To: UTC
From: Richard Cook and Deborah Anderson
RE: Comments on the Tangut radicals and strokes proposal (N3495 = L2/08-335)
Date: 29 October 2008

I Tangut Radicals and Components
L2/08-187 = N3467 outlined a path forward on two blocks of Tangut radicals, one based on the HXM (2004) set, and the other a supplement of various other radicals and components. The UK/IE proposal (N3495 = L2/08-335) defines a single block, interspersing radicals and components from various sources. Specific comments follow on the Tangut radicals and components portion of L2/08-335:

1. Ordering
    Although the ordering principles are mentioned briefly in N3495:S12, the ordering data is not explicitly given in the mapping table (S[ection]13).

2. Subsidiary Radicals and Non-Significant Glyph Variants
    So-called "subsidiary radicals" (mainly those of Kychanov?) are unified, but it is also suggested they be separated by means of IDS? Note, however, that IDS is applicable to CJK, and Tangut is not CJK. Likewise, so-called "non- significant" glyph variants are ignored by caveat, through it is not clear what "non-significant" might mean if it does not mean "unifiable". Unifications in the main table (S13) may generally be right, but careful checking in a multi-column *and* multi-row table framework will be required.

3. Formatting
    The glyph and mapping data in S13 should be formatted with the mappings in cells below each glyph. The main table breaks some rows across page breaks, and so is illegible in places.

4. Glyphs
    The representative glyphs for the proposed single-column chart derive from the old "Mojikyo" style glyphs (no longer proposed for the main Tangut block code charts). It would be preferable to use the Jing Yongshi font data, and incorporate a column for this in the S13 charts.

5. Blocks
    It may be preferable to encode two blocks (as outlined in L2/08-187 = N3467), a main

block with the subset of HXM radicals, unless it can be shown that the there is no possible need for a future radical supplement. If a radical supplement is anticipated, at very least we might choose to clearly distinguish components or minor radicals from the main block.

6. Approach for Unification/Separation Issues
   A UVS solution (as in L2/08-349 = N3521) might be the best way to address unification/separation issues.

Thus, significant work remains to be done in re-working, re- tabulating, and verifying the mapping data in S13, to prepare a finished proposal.

II Tangut Strokes

1. There are several problems with augmenting the CJK Strokes block for Tangut, not the least of which is the fact that Tangut is (still) not CJK. The principles governing encoding of strokes in this block are clearly set forth: strokes occurring in representative glyphs in CJK code charts are suitable for this block. To add Tangut strokes to this block, we would need to broaden the purpose to encompass "Siniform" strokes beyond CJK. This might not be objectionable, if a clear set of candidate strokes is found. However, from the presentation in S9, it is evident that the current proposal does not take into account the important principles underlying CJK stroke definitions nor the explicit unifications in CJK Stroke block work. The proposal does not clearly demonstrate the relation of encoded CJK strokes to proposed new strokes.

2. Use of "P" in three proposed names seems incorrect or at any rate inconsistent with CJK stroke block naming conventions.

3. UK's L2/08-175 (= N3448) contribution identifies 24 basic Tangut strokes, misreading a table copied from Li 1997 (also give here as Fig. 13). However, that table is not a set of strokes, but relates to Li's Four- Corner system, including stroke complexes. This misunderstanding is not corrected in the current document, though it seems not to be perpetuated either. A clear tabulation of Tangut strokes is given in Han (2004), given in L2/07-289 = N3307, and in UK's current contribution as Fig. 11a. This does in fact seem to reflect the state of the art for determination of the set of Tangut strokes. However, there is no clear authoritative study determining the relation between Tangut strokes and CJK strokes, so the current proposal is not following existing scholarship, but is undertaking a novel analysis that is not supported by the work of current experts.

4. Of the four strokes currently proposed, two of these "HPZ" and "HPZP" are already encoded and not already mapped in the proposal: HZZ and HZZP. It is not clear that the other two proposed strokes "HPDT" and "PDT" are not simply variants of encoded strokes, or that existing un- mapped CJK strokes might not be used or simply repurposed.

5. If it is deemed appropriate to extend the CJK Strokes block for non-CJK, the proposers should consult with the original CJK Strokes team in IRG, to determine handling of Tangut strokes.

6. However, it might be preferable to propose a full independent block of Tangut Strokes, for the following reasons:
   o Tangut is under ballot for its own block and is (still) not generally considered to be CJK (though it is largely Siniform);
   o The set of strokes for Tangut is relatively few (compared to CJK);
   o The range of variation in Tangut strokes and explicit CJK Strokes block unifications might be problematic from a Tangut perspective.