

Unicode request to add capital U with left hook

Ismael Robles Haloui, independent researcher in psycholinguistics

arbadac.arba65@gmail.com

2025 May 3

An uppercase letter U with left hook, U, was used as a case pair with lowercase u (already encoded as U+AB52 LATIN SMALL LETTER U WITH LEFT HOOK) in the following works:

- Noah Webster's 1828 dictionary of American English, where it represents the diphthong [ju] as in *union*
- Samuel Noory's 1981 pronunciation dictionary, where it represents the open-mid back unrounded vowel [ʌ]

Both of these dictionaries employed phonetic notations that employed uppercase glyphs in one way or another; in particular, Webster incorporated all-caps phonetic letters in his entries for indicating pronunciation, while Noory used capitals for proper nouns. In light of this, the addition of capital U is required for the proper digitization of the dictionaries in question.

Character

U U+AB6E LATIN CAPITAL LETTER U WITH LEFT HOOK

UnicodeData.txt

```
AB52;LATIN SMALL LETTER U WITH LEFT HOOK;Ll;0;L;;;;N;;;AB6E;;AB6E
AB6E;LATIN CAPITAL LETTER U WITH LEFT HOOK;Lu;0;L;;;;N;;;AB52;
```

Annotations

The following annotations are recommended:

@ Addition for case-sensitive phonetic notations

AB6E LATIN CAPITAL LETTER U WITH LEFT HOOK

* English pronunciation (Webster, Noory)

* lowercase is AB52

The following annotation should be added in NamesList.txt:

AB52 LATIN SMALL LETTER U WITH LEFT HOOK

*** uppercase is AB6E**

References

- Noory, S. (1981). *Dictionary of pronunciation: guide to English spelling and speech* (4th edition). Cornwall Books. <https://archive.org/details/dictionaryofpron0000noor>
- Webster, N. (1828). *An American dictionary of the English language*. University of California Libraries.
 - First volume (comprising entries starting with letters A through I): <https://archive.org/details/americandictionarya01websrich>
 - Second volume (comprising entries starting with letters J through Z): <https://archive.org/details/americandictionarya02websrich>

Figures

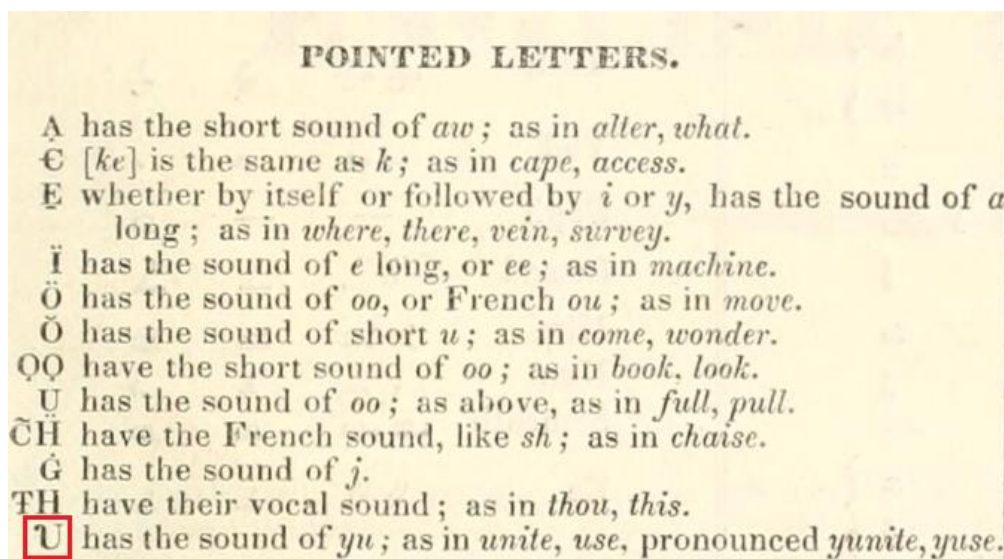


Fig. 1. Capital letters with diacritics, including U, in Webster (1828, vol. 1, p. 85).

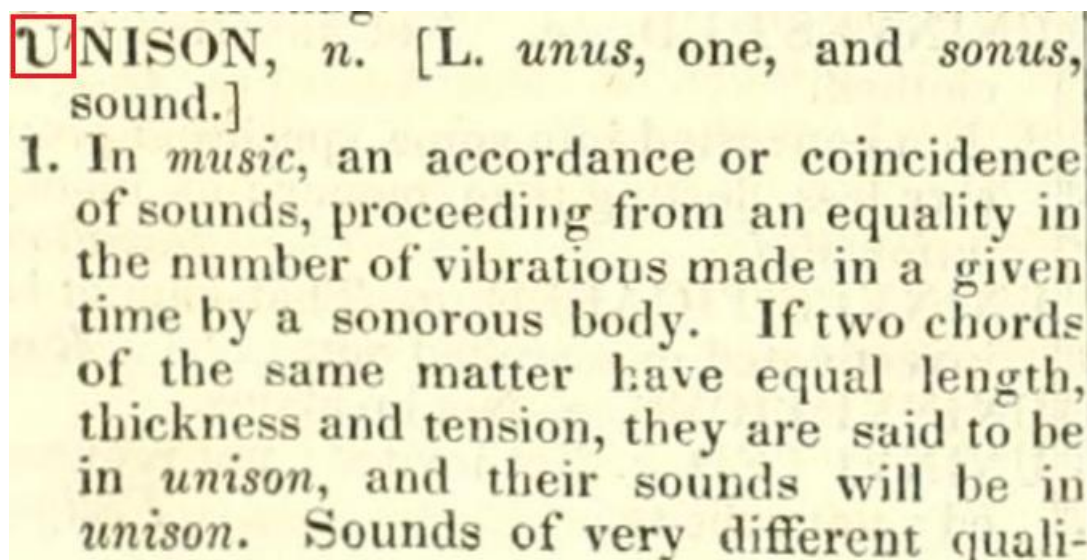


Fig. 2. The entry U'NISON in Webster (1828, vol. 2, p. 843).

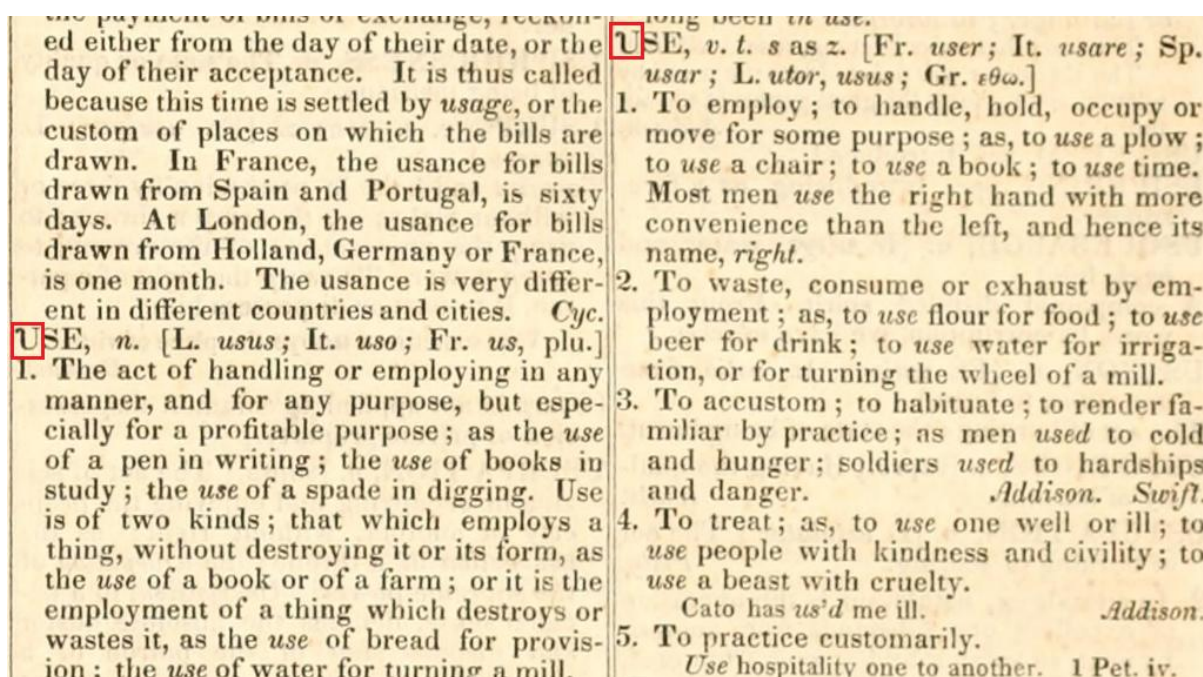


Fig. 3. The respective entries for U'SE as a noun and as a verb in Webster (1828, vol. 2, p. 843).

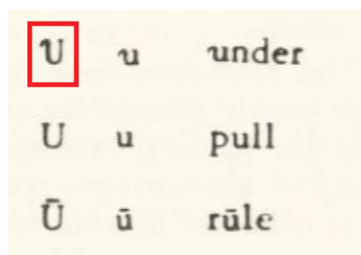


Fig. 4. U u (alongside U u and Ū ū) in Noory (1981, p. xxxvii).

U, u (Yü)

§ 66. A short, open-back vowel (formed with the mouth markedly open, and the back of the tongue away from, but pointing toward, the palate) corresponding to the long o (as in *odd*, mentioned in § 45). The sound is commonly indicated in the ordinary spelling by *u* as in *but* and by *o* as in *ton*. Less commonly, the sound is indicated by *oe* as in *does*, by *oo* as in *flood*, by *ou* as in *double*, by *wo* as in *twopence* (tu'pens) and by *e* as in *fellow* (fu'le, as sometimes heard in colloquial speech).

§ 67. Accented *u*, *o*, *ou*, when followed by *r* or *rr*, plus a vowel, are in some words variously pronounced *u* (as in *but*), or *e* (as in *err*). Thus, *burry*—hu'ri or he'ri; *borough*—bu'rō or be'rō; *flourish*—flu'rish or fle'rish. In order not to deviate unnecessarily from letters in the ordinary spelling, and to avoid ambiguities, the *u* (as in *but*) sound is invariably assigned in the DICTIONARY to *u's*, *o's* and *ou's* in the above category.

Fig. 5. Description of the properties of U u in Noory (1981, p. xlix).

U-ban-gi, river	U-bawn'gi	Un-ley, g	→ Un'li
U-ban-gi - Sha-ri, Cen- tral African Republic*	U-bawn-gi - Shaw'ri	Un-ter-mey-er, poet	→ Un'ter-my-er
U-cay-a-li, river	Ū-ky-aw'li	Un-ter-wal-den, p	Un'ter-vawl-den
Uck-field, g	→ Uk'feld	U-nyam-we-zi, region	Ū-nyawm-way'zi
U-dai-pur, g	Ū-dy-pūr'	U-per-ni-vik, g	Ū'per-ni-vēk
U-dall, p	Yū'dawl	Up-ham, p	→ U'pem
U-den, g	Ū'den, 73	U-phar-sin, p	Yu-far'sin
U-di-ne, g	Ū'di-ne	Up-ing-ton, g	→ U'ping-ten
Ud-murt, var of Votyak	Ūd-mūrt'	Up-land, g	→ Up'lend
Uel-e, Wel-le*, river	We'le	U-po-lu, g	Ū-pō'lu
Uel-zen, g	Vel'tsen	Up-per Aus-tri-a	U-per Aws'tri-e
U-fa, in U.S.S.R.	Ū'fe	Up-per Ba-var-i-a	U-per Be-vay'ri-e
U-gan-da, g	Yu-gan'de	Up-per Hesse	U-per Hēs'
U-gines, g	Ū-zhēn', 73	Up-per Vol-ta	U-per Vōl'te
Uhl, English surname	Yūl	Up-per Si-le-sia	U-per Si-lē'zhe
Uh-land, German poet	Ū'lawnt	Up-ping-ham, g	U'ping-hem
Uh-ler, English surname	Yū'ler	Up-sa-la*, Upp-, g	Up'se-le
U-in-ta Moun-tains	Yu-in-te Mown'tenz	Up-shur, p	Up'sher
U-ji-ji, g	Ū-jē'ji	Up-ton, g, p	Up'ten
		Ur, a	Er
		U-ral Moun-tains	Yū-rel Mown'tenz
		U-ralsk, g	Ū-rawlsk'
		U-ra-ni-a, m	Yu-ray'ni-e
		U-ra-nus, m	Yū're-nes; Yu-ray'

Fig. 6. Various proper names whose phonetic spelling starts with U, in Noory (1981, p. 498).

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹.**

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:	Unicode request to add capital U with left hook		
2. Requester's name:	Ismael Robles		
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution		
4. Submission date:	2025-05-02		
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:	Latin Extended-E		
2. Number of characters in proposal:			
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	A	B.1-Specialized (small collection)	
C-Major extinct		B.2-Specialized (large collection)	
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?			
Ismael Robles Haloui			
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):			
Google (Noto release)			
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.)?	No
If YES, with whom?	
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?	Linguists, lexicographers
Reference:	See proposal
4. The context of use for the proposed characters (type of use; common or rare)	Orthographic
Reference:	See proposal
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference:	
See proposal	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	Yes
If YES, is a rationale provided?	
Casing of BMP characters	
If YES, reference:	
See proposal	
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?	No
If YES, is a rationale for its inclusion provided?	
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?	No
If YES, is a rationale for its inclusion provided?	
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:	