Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: ISO/IEC JTC1/SC2/WG2/IRG

Title: Proposal to disunify U+5F50 into three separate ideographs—redux

Source: Ken Lunde

Status: Member Body Contribution

Action: To be considered by the IRG and UTC

Date: 2022-02-18

This document is a proposal to disunify U+5F50 彐 into three separate ideographs, and to treat the two disunified forms as Urgently Needed Characters (UNCs) for encoding in the next version of the Unicode Standard. Disunifications are treated as *de facto* UNCs.

Background

IRG N2509 (aka L2/21-152) from IRG #57 proposes that U+5F50 彐 be disunified into three separate ideographs, 彐, 彐, and 彐. The first form would remain encoded at U+5F50 彐. The second form would be assigned the U-Source source reference UTC-01005 that is currently assigned to U+5F50 彐. The third form was proposed in L2/21-179 as a new U-Source ideograph to be added to UAX #45 as UTC-03251, and was accepted during UTC #169.

IRG N2538 from IRG #58 proposes that the second disunified form of U+5F50 彐, 彐, be treated as a UNC.

Proposal

Given that there is already a proposal on the table to disunify U+5F50 彐 into three separate ideographs, which is in reference to IRG N2509 (aka L2/21-152), it doesn't make much sense to treat only one of the disunified forms as a UNC. The formal proposal is as follows: **Disunify U+5F50 彐 into three separate ideographs—彐, 彐, and 彐—and treat the two disunified forms—彐 and 彐—as Urgently Needed Characters**.

The next two available CJK Unified Ideograph code points are **U+2B739** and **U+2B73A** at the end of Extension C, and we further propose that they be assigned the following normative and informative Unihan database property values:

```
U+2B739    kIRG_MSource    MC-00140
U+2B739    kIRG_USource    UTC-01005
U+2B739    kRSUnicode      58.0
U+2B739    kTotalStrokes   3
U+2B73A    kIRG_USource    UTC-03251
U+2B73A    kRSUnicode      58.0
U+2B73A    kTotalStrokes   3
```

This entails moving UTC-01005 from U+5F50 to U+2B739, and changing the UAX #45 *USourceData.txt* records for UTC-01005 and UTC-03251 as follows:

```
UTC-01005;C;U+2B739;58.0;;彐 ̄二;UTCDoc L2/12-333 37, UTCDoc L2/21-179 2;;3;0
UTC-03251;C;U+2B73A;58.0;;彐 ̄一一;UTCDoc L2/21-179 1;;3;0
```

That is all.