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

Doc Type:	**Ideographic Research Group Document**
Title:	**Feedback on IRGN2551 Rules to Restrict the Encoding of Modern "Self-created Characters"**
Source:	**Eiso Chan (**陈永聪**, Culture and Art Publishing House)**
Status:	**Individual Contribution to IRG #59, online meeting**
Action:	**For consideration by IRG**
Date:	**2022-09-26**

Dr. Lu submitted IRGN2551 to clarify the encoding rules of the so-called modern "self-created characters". I basically agree with Dr. Lu, but I think I also need to make something clearer in this feedback.

This issue is pointed out by Mr. Wang Xieyang on IRGN2482, and Toby and Prof. Kim submitted their feedbacks. Andrew provided IRGN2521 to respond in general, and Mr. Wang also submitted his feedback. In these discussions, the main object is the term "self-created character" (生造字，自造字，自创字), so we need to clarify what the meaning of "self-created character" for Chinese usages in mainland China first.

1. *History*

This section is a literature review on terms, but it does not list all the documents that use these terms. Some documents that do not provide actual use cases have been excluded.

1.1. "Shengzaozi" (生造字)

1.1.1. In 1983, People's Government of Sangzhi County, Hunan Province (湖南省桑植县) published the book 《湖南省桑植县地名录》. The Annex 3 is related to the "shengzaozi" used for the geographical names. The compliers advocated to replace all the "shengzaozi" to make the geographical names look more standardized. (p. 300)

1.1.2. In 1991, 田元贞 published the article 《"文字病"及其治疗随感四题》. In this article, the author treated "shengzaozi" as the same as "错别字", and they must be eliminated in the daily life.

1.1.3. In 1999, 黄家礼 published the article 《"遒"不能写作"迳"》. The author advocated the proper form of 遒 should be 径, and 迳 is a "shengzaozi". However, 迳 is included in Telegraph Code Book, GB/T 2312-1980, and this character is still kept in TGH-2013.

1.1.4. In the same year, 李耕拓 published the article 《何"呯"之有》. The author advocated 呯 is a "shengzaozi", and the user should replace it with 砰.

1.1.5. In 2007, the book named 《现代校对实用手册》 written by 杜维东 and 陈瑗 was published. This book shows the meaning of "shengzaozi" is "the ones which can't be found in

the dictionaries". They provides the following examples: 1) the unstandard and the obsolescent simplified forms, 2) the Japanese "kokuji", such as 円, 3) the characters which the writers didn't know, such as 吔, 俬, 篱, 4) error forms (错字) and so on. They also show several examples for their elaboration analysis, such as 玥 is not a "shengzaozi", but some editors make it as two characters 王月; 呼 is not included in《现代汉语词典》, so the authors trusted this character must not be included in any font. They also clarified the ancient forms are not the "shengzaozi". (p. 111-113)

1.1.6. In 2010, 罗櫶宇 published the article 《网络生造字的创意与生成：概念整合视角》. The author reported that the Internet users reinterpreted some characters which had been encoded in URO or GBK. The "shengzaozi" here means the newly created meanings.

1.1.7. In 2011, 陈怡静 published the article 《从点校〈（嘉靖）衡州府志〉看纠正错字、生造字的重要性》. The author treated the characters which can't be found in 《辞海》, 《康熙字典》 and 《说文解字》 are the "shengzaozi". Therefore, so many characters used for the geographical names and person names belong to "shengzaozi", and they should be replaced to other proper ones, such as 埠. In fact, 埠 has been included in TGH-2013, and this character is used for at last three geographical names in Shandong Province (山东省). On the other hand, the traditional form 墠 is included in 《说文解字》 and 《康熙字典》, and 《东门之墠》 is one poem in 《诗经》. The author advocated replacing the "shengzaozi" to prevent future generations from following suit.

1.1.8. In 2014, 李丽群 published the article 《"火星文" 的结构类型及其对规范汉字的影响》. The author clarified the scripts for the so-called "Martian language" include non-Han and Han, and she treated the Han elements as "shengzaozi". However, the Han elements used for "Martian language" are all included in URO or GBK.

1.1.9. In 2018, the book named 《淮剧知识 300 问》 written by 陆连仑 was published. This book includes a list of the "shengzaozi" used for the dialect of Yancheng City, Jiangsu Province (江苏省盐城市), and the author call the dialect as 盐阜方言. The author doesn't advocate removing or replacing these characters because they are useful for Huaiju Opera. (p. 378-379)

1.1.10. In 2019, 魏书珺 published the article 《"穴丿丑" 字初探》. "穴丿丑" means "poor and ugly" (又穷又丑) as the self-mockery of contemporary young people, which is only a game for fun.

1.1.11. In 2021, 林升栋, 斯柯 and 赵广平 published the article 《生造字广告传播效果研究》. "Shengzaozi" in this article means the logotypes without the stable meanings and readings.

1.2. "Zizaozi" (自造字)

1.2.1. In 1941, 葉勁風 published the article 《怎樣研究英文（三）》 in *Shun Pao* (《申報》). "Zizaozi" means the unstandard English word in the article, because the term "word" was translated as "字" at that time. Please also see the real terms in 《馬氏文通》.

1.2.2. In 1981, the book named《出版工作手册》written by 王天恩 and so on was published. This book shows a list of "zizaozi", and the "zizaozi" means the error forms. (p. 421)

1.2.3. In 1982, 陈文升 published the article 《村名、地名用字要规范化》. The author treated "zizaozi" as the error forms as well.

1.2.4. In 1984, 李宁国 published the article 《微型机汉字输出》. This article introduced that "zizaozi" could be added to the font used for the computer MCZ-1/50 produced by ZILOG.

1.2.5. In 1985, the book named 《工程字法》 written by 中国工程图学学会科普工程委员会 and 四川省工程图学学会. This book defined that the "zizaozi" means the characters

fabricated casually by anyone or accepted by State Language Affairs Commission explicitly.

1.2.6. In 1991, the book named 《2.13-HCCBIOS 汉字操作系统使用指南》 written by 余素光 and 于强 was published. This book introduced how to create the "zizaozi" in one old BIOS, and the author suggested including the created characters since Row 88. (p. 35-36)

1.2.7. In 1994, 邓中明 published the article named 《为金山 SPDOS 建立小型汉字库》. The article introduced how to create the "zizaozi" by SUPER-SCW program in Kingsoft SPDOS.

1.2.8. In 1995, 龙启群 published the article named 《"屵" 是个自造字》. The author treated the simplified forms of the variants included in 《一异表》 as the "zizaozi". This is the first one mentioned "zizaozi" in the title.

1.2.9. In 1996, the book named 《最新 DOS 中文平台实用指南》 written by 张锺澍 was published. This book introduced how to create the "zizaozi" in DOS, and the example is Row 88 of FounderType 748 encoding. (p. 148-150)

1.2.10. In the same year, 潘晓虹 published the article 《金山系统自造字的输入与打印》. The article introduced how to create the "zizaozi" with different sizes and typefaces in Kingsoft SPDOS.

1.2.11. In 1998, the book named 《最佳电脑技能培训快通教程》 written by 刘亮 was published. This book introduced how to use the bitmap program in Windows to draw the glyph as EUDC. The Han example is 塑, and the musical symbols and the Tibetan letters are both treated as "zizaozi". (p. 68-69)

1.2.12. In the same year, the book named 《CCED 6.0 使用指南》 written by 郭志强 and 朱崇君 was published. This book introduced how to create and remove the "zizaozi" in CCED 6.0. (p. 71-72)

1.2.13. In the same year, the book named 《希望汉字系统 UCDOS/WIN 7.0 超级手册 UCDOS 7.0 使用手册》 written by 北京希望电脑公司 was published. This book introduced how to create the "zizaozi" in UCDOS 7.0. (p. 123)

1.2.14. In 1999, 李春燕 and 刘运飞 published the article 《为 "砼" 字正名——读《编辑学报》1999 年第 1 期一文的意见及建议》. The authors opposed that 刘仲桂 treated 砼 and other characters as artificial "zizaozi" (人工自造字) and also denied its necessity. They said the comment from 刘仲桂 had caused confusion to the standardization of editing works undoubtedly. They also said that all Hanzi were created by people to meet the needs of social production practices, and some Hanzi are recognized gradually and eventually became the specific characters. As a new character, it is inevitable that 砼 is unknown to a certain extent and even considered to be an error one, but the simplicity of use is obvious. Therefore, the authors supported to use the new Hanzi with realistic bases and necessities. This is the most sober cognition on this issue.

1.2.15. In 2003, 刘保全 published the article 《慎重处理地名用字 保护地名文化资源》. The author focused on the standardization of geographical names. He said the common requirement is to eliminate the traditional forms, variants and "zizaozi" and he defined "zizaozi" as the Hanzi without the nationwide Putonghua readings, which the meanings were commonly used and accepted locally. However, he said the original intension was so idealistic that they could not get people's understanding. At last, the author said we should face the fact of different needs and protect the geographical name cultures with local characteristics.

1.2.16. In 2006, the book named 《电脑医院（2006 全新版）》 written by 《电脑报》, the book named 《重装系统一条龙》 written by 彭爱华 and the book named 《2006 招征服 Windows XP》 written by 向光祥 were published. These books introduced how to create the "zizaozi" by EUDC program in Windows XP. (p. 229, p. 242-243, p. 229-230)

1.2.17. In 2008, the book named 《错别字例释》 written by 赵克勤 was published. The author pointed out 4 types of unstandard Hanzi: 1) traditional forms which has the corresponding simplified forms, 2) obsolescent variants, 3) "zizaozi" and 4) error forms. The author defined "zizaozi" as "自己生造的字" and 2nd stage simplified forms. (p. 3) From the examples provided by the author, "zizaozi" here means the vulgar simplified forms, such as 朳 (U+233E5) is treated as the "zizaozi" of 楼.

1.2.18. In 2010, 刘洋 published the article named 《生僻汉字计算机录入方法研究》. The author treated the elements of the some fonts as "zizaozi", which are 今昔文字镜, 汉达甲骨文字库, 华东师范大学商周金文数字化处理系统金文字库, 华东师范大学战国楚文字数字化处理系统战国楚文字字库 and 北京师范大学说文解字小篆字库. And then, he introduced the method on Windows EUDC program, which the Hanzi example is 玥. The author also pointed out other method to show the characters he was able to input, but those were not treated as "zizaozi".

1.2.19. In 2013, China Education Television (中国教育电视台) published the report 《网络用字急需规范 专家呼吁使用规范汉字》. This article said Prof. 王宁 said that some "strange" Hanzi and "zizaozi" were not encouraged to use in the common daily life, and she hoped the people use "规范字". This report is included in the book 《信息时代汉字规范的新发展——〈通用规范汉字表〉文献资料集》 published in 2015. "Zizaozi" here looks like the characters mentioned in 1.1.6 and 1.1.8.

1.2.20. In 2015, the book named 《戏人戏语 粤剧行话俚语》 written by 蔡孝本 was published. The author listed 奀 (U+5940) and 攰 (U+2512B), but he thought 攰 is "zizaozi" and 奀 is not, because he didn't find 攰 in any dictionary. (p. 9) If his theory is consistent, these two characters are both "zizaozi". 《康熙字典》 included 奀, but the reading and the meaning are totally different from Cantonese use. 攰 had been included in the Vietnamese dictionary 《五千字》 (*Ngũ Thiên Tự*). Therefore, the definition of the author is not rigorous enough.

1.2.21. In 2017, my teacher Prof. 侯兴泉 and 吴南开 published a book named 《信息处理用粤方言字词规范研究》. They defined 4 types of characters used for Cantonese, "zizaozi" is the main part, which is different from ancient Hanzi (沿用古汉字), phonetic loan Hanzi (假借字) and kunyomi Hanzi (训读字). "Zizaozi" here means the specific Hanzi created by the Cantonese speakers to record some Cantonese words extendedly, which most of them are not Sino words. (p. 18) This is the most accurate definition of "zizaozi" related to the dialect used Hanzi based on linguistics. The examples shown in the book are 嘢, 㰤, 嚟/唻 and so on. 撠/撠 has also been included in TGH-2013, so some people would disagree for this case. For 唔, 佢 and so on, they are also used for Hakka, so the authors thought it is very hard to define if they are "zizaozi" for Cantonese.

1.2.22. In 2018, the book named 《阳江文化濒危的瑰宝 咸水歌》 written by 阳江市阳东区文化馆 and 阳东区非物质文化遗产保护中心 was published. The song 《白鹤仔》 used 氹 (U+2057D), and the author treated it as "zizaozi" for the local dialect, which the meaning is to fetch or ladle water. (p. 67) This use is similar to the definition by Prof. Hou.

1.2.23. In the same year, the book named 《〈金瓶梅词话〉〈醒世姻缘传〉〈聊斋俚曲集〉语言词典》 written by 徐复岭 and 王永超 was published. The author said that 搞 (U+22D5E) is the "zizaozi" of 拍, and he said 喴 used in the words "狗喴黄/狗喴荒" is the 借音字 of 嗌. The author further explained "zizaozi" here means 异写字, such as 葶 used in 葶箩 is the "zizaozi" of 簸 or 筲; 攞 (U+22E7C) meant to impact with a heavy object is the "zizaozi" of 砸. (p. 1050-1051) In the theory of this author, "zizaozi" only means the ununfiable variant,

and 俗字 is different from "zizaozi" here.

1.3. "Zichuangzi" (自创字)

1.3.1. In 1992, 杨承任 published the article 《要提高会计人员的书写技能》. In this article, "zichuangzi" is similar to error forms. In 1993, there was other article with highly similar contents, named 《和农业会计侃写字》 written by 安树义.

1.3.2. In 1993, the book named 《历史文化名城濮阳》 written by 河南省濮阳市人民政府 was published. This book mentioned one local opera {⿰口亻⿰口山西}梆, which ⿰口亻⿰口山西 (aka WS2021-00169) was treated as a "zichuangzi". (p. 99-100)

1.3.3. In 1997, the book named 《中国电影名片鉴赏辞典》 written by 程树安 was published. The author mentioned ⿰口吉囍 was a "zichuangzi" created by Director And Lee (李安) which was used for one of the movie post titles of the movie *Wedding Banquet* (《喜宴》). "Zichuangzi" means logotype here. (p. 890)

1.3.4. In 2002, the book named 《趣闻英语手册》 written by 夏之秋 was published. The author said 袦 is a "zichuangzi" created by the Hong Kong businessmen. (p. 14-15)

1.3.5. In 2003, 吴英才 published the article 《〈金瓶梅〉是正定方言——兼谈其方言研究》 which was also included in the book named 《古圃文汇》. The author doubted �/� is a "zichuangzi" created by 兰陵笑笑生, because he did not find it in 《康熙字典》. In fact, 《康熙字典》 included this character and provided the detailed explanation. (p. 355)

1.3.6. In 2005, the book named 《随缘沉浮》 written by 蔡澜 was published. The article 《恂愁》 showed the discussion on the Cantonese word ngau4 dau6 with 黄霑. Mr. Wong said 绚 (綢) was a "zichuangzi" because the dictionary didn't include it. This running text looks questionable. It looks there is something errors for editing.

1.3.7. In 2006, the book named 《广州话点解》 written by 周无忌 was published. The author said people of Guangzhou paid more attention to diet, and so many cooking techniques of Cantonese Cuisine used the dialectal inherence words, so they needed several "zichuangzi" to record the words, such as 焗, 炆, 煀, 渌 and so on. Some taboo words which used the antonym to refer in Cantonese, so these words wouldn't be written as the original Hanzi, and the "zichuangzi" are needed. (p. 291-293)

1.3.8. In the same year, the book named 《潮乐问》 written by 余亦文 was published. The author mentioned so many "zichuangzi" were used in 潮州歌册, which were called as 歌册字 in Chaozhou. However, it was a pity that the author had not provided any example yet. (p. 389)

1.3.9. In 2007, 韩士奇 published the article 《现代文人的自创字》. The author shows some examples of "zichuangzi": 垮 and 搞 were created by 夏衍 during 1939 to 1941; 她 was created by 刘半农 in 1917; 妳 was created by 钱玄同; 猹, ⿰女人 and ⿰虫迁 were created by 鲁迅; 烤 was created by 齐白石; 閜/闪 was created by 胡适 for Huizhou dialects.

1.3.10. In 2009, the book named 《彩云南文史探究》 written by 廖德广 was published. The author said that 龚/夔 is a "zichuangzi". (p. 311)

1.3.11. In 2011, 董弈伯 published the article 《徐州小吃——饦汤》. The author treated 饦 as a "zichuangzi" created by Xuzhou people. The author was a pupil at that time, so I don't want to guess what he meant. It is clear that the term here is related to dialect.

1.3.12. In 2014, 邓景滨 and 董月凯 published the article 《从粤方言字的优化看现行汉字的优化》 which was included in 《中国方言学报》（第四期）. The comment on "zichuangzi" is similar to 1.2.21.

1.3.13. In 2015, the book named 《千年望族慈城冯家——一个宁波氏族的田野调查》 written by 王静 was published. The author mentioned one "zichuangzi" created by 冯孙眉, which was ⿱⿰扌⿰⿴口覀. 扌 comes from 捲, 覀 comes from 煙, 口 means cigarette case, so this is only a logotype. (p. 696)

1.3.14. In 2016, the book named 《高曰解红楼》 written by 高曰 was published. The author mentioned that he found 曹頫's "zichuangzi" should be called as 融旨字 in 2014. However, the author did not explain what "zichuangzi" and 融旨字 mean here. (p. 420)

1.3.15. In 2017, the book named 《邵武便览》 written by 邵武市地方志编纂委员会 was published. The book shows one kind of ICH named 长门, which they treated 门 as a "zichuangzi". (p. 24)

1.3.16. In the same year, the book named 《汉字与中国文化》 written by 王宝珍 was published. The author mentioned that 罍 and 龚/龔 belong to "zichuangzi", but they have been included in 《新华字典》, and "zichuangzi" is accepted by people. (p. 112-114)

1.3.18. In the same year, the book named 《两岸语言文字调查与语文生活》 written by 李宇明 was published. The author mentioned the education competent department of Taiwan Province, China released a 700 Hanzi list in 2009, and revised in 2014. These 700 Hanzi included original forms, kunyomi forms, phonetic loan forms and "zichuangzi" in Fujian-Taiwan region for Min-nan dialect. (p. 460-461)

1.3.19. In 2018, the book named 《永强方言实用词汇》 written by 龙湾区档案局 was published. The book showed ⿱⿴从工 or ⿱⿴从土 (aka WS2021-00791) was a "zichuangzi". (p. 50)

1.3.20. In the same year, the book named 《石岐方言故事》 written by 舒饭 and Julie Tong was published. The authors treated 刱 as a "zichuangzi". (p. 257)

1.3.21. In the same year, the book named 《孤山镇志》 written by 赵铭 was published. The author treated ⿱米叉 (aka WS2021-03000) as a "zichuangzi". (p. 7)

1.3.22. In 2019, the book named 《闽北记忆》 written by 南平市文化旅游局 and 中共南平市委党史和地方志研究室 was published. This book introduced one kind of local folk music named 邵武长门. The author treated 门 as a "zichuangzi". (p. 12)

1.4. Other similar terms
We will meet other terms for handling the materials, I list two here.
1.4.1. "Shengpizi" (生僻字)
"Shengpizi", maybe we could translate this word as subjective "unfamiliar Hanzi" for one concrete cognitive agent or objective "uncommon Hanzi" for one language or dialect speakers, is a common word in our daily life, but its meaning has split into at least two different definitions.
1.4.1.1. "Shengpizi" for human
This is the basic use for the term. There is a small-sized reference book named 《常见文物生僻字小字典》 written by 四川博物院 and 四川省凉山彝族自治州博物馆. "Shengpizi" in this book means the uncommon Hanzi used in the names of cultural relic for the ordinary people. The author define that the "shengpizi" included in the book are all outside 《简化字总表》 or the meanings or readings are unfamiliar for ordinary people.
1.4.1.2. "Shengpizi" for machine processing
For a long time, it is difficult to distinguish the detailed associations and differences of natural Hanzi or other symbol, character lists, character sets, encodings, glyph collections, type fonts, rendering, typesetting software and other relative programs for so many people. Therefore,

when they meet a Hanzi which can't be used as the most common one, they will mistakenly think this one is a "shengpizi" so that it is rejected to use in computer. In the China's financial industry standard JR/T 0253-2022, they defined "shengpizi" as outside the range of U+4E00 to U+9FA5, and they said this is the range in GB/T 13000, which is also a common misunderstanding. (Section 3.6) In fact, after Microsoft CP936 and GBK were used widely in Windows, so many programmers and type designers think they have been included all Hanzi, and they set up many obstacles for the users and make their daily life difficult. The original aim of JR/T 0253 is to overcome this difficulty, but it looks that the developers should consider more practical situations in future.

The implementers of JR/T 0253 often confuse these two uses of the term, so the rectification and upgrading of so many banks and their branches have not reached the expectation. For example, one branch of Harbin Bank post one short video to their 微信视频号 account to promote the implementation of JR/T 0253, but they think 龘 (U+9F98) and 朣 (U+6724) are the "shengpizi" mentioned in JR/T 0253, which is the common misunderstanding.

1.4.2. "Jiwaizi" (集外字)

This term is used for the book named 《IDS 与集外字处理方法研究》 written by 肖禹, which the term is similar to the definition of Gaiji in Japanese. The specific definition of this term is not absolute, so I think the best term for JR/T 0253 is the "jiwaizi" (external characters) of URO (URO 集外字) not "shengpizi" (the standard translated as rarely used Chinese characters, which should be 罕用汉字 not 生僻字).

2. Uses

The terms mentioned in Section 1 are often used confusingly.

2.1. For the daily use, so many people treated their unfamiliar Hanzi as "self-created characters", at this time, what they means is the same as "shengpizi" for human. This use is often derogatory. In 2007, 吴欣鸿 developed a converter of 火星文. 火星文 here means to replace the characters to the traditional ones, the similar ones or the homophonic ones in the range of CP936 for the modern Chinese used in mainland China. At that time, 火星文 was very popular among young people to express their private feelings in QZone, BBS forums and so on, but so many parents, teachers and cultural workers were very resistant to this phenomenon, and said the young people were creating the strange and nonexistent characters. Also see 1.1.8. In the next year, 吴欣鸿 changed the track to develop the later Meitu software. On the other hand, some parents would name their children with a euphonious reading for the phonetic element and a suitable semantic element. In fact, they are creating a Hanzi, but some results are also included in some dictionaries or character sets. For example, the TVB artist Verana Wong's Chinese name is 黃瀅仴, and her full name for Cantonese reading is Wong Ying Yuet (aka Wong4 Jing4 Jyut6) on the TVB artist page, that means 仴 in her name reads jyut6 in Cantonese, yuè in Putonghua, not wut6 and wò like the dictionaries show. However, she calls herself Wò·wo in Putonghua, see her vlog with Brian Chow (01:50; 05:22 also shows the Cantonese and Putonghua readings together). This situation is very common and some thinking motivations are like 火星文.

2.2. For typesetting works, the external characters of movable plates (活字字盘) or character sets are called as "self-created characters". The object of the verb "create" here means the glyph designing result not the realistic Hanzi in the real world. According to the interview with Mr. 朱志伟 and Ms. 杨雁 by Prof. 小宫山博史 (Komiyama Hiroshi) and Dr. 孙明远, it was not very difficult to create the external movable types by the skilled engravers, please see the

relative details on the forthcoming book 《方寸之间——汉字文字设计文集》. And we all know it is also easy to create the digital bitmap glyphs, even outline glyphs, in Windows.

2.3. For type designing, the type designers often call the type design tools as "造字工具" in Chinese, and the design activities are also called as "造字" or "做字", but what they means is just to design the digital outline results of glyph in a digital font, especially those designers who don't know much about encoding and linguistics.

2.4. For some program developers, they used to keep their programs stable, and the easiest and most common way is to use an outdated regex to intercept the characters out of the original URO. Therefore, they will treat all characters encoded after 1993 as the characters created by anyone.

All in all, the use of the terms "shengzaozi", "zizaozi", "zichuangzi", "shengpizi" are all confusable. The corresponding concepts are unclear on different running texts for the limitations of the users and the times. The corresponding concept of "jiwaizi" is relatively clear, but not stable for different character sets, because this is only a technical term. It is OK to use these terms in the specific and definite verbal contexts with the clear definitions, but it will also lead to confusion in our encoding works because of the evidence and theories cited from multifarious verbal contexts.

3. Solution

When a person meets a character which is hard to handle for digitization, he/she will do as below by himself/herself or with other persons. We assume that this person knows little about encoding.

1) handwriting on paper,
2) casting a new movable type,
3) using a picture,
4) using EUDC program on Windows,
5) using common type designing programs,
6) using GlyphWiki to get single-glyph font,
7) disassembling into multiple components,
8) using the replacing character.

For 1), 2) and 3), these are more traditional and burdensome solutions.

For 4) and 5), these will cause further difficulties in future use because of PUA or substantive duplicate.

For 6), digital text rendering relies heavily on fonts.

For 7), it will make confusion of text information without any additional mark. For example, 倪征燠 is recorded as 倪征日奥 by so many websites, because the last character is not included in GBK. And we can meet a regrettable question "倪征日奥是中国名吗？" (Is 倪征日奥 a Chinese name?) on Baidu searching, because the name 倪征日奥 looks not like a common name for a Han person in China. Some website added pair of brackets with 日奥, but when we see 倪征（日奥）, we also think the person's name is 倪征 and his or her 字 is 日奥. We assume there is a girl named 语嶴契, maybe her name will be recorded as 讠工白日奥 丰刀女 regretfully.

For 8), its harm is more serious than 7). For example, 氹 is a common character used in Cantonese, which means a pool and reads as tam5. However, some people use this character to record the other words like tam3 and dam4/6. On the other hand, there are two meanings for 仔 in Cantonese. 氹仔 (Taipa) is a part of Macao SAR. If three readings are accepted to

add to 氹, we will get the following meanings for Taipa: the first one is a pool (tam5 zai2), the second one is to coax a child or boyfriend (tam3 zai2), the third one is to stamp on a child's or boyfriend's foot (dam4/6 zai2), and the most absurdity one is to trap a child or boyfriend (also tam5 zai2, but tam5 is treated the corresponding Putonghua word as 坑). The same misunderstanding also occurs in the discussion of the song 《氹我》 sang by JW 王灝兒 (Joey Wong), which the original meaning of the title is to coax me (tam3 ngo5). Therefore, it is better to keep 氹 only reads tam5, and dam4/6 should be 踠 (U+280BD), tam3 should be � (U+2C8A9) or 􏿿讠凶 (WS2021-03791).

4. Analyses

In the language activities, the most basic level should be "concept" which is directly related to the practical activities how to understand and transform the real world for people; the more advanced one should be linguistic "word" for languages and their dialects, and almost all elements of words are stable. When "word" used with speech sound, it is reading what we can speak and hear; when "word" used with symbol / sign to follow some rules, it is script what we can write and read. Based on these, I list the following situations and the appropriate encoding methods.

4.1. Core Hanzi

The core Hanzi could be used for the basic concepts and words and are known by so many people. This kind must be encoded. The core Hanzi for the Sino words in modern Chinese, Japanese, Korean and Vietnamese are all been encoded for a long time, but maybe so historical ones and the variant ones should be encoded in future.

Suggestion: Accept to encode.

4.2. Extended Hanzi for inherent words

Different languages and Chinese dialects always include so many inherent words which are different from the basic Sino words, or the readings have been changed. And the people need different characters to distinguish from the basic Sino words, because two kinds of words are both used and needed in the daily life. Some of this kind have not been regulated, so it is common that one word have different forms. On the other hand, it is not easy to distinguish if they are the same word when the readings are the same. The results that cannot be easily obtained from formal linguistic discussions should not be used to embarrass encoding experts. The main object of IRG is to make users convenient, not set barriers at the basic encoding level. Therefore, if one official publication provides the inherent words, the characters could be acceptable for encoding. So many characters used for the geographical names (地名), ethnic languages, dialects, music scales and so on belong to this one.

Suggestion: Accept to encode.

4.3. Extended Hanzi for loan words

When people lived in Hanzi used regions meet the persons, places and unclear syllables out of this region, we often write some proper forms subconsciously to express those concepts. When the forms have been used in the historical running texts, we should encode them for the digitization.

Suggestion: Accept to encode.

4.4. Extended Hanzi for new words or concepts

In general, there will not be completely new words or concepts in daily life, but the modern technology is changing our work and life day by day, and the new concepts, even completely new words, will be used for some time. The names of new chemical concepts need new characters. In this part, we will meet the gendered and species variants possibly, such as 1.3.9.

Suggestion: Accept to encode.

4.5. Extended Hanzi only for person names

In fact, some people use the Hanzi without the clear concept and even not belong to 4.2, 4.3 and 4.4 in their names. If the household registration systems accept the names, and no objection from the holders, the extended Hanzi should be accepted to encode, and the direct concepts are the holders, because the digital real name system (实名制) is very common in mainland China and Taiwan Province. When we reject one extended Hanzi used in the household registration system, the corresponding person will continue to meet the difficulties in daily life. On the other hand, the submitters should understand, unifications don't mean to reject based on the evidence and use. And all the submitters should consider handling the IVD collection for the unified ones.

Suggestion: Accept to encode.

4.6. Complex Hanzi

This is an issue I once mentioned, but the researching works are ongoing. I just list them to make the description here more complete.

Suggestion: Accept to encode the elements, not the complex results.

4.7. Extended Hanzi for artificial languages

As I know, [Toki Pona](), [Rangyayo]() and [Shidinn]() could use Hanzi to write. Toki Pona and Rangyayo just use the basic Hanzi, but there are some new extended Hanzi used for Shidinn. All the elements used to write Shidinn have been included in [UCSUR PUA project](). If an artificial language could be accepted by any part of ISO 639 and the language rules are stable, the extended Hanzi for it should be considered to encode. Note that there have been 24 artificial languages included in ISO 639-3.

Suggestion: Consider encoding.

4.8. Pseudo-Hanzi without words and concepts

The pseudo-Hanzi used in [余少镭's book](), 王左中右's works and so on don't express the stable words and concepts in any language and dialect, which are all related to some hot spots of society, so we can't call them as Hanzi, which are like logotypes or behavioural arts more. The pseudo-Hanzi used in 徐冰's 天书 works (aka 《析世鉴 天书》) don't express any word and concept, which just only the Hanzi-liked pure work of art, so we can't call them as Hanzi as well. There are also other situations related to Internet fraud or just for fun.

This kind is totally different from the above ones. The ones mentioned in 4.1 to 4.7 have the actual semantics bases, and the pseudo-Hanzi is a pseudo proposition for encoding based on the use. If any expert can't distinguish from them, it is better to learn the relative linguistics course not language courses again.

Suggestion: Reject based on the use.

Please also note that the Cantonese word "chok" sometimes is treated as the newly developed word because of [Raymond Lam](林峯)'s role image in the TVB drama *The Mysteries of Love* (《談情説案》), who imitated [Fukuyama Masaharu](福山雅治)'s role image of Yukawa Manabu (湯川学) in the drama *Galileo series* written by [Higashino Keigo](東野圭吾). However, some researchers think the origin of the word "chok" comes from the classical Japanese game *Street Fighter* or the English word "choke". All in all, the word "chok" is not a "new" word for Cantonese in daily life, which is like "fajita", "ninja", "samurai", "dim sum" and so on, not "covfefe" in English. However, there is no Hanzi form of the word "chok" now.

5. Advanced suggestion

There is no need to introduce a term "self-created characters" in our encoding works because the uses and the corresponding meanings are confused. If some experts think we must need a term used for 4.2 to 4.7, I suggest use "later developed Hanzi/ideograph" (后起字/後起字), which the use and meaning is relatively stable.

And when we meet the submitted evidence, we should prohibit new words and concepts created by any individual and reject pseudo-Hanzi without words and concepts. For the covfefe-like misprint characters in the modern digital re-typesetting versions of ancient books, we should have other rules to prevent encoding them.

6. Acknowledgement

Jerry You provides some comments on the translations and definitions of the terms.
Dr. Sun Mingyuan provides some comments on the external movable types and Gaiji.
Clerk Ma provides some helps on ZILOG.
Kushim Jiang provides some helps on the terms.

(End of Document)