

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

Doc Type: ISO/IEC JTC1/SC2/WG2/IRG
Title: On the encoding of modern self-created characters
Source: Andrew West
Status: Expert Contribution
Action: For consideration by IRG
Date: 2022-01-01

We recognise that some modern self-created characters are not appropriate for encoding, but we do not believe that it is necessary to modify the IRG Principles and Procedures to prohibit or avoid the encoding of all so-called self-created characters (生造字 *shēngzàozi*). We note that all characters were created by someone at some time, so self-created characters is not a particularly useful term. Many characters which IRG has accepted for encoding in recent working sets could be considered to be modern self-created characters, but are nevertheless acceptable for encoding. In particular many or most of the personal name characters proposed by China, TCA, and Japan may be considered to be personal-use self-created characters, but they are acceptable for encoding because of their use in government databases.

We consider that a rule in the IRG Principles and Procedures specifically excluding the encoding of modern self-created characters that are only used by their creator is not helpful, and that the suitability of encoding modern self-created characters should be determined on a case-by-case basis. To assist in determining whether self-created characters are appropriate for encoding or not, we provide the following guidelines.

A. Factors which argue for encoding:

- Created in order to fulfil a specific need (e.g. representation of spoken morphemes in dialect usage; transcription of Old Hanzi character forms; naming of chemical elements, animal species, or plant species)
- Created by an acknowledged expert in the field for which the characters are intended to be used
- Published in print by a reputable publishing house
- Occur as part of a larger corpus of related characters (e.g. in a dictionary of words or characters used in a particular dialect)
- Adopted in print by other users
- Required for use in government databases

B. Factors which argue against encoding:

- Ephemeral (e.g. characters created for special events, competitions, or puzzles)
- Personal use (e.g. personal name characters)
- Only used on the creator's web site or in a self-published book
- Created for fun or amusement only
- Graphic variants or alternative forms for existing encoded characters

Case Study 1

The character 𪛗 / 𪛘 anonymously created to represent the neologism *duāng* derived from a 2004 advertisement for shampoo spoken by Jackie Chan (stage name 成龍).

- Ephemeral (widely used internet meme from 2015, but no indication that the character has long-term significance)
- Created for fun only
- Few or no attestations in printed usage

Decision: Not suitable for encoding at the present time

Case Study 2

Characters created by Yú Shǎoléi 余少镛 (see IRGN2482 pp. 2–3)

- Created for amusement (do not fulfil any specific need)
- Not adopted by other users or in other printed publications

Decision: Not suitable for encoding at the present time

Case Study 3

Cantonese-usage characters in the *Concise Cantonese-English Dictionary* compiled by Yáng Míngxīn 杨明新.

- Created for the specific purpose of representing Cantonese morphemes for which no written character previously existed
- Published by a reputable academic publishing house (广东高等教育出版社)
- Characters adopted (and corresponding traditional forms created) by Cheung Kwan-hin and Robert S. Bauer in "The Representation of Cantonese with Chinese Characters" (*Journal of Chinese Linguistics*, Monograph Series Number 18, 2002)

Decision: Suitable for encoding

Case Study 4

Characters created by Yuen Ren Chao (趙元任) for the translation of Lewis Carroll's poem *Jabberwocky*,

- Created for the specific purpose of representing nonsense words created by Lewis Carroll (e.g. 炸脖龍 *zhábówò* for *Jabberwock*; 亮裡 *béilǐ* for *brillig*; 鵲鵲鴿子 *bólōgōuzi* for *borogoves*; 歪妙 *biǎmiào* for *beamish*)
- Chao was a renowned linguist and translator
- Quoted in other works, e.g. in the 1996 translation of Douglas Hofstadter's *Gödel, Escher, Bach: an Eternal Golden Braid* (哥德爾·艾舍爾·巴赫：集異璧之大成)

Decision: Suitable for encoding