6.2.1 Ideographic and phonetic characters in Data Elements with VR of PN

The first component group (identified by DICOM as “alphabetical”) shall be encoded using the character set specified by the Attribute Specific Character Set (0008,0005), value 1. If Attribute Specific Character Set (0008,0005) is not present, the default Character Repertoire ISO-IR 6 shall be used. ISO 2022 escapes for Code Extension shall not be used in this component group. When Specific Character Set (0008,0005) value 1 specifies a multi-byte character set without Code Extension (i.e., Unicode in UTF-8, GB18030, or GBK), the characters of this component group may be encoded with multiple bytes, but shall be drawn from the code points U+00200 through U+1FFF of ISO/IEC 10646, or the following ISO/IEC 10646 code points:

U+3001, U+3002, U+300C, U+300D, U+3099 through U+309C, and U+30A0 through U+30FF