

|  | Arabic   | Chinese  | Cyrillic  | Devanagari  | Greek                                | Latin   |
|--|--|--|---|---|--------------------------------------|---|
| <b>I. Code point substitutability</b>                              |  |  |   |   |                                      |   |
| <b>A. Intra-script</b>   |  |  |   |   |                                      |   |
| <b>1. Same abstract character</b>                                  |  |  |   |   |                                      |   |
| <b>1a. Abstract characters that have more than one encoding</b>    |  |  |   |   |                                      |   |
|  | <p>Identical: decomposables, esp. if not normalized (6.1)</p> <p>Optional diacritics (notably vowels) (6.4)</p>  | <p>Non-identical: Generic variant characters (5)</p> | <p>Identical: decomposables, esp. if not normalized (3.8)</p> |   | <p>TONOS may be absent (11)</p>      | <p>SMALLER TURNED E, SMALL LETTER SCHWA (6.2)</p> <p>(Discussed at 6.5: Precomposed characters)</p> |
| <b>1b. Apparently the same abstract character in some contexts</b> |  |  |   |   |                                      |   |
|  | <p>Identical: ZWNJ/zero (5.21), KAF HEH YEH (6.1.abc), TEH MARBUTA (6.1g) HEH with HAMZA (6.1h) NOON (6.1j);</p> <p>Non-identical: KAF YEH ALEF with HAMZA (above, below, wavy); ALEF with MADDA; YEH with HAMZA; variants of dot orientation (all in 6.2)</p> | <p>Japanese-only Kanji (6.3)</p>                     |   | <p>Correct spelling of inflected forms in Nepali requires use of ZWNJ (4.3.2)</p> | <p>SIGMA and FINAL SIGMA (5, 12)</p> | <p>(Discussed at 6.4: decorative and contrastive variants)</p>                                      |

|  | Arabic   | Chinese  | Cyrillic   | Devanagari   | Greek                    | Latin  |
|--|--|--|--|--|--------------------------|--|
| <b>2. Different abstract characters</b>                            |  |  |  |  |                          |  |
| <b>2a. Different abstract characters interchangeable for users</b> |  |  |  |  |                          |  |
|  | vowels with/without HAMZA, ALEF with/without composed form (6.3ab)                               | Simplified characters (SC) / Traditional characters (TC) (5) |  |  |                          | (Discussed at 6.6: combining marks)  |
| <b>2b. Cases and lost information</b>                              |  |  |  |  |                          |  |
|  | Optional diacritics (notably vowels) (6.4)   |  | IE/IO in Russian (3.1); SMALL LETTER I WITH GRAVE (3.3); APOSTROPHE/zero in Ukrainian (3.6)      | MODIFIER LETTER APOSTROPHE/zero in Boro/Dogri/Maithili languages (3.4); EYELASH RA represents a different phoneme from RA, and is currently only rep'd by combination of RA with ZWJ (4.3.1) | TONOS may be absent (11) | Case folding (6.3)<br><br>(Discussed at 6.7: Punctuation: Latin characters correctly substituted for APOSTROPHE) |
| <b>2c. Simple visual confusability</b>                             |  |  |  |  |                          |  |
|  | Identical: BEEH/E FEH VEH (6.1def), TTEY/RNOON NOON/PEH (6.1ik);<br><br>Non-identical: GHAIN/FEH |  | GHE/GHE WITH UPTURN in Ukrainian (3.2); Old Letters, e.g YAT/SEMISOFT SIGN (3.4); ZE/DIGIT THREE | GHA/DHA BHA/MA etc (3.2.1 + Appx III); composite characters DGA/DNA/DRA etc (3.2.2 + Appx IV); Potential mistaken  |                          | (Discussed at 6.7: Punctuation: Latin characters confusable with APOSTROPHE)                                     |

|   | Arabic  | Chinese | Cyrillic  | Devanagari  | Greek   | Latin  |
|---|---|---------|---|---|---|--|
|   | QAF/AIN with 2 DOTS ABOVE<br>AIN/FEH/QAF with 2 DOTS ABOVE<br>(6.2) |         | (3.7); various forms of GHE, KA and EN (3.9); Appx I passim   | characters due to rendering errors of poor font design (3.3.1)  |   |  |
| <b>2d. Compatibility mappings</b>                             |   |         |   |   |   |  |
|   | TEH MARBUTA/HE<br>(6.3c)  |         | Rarely used character/sequence of characters, e.g EN WITH HOOK/EN + GHE (3.5); C WITH ACUTE/ C + J in Montenegrin (9.1) |   |   | (Discussed for Swedish and German umlaut in 6.3)   |
| <b>3. Code point substitutability dependent on the string</b> |   |         |   |   |   |  |
|   | digits?? (6.3d)   |         |   | Homophonous spellings (4.2)   |   |  |
| <b>B. Inter-script</b>  |   |         |   |   |   |  |
|   |   |         | Especially of concern for WHOLE-SCRIPT CONFUSABLE strings with Greek and Latin (4)                                      | Especially of concern for WHOLE-SCRIPT CONFUSABLE strings with other Brahmi scripts, e.g. Gujarati (3.5, 4.1) | Especially of concern for WHOLE-SCRIPT CONFUSABLE strings with Cyrillic and Latin | (Discussed at 6.1., 7 and 8.)<br>Especially of concern for WHOLE-SCRIPT CONFUSABLE strings with Greek and Cyrillic |
| <b>II. Linguistic variants</b>                                |   |         |   |   |   |  |

|  | <b>Arabic</b> | <b>Chinese</b> | <b>Cyrillic</b> | <b>Devanagari</b> | <b>Greek</b>   | <b>Latin</b> |
|--|---------------|----------------|-----------------|-------------------|--|--------------|
|  |               |                |                 |                   | Equivalence between corresponding words in Dimotiki and Katharevousa dialects (13) |              |