

Proposal to encode three combining diacritical marks for Low German dialect writing

2008-01-18 ===DRAFT===

Karl Pentzlin, karl-pentzlin@europatastatur.de

The following three diacritical marks are proposed:



U+1Dxx COMBINING STRAIGHT RIGHT-POINTING HOOK BELOW

→ 0238 combining ogonek

- the stem gets slanted when used within an italic font



U+1Dyy COMBINING LONG VERTICAL LINE BELOW

→ 0329 combining vertical line below

- touches the base character, unlike U+0329
- gets slanted when used within an italic font



U+1Dzz COMBINING TRIPLE BREVE BELOW

→ 035C combining double breve below

- to be applied to the first letter of any three-letter combination
- used with "sch", "Sch", and "SCH" in Low German dialect writing

The **straight right-pointing hook below** is used in several writing systems developed for Low German dialects throughout the 20th century. It differs from the ogonek

- in shape (see fig. 1): while the vertical part of the ogonek is usually bound or inclined to the right and touches the base letter at the right bottom, the straight hook proposed here shows a straight vertical part (which inclines accordingly when used within an italic font) and touches the base character at the middle bottom.
- in function: while the ogonek usually indicates nasalization, the straight hook indicates "openness" of a vowel (pronunciation of ϵ when applied to e; ɔ when applied to o; œ when applied to ö).

Thus, the straight right-pointing hook below is a character different from the ogonek and deserves its own encoding.

The two other characters, together with the first one, are used in the "Rheinische Dokumenta" [2], an orthography promoted by the "Landschaftsverband Rheinland" (a cultural authority of the southern part of German federal state Nordrhein-Westfalen) during the 1980s for the local Low German dialects.

There, the **combining long vertical line below** is used to mark an indistinct pronunciation of a vowel, especially in combination with e to mark a schwa. Unlike U+0329, it is attached to the base letter (or at least touches it almost) and is distinctly long (see fig. 2). It is productive (fig. 8 shows it attached to an ö).

The **combining triple breve below** is to mark the trigraph "sch" (which is used to denote the voiceless ʃ sound, like in standard High German orthography) denoting the voiced ʒ sound. Thus, it may be applied to "sch" in lower case, "Sch" in title case, and "SCH" in upper case.

Obviously, these three marks are selected for the "Rheinische Dokumenta" as they are easily applicable by pen in texts written by typewriter (the usual tool for scientific and private writing in the 1980s).

The proposed properties are:

1Dxx:COMBINING STRAIGHT RIGHT-POINTING HOOK BELOW;Mn;220;NSM;;;;;N;;;;;

1Dyy:COMBINING LONG VERTICAL LINE BELOW;Mn;220;NSM;;;;;N;;;;;

1Dzz;COMBINING TRIPLE BREVE BELOW;Mn;233;NSM;;;;;N;;;;;

The triple breve below is to be input after the first character of the combination it applies to, analogous to the double diacritics like U+035C COMBINING DOUBLE BREVE BELOW. Thus, the correct way to enter a "sch" trigraph with triple breve below is: U+0073 U+1Dzz U+0063 U+0068.

Bibliography:

- [1] CORNELISSEN, Georg, HONNEN, Peter , LANGENSIEPEN, Fritz (ed.):
Das rheinische Platt · Eine Bestandsaufnahme
(Handbuch der rheinischen Mundarten, Teil 1: Texte)
Köln 1989, ISBN 3-7927-0689-X
 - [2] HONNEN, Peter:
Rheinische Dokumenta. Lautschrift für rheinische Mundarten
Pulheim 1987, ISBN 3-7927-0947-3
 - [3] LINDLOW, Wolfgang, et al.:
Niederdeutsche Grammatik
Leer 1998, ISBN 3-7963-0332-3
 - [4] http://de.wikipedia.org/wiki/Rheinische_Dokumenta as of 2008-01-18 (*German*)
(that page displays the e with long vertical line below incorrectly (like an e with U+0329), i.e. not in a way consistent with all found printed examples).
-

ǣǣǣ ǣǣǣ ǣǣǣ ǣǣǣ
ǣ ǣ ǣ ǣ

Fig. 1: Comparison of ogonek (first line, fonts Arial as found on Microsoft Windows XP, and Doulos SIL as found on www.sil.org, each straight and italic) with samples of the straight right-pointing hook below (second line, as found in the figures below).

ǣǣǣ ǣǣǣ ǣǣǣ ǣǣǣ
ǣ ǣ ǣ ǣ

Fig. 2: Comparison of U+0329 COMBINING VERTICAL LINE BELOW (first line, fonts Arial as found on Microsoft Windows XP, and Doulos SIL as found on www.sil.org, each straight and italic) with samples of the long vertical line below (second line, as found in the figures below).

Als Weiterführung der LÜBECKER RICHTLINIEN, an deren Erarbeitung er selbst maßgeblich beteiligt gewesen war, begreift BORCHLING seine »Recht-schreibungslehre« aus dem Jahre 1935. Das vorstehend skizzierte Prinzip der Längenbezeichnung durch Verdopplung wird jetzt auch auf die Umlaute ausgedehnt (*Dääk* ›Decke‹, *Bööm* ›Bäume‹). Analog zu ǣ empfiehlt er das Zeichen ö für tonlanges ö; da diese Laute stets lang sind, unterbleibt die Quantitätsandeutung (*eten* ›essen‹, *över* ›über‹). Auf das Zeichen ǣ meint

Fig. 3: Sample from [3], p. 32, mentioning letters with the straight right-pointing hook below in a Low German orthography of 1935.

¹³ Aß Jesus dat höört har, sail hee mit en Boot fun dōor wēch noo 'n afgelee-
gen Sfteet hen; hee wul för sik aleen ween. Dē Lüüt kreegen dat awerß liikerß
to weetē un keemen em tofōotß noo uut dē Dörper. ¹⁴ Aß hee an Lant ßteech
un dōor aal dē Minschen seech, dōo dēen see em banich leet; un hee mook eer
Krangken gesunt. ¹⁵ Aß dē Dach to En güng, keemen siin Jüngerß to em un
sēen: »Dat is hiir en eensoome Geegent, un dat is nuu al wat loot worn. Loot
dē Lüüt leewer goon un schik eer wēch in dē Dörper; dōor köönt sē sik noch
wat to eetē kööpen.« ¹⁶ Jesus awerß sē: »See bruukt doch goor nich wēch-

Fig. 4: Another sample from [3], p. 35, showing a biblical text (Mt. 14, 13 ff.) in a Low German dialect from the German federal state Schleswig-Holstein, recorded 1984, using the "Bremer Schreibung" orthography devised in [3], also containing several examples of the straight right-pointing hook below.

o, ö stehen für offenes o, ö: *nōor* (nach), *ööver* (über), *Kōp* (Kopf), *kōchē* (kochen), *Schlachōora* (Schlagader).

ē steht für den sogenannten Murmelvokal (Schwa): *gefūrt* (gefeiert), *loogen* (lagen), *dē* (die), *erop* (herauf), *Mōorjē* (Morgen), *sē* (sie).

Fig. 5: Sample from [2], p. 15, introducing the straight right-pointing hook below and the long vertical line below.

Mundart: Mutterschied (Hunsrück)

*Im Schpäätjoa, wän dē Äärē in waa un di Sou fāt genuuch schiin, is bai dē Doorfmätska en Tärmiin foor dē Schlachtdaach ousgemach woa. Am Qōment foorhea is di Muul med-em Schupkarē gehuult woa, dē Wäschkäşel met Waşa gefilt un Haitsmadriaal baraat geläät, das am friiē Mōorjē angehaitst weerē kont. Wän dē Mätska kom is, hot mişē dēs Waşa kōchē. Wän-s dan so weit waa, is di Muda oore dē Fada in dē Schdal gang un hot de Sou en Schtrik aan en Hinabään gebun un sē foa di Deer gefoort, wo sē dan koats an-ē Hooge oore Ring draan gebun is woa. Intwischē hot şisch dē Mätska met dē Aks bewavent un hot dōomet di Sou uf dē Schdeere geschlaan. Un wän sē nit gleisch umgefäl is, hōr-a noch-emōo noogeschlaan. Dan hōr-a das groos Mäşa genom un hot di Sou in dē Hals eningeschdoch un di Schlaachōora doorischgeschniit.**

Fig. 6: Sample from [2], p. 14, showing all three proposed diacritical marks.

495 Mandel (Vbg. Rüdesheim)

S Enck, Helmut, Jg. 1928, geb. in Mandel, lebt auch dort, Winzer.

A 1982, 3.54 min.

T Weihnachten, Mandel um 1933, spontan gesprochen.

Doo waa dē alt Qōba noch, dē °Änggē, also maim Fader şai Fader, deer hat doo ove şai Schdup gehat, un dē hot doo gewōont und also das Şäkrädärschē schdään [. . .], un isch waa doo şälischēmōō¹ finēf, şäks Joer alt. Ich kond-ē bisjē friier lääşē wii dii Nōrmaale, also ich kont mit finf Joer lääşē ooder, dāsch-ē²: un lääşē, das waa drum, isch hōn halt gelääşē, şoo mach ich-s hōit noch, also isch lääşē tsäntnerwais Bischer

Fig. 7: Sample from [1], p. 495, also showing all three proposed diacritical marks.

Öt ös nou öwër tweevoondert Jöör aalt. Wat nei draan ös, dat şint dii körtę Wolängskęs an dę Gadiineschtangęn. Fan Tiit tu Tiit wört öt Bet ouk blaangk gęschtreekęn; meistęns för Pöschęn af Peengstęn, Grosfaadęrsch Gębortsdaach af Keerschmęs. Dat Osęnbluut ös met aal dii Jööręn ömęr doongklęr bruun in dę

Fig. 8: Sample taken from [2], p. 34, showing a sample from:

Katharine Specht: Os Hemelbet . . . döş sçteit öt aafgęmpölt, in: Heimatkalender des Kreises Wesel 1986 (resident's calendar for the Wesel county) p. 86, showing the long vertical line below applied to an ö, and showing a comma-like font variant of the straight right-pointing hook below (presumably due to limitations of the typesetting system used).

- Die meisten Mundartarbeitskreise und die Sprachabteilung des Amtes arbeiten mit der Zusatzzeichenversion. Sie kann problemlos mit jeder herkömmlichen Schreibmaschine geschrieben werden. Die Zusatzzeichen lassen sich mit der Hand in einen bereits getippten Text nachtragen:

Im April nõinşęhunętelęf koom isch eę dę Schuol. Hou mein isch nõch, ęt wäär geęstęrn erscht gęwääst. Isch durfdę tsęgaa de şoojęssçhoo aadoo, un dat wol wat haişę. Duteę bet alęrlei Schnuptsich wi hou, dat goov-ęt duu nõch net.*

Fig. 9: Sample from [2], p. 25/26, showing a typewritten text with the diacritical marks added by hand.

The explanatory text translates:

Most dialect working groups and the language department of the authority work with the version [of the Rheinische Dokumenta orthography] using the additional marks. It can be written with any usual typewriter without problems. The additional marks may be added by hand into an already typewritten text.

**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:	<i>Proposal to encode three combining diacritical marks for German dialect writing</i>
2. Requester's name:	<i>Karl Pentzlin</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual Contribution</i>
4. Submission date:	<i>2008-xx-xx</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:	<i>no</i>

B. Technical – General

1. Choose one of the following:					
a. This proposal is for a new script (set of characters):	<i>no</i>				
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>1DC0-1DFF Combining Diacritical Marks Supplement</i>				
2. Number of characters in proposal:	<i>3</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"/>	B.2-Specialized (large collection)	<input type="checkbox"/>
C-Major extinct	<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?	<i>yes</i>				
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	<i>- tbd -</i>				
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	<i>- tbd -</i>				
6. References:					
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?					
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 <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: N3102-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)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>no</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.)? 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? Reference:	
4. The context of use for the proposed characters (type of use; common or rare) Reference:	
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>yes</i> <i>see proposal text</i>
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>yes</i> <i>keeping them together with all other diacritics for the Latin script</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>no</i>
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:	<i>no</i>
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:	<i>no</i>
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>yes</i> <i>yes</i> <i>see proposal 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:	<i>yes</i> <i>yes</i> <i>see proposal text</i> <i>yes</i> <i>see proposal text</i>
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)	<i>no</i>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	<i>no</i>