

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

Doc Type: Working Group Document
Title: Proposal to add kFanqie to UniHan Database
Source: Eiso Chan (陈永聪, Culture and Art Publishing House)
Reference: [L2/22-135](#), [L2/22-127](#), [L2/23-002](#), [L2/23-011](#)
Status: Individual Contribution
Action: For consideration by UTC
Date: 2023-06-26

0. Introduction

Fanqie (反切) is a traditional method to indicate the pronunciations of a syllable for Chinese character (Hànzì, Kanji, Hanja, Chū Hán) by using two characters and ending with “反” or “切”. The first character means the initial (聲母), which is called as “反切上字”, and the second character means the final (韻母) and tone (聲調), which is called as “反切下字”. In some early usages, the first character and the second character must share the same “等” and “開合”, but this practice had not been continued in the later time.

The following two tables show the fanqie analyses for U+6771 東 (德紅切) and U+4EBA 人 (如鄰切) based on 《廣韻》 (Guǎngyùn).

Table 0.1 Fanqie analysis for U+6771

UCS	Char.	fanqie	聲	攝	韻	等	開合	調
U+6771	東	德	端	曾	登	一	開	入
		紅	匣	通	東	一	開	平

Table 0.2 Fanqie analysis for U+4EBA

UCS	Char.	fanqie	聲	攝	韻	等	開合	調
U+4EBA	人	如	日	遇	魚	三	開	平
		鄰	來	臻	真 A	三	開	平

In the IRG encoding review works, the evidence for so many submitted characters shows the fanqie, or the submitted characters are used in a fanqie. The possible kFanqie property will be useful for the encoding works in future to get one form of Chinese readings to help the unification works, or judge if the unencoded character is the variant of encoded one when the character is only used in a fanqie based on the submitted evidence.

1. kFanqie

The proposed property values for kFanqie are collected from different versions of 《廣韻》 basically, which is the most important use of fanqie in the history, and I add some values for CJK Ext. H characters based on the IRG ORT. The current suggested values in the data file are only collected and reviewed by me. It is better to be reviewed by other encoding experts later.

We could add more values based on other sources in future if possible. The fanqie used for Chinese dialects and other languages are all OK to include in `kFanqie` property. If the characters used in the suggested values are unified with the characters used in the existing values, or they are the undoubted variants anytime and anywhere, the suggested values will not be accepted.

All the fanqie in the real publishing sources are ended with “反” or “切”, and “反” and “切” could be interchangeable, so there is no need to include these two characters in the `kFanqie` property, that means all the values include and only include two CJK Unified Ideographs.

As Fig. 1.1 shows, the fanqie of 東 is 德紅切; as Fig. 1.2 shows, the fanqie of 東 is 德紅反. We should treat them as the same fanqie, that is the reason why I propose to include and only include two Unified Ideographs in the property values without ending with “反” or “切”. Therefore, the `kFanqie` property value of 東 (U+6771) should be 德紅. In fact, some characters have multiple readings, and hence multiple fanqie, so the values should be separated by space.

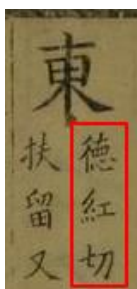


Fig. 1.1 陳彭年, 邱雍 等: 《廣韻》, 符山堂藏板, 上平聲卷第一



Fig. 1.2 呂大臨: 《考古圖釋文》, 四庫全書本, 上平聲

The following table provides the fields for this property as reflected in UAX #38.

Table 1 Fields as reflected in UAX #38

Field	Text
Property	<code>kFanqie</code>
Status	Provisional
Category	Readings
Introduced	<i>TBD</i>
Delimiter	space
Syntax	<code>[\x{3400}-\x{4DBF}\x{4E00}-\x{9FFF}\x{FA0E}\x{FA0F}\x{FA11}\x{FA13}\x{FA14}\x{FA1F}\x{FA21}\x{FA23}\x{FA24}\x{FA27}-\x{FA29}\x{20000}-\x{2A6DF}\x{2A700}-\x{2EBEF}\x{30000}-\x{323AF}]{2}</code>
Description	The fanqie reading, which is expressed as two CJK Unified Ideographs. The first one means “initial” (聲母), and the second one means “final” (韻母) and the tone (聲調). Complete fanqie readings always end with U+53CD 反 or U+5207 切 as a third character, which are interchangeable and therefore not included as part of the property value.

Note: USNB submitted SC2 N4851=[L2/23-114](#) to include 603 UNC to support the additional repertoire in GB 18030-2022 DAM.1 Draft 2 as CJK Ext. I, but this part has not been included in UCS / Unicode officially, so we don't include \x{2EBF0}–\x{2EE4F} in the syntax at this stage.

2. Exclusion

There are some kinds of sequences like fanqie used in the books and our daily life, but it is not suitable to include them as the `kFanqie` property values.

2.1. Pronunciations used only for stage performances

In Peking Opera (京剧), some pronunciations are only used for stage performances, not for the daily life in common Pekingese (北京话), and the phonology for those pronunciations are larger than the one for Pekingese and Putonghua, that means the Hanyu Pinyin can't cover all the syllables and their tones are not stable. This kind of linguistic phenomenon is called as “上口字” (shàngkǒuzì). In the development history of Peking Opera, the performers tried to imitate some so-called archaic pronunciations and different Chinese dialects (such as 中州音, 湖广音, 楚音, 徽音, 吴音 and so on) to pursue elegance or enrich the characters images.

Some researchers still use two Hanzi to record those special pronunciations. In Fig. 2.1.1, the initial is not included in the phonology for Putonghua, and the red squares in Figs. 2.1.1 and 2.1.2 shows the similar readings by two characters.

2. 方言声母 ㄋ̌ (读似“妮”，近似拼音“ni”)					
例字	北京音		上口字		
	拼音	辙名	拼音	辙名	读似
虐	nüe	乜斜	ㄋ̌io	梭波	妮哟
浓	nong	中东	ㄋ̌ong	中东	妮翁

Fig. 2.1.1 《京剧上口字尖团字实用手册》，p. 5

4. 方言韵母 io (读似“哟”)					
例字	北京音		上口字		
	拼音	辙名	拼音	辙名	读似
略	lüe	乜斜	lio	梭波	乐哟
觉	jue	乜斜	jio	梭波	鸡哟
却	que	乜斜	qio	梭波	七哟
学	xue	乜斜	xio	梭波	西哟
岳	yue	乜斜	yio	梭波	哟
药	yao	遥条	yio	梭波	哟

Fig. 2.1.2 《京剧上口字尖团字实用手册》，p. 6

The sequences shown in the red squares above could be treated as fanqie for the corresponding characters, but I still suggest excluding them from the property values because they are not used for the daily life and the specific values are under the certain degree of flexibility, that means the performers could choose the common pronunciations of Pekingese or this kind of pronunciations based on the specific content of the performance scripts.

On the other hand, maybe we could consider adding `kPekingOpera` property or other name based on the pronunciations of “上口字” in future.

2.2. Sequences for joking purposes only

In 2016, the boy-band SWIN-S released a mini album, and there was one song named [《只因你太美》](#) (*Just Because You're So Beautiful*). In 2018, one of the members of SWIN-S, 蔡徐坤 (KUN), participated in one variety show [《偶像练习生》](#) (*Idol Producer*). He introduced himself and showcased his talents and skills, which the background music was this song at that time. Unfortunately, the sentence “只因你太美” was heard as “鸡你太美” (chicken, you're so beautiful). In 2019, NBA announced to hire KUN as their image ambassador. And then, many netizens used “只因” to replace “鸡” and other characters related to the Putonghua syllable “ji” only for joking purposes on the simplified Chinese Internet. Please also see more details [here](#). This kind of phenomenon is common in our daily life but not encouraged at least in the standards.

It is not hard to know the sequence is the vulgar fanqie to the similar reading in the daily life in the above case. The following table shows other similar vulgar fanqie, that means the followings and the similar ones we will meet in future should be rejected for the `kFanqie` property values.

Table 2 Examples of vulgar fanqie

Character	Putonghua reading	Vulgar fanqie	Putonghua readings	Related events
鸡	jī	只因	zhǐ + yīn	蔡徐坤 《只因你太美》
宣	xuān	喜欢	xǐ + huān	卓文萱 (池早早) 《爱情是从告白开始的》
造	zào	知道	zhī + dào	ditto
酱	jiàng	这样	zhè + yàng	unclear
酿	niàng	那样	nà + yàng	unclear

2.3. Sequences for personal names

杨绛 (Yáng Jiàng) is an important writer and translator in modern Chinese literature. Her original name is 杨季康 (Yáng Jíkāng). Some researchers guess 绛 (jiàng) means 季康切 (jì + kāng), but there is no clear confirmation from her own. The relationship is actually unclear for this kind of usage, so I suggest reject it and the similar ones at this stage.

3. Coverage

A total of 20,225 CJK Unified Ideographs are assigned this proposed property. The following table indicates the number of CJK Unified Ideographs that are assigned this proposed property based on Unicode Version 15.1. The characters in Ext. C, D, E, F, G and the compatibility characters have not been assigned this proposed property currently. Any expert could request to add the property values for them in future.

	URO	Ext. A	Ext. B	Ext. H
Characters	13,340	2,702	4,160	23
Percent	63.55%	40.99%	9.74%	0.55%

4. Data file

The `kFanqie.txt` data file, which is a PDF attachment, provides everything that is necessary for adding this property to the UniHan Database, and for adding a full description of the property to UAX #38, Unicode Han Database (UniHan). The examples are shown as below.

U+4EBA	<code>kFanqie</code>	如鄰
U+6771	<code>kFanqie</code>	德紅

In the data file, the number of the value(s) for one character is 1 through 7. The following table shows the situations.

	1	2	3	4	5	6	7
Characters	15,461	3,813	789	137	22	2	1
Percent	76.44%	18.85%	3.90%	0.68%	0.11%	0.01%	0.00%

For U+24865 (𪛗), I listed two values in the preliminary proposal, but one has been removed now. IRG had decided to encode [WS2021-02424:GKJ-00588](#) (𪛗 𪛗間) separately during IRG #59, so the removed value should be moved to WS2021-02424 in future.

5. Bibliography

陳彭年, 邱雍 等: 《鉅宋廣韻》, 宋乾道五年刻本
 陳彭年, 邱雍 等: 《廣韻》, 符山堂藏板
 陳彭年, 邱雍 等: 《廣韻》, 龍谷大學藏至正南山書院刊本
 陳彭年, 邱雍 等: 《廣韻》, 覆元泰定本
 陳彭年, 邱雍 等: 《廣韻》, 澤存堂本
 陳彭年, 邱雍 等: 《廣韻》, 古逸叢書覆宋本重修
 陳彭年, 邱雍 等: 《廣韻》, 四部叢刊巾箱本
 呂大臨: 《考古圖釋文》, 四庫全書本

IRG ORT

張文江: 《錢鍾書傳——營造巴比塔的智者》, 上海: 上海人民出版社, 2016.1, ISBN 978-7-208-13387-7/K·2450
 咎紅英, 許鴻飛, 張坤麗, 穗志方: 《網絡用語詞典的構建及問題分析》 (*The Construction of Internet Slang Dictionary and Its Analysis*), 《中文信息學報》 (*Journal of Chinese Information Processing*), 2016.11, pp. 133-139
 賈啟艾: 《人際溝通 (案例版·第4版)》, 南京: 東南大學出版社, 2019.6, ISBN 978-7-5641-8441-4
 孫雲龍: 《互聯網語境中明星形象污名化問題研究——以蔡徐坤為例》, 西南大學公共關係學碩士學位論文, 2020.4
 費玉平: 《京劇上口字尖團字實用手冊》, 北京: 文化藝術出版社, 2023.5, ISBN 978-7-5039-7364-2

6. Acknowledgement

Prof. Hou Xingquan (侯興泉) provided some helps on this document.
 Dr. Ken Lunde (小林劍), Mr. Stephan Hyeonjun Stiller (尹賢儒/尹賢俊/윤현준) and Mr. Kushim Jiang (姜兆勤) provided some comments.
 Mr. Lee Collins (康立論) helped me check the data file to match kSBGY property and the sources.

(End of Document)

Appendix Main updates of the data file

In this revised version, I updated the data file based on the experts' comments and my studies. For (1) through (4), these updates are to match kSBGY property and the sources; for (5), these updates are to eliminate the duplicates.

(1) Removal (23)

The following 23 characters have been removed directly, because the values are duplicate with the variants, and we only keep the ones to match kSBGY property.

U+5156, U+5415, U+5F66, U+613D, U+6236, U+6237, U+654E, U+6CA1, U+6F11, U+7151, U+72F6, U+771E, U+7984, U+7B11, U+8017, U+885B, U+886E, U+8943, U+8CD3, U+8CFA, U+9332, U+9854, U+985A
--

(2) Movement (15)

For the first column, that means the property value of U+5019 in the previous version has been moved to U+202EB, and there is no value for U+202EB in the previous version. This is to match kSBGY property and the sources.

U+5019 → U+202EB	U+5379 → U+460F	U+53F1 → U+20B9F
U+59CA → U+216F7	U+5EBE → U+22214	U+6052 → U+6046
U+69E9 → U+3BA3	U+6D93 → U+3CD9	U+6F97 → U+3D4E
U+7358 → U+21681	U+7A3D → U+2583B	U+7C21 → U+25CD1
U+8947 → U+8949	U+8CD2 → U+8CD6	U+9B42 → U+4C1F

(3) Replacement (1)

The property value for U+26AF3 in the previous version is not right, but it is really included in the sources.

U+20065 → U+26AF3

(4) Merger (11)

For the first column, there is the property value for U+4E3C in the previous version, which there is also the kSBGY property value, and there is no kSBGY property value for U+4E95. In the sources, the different fanqie used for the same Hanzi were allocated in the previous version. I merge the fanqie values and only allocate them to U+4E3C to match the sources.

U+4E95 → U+4E3C	U+522B → U+5225	U+54E1 → U+8C9F
U+66C1 → U+66A8	U+6717 → U+34EA	U+67FB → U+67E5
U+6A23 → U+6837	U+6BC1 → U+6BC0	U+758F → U+47FD
U+865A → U+865B	U+8846 → U+773E	

(5) Reduction (48)

There are the duplicate values for the following 48 characters in the previous version. I have corrected them in this version.

U+412A, U+450C, U+4C71, U+4EE0, U+4F02, U+51DC, U+5857, U+5858, U+5D0E, U+5DBE, U+625C, U+637B, U+6398, U+6778, U+799A, U+79B8, U+79C6, U+7A08, U+7B08, U+7B45, U+7C0F, U+7C5C, U+7E1D, U+8028, U+8197, U+845F, U+84EF, U+8599, U+85D9, U+85EE, U+8691, U+871D, U+8A3E, U+8A6D, U+8B7D, U+8CA4, U+920B, U+923C, U+96DA, U+97A4, U+9B92, U+9C6E, U+9D42, U+9D99, U+20BD7, U+22F1B, U+23014, U+24D31
--

(6) Special case (1)

There is the kSBGY property value for U+555F (啟), but it is not included in the sources. I have requested to update this one in [L2/23-128](#). The kFanqie property value for U+5553 (啓) should be kept.