Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: ISO/IEC JTC 1/SC 2/WG 2/IRG
Title: US/Unicode Urgently Needed Character Proposal for Two Ideographs
Source: Ken Lunde
Status: Joint National/Member Body Contribution
Action: To be considered by IRG
Date: 2025-02-28

This document is a proposal to encode two ideographs as urgently needed characters.

Proposed Ideographs

The table below provides the representative glyphs and attributes for the two ideographs in this UNC proposal:

Glyph	Source Reference	IDS	RS	TS	FS	Reading (J)	Variants
庁	UTC-03567	⿸广K	53.3	6	0	ケイ (kei)	U+6176 慶
応	UTC-03568	⿸广O	53.1	4	0	オウ (ō)	U+5E94 应 U+5FDC 応 U+61C9 應

The two ideographs are abbreviated forms of the ideographs 慶 and 應 (応 is its official Japanese simplified form) that, when used together, represent an unofficial abbreviated form of 慶應 (*keiō*) in 慶應義塾大学 (*keiō gijuku daigaku*, meaning *Keio University*, which is a private research university located in the Minato ward of Tokyo, Japan). In other words, these abbreviated ideographs are not considered logos.

The proposed code points are U+2B81E (UTC-03567) and U+2B81F (UTC-03568), which are at the end of the CJK Unified Ideographs Extension D block. If accepted, they would fill that particular block.

Urgently Needed Rationale

The rationale for the urgency of encoding these two ideographs is mainly for the IRG to go on record to accept such ideographs prior to the announcement of the next IRG working set, which is at least two years from now. In addition, among the small number of known ideographs that include Latin components, these two are by far the most prominent, have a history that dates back nearly 100 years, have more than sufficient evidence, and are frequently used together as a pair. These two ideographs also missed the opportunity of being included in IRG Working Set 2024.

IRG Meeting #63 Documents & Discussions

Document IRG N2717, along with several feedback documents (IRG N2731, IRG N2738R, IRG N2741, IRG N2742, and IRG N2744), were discussed at length during IRG Meeting #63, culminating as Recommendation IRG M63.19 (*Script-Hybrid Han Ideographs*) in document IRG N2702:

The IRG discussed the proposal to accept script-hybrid Han ideographs, which would overturn the second recommendation of Recommendation IRG M61.18 in document IRG N2620, along with its five feedback documents. This topic will be included in the agenda for IRG Meeting #64 for further discussion, along with any additional documents on this topic.

Included in the discussions was a brief tour of UTN (Unicode Technical Note) #43 that documents the provisional kStrange property of the Unihan database, which made it abundantly clear that the ideographs in this proposal are no more "strange" than many ideographs that have already been encoded.

Open Issues & Questions

The following series of questions and answers are intended to address all of the open issues and questions that were discussed during IRG Meeting #63:

Q: *Are script-hybrid Han ideographs in common use?*

A: **Yes.** However, it depends on the language and region, which is a topic that I researched many years ago. Script-hybrid Han ideographs are very rare in Chinese-speaking regions due to the heavy use of the Han script. Among the Han ideographs that were coined in Korea, the vast majority use Hangul components whose forms are virtually identical to Han ideograph components. Given the extent to which the Hangul script is used in Korean-speaking regions, this makes sense. Among the Han ideographs that were coined in Japan, several hundred of which are already encoded, nearly all of them include only Han ideograph components. Among the Han ideographs that were coined in Vietnam, several thousand of which are already encoded, virtually all of them include only Han ideograph components. In other words, script-hybrid Han ideographs exist and are used together with normal Han ideographs, but the extent that they are considered common completely depends on the language and region.

Q: *Should the Latin components of Han ideographs be represented using Han ideograph strokes?*

A: **No.** At least for the two script-hybrid Han ideographs in this UNC proposal, every print-based evidence image that has been collected for this proposal clearly shows that the components are intended to be Latin, not Han. For example, if the enclosed component of 庇 were to be substituted with U+30020 𰀀 that uses Han ideograph strokes, the user community (aka Japan) would immediately recognize it as inappropriate.

Q: *Should Latin displacements of Han ideographs be treated as Han ideographs?*

A: **No.** As raised on page 2 of document IRG N2731, when Ｅ (U+FF25 Ｅ FULLWIDTH LATIN CAPITAL LETTER E) displaces 医, such as in Ｅ院/医院, it should simply be treated as a full-width Latin character, which are supported by virtually all East Asian fonts, and generally share the same glyph metrics as Han ideographs.

Q: *Should script-hybrid Han ideographs that include an uppercase or lowercase Latin component be treated as a case pair?*

A: **No.** Unless there is actual evidence that a form that includes a lowercase or uppercase Latin component exists, they should not be encoded. In other words, case pairs should be considered only when actual evidence of their use exists. If such case pairs exist, the *kSemanticVariant* property can be used to relate them.

Q: *Can Latin components be supported in IDSes?*

A: **Yes.** The current IDS syntax, as shown in Section 18.2, *Ideographic Description Characters*, of the *The Unicode® Standard Version 16.0 – Core Specification*, already allows U+FF1F ？FULLWIDTH QUESTION MARK to be

used for unknown components, and given the full-width nature of Han ideographs, the full-width Latin characters — U+FF21 through U+FF3A (Ａ through Ｚ) and U+FF41 through U+FF5A (ａ through ｚ) — can be added to the syntax as follows (additions are shown in **red**) to accommodate Latin components:

```
IDS := Ideographic | Radical | CJK_Stroke | Private Use | U+FF1F
  | U+FF21 | ... | U+FF3A | U+FF41 | ... | U+FF5A
  | IDS_UnaryOperator IDS
  | IDS_BinaryOperator IDS IDS
  | IDS_TrinaryOperator IDS IDS IDS
CJK_Stroke := U+31C0 | ... | U+31E5
IDS_UnaryOperator := U+2FFE | U+2FFF
IDS_BinaryOperator := U+2FF0 | U+2FF1 | U+2FF4 | ... | U+2FFD | U+31EF
IDS_TrinaryOperator := U+2FF2 | U+2FF3
```

Q: *Can the strokes of Latin components be reliably counted?*

A: **Yes.** Although there is no universal methodology for counting the strokes of Latin characters, if each line and curve were to be counted as a single stroke, the following table provides the recommended stroke counts for the upper and lowercase Latin characters (the values highlighted in yellow appear to have a universally-accepted number of strokes):

Uppercase	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Strokes	3	3	1	2	4	3	2	3	3	1	3	2	4	3	1	2	2	3	1	2	1	2	4	2	3	3
Lowercase	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
Strokes	2	2	1	2	2	2	2	2	2	2	3	1	3	2	1	2	2	2	1	2	2	2	4	2	2	3

See the section entitled **Counting Latin Character Strokes** on pp 9 and 10 of this document for examples of Latin character stroke-counting diagrams.

Counting the strokes of the characters of other East Asian scripts, such as Katakana, Hiragana, and Hangul, that may also serve as components of script-hybrid Han ideographs, can be reliably counted as well, and perhaps more reliably, because there are stroke-counting standards for such characters. In particular, the strokes of the characters for the Katakana and Hangul scripts closely mimic those of the Han script. The strokes of the characters for the Hiragana script are largely cursive, yet still have stroke-counting rules. In other words, counting the strokes of non-Han components should be considered a non-problem.

Q: *Can first residual stroke (aka FS) values be specified for Latin components?*

A: **Yes.** As proposed in document IRG N2715, if an ideograph has no residual strokes, or if its residual stroke value is unclear, its first residual stroke value shall be set to 0 (*zero*). In other words, if the proposal in document IRG N2715 is accepted, the value 0 (*zero*) should be used when the first residual component of a Han ideograph is Latin.

Q: *Should script-hybrid Han ideographs be encoded as ordinary CJK Unified Ideographs?*

A: **Yes.** Such ideographs are used like ordinary Han ideographs: their primary component (aka Kangxi Radical) is Han, they have well-established readings and meanings, and they are used in text with other Han ideographs. As demonstrated by the *kStrange* property and its documentation in UTN #43, the various CJK Unified Ideographs blocks already include Han ideographs that include Katakana and Hangul components. An example that includes a genuine Hangul component is U+2D939 𭤹 whose code chart excerpt is shown below:

2D939
方 70.6

KC-07034

Q: *Should script-hybrid Han ideographs be encoded in a separate CJK Unified Ideographs block?*

A: **No.** As the *kStrange* property and its documentation in UTN #43 clearly demonstrate, a non-trivial number of script-hybrid Han ideographs have already been encoded, and are present in various CJK Unified Ideographs blocks. In other words, there is no precedent for encoding such Han ideographs in a separate block. Furthermore, encoding script-hybrid Han ideographs in a separate block would entail an omnibus proposal whereby all known script-hybrid Han ideographs would be included. Such an approach works for CJK components, because all of the characters serve a specific purpose. However, each script-hybrid Han ideograph should be considered on its own merits, thus processed in IRG working sets.

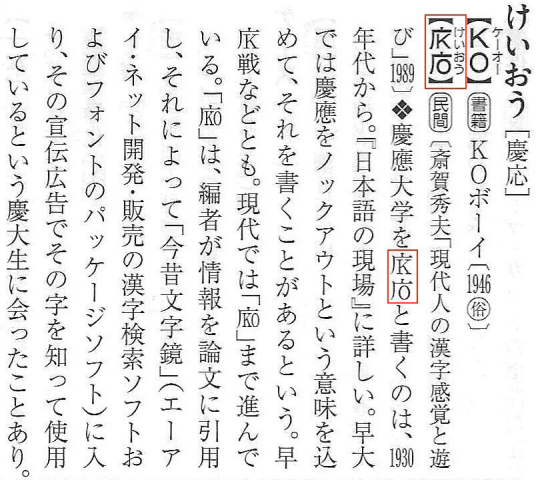
Evidence

This section includes six images that provide printed evidence for both ideographs. Note that some of the evidence images include ⿸广⿰KO that represents a single abbreviated form of both ideographs, probably because both ideographs share Kangxi Radical #53. Its use is not nearly as widespread as the two separate abbreviated forms, so it is not included in this UNC proposal.

The first evidence image is from the ISO/IEC TR 10036:2020 standard, specifically the glyphs with serial numbers 10066927 and 10066928, which can also be viewed on this web page and on page 1161 of this PDF file:

10066927	10066928
	

The second evidence image is an excerpt from page 247 of the dictionary entitled “当て字・当て読み 漢字表現辞典,” edited by 笹原宏之, and published by 三省堂 in 2010 (ISBN-13 978-4-385-13720-9):



けいおう［慶応］
【KO】（ケーオー）（書籍）KOボーイ〔1946㊎〕
【庆应】（けいおう）（民間）〔斎賀秀夫「現代人の漢字感覚と遊び」1989〕❖慶應大学を庆应と書くのは、1930年代から。『日本語の現場』に詳しい。早大では慶應をノックアウトという意味を込めて、それを書くことがあるという。早庆戦などとも。現代では「庆」まで進んでいる。「庆」は、編者が情報を論文に引用し、それによって「今昔文字鏡」（エーアイ・ネット開発・販売の漢字検索ソフトおよびフォントのパッケージソフト）に入り、その宣伝広告でその字を知って使用しているという慶大生に会ったことあり。

The third evidence image is page 70 of the book entitled “奇妙な漢字,” written by 杉岡幸徳, and published by ポプラ社 in 2023 (ISBN-13 978-4-591-17603-0):

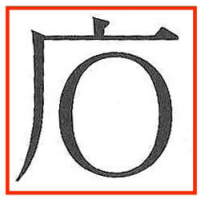
COLUMN

謎めいたスラング

広部3画（6）

読み
ケイ

意味
慶應の慶



広部1画（4）

読み
オウ

意味
慶應の應

広部4画（7）

読み
ケイオウ

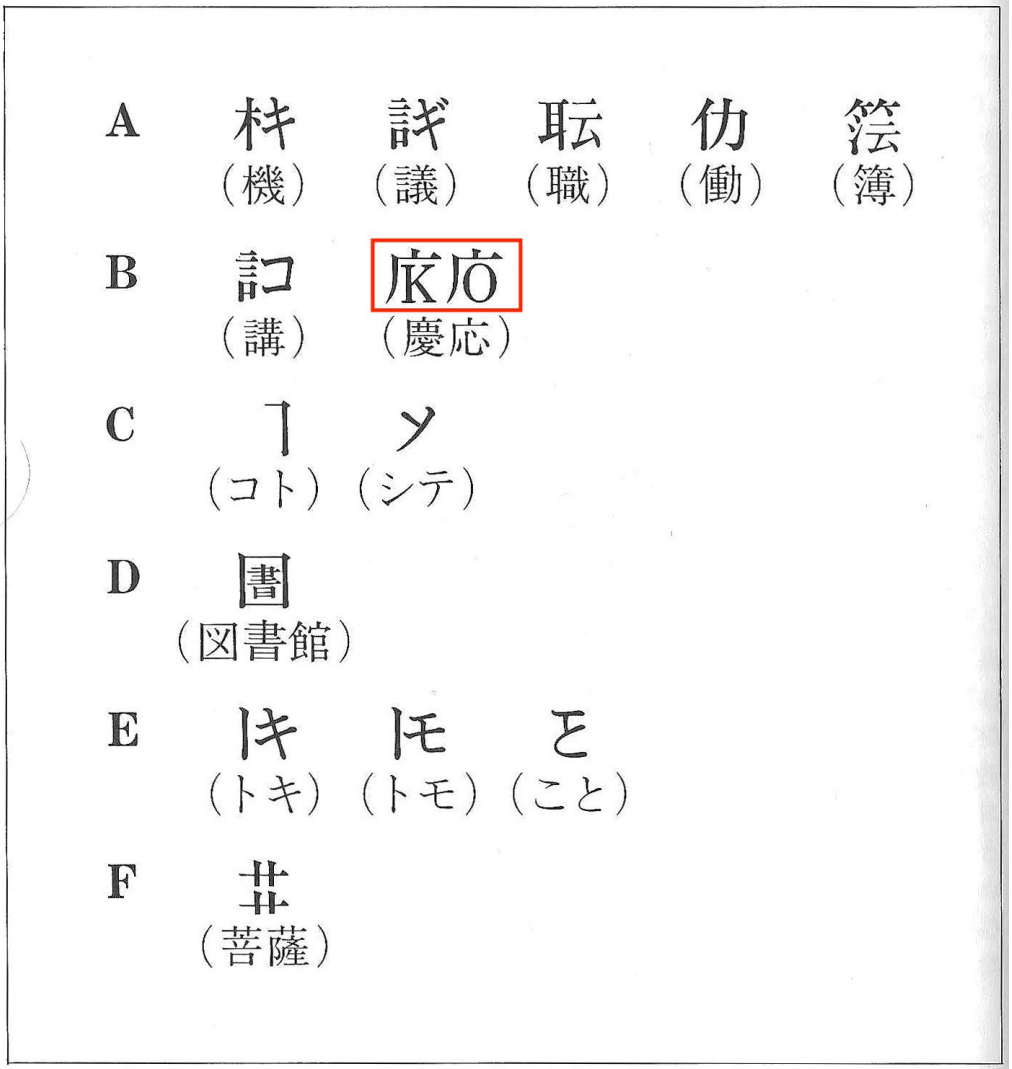
意味
慶應

完全にふざけた字だが、これは慶應義塾の学生だけが密かに使っているという漢字。「慶」は广と書き、「應」は心と書く。「慶應」で㢧という字まであり、「㢧大」「广早戦」のような形で使われる。おそらく、「慶應」という漢字はあまりに画数が多すぎて、書くのが面倒なのだろう。

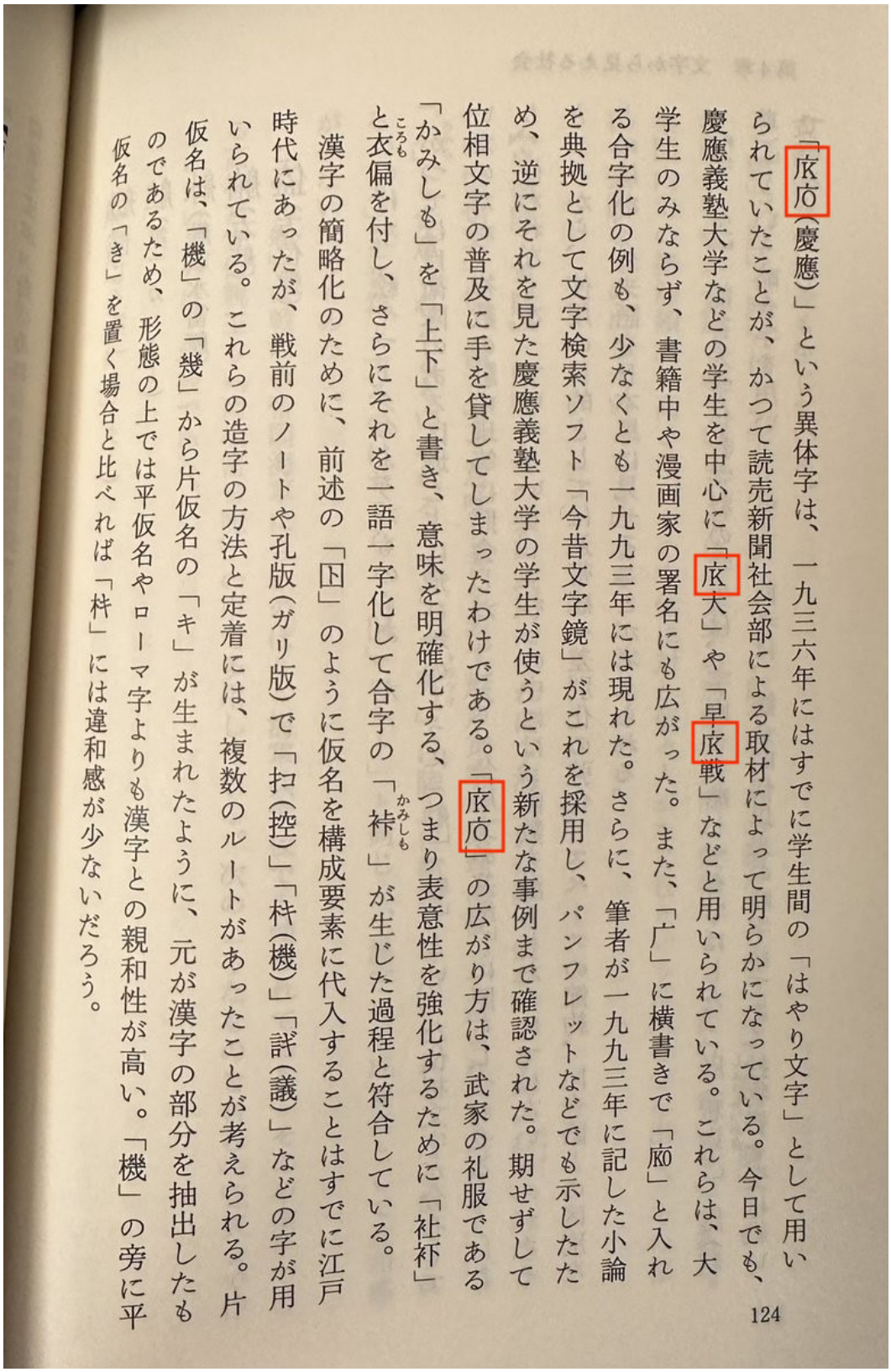
また、この漢字が一種のスラングになっていて、メンバー間の結束を高めるために役立っているのかもしれない。

The fourth evidence image is page 68 of the book entitled “事典 日本の文字,” edited by 樺島忠夫 et al., and published by 大修館書店 in 1985 (ISBN-10 4-469-01209-2):

略字・合わせ字一覧



The fifth evidence image is page 124 of the book entitled “日本の漢字,” written by 笹原宏之, and published by 岩波新書 in 2006 (ISBN-13 978-4-004-30991-8):



「㕣応（慶應）」という異体字は、一九三六年にはすでに学生間の「はやり文字」として用いられていたことが、かつて読売新聞社会部による取材によって明らかになっている。今日でも、慶應義塾大学などの学生を中心に「㕣大」や「早㕣戦」などと用いられている。これらは、大学生のみならず、書籍中や漫画家の署名にも広がった。また、「广」に横書きで「庶」と入れる合字化の例も、少なくとも一九九三年には現れた。さらに、筆者が一九九三年に記した小論を典拠として文字検索ソフト「今昔文字鏡」がこれを採用し、パンフレットなどでも示したため、逆にそれを見た慶應義塾大学の学生が使うという新たな事例まで確認された。期せずして位相文字の普及に手を貸してしまったわけである。「㕣応」の広がり方は、武家の礼服である「かみしも」を「上下」と書き、意味を明確化する、つまり表意性を強化するために「衽衿」と衣偏を付し、さらにそれを一語一字化して合字の「裃（かみしも）」が生じた過程と符合している。

漢字の簡略化のために、前述の「囚」のように仮名を構成要素に代入することはすでに江戸時代にあったが、戦前のノートや孔版（ガリ版）で「扣（控）」「杵（機）」「訛（議）」などの字が用いられている。これらの造字の方法と定着には、複数のルートがあったことが考えられる。片仮名は、「機」の「幾」から片仮名の「キ」が生まれたように、元が漢字の部分を抽出したものであるため、形態の上では平仮名やローマ字よりも漢字との親和性が高い。「機」の旁に平仮名の「き」を置く場合と比べれば「杵」には違和感が少ないだろう。

124

The sixth evidence image is page 208 of the paper entitled “一九六五〜一九七五年度頃の略字,” written by 蜂矢真郷, in the journal entitled “国語文字史の研究 八,” edited by 国語文字史研究会, and published by 和泉書院 in 2005 (ISBN-10 4757603045):

「藤」、6「導」)、〜イのもの(3「勢」)については、韻尾を切り離して用いること(2)cKのものと同様である。

六

そして、(2)cRの例である。これらは、ローマ字が声符となる新形声文字である。

1慶→广K　例、广K广O(慶応)

2応→广O［应］　例、同右

1は「K(ケイ)」が声符、2は「O(オウ)」が声符であるので、新形声文字と見ることに問題はない。

3確→石K［确］　例、石K言N(確認)　前掲(2)cKの1参照

4認→言N［认］　例、同右　前掲(2)bの2参照

3「確」は、頭子音がkで、かつ、韻尾もkであるので、その意味において、カクを表す声符に「K」を用いるのは、それなりに理由があると言うことができる。同様に、4「認」は、頭子音がnで、かつ、韻尾もnであるので、その意味において、ニンを表す声符に「N」を用いるのは、それなりに理由があると言うことができる。

5還→辶K［还］　例、返辶K(返還)　前掲(1)aの1参照

6館→飠K［馆］　例、本飠K(本館)

7韓→龺K［韩］　例、龺K囗丶(韓国)「囗丶」は前掲(1)cの1参照

8軍→冖G　例、冖G事(軍事)

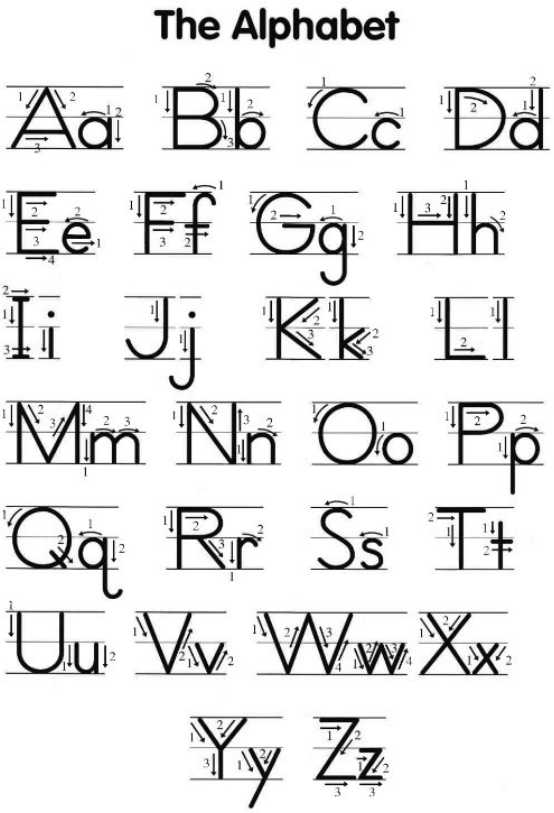
9寮→宀R　例、〇〇宀R(〇〇寮)

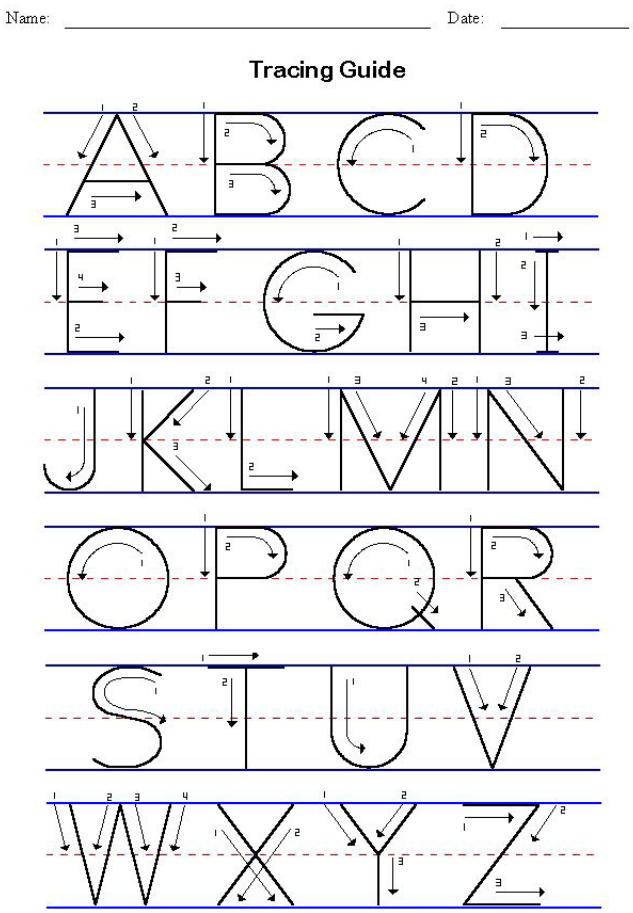
10療→疒R［疗］　例、医疒R(医療)　前掲(2)aの5参照

11縄→糸N　例、沖糸N(沖縄)

Counting Latin Character Strokes

This section includes for reference the Latin character stroke-counting diagrams that were used as the basis for the uniform methodology for counting the strokes of Latin characters when used as components of Han ideographs (source URLs are not provided, because these stroke-counting diagrams serve only as evidence that the counting of Latin strokes is not consistent across sources):





That is all.

ISO/IEC JTC 1/SC 2/WG 2/IRG
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITION OF CJK UNIFIED IDEOGRAPHS TO THE REPERTOIRE OF ISO/IEC 10646
Submitters are reminded to:
1.Fill in all the sections below.
2. Read the Principles and Procedures Document (P & P) available at
https://appsrv.cse.cuhk.edu.hk/~irg/irg/irg56/IRGN2424Confirmed.pdf
for guidelines and details before filling in this form.
3. Use the latest Form from
https://appsrv.cse.cuhk.edu.hk/~irg/irg/irg56/IRGN2424SubmissionForm.xlsx
See also http://appsrv.cse.cuhk.edu.hk/~irg/irgwds.html **for the latest *Unifiable Component Variations*.**

Administrative

1. **IRG Project Code:**	IRG N2789
2. **Title:**	US/Unicode Urgently Needed Character Proposal for Two Ideographs
3. Submitter's Region/Country Name:	US / Unicode Consortium
4. Submitter Type (National Body/Individual Contribution):	Joint National / Member Body
5. Submission Date:	2025-02-28
6. Requested Ideograph Type (Unified or Compatibility Ideographs)	Unified
If Compatibility, the submitter is strongly encouraged to instead register them as IVS in a new or an existing IVD collection(See UTS #37) with the IRG's approval (Registration fee will not be charged if authorized by the IRG.).	
7. Proposal Type (Normal Proposal or Urgently Needed)	Urgently Needed
8. Choose one of the following:	
This is a complete proposal.	Yes
(or) More information will be provided later.	

B. Technical – General

1. Number of ideographs in the proposal:	*2*
2. Glyph format of the proposed ideographs is in TrueType?	*Yes*
Are all the proposed glyphs put into BMP PUA area?	*Yes*
Are data for source references vs. character codes provided?	*Yes*
3. Source references:	
Do all the proposed ideographs have a unique, proper source reference (member body/international consortium abbreviation followed by no more than 9 alphanumeric characters)?	Yes
4. Evidence:	
a. Do all the proposed ideographs have a separate evidence document which contains at least one scanned image of printed materials (preferably dictionaries)?	Yes
b. Do all the printed materials used for evidence provide enough information to track them by a third party (ISBN numbers, etc.)?	Yes
5. Attribute Data Format: (Excel file or CSV text)	*Excel*

C. Technical - Checklist

Understanding of the Unification Principles	
1. Has the submitter read ISO/IEC 10646 Annex S and does the submitter understand the unification principles?	*Yes*
2. Has the submitter read the "Unifiable Component Variations" (contact the IRG technical editor through the IRG Convenor for the latest version) and does the submitter understand the unifiable variation examples?	*Yes*
3. Has the submitter read the IRG PnP document and does the submitter understand the 5% Rule?	*Yes*
Character-Glyph Duplication (http://www.itscj.ipsj.or.jp/sc2/open/pow.htm contains all the published ones and those under ballot**)**	
4. Has the submitter checked that the proposed ideographs are ***not unifiable*** with any of the unified or compatibility ideographs of the latest version of ISO/IEC 10646?	*Yes*
If the checking has been done against an earlier version of ISO/IEC 10646, please specify the version. (e.g. 10646:2012)	
5. Has the submitter checked that the proposed ideographs are ***not unifiable*** with any of the ideographs in the amendments, if any, of the latest version of ISO/IEC 10646?	*Yes*
If yes, which amendment(s) has the submitter checked?	6th Edition Amd 1
6. Has the submitter checked that the proposed ideographs are ***not unifiable*** with any of the ideographs in the proposed amendments, if any, of ISO/IEC 10646?	*Yes*
If yes, which draft amendment(s) has the submitter checked?	6th Edition CDAM2
7. Has the submitter checked that the proposed ideographs are ***not unifiable*** with any of the ideographs in the current working M-set and D-set of the IRG? (Contact IRG chief editor and technical editor through the IRG Convenor for the newest list)	*Yes*
If yes, which document(s) has the submitter checked?	WS2024
8. Has the submitter checked that the proposed ideographs are ***not unifiable*** with any of the over-unified or mis-unified ideographs in ISO/IEC 10646? (See Annex E of the IRG PnP document)	*Yes*
9. Has the submitter checked whether the proposed ideographs have any ***similar ideographs*** in the current standardized or working sets mentioned above?	*Yes*
10. Has the submitter checked whether the proposed ideographs have any ***variant ideographs*** in the current standardized or working sets mentioned above?	*Yes*
Attribute Data	
11. Do all the proposed ideographs have attribute data including the Kangxi radical code, stroke count, and first stroke(primary)?	*Yes*
12. Do the proposed ideographs contain secondary radical code and their stroke count and first stroke are also provided?	*No*
13. Do all the proposed ideographs have the document page number of evidence documents in the attribute data?	*Yes*
14. Do all the proposed ideographs have the proper Ideographic Description Sequence (IDS) in the attribute data?	*Yes*
If no, how many proposed ideographs do not have the IDS?	
15. If the answer to question 9 or 10 is yes, do the attribute data include any information on similar/variant ideographs for the proposed ideographs?	*Yes*
16. Do all the proposed ideographs contain the total stroke count (kTotalStrokes)[1]?	*Yes*

[1] The IRG understands that kTotalStrokes can be ambiguous and subject to different interpretations. The IRG takes no responsibility to check the correctness of the submitted attribute data.