

JTC1/SC2/WG2 N3917

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

Doc Type: Working Group Document**Title: Revised Proposal to encode a punctuation mark "Double Hyphen"****Source: Karl Pentzlin****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2010-09-28****Dashes and Hyphens**

- = U+2E4E DOUBLE HYPHEN
- 2010 hyphen
 - 2E17 double oblique hyphen
 - 003D equals sign
 - A78A modifier letter short equals sign
 - used in transcription of old German prints and handwritings
 - used in some non-standard punctuation
 - not intended for standard hyphens where the duplication is only a font variant

Properties:

2E4E;DOUBLE HYPHEN;Pd;0;ON;;;;;N;;;;;

Entry in LineBreak.TXT:

2E4E;BA # DOUBLE HYPHEN

1. Introduction

The "ordinary" hyphen, which is representable by U+002D HYPHEN-MINUS or U+2010 HYPHEN, usually is displayed by a single short horizontal dash, but has a considerable glyph variation: it can be slanted to oblique or doubled (stacked) according to the used font. For instance, in Fraktur (Blackletter) fonts, it commonly is represented by two stacked short oblique dashes.

However, in certain applications, double hyphens (consisting of two stacked short dashes) are used as characters with semantics deviating from the "ordinary" hyphen, e.g. to represent a definite unit in transliteration.

For such a special application, in this case for transliteration of Coptic, U+2E17 DOUBLE OBLIQUE HYPHEN was encoded ([1], example on p. 9).

However, there are other applications where the double hyphen is usually not oblique. For such applications, here a "DOUBLE HYPHEN" is proposed, which consists of two stacked short dashes which usually are horizontal.

In most cases described below, the "double hyphen" is essentially used like a hyphen. Thus, it inherits its properties (including its line breaking behavior) from U+2010 HYPHEN.

However, there are cases where the "double hyphen" is used to indicate a specific pronunciation (to denote the single pronunciation of consonants which usually constitute a digraph).

In such use, as well as in cases where an ordinary hyphen also were non-breaking, the double hyphen is non-breaking.

Instead of proposing a different character NON-BREAKING DOUBLE HYPHEN here, it is referred to the possibility to use an U+2060 WORD JOINER after the DOUBLE HYPHEN to achieve the same effect (for examples, see fig. 4).

It shall be noted that the superficially similar true letter U+ A78A MODIFIER LETTER SHORT EQUALS SIGN is not suitable for use as a "double hyphen" punctuation mark, as (besides the fact that the glyph is not necessarily the same in well-designed fonts) a letter (with its line breaking behavior and its suitability for searching) is not appropriate here.

1. Use in Transcriptions

The double hyphen is commonly used in transcriptions of German handwritten texts (fig. 1) and texts printed in Fraktur (Blackletter) into text of modern typographic appearance (fig. 2, 3), to represent any double hyphen found there, independent of the meaning of that punctuation mark in the original text (where its function could be equal to the ordinary hyphen).

This may be due to the fact that the hyphenation rules valid during the creation of the original text are different from the modern German hyphenation rules, and the use of modern (i.e. single) hyphens may be regarded as a distortion of the original text.

A more important reason is the desire to preserve the original hyphenation by denoting it by the double hyphen, while allowing the transcribed text to be hyphenated according to modern rules when presented as continuous text by modern (single) hyphens, without these hyphens to be misrepresented as part of the original text (fig. 3).

As the examples show, in absence of an encoded double hyphen, equals signs are used in spite of the fact that their typographic appearance is inferior in this use, as they are usually too long and thin to represent hyphens.

However, they are preferred over the the encoded double oblique hyphen, as this is considered not matching the modern font (especially when the original double hyphens are not explicitly oblique, as it is common for handwritten originals). In fact, oblique hyphens (single or double) look strange or peculiar to a German reader when found in German text set in Roman type.

2. Use as Phonetic Indicator

Fig. 4 shows an example where the DOUBLE HYPHEN is used to emphasize the dialectal pronunciation [st] for the German digraph "st", which is pronounced [ʃt] in standard German, by writing "s=t". (Emphasizing this pronunciation in literature for a common audience also means emphasizing someone's origin from the city of Hamburg or its rural surroundings.)

3. Use in Non-Standard German Punctuation

In modern German literature, the double hyphen is heavily used by the author Arno Schmidt (1914-1979, [2]), who developed his own orthography deviating from the German standard one. He uses the double hyphen for some kinds of compound words, contrasting to other uses of the hyphen where he uses the standard (single) hyphen (fig. 4, 5). Also, he uses this to let words look like a compound word which in fact are none, thus emphasizing special connotations, or to mark special pronunciation or word emphasis (fig. 4).

(In modern standard German orthography, compound words usually are written by concatenating the constituents without gaps, spaces, or hyphens).

The author, due to his importance for German post-World-War-II literature and his challenging dealing with the German language, is often discussed and cited in germanistic, linguistic, and

literary critical works. In such texts, he usually is cited exactly, i.e. using his special orthography including the double hyphen (fig. 6).

While the use of Arno Schmidt's deviating orthography could be considered idiosyncratic as long as the author only had used it, (*the rest of this paragraph is cited from [11]*) but the use of it in studies of his work is not. All notation is originally idiosyncratic. Nobody argues that this is a case of a mainstream character, it's more in line with the symbols and letters we've added for specialized notations in mathematics and linguistics, or, more recently, the editing of archaic texts.

Moreover, the double hyphen, invented as it is for non-standard orthography, is now used elsewhere in the "Arno Schmidt style", especially to emphasize the components of German compound words (fig. 7). Thus, even when only its use as non-Standard German punctuation mark is taken into account, it has left the realm of idiosyncrasy anyway.

Acknowledgements

Thanks to Asmus Freytag for his insightful comments when discussing the double hyphen on the Unicode mailing list during January 2007.

References

- [1] EVERSON, Michael: Revised proposal to add the Coptic alphabet to the BMP of the UCS. ISO/IEC SC2/WG2 document N3626, 2003-10-01.
- [2] http://de.wikipedia.org/wiki/Arno_Schmidt (German; as of 2010-05-03)
- [3] Transcriptions from the Brandscheid Kirchenbuch Z2 Gestorbenen 1835-1873 (i.e. transcriptions from the parish register of Brandscheid: Z2, persons died 1835-1873) <http://home.hccnet.nl/j.hastrich/bd-z2-toten-1835-1873.htm> (as of 2010-05-03)
- [4] WISCHMANN, Johann: Der Unteutsche Opitz. Oder Kurtze Anleitung zur Lettischen Dicht-Kunst. (The Not-German [i.e. Latvian] Opitz. Or, Short Guidance to the Latvian Art of Poetry.) Rīga 1697
- [5] WISCHMANN, Johann: Der Unteutsche Opitz. Ediert und kommentiert von Stephan Kessler (edited and commented by KESSLER, Stephan). Wiesbaden 2008. ISBN 978-3-447-05788-2 (German).
- [6] SCHMIDT, Arno: Erzählungen (Stories). Frankfurt am Main 1994. ISBN 3-10-373505-7 (German).
- [7] SCHMIDT, Arno: Abend mit Goldrand. Frankfurt am Main 1993. ISBN 978-3100706058 (German).
- [8] ZIMMER, Dieter E: Sprache in Zeiten ihrer Unverbesserlichkeit (Languages in the times of their incorrigibility). Hamburg 2005. ISBN 978-3455094954 (German).
- [9] <http://faz-community.faz.net/blogs/stuetzen/archive/2010/04/06/totentanz-und-technikfeindlichkeit.aspx> (German; as of 2010-04-07; see also legend of fig. 7).
- [10] <http://faz-community.faz.net/blogs/deus/default.aspx?PageIndex=2> (German; as of 2010-09-26).
- [11] FREYTAG, Asmus: Re: Proposing a DOUBLE HYPHEN punctuation mark. <http://www.unicode.org/mail-arch/unicode-ml/y2007-m01/0113.html> (Unicode Mailing List archive, 2007-01-22).

Figures

Verze der Gestorbenen in dem Kirchspiel							i ch n i ß	
N u m m e r .	Zeit des Sterbens im Jahr 18			Zeit des Begräb= nisses.		Des Gestor		benen
	Monat.	Tag.	Stunde.	Monat.	Tag.	Familien=Name.	Tauf=Name.	Ort, wo er gest
#	date				name		place	

Fig. 1: A screenshot from [3], representing double hyphens found in the original manuscript transcribed here. The headline translates "Register of the persons died in the parish ...". The modern orthography for "Familien=Name" and "Tauf=Name" ("family name" = surname and "baptizing name" = Christian name) is "Familiennamen" and "Taufnamen".

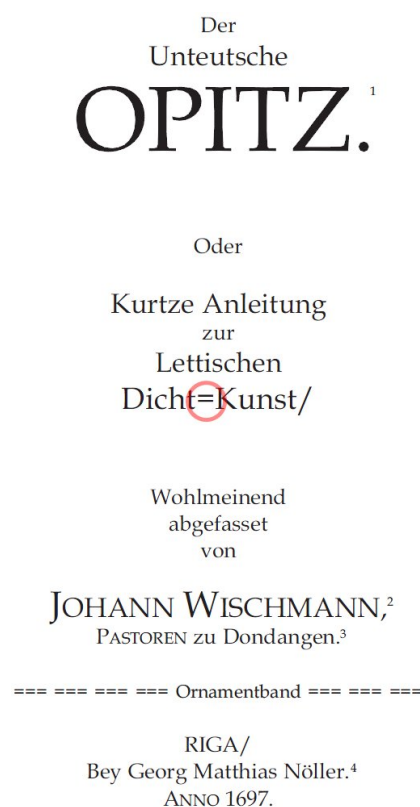
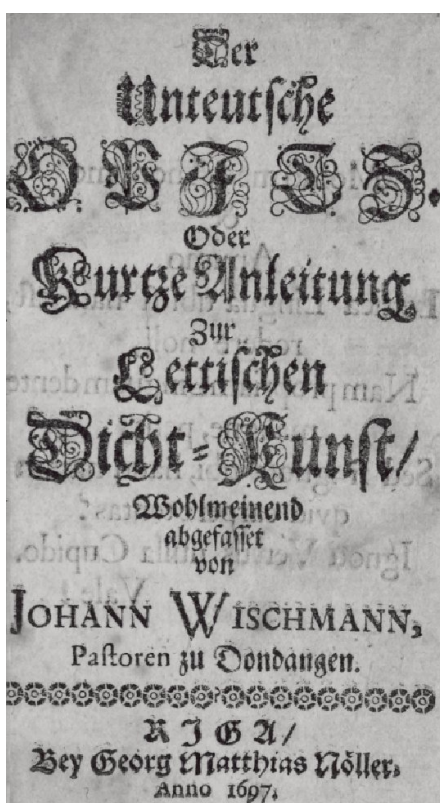
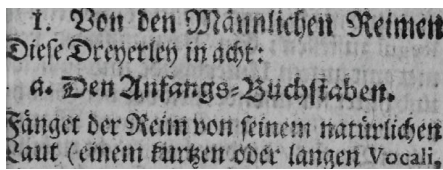


Fig. 2: Title page of [4] (a Fraktur print of 1697) and [5] (its modern transcription of 2008), showing the transcription of the historic "ꝛ" by a (non-oblique) double hyphen when using a Roman font in the modern transcription. As a true double hyphen is not available, it is represented here by an equals sign in spite of its typographical shortcomings, rather than by an oblique double hyphen which looks even more strange with a Roman font.



I. Von den Männlichen Reimen

Diese Dreyerley in acht:

«a. Den Anfangs=Buchstaben. Fängt der Reim von seinem natürlichen Laut (einem kurtzen oder langen VOCALI, oder von einem DIPHTHONGO) an/ so müssen ein oder

Fig. 3: Another example from [4] (p. 19) and its modern transcription [5] (p. 31), showing a double hyphen for representing a hyphen from the original, besides an ordinary (single) hyphen for a word division in the transcription text..

: »Die ham rausgekrickt – durch irgndwe'che Magnets=traaln –« (cirkelzuendige: GOUE!) – »daß da, wo die Zoon-grenze über'e Elbe geht, im Fluß=Bett n gans lebhafter Verkehr s=tadtfindet: Ah=Gentn!«. (Entn; in Taucher=Überzügen.) »Und zwa nich nur Vondrüb'm Nachhierher. Sonndern Pann=Koukn hat sofort auch'n Weis=Tum veröf-fentlicht: daß die Bonna Ehe=Maille Eß=Eß=Leudte ebn-falls durch schläußtn. In umgekehrter Richtunk.« Und sah mich erwartungs=folll an:?. Ich nikköpfte stumm. / (Dolle

Fig. 4: Scan from [6] (an 1994 edition of earlier stories of Arno Schmidt), p. 522 (an excerpt from the story "Caliban über Setebox" from 1958).

Red circles denote the use of the double hyphen to emphasize composite words.

Green circles denote the use to emphasize the pronunciation, e.g. s-t to denote the Northern German pronunciation of the "st" composite sound. In this case, obviously the use of the non-breaking variant of the double hyphen is appropriate (as illustrated by the second example "s=tadtfindet").

Note that the first example (which is "Magnetstrahlen" in standard orthography, compound of "Magnet" + "Strahlen") had been written "Magnet-Straaln" with capital S if the author had intended to emphasize the composite word here instead. It is presumed that the line break is wrong here. Correctly, the double hyphen is non-breaking (as there is no syllable break within the digraph "st"), and the line break has to be:

"Magnet-s=traaln".

Using Unicode, this can be achieved by inserting a "soft hyphen" after "Magnet" (unless you had an automatic hyphenation system applicable to dialectal orthography), and a "word joiner" after "s + double hyphen".

Cyan squares denote ordinary hyphenation (note in the first example the composite word "Zoongrenze" is not emphasized).

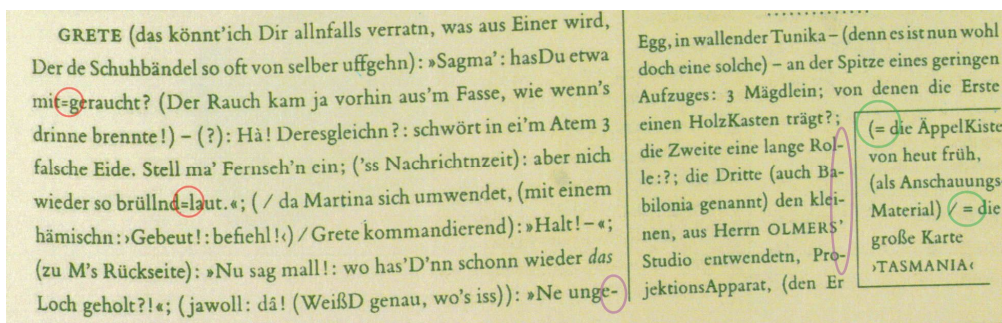


Fig. 5: A scan from [7] (a fine edition of a late work of Arno Schmidt), p.143. Double hyphens are shown within red circles, together with some single hyphens (encircled purple) and (differently looking) equals signs (encircled green) on the same page, showing that the double is not a font variant of the latter two.

graphie. Der Schriftsteller Arno Schmidt hatte es: »Achdu wie man de Milch=Schtraße sieht! → (Die tüüpische Groß-Schtädterinn. Aber der Himmel zitterte & schütterte tatsächlich vor Geschirtheit.) / Wie Boogn-Lampn drangeschellt: linx Prokyon; rechts M'sjö Sie-Riuß.« Aber so etwas ist an-

Fig. 6: A scan from [8], p. 169. This book dealing with German language uses the double hyphen (red circle) as well as the ordinary hyphens when citing Arno Schmidt, preserving the use of the original. Besides this, the ordinary hyphen is used for word division within the commenting text, as expected.

Kalchas

07. April 2010, 10:22

http://www.sattel.info/artikel_lesen--rennsattel--10000015.html

Diesen Text, lieber Don, habe ich denn mal kurz zum Anlaß genommen, bei Google zu recherchieren, und was muß ich lesen, Rennsättel, also richtige Rennsättel, werden heute auch aus Kunst=Stoff gemacht, d. h. aus Plastik! Sind die zugehörigen Pferde auch schon gen-manipuliert? Ich warte mal ab, was MvB-B dazu sagt, beim Polo kann ich mir solche

Fig. 7: Screenshot from a German blog [9], hosted by a leading German newspaper, the FAZ (Frankfurter Allgemeine Zeitung). In this blog, a somewhat sophisticated attitude prevails. The author of the blog entry writes the German word "Kunststoff" which means plastic, using the double hyphen to emphasize the meaning of the constituent parts of the compound word ("Kunst=Stoff" = *artificial* material), using the style which is known to well-read Germans in the subsequence of Arno Schmidt. As no real double hyphen is available, he uses the equals sign as substitute. For other text, he uses the ordinary (single) hyphen according to the standard German orthography rules.

Alle Blogbeiträge

1 | 2 | 3 | 4

Deus ex Machina

28. August 2010, 16:03

Ansichten eines Codierknechts

61 Kommentar(e)

worin der hochverehrte Gastautor dieses Monats allhier von den Widrigkeiten und Beschwerissen berichtet die im Betrieb wundersamer Computer=Apparati ihm beim Systemadministrantentum wiederfahren sind. >

von Der hochverehrte Gast

welches ein Blog vorstellt, worin getreulich und mit Plaisir berichtet wird von den göttlichen Internetmaschinen neuester Art, ihren Berechnungen unberechenbarer Weis- und Torheiten, und ihrem Wirken zum Gedeih und Verderb der menschlichen Art hienieden auf Erden, als da ist in Philosophieerei, Wirtschaft,



Fig. 8: Screenshot from another German blog [10], also hosted by the FAZ (see legend of fig. 7). Here, the German orthography of the 18th century is mimicked to yield an ironic or sophisticated appearance. Also here, the double hyphen is used in a place where modern orthography would require a composite noun without an interruption.

**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://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title: Propose to encode a punctuation mark "Double Hyphen"

2. Requester's name: Karl Pentzlin

3. Requester type (Member body/Liaison/Individual contribution): Individual contribution

4. Submission date: 2010-05-04, revised 2010-09-28

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: Supplemental Punctuation

2. Number of characters in proposal: 1

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary	B.1-Specialized (small collection)	<input checked="" type="checkbox"/>	B.2-Specialized (large collection)
C-Major extinct	D-Attested extinct	<input type="checkbox"/>	E-Minor extinct
F-Archaic Hieroglyphic or Ideographic	G-Obscure or questionable usage symbols		

4. Is a repertoire including character names provided? Yes

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? Karl Pentzlin

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): <http://pentzlin.com/DoubleHyphenProForma.ttf>; the font is released into the Public Domain

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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N3702-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)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain		No
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.)? If YES, with whom? If YES, available relevant documents:	<i>The character is ubiquitous, thus the author is a member of the user community himself</i>	N/A
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>see text</i>	<i>see text</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>see text</i>	<i>see text</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>see text</i>	Yes
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>contemporary use; keeping in line with similar encoded characters</i>	Yes Yes
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? 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? If YES, is a rationale for its inclusion provided? If YES, reference:		no
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>see text</i>	<i>see text</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:		no
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)		no
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:		no