictionary* (Charles E. Tuttle Company, 1974), and serves to reinforce the evidence from the earlier edition.

In the left column of the table below, UNC #025 (㕕) is shown to be a variant form of U+5354 (協). In the right column of the table below, U+43EE (胘) is shown to have a virtually identical relationship with U+8107 (脇), yet these ideographs were not unified. Considering that U+43EE is in Extension A, the Source Separation rule can be ruled out. For this reason, I propose that UNC v3.0 #025 be moved into the "M" set.

UNC #025 versus U+5354	U+43EE versus U+8107
㕕 $\frac{780}{F-X}$ Nonstandard for 協 774.	胘 $\frac{3756}{F-X}$ Nonstandard for 脇 3765.
㕕 $\frac{613}{J-X}$ $\frac{}{M-X}$ Nonstandard for 協 617.	胘 $\frac{4795}{J-X}$ $\frac{}{M29494}$ Nonstandard for 脇 4807.

There are other similar ideograph pairs in the BMP that follow this simplification pattern, specifically that the bottom two of three identical elements are reduced to four diagonal strokes, but their lack of unification can be attributed to the Source Separation rule. These example pairs are 塁/壘 (U+5841/U+58D8), 摂/攝 (U+6442/U+651D), 渋/澁 (U+6E0B/U+6F81), 畳/疊 (U+7582/U+758A), and 轪/轟 (U+8EE3/U+8F5F).

Additional UNC #069, #072 & #111 Evidence

A recently-published kanji dictionary entitled『新潮日本語漢字辞典』(新潮社, 2007) provides additional evidence that supports unification for the following "D" set ideographs: #069, #072, and #111. The table below includes scans from this reference.

In the first column, UNC #069 (互) is shown with the ideograph with which unification has been proposed, specifically U+5DE5 (工).

In the second column, UNC #072 (座) is shown with the ideograph with which unification has been proposed, specifically U+5EA7 (座). Of course, the glyph that is shown in the table below is not identical to the glyph for UNC #072, but it is identical to the glyph for UNC #071. UNC #071 and #072 use the same components, and differ only in the relative length of the tenth stroke.

Finally, in the third column, UNC #111 (涼) is shown with the ideograph with which unification has been proposed, specifically U+6DBC (涼).

UNC #069	UNC #072	UNC #111

This additional evidence helps to reinforce the variant relationship with the ideographs with which unification has been proposed.