

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Ideographic Rapporteur Group Document
Title: Feedback on IRGN2373
Source: Eiso Chan (陈永聪)
Status: Individual Contribution to IRG #52, Hong Kong
Action: For consideration by IRG
Date: 2019-05-12

In IRGN2373, TCA pointed out the publishing source of all the SAT-Submitted characters is questionable. And, in the previous round of the review of WS2017, Ming Fan suggested SAT withdraw all the characters for the further researches. In IRG #51, Yifan Wang said the error forms in the SAT-Submitted characters are the stable errors. In the review works of WS2015 and WS2017, I pointed out some RS issues for the SAT-Submitted characters by comparing with different versions of the same books or the same sentences. I realized the issue how to handle the error forms is necessary and urgent to discuss for us.

So, I request the SAT experts and others consider the following questions carefully.

- 1) How do we define the “stable errors”?
- 2) Is it suitable to encode the error forms in IRG step? If not, which way is better to encode the error forms? I’m sure it’s valuable to digitize all the characters in 大正藏, but not by only one way.
- 3) Are all the SAT-Submitted characters the error forms in the ongoing working set and the future working set, or only the characters cited from 一切經音義?
- 4) Did any SAT expert know that there were so many error forms in the SAT submission when they submitted the proposal to IRG?
- 5) Is it okay to use the publishing source outside of 大正藏 for SAT?

By the way, some encoded SAT-Submitted characters in Ext. F can be also used for the Chinese dialects, but the pronunciations and meanings are totally different. For example, U+2CFD5 僮 can be also used for the Huadu (花都) dialect which is a sub-dialect of Yue-dialects, please see 珠江三角洲方言調查報告 mainly written by Prof. Zhan Bohui (詹伯慧); U+2D25D 啲 can be also used in Hong Kong, please see 胭脂扣 written by Lilian Lee (李碧華); U+2DFEB 瞞 can be also used for the old Suixi (遂溪) dialect which is a sub-dialect of Min-dialects, please see

The Suixi Dialect of Leizhou written by Anne O. Yue-Hashimoto (橋本-余霽芹).

If the SAT experts or other expert can provide the sufficient *heterogenous* evidence for the pending character which has been confirmed as the error form by IRG according to the submitted evidence, the new evidence should be accepted. If the new evidence is the *homogenous* evidence for the pending character which has been confirmed as the error form by IRG according to the submitted evidence, the new evidence should be rejected, and the character should be kept pending for the new heterogenous evidence. The *heterogenous* evidence here means two or more evidences provide the different usages, meanings or pronunciations for the same shape as what I mentioned above; the *homogenous* evidence here means two or more evidences only provide the same usage, meaning or pronunciation.

(End of Document)