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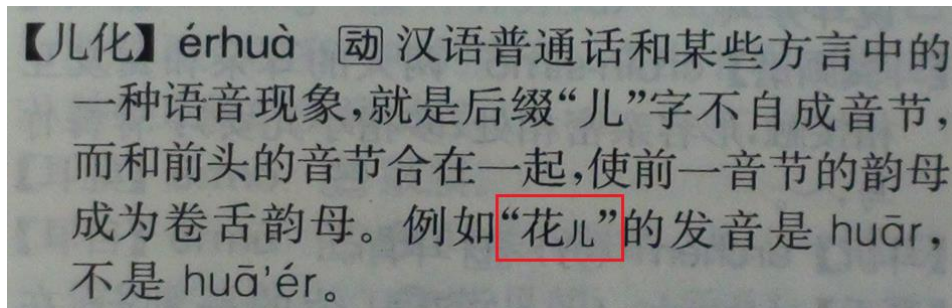Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Proposal to encode two small form CJK characters for Chinese**
Source: **Andrew West, Eiso Chan (陈永聪)**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2023-12-12 (replaces WG2 N4720 = L2/16-109)**

1. Background

The characters U+5152 兒 (traditional form) and U+513F 儿 (simplified form), pronounced *ér* in Mandarin Chinese, mean 'son' as a noun. However this character pair is also used as a non-syllabic suffix representing the linguistic phenomenon known as *érhuà* 儿/兒化 (rhotacization), whereby some words are given an r-coloured pronunciation in some forms of Chinese, notably the Beijing dialect of Mandarin Chinese. For example, in Beijing Mandarin the word 'flower' 花 *huā* is pronounced as *huār* (pinyin romanization), and is normally written 花儿 (simplified form) or 花兒 (traditional form).

In some publications, particularly dictionaries and linguistic texts, the non-syllabic *érhuà* usage of this character is indicated by writing the character 儿/兒 in a smaller size than the character 儿/兒 used as an ordinary noun (see Fig. 1).

Fig. 1: Entry for *érhuà* in *Xiàndài Hànyǔ Cídiǎn* 现代汉语词典 (6th ed. 2012) p. 343

花儿 is read *huār*, not *huā'ér* (which would be written 花儿)

The practice of distinguishing *érhuà* with a small-sized 儿 was perhaps introduced by the Chinese linguist Yuen Ren Chao 趙元任 (1892–1982), who uses a tiny 儿 to represent *érhuà* in his 1969 translation of Lewis Carroll's *Through the Looking-Glass* (see Fig. 2). In usage within China from the 1980s onwards, a medium-sized 儿 is more normal, and books set in traditional Chinese characters normally use a small-sized 兒 for *érhuà* (see Fig. 7).

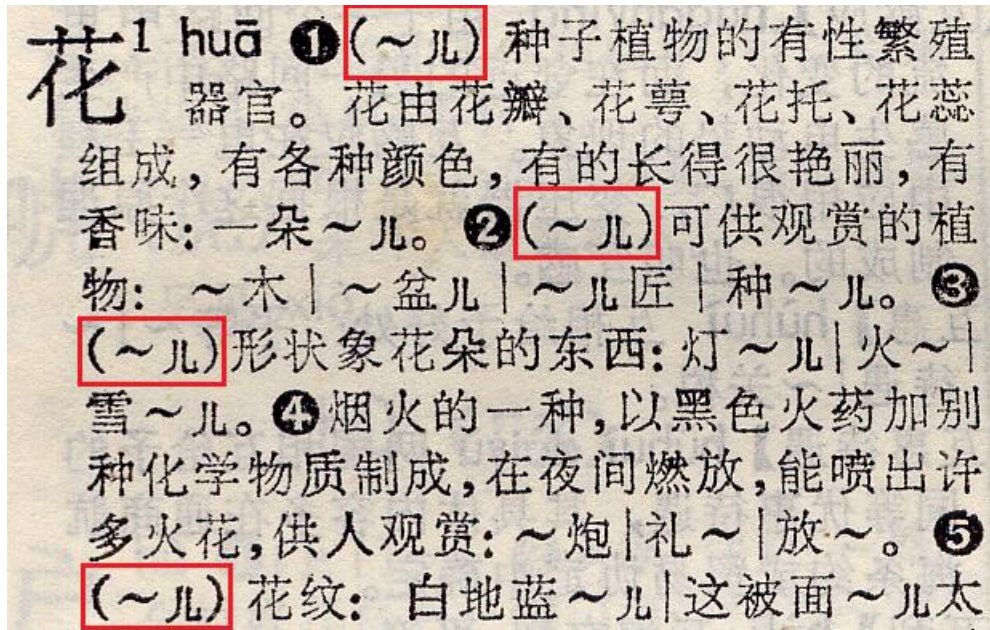**Fig. 2: Page from Yuen Ren Chao's translation of *Through the Looking-Glass***

第一章

鏡子裡的房子

廿[1] ，不是那小白貓，一定全是那
小黑貓做的壞事。因為小白貓
讓大貓給他洗臉來着，一直洗了一刻
鐘，總算沒很鬧；所以你瞧剛才
5　那淘氣的事不會有他的份[2]的。
　　黛那給他的孩子們洗臉是這麼洗的：
他先拿一個爪子把那可憐的小東西的
耳朵摁着，再用那個爪子解鼻子尖起頭
望上，給他滿臉那麼和弄[3]；剛才
10　不是我說他正在那忙着弄那個小白
貓嗎？這個就乖乖的獃着，還想打呼嚕[4]
—彷彿覺得出這都是為他好似的。
　　可是小黑貓在下午已經早洗
好了，所以一邊阿麗思坐得圈身椅[5]的
15　犄角上團成一團[6]似的那麼一半跟
自己說話一半睡着了，一邊那小貓
就拿阿麗思想要繞起來的絨線球
大玩大瘋[7]，給他滾上滾下的又
都滾散了，你瞧現在弄的一地毯的
20　瞎結子亂團子，當間還有個小貓
待那[8]拼命追他自各的尾巴。
　　「嗐[9]，你這小壞東西壞透了[10]！」
阿麗思說着就把那小貓提遛[11]起來親
一下，讓他明白這親他是給他丟臉
25　的。「真的，黛那應該給你教點好樣子
的！應該的嘜，黛那，你知道你應該的嘜！」他
一頭說着，一頭對老貓做出責備他的
樣子，可是他的聲音要兌也兌的不大
像——會他轉身又爬回到大椅子
30　上，連小貓帶絨線都抱了上去，

—14—

Sayable Chinese **vol. II (San Francisco: Asian Language Publications, 1969) p. 14**
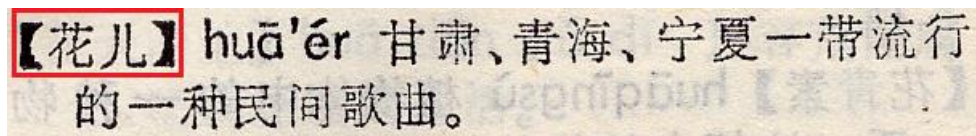
Although most books published in China use a standard-sized 儿/兒 to represent *érhuà* usage, small-sized 儿 is commonly used in publications relating to Chinese linguistics (see Fig. 6). Moreover, the practice of distinguishing *érhuà* with a small-sized 儿 is adopted in the authoritative concise dictionary of Chinese, *Xiàndài Hànyǔ Cídiǎn* 现代汉语词典 [Dictionary of Contemporary Chinese], published in Beijing by the Commercial Press since 1978. See Fig. 3 and Fig. 4 where the dictionary entry for huā 花 'flower' shows *huār* 花儿 and *huā'ér* 花儿 used constrastively.

Fig. 3: *Xiàndài Hànyǔ Cídiǎn* 现代汉语词典 (2nd ed. 1983) p. 478

Entry for huā 花 'flower' showing *huār* 花儿 with small *ér* 儿

Fig. 4: *Xiàndài Hànyǔ Cídiǎn* 现代汉语词典 [Contemporary Chinese Dictionary] (2nd ed. 1983) p. 479

Entry for the word *huā'ér* 花儿 (a type of folk song) with full-sized *ér*

2. Proposal

There is a need to be able to preserve the small 儿/兒 character representing *érhuà* in plain text, and to distinguish between ordinary 儿/兒 and small 儿/兒 in text processing. The small-sized 儿/兒 are essentially presentation forms of the characters U+513F 儿 and U+5152 兒, but they are not simple stylistic variants as they have a specific linguistic function, and are pronounced differently (*-r* vs *ér*). In 2016 we proposed to represent the characters as standardized variation sequences of U+513F and U+5152 (see WG2 N4720 = L2/16-109), but the UTC rejected this proposal, and suggested that registering Ideographic Variation Sequences (IVS) in the Ideographic Variation Database (IVD) would be a more suitable solution (UTC Meeting 147 B.15.3).

The Ideographic Variation Database allows for the registration of stylistic variants of CJK unified ideographs, but as the small-sized 儿/兒 characters are not simply stylistic variants, we do not consider that they are appropriate to be registered as Ideographic Variation Sequences. After recent discussion with CJK experts, including Dr Ken Lunde, we now propose to encode two separate characters representing small-sized 儿 and 兒 in the Ideographic Symbols and Punctuation block at U+16FF2 and U+16FF3.

Table 1: Proposed Characters

Code point	Glyph	Character name
16FF2	儿	CHINESE SMALL SIMPLIFIED ER
16FF3	兒	CHINESE SMALL TRADITIONAL ER

Character properties:

Script = Han
General Category = Lo [Letter, Other]
Canonical Combining Class = 0
Bidi Class = L
Line Break = NS [Nonstarter]
Ideographic = Y
Unified Ideograph = N
Extender = Y

Code chart annotations:

* small form of 513F representing erhua pronunciation
x 513F
* small form of 5152 representing erhua pronunciation
x 5152

---

3. Additional Evidence of Usage

Fig. 5: Sū Xīnchūn 苏新春,《〈现代汉语词典〉中的"儿"尾词该如何处理？》 [How should the Er-suffix in Xiandai Hanyu Cidian be treated?]; *Yǔwén Jiànshè* 语文建设 2000.12: 20

《现代汉语词典》中的『儿』尾词该如何处理？

□苏新春

《现代汉语词典》对"儿化"的解释是这样："汉语普通话和某些方言中的一种语音现象，就是后缀'儿'字不自成音节，而和前头的音节合在一起，使前一音节的韵母成为卷舌韵母。例如'花儿'的发音是 huār，不是 huāér。"这个解释很清楚，"儿化"不是音节，只是一个卷舌动作，依附在前一音节的后面，改变其性质，成为卷舌韵。声调也只有一个，标在前面的主要元音上。

《现代汉语词典》对儿化词的书写是在词的右下方标上一个小"儿"字，如"鼻儿"。这种标示是符合儿化性质的，"儿"不是独立的音节，也不是独立的汉字，它表示的只是一种附加性的词义色彩，或是含有"小"的意思，如"【鼻】① 器物上面能够穿上其他东西的小孔：门~｜针~。② [方] 象哨子的东西：用苇子做了一个小~"；或是含有"亲切"义、"名词"义等，如"哥儿""爷儿""今儿""块儿"等。儿化韵的"儿"与"宠儿""孤儿""健儿""胎儿"的"儿"，在语音性质和语义传递上都是完全不同的。

但细看《现汉》对"儿"类词的排列，人们却会生出儿化韵算不算一个独立音节的疑问。这是由《现汉》对"儿"化词的排列引起的。"白面儿""被窝儿"该算两个音节还是三个音节，"捡破烂儿""耍心眼儿"该算三个音节还是四个音节，人们一般不会细究，反正它们都是复音节词，但象"鼻儿""哥儿""爷儿""今儿""块儿"这样的词就含糊不了。根据儿化韵的性质来看，"鼻儿"类的词应该算单音节词，它们应该归到单字字头下，成为一义项义。而现在列入复音节词目序列的做法却很容易使人把它们当作双音节词来看。这是一个矛盾，不能小觑的矛盾，因为1996年的第三版《现代汉语词典》二音节以下的"儿"尾词共有74例，其中大部分是"儿"尾，达55例，大"儿"尾词才19例。在1983年的第二版中这样的词语也有70例。

将单音节的儿化词归入复音词词目的做法，会引起不必要的误解，或是会模糊人们对"儿化"语音性质的认识，或是会给人造成词典在编纂上自乱其例的印象。而将它们归入单字词目，成为其中的一个义项，在该义项前再标上"儿"字，对诸多方面都有好处。

Discussion of Small 儿, with small 儿 marked in red, and full-sized 儿 marked in blue

Fig. 6: Yīng Yǔtián 应雨田，《湖南安乡方言的儿化》 [Erhua in Hunan Anxiang dialect]; *Fāngyán* 方言 1990: 52

壹   儿化的作用和范围

　　安乡话的儿化主要有两方面的作用。一是构成名词,即将名词性语素或动词、形容词、量词等儿化成名词; 二是增添一定的附加意义,或表小称,或增添喜爱的感情色彩,或增添口语色彩,有时是三者兼而有之。具有构词作用的儿化大都增添了一定的附加意义,但具有修辞作用的却不一定具有构词作用。

　　从儿化的范围而言,能儿化的主要是名词(包括名词性语素)和数量短语,少数形容词、动词、量词也可儿化。下面按词类分别举例说明。

1·1　名词(包括名词性语素)的儿化

1·11　名词性语素儿化成名词

箉	箉箉儿 p'aɹ p'ərɬ 箉子: 竹~		杯	杯杯儿 pei˥ pərɬ 茶~｜酒~｜瓷~
码	码码儿 maɹ mərɬ 码子:数~｜价~｜画个~		嘴	嘴嘴儿 tseiɹ tsərɬ 茶壶~｜烟~
法	法法儿 fa˥ fərɬ 办法:巧~｜想不出~		苞	苞苞儿 pauɹ pərɬ 花~｜桃花结~
桠	桠桠儿 ŋaɹ ŋərɬ 桠杈:树~｜竹~		泡	泡泡儿 p'uɹ p'ərɬ 泡儿:水~｜肥皂~
渣	渣渣儿 tsaɹ tsərɬ 渣子:煤~｜油~｜豆腐~		泡	泡泡儿 p'auɹ p'ərɬ 电灯~｜~肉
末	末末儿 moɹ mərɬ 煤~｜茶叶~		帽	帽帽儿 mauɬ mərɬ 笔~｜螺丝~
垛	垛垛儿 toɹ tərɬ 垛子:门~｜墙~		套	套套儿 t'auɹ t'ərɬ 布~｜手~
坨	坨坨儿 t'oɹ t'ərɬ 泥巴~｜肉~		篙	篙篙儿 kauɹ kərɬ 篙子:竹~｜晒衣~
络	络络儿 loɹ lərɬ 斗笠里面的头箍:斗笠~		罩	罩罩儿 tsauɬ tsərɬ 罩子:灯~
盒	盒盒儿 xoɬ xərɬ 盆子:饭~｜火柴~		槽	槽槽儿 ts'auɹ ts'ərɬ 槽子:猪食~｜水~
座	座座儿 tsoɬ tsərɬ 座子:自行车~		兜	兜兜儿 touɹ tərɬ ①口袋一样的东西: 网~
格	格格儿 ke˥ kərɬ 格子:方~			②兜肚:系一个~｜~裤
枝	枝枝儿 tsʅ˥ tsərɬ 树~｜竹~｜柳~		斗	斗斗儿 touɹ tərɬ 柜、桌的抽屉
牌	牌牌儿 p'aiɹ p'ərɬ 门~｜纸~｜灵~		头	头头儿 t'ouɹ t'ərɬ 山~｜笔~｜起个~｜单位的~
袋	袋袋儿 taiɬ tərɬ 布~			
带	带带儿 taiɬ tərɬ 鞋~｜松紧~｜裤腰~		篓	篓篓儿 louɹ lərɬ 箕~｜字纸~｜筷子~

Full-sized 儿 [ér] marked in blue; small 儿 [-r] marked in red

大概是因爲忙着去考的緣故，他考中過舉人的。那時代平常總是先念《三字經》、《百家姓》、《千字文》。可是祖父一起頭兒就教我跟我哥哥兩個人念《大學》。我念念念不好就改了念朱子的《小學》，覺着好念多了，后來《小學》沒念完又回頭念《大學》了。我到七歲才開始正式上書房，天天早晨上學，晚上才回家。那時候我祖父做冀州的知州，書房在衙門的一個跨院兒里，所以多半在書房吃午飯。書房好像就只有我跟我哥哥跟一個親戚家的小孩兒仁人兒，因爲那時候我兩個堂房姊姊他們女孩子們都得躲得家里念書，不能跟男孩子們一塊兒念的。

Traditional form small 兒 marking *érhuà*

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html **.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Proposal to encode two small form CJK characters for Chinese*
2. Requester's name: *Andrew West and Eiso Chan (陈永聪)*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *2023-12-12*
5. Requester's reference (if applicable):
6. Choose one of the following:
    This is a complete proposal: *Yes*
    (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
    a. This proposal is for a new script (set of characters): *No*
        Proposed name of script:
    b. The proposal is for addition of character(s) to an existing block: *Yes*
        Name of the existing block: *Ideographic Symbols and Punctuation*
2. Number of characters in proposal:
3. Proposed category (select one from below - see section 2.2 of P&P document):
    A-Contemporary **X** B.1-Specialized (small collection)   B.2-Specialized (large collection)
    C-Major extinct   D-Attested extinct   E-Minor extinct
    F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided?
    a. If YES, are the names in accordance with the "character naming guidelines"
        in Annex L of P&P document? *Yes*
    b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Fonts related:
    a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
    Andrew West
    b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
    Andrew West
6. References:
    a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
    b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes*
7. Special encoding issues:
    Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *No*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

---

[1]. Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	*No*
If YES explain	
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	*Yes*
If YES, with whom?	*CJK and Unihan experts*
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	*No*
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	*common*
Reference:	
5. Are the proposed characters in current use by the user community?	*Yes*
If YES, where? Reference:	*Dictionaries, linguistic works, etc.*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	*No*
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	*Yes*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	*Yes*
If YES, is a rationale for its inclusion provided?	*Yes*
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	*No*
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	*Yes*
If YES, is a rationale for its inclusion provided?	*Yes*
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	*No*
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	*No*
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility characters?	*No*
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	