

Proposal to encode COMPOSE

Author: Doruk K.

Date: 20 April 2026

1. Introduction

The COMPOSE character is a non-printable character that comes in between two printable characters to compose a new glyph. The resulting glyph is the second character visually overlaid (“overstruck”) on top of the first one, centered horizontally on the first glyph; and has the same width of the glyph widest of the two. This does not affect the vertical placement of either.

m + | → m̂
 ! + - → !̂
 2 + . → 2̂

Figure 1.1: Representation of the "overstriking" process

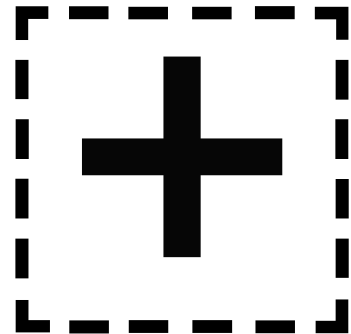


Figure 1.2: Proposed glyph for representing COMPOSE

1.1 Overstriking

Overstriking, the process described above, is a common way to obtain more glyphs on a typewriter using the existing base glyph set available. The result should be treated as one grapheme cluster, and should be treated as one unit when it comes to combining diacritics placement.

This process is not clearly markup unlike *italic* or *bold* styles, as the result is a distinct glyph that isn't necessarily related to the meaning of the two composing characters. Therefore, the result is in essence, a plain-text transformation process that most mechanical typewriters and many electric typewriters can perform easily, that the current encoding cannot encode reliably and requires special software or manual alignment.

This process also allows the user to derive and form almost infinite combinations of different glyphs without having to individually including each needed glyph in the Unicode standard. It is already used for constructing mathematical notation and symbols, e.g. \int U+2A0B

SUMMATION WITH INTEGRAL, \int U+2A16 QUATERNION INTEGRAL OPERATOR. Making this process standardized would facilitate using new notation.

This character resembles the former ^{APL} COMPOSE which used to be encoded as U+2300 until Unicode 1.0.1, but unlike APL COMPOSE, COMPOSE is general purpose and is not specifically for APL symbols.

1.2 Confusability Issues

Some of the composed glyphs can be confusable with other glyphs in some fonts.

Combined		Result	Resembles
≡	L	E	E
S		\$	\$

This property of overstriking is frequently utilized by typewriters to omit characters and construct glyphs that are otherwise not typable. And like typewriters, it can be used to construct glyphs that resemble duplicates of other Unicode characters in certain fonts. While this is an issue that cannot be ignored, it should be noted that many of these potential duplicates can already be replicated using existing combining characters, such as U+0046 F LATIN CAPITAL F and either of U+0332 ◌ COMBINING LOW LINE or U+0331 ◌ COMBINING MACRON BELOW can resemble U+0045 E LATIN CAPITAL LETTER E in fonts. Combining diacritics are comparatively worse in this issue when it's considered that confusability with overstriking varies by font style and spacing, and therefore not generally universal. In contrast, some combining diacritic confusabilities like U+0307 ◌ COMBINING DOT ABOVE used together with U+006A j LATIN SMALL LETTER J or U+0237 j LATIN SMALL LETTER DOTLESS J, do not only generally look like U+006A j LATIN SMALL LETTER J, but they are usually expected to look similar to it in all normal conditions.

Some combinations with modified letters may also cause it to resemble combining character combinations, for example the combination of U+02DC ~ SMALL TILDE or a simple combining U+0303 COMBINING TILDE with LATIN SMALL LETTER A appears similar on certain font styles: ã ã. Despite this, the placement of diacritic-like combinations are often different than their combining cousins.

The behavior of this character can also be compared to some implementations of U+0008 ^B BACKSPACE on physical teletypes, although that character's meaning is varying and ambiguous, and virtually no modern implementation treat it like COMPOSE.

1.3 Encoding

Proposed line in UnicodeData.txt:

```
[CODEPOINT];COMPOSE;Cf;0;BN;;;;;N;;;;;
```

It would in theory not be a problem to place it outside the BMP, but since there is currently no suitable block for it outside of the BMP, it appears that putting it in Combining Diacritical Marks Extended is currently the best option.

It functions similar to U+200D ZERO WIDTH SPACE in behavior outside of visuals.

2. Uses (examples with emojis are only demonstrative; U+FE0X are omitted)



Figure 2.1: "Calls are OK." (2606 ☒ 2714)



Figure 2.2: "No calls." (2606 ☒ 274C)



Figure 2.3: "Calls please." (2606 ☒ 203C)



Figure 2.4: APL character \mathcal{O} , composed using the characters O, U and T, which was encoded as U+2301 \mathcal{O} APL OUT until Unicode 1.0.1.

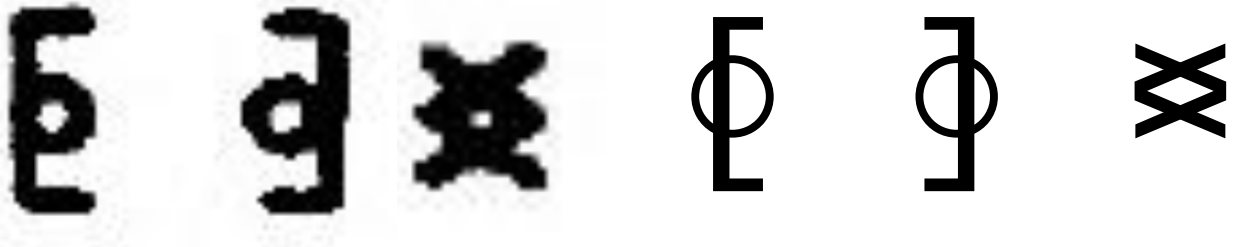


Figure 2.5: APL characters ϕ ϕ ∞ in EBCDIC terminals constructed using overstriking, used for substitutes for {, } and] respectively.



Figure 2.6: Nonstandard punctuations ! “exclamation comma”, ? “question comma” and ⌞ “friendly period” (203F + 002E) constructed using overstriking.

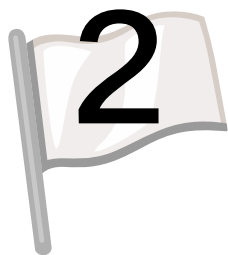


Figure 2.7: Flag with number 2 (1F3F3 + 00B2)



Figure 2.8: Speech bubble with cross, expressing denial (1F5E8 + 00D7)

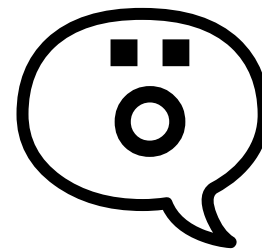


Figure 2.9: Speech bubble expressing surprise (1F5E8 + 2364)



Figure 2.10: Encrypted mail (2709 + 1F512)



Figure 2.11: Laptop overheating (1F4BB + 1F525)



Figure 2.12: Reminder ribbon on black background, representing mourning/black ribbon (2B1B + 1F397)

3. Conclusion

Overstriking cannot be completely ruled out as markup or formatting, as it produces a distinct grapheme. The same hypothetical criteria that could rule it out as markup can be likewise used to rule out many other characters, such as many of the APL characters in Miscellaneous Technical, or the characters in the blocks Variation Selectors or Mathematical Alphanumeric Symbols, which are clearly not markup.

In the case that this proposal gets rejected, it would still be important to find a way to encode overstriking in some way given its plaintext nature.

Sources

- Former characters APL COMPOSE U+2301 APL COMPOSE, **⊖** U+2300 APL OUT, listed in: [The Unicode Consortium. \(1991\). *The Unicode Standard, Version 1.0, Volume 1*. Addison-Wesley Developers Press.](#)
- The characters in Figure 2.5: ϕ ϕ \approx are listed in [Honeywell Software Multics APL User's Guide](#), December 1985.
- Nonstandard punctuation marks in Figure 2.6: $!$ $?$ \cdot are widely mentioned on the Internet. They are included in the [PUA of the font Fairfax HD](#), (U+F814-5, U+FF10A), [PUA of the font Catrinity](#), (U+EB88-9, —), [PUA of the font Nishiki-Teki](#) (U+E9F8-9, —).
- Emoji images are sourced from Noto Emoji, © Google LLC, licensed under [Apache License v2.0](#)

**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:	<i>Proposal to encode COMPOSE</i>
2. Requester's name:	<i>Doruk K.</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>20 April 2026</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>Yes</i>
(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):	
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<i>Yes</i>
Name of the existing block:	<i>Combining Diacritical Marks Extended</i>
2. Number of characters in proposal:	<i>1</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input checked="" type="checkbox"/>
C-Major extinct <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>
4. Is a repertoire including character names provided?	<i>Yes</i>
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?	<i>Yes</i>
b. Are the character shapes attached in a legible form suitable for review?	
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Doruk K.</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Doruk K. (insertnamehere1923 AT outlook DOT com)</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>Yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>Yes</i>
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)?	<i>Yes</i>

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?	Yes
If YES explain	<i>Expanded usage</i>
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?	Yes
Reference:	<i>Proposal document</i>
4. The context of use for the proposed characters (type of use; common or rare)	rare
Reference:	
5. Are the proposed characters in current use by the user community?	Yes
If YES, where? Reference:	<i>Users of mechanical and many electrical typewriters; versions of APL</i>
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?	Yes
If YES, reference:	<i>Proposal document</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	N/A
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?	Yes
If YES, is a rationale for its inclusion provided?	Yes
If YES, reference:	<i>Proposal document</i>
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?	Yes
If YES, reference:	<i>Proposal document</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	Yes
If YES, describe in detail (include attachment if necessary)	<i>Proposal document; non-printable character that comes in between two printable characters to compose a new overstruck glyph</i>
13. Does the proposal contain any Ideographic compatibility characters?	No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	