

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

Doc Type: Working Group Document

Title: Proposal for extending the annotation of the character “Y WITH DIAERESIS” in the BMP of the UCS

Source: Dr. Gábor Hosszú and Péter Krauth

Status: Individual Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2009-08-12

This proposal contains the proposal summary form.

Please send any responses to this proposal to Dr. Gábor Hosszú (email: hosszu@eet.bme.hu).

Contents

1. Introduction.....	1
2. Detailed description of the character annotation of this proposal.....	1
3. Unicode Character Properties.....	2
4. Acknowledgement.....	2
5. Bibliography.....	2
6. Proposal Summary Form.....	3
7. Appendix – Example of the use of character “ÿ” in medieval Hungarian texts.....	5

1. Introduction

This document proposes to extend the annotation of the 00FF ÿ character (Y WITH DIAERESIS) in the ISO/IEC 10646 standard with the term “*Old Hungarian orthography*”, since this character was used extensively in the ancient Hungarian Latin script. In the current Hungarian Latin orthography it is not used.

Many examples of the use of the character ÿ in ancient Hungarian scripts are presented in the *Appendix*.

2. Detailed description of the character annotation of this proposal

The annotations in the ISO/IEC 10646 standard describe that specific characters are used in certain languages. In case of the 00FF ÿ character from the C1 Controls and Latin-1 Supplement block, the “French” language is denoted only. However, as the examples demonstrate the Hungarian orthography from the Middle Ages extensively used the letter ÿ. In order to write the large amount of Hungarian medieval texts in current documents the character 00FF ÿ is necessary.

In order to improve the accuracy of the annotations in the ISO/IEC 10646 standard, the modification of the annotation of the character 00FF ÿ is proposed.

The terminology “*Old Hungarian*” is widely used in Hungarian linguistics referring to medieval Hungarian orthography. Fig. 2-1 demonstrate this fact.¹ The term “*Old Hungarian orthography*” is also known.

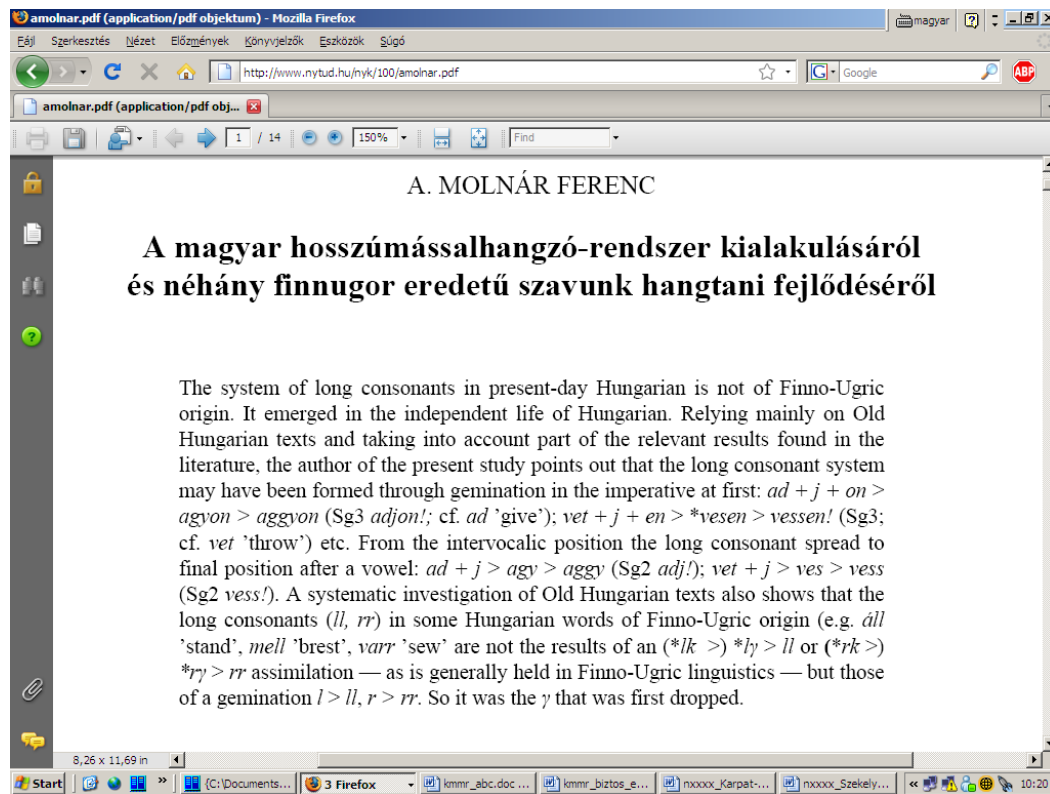


Figure 2-1: The use of the term “*Old Hungarian*” for medieval Hungarian orthography²

3. Unicode Character Properties

In the followings the proposed annotation of the character *ÿ* is presented.

00FF *ÿ* LATIN SMALL LETTER Y WITH DIAERESIS
 French, Old Hungarian orthography
 → 0178 *Ÿ* latin capital letter *y* with diaeresis
 ≡ 0079 *y* 0308 *̈*

4. Acknowledgement

The authors express their gratitude to Prof. Dr. Ferenc Kovács for his useful advice and Mr. Árpád Zubrits for proof-reading the document.

5. Bibliography

Dékány, Csaba: A Dékány család története (*History of the Dékány Family*) The home page of the Dékány Family, Retrieved in 2009 from <http://www.dekanycsalad.extra.hu/01.D.doc>

¹ Molnár, A. Ferenc, 2003.

² Molnár, A. Ferenc, 2003.

Lázs, Sándor: A megkerült Pozsonyi Kódex és Szegedi Antifonale, valamint két elkallódott kézirat. (“*The re-discovered Pozsony Codex and Szeged Antifonale, furthermore two lost manuscripts*”) In Magyar Könyvszemle, Vol. 122, No. 2., 2006, pp. 146-161. Retrieved in 2009 from <http://epa.oszk.hu/00000/00021/00049/Ksz2006-2-02.htm>, from the Hungarian Electronic Library.

Molnár, A. Ferenc: A magyar hosszúmássalhangzó-rendszer kialakulásáról és néhány fginnugor eredetű szavunk hangtani fejlődéséről (“*About the development of the system of Hungarian long consonants and the phonetic development of some of our words of Finno-Ugric origin*”). In: Nyelvtudományi Közlemények (*Linguistics Publications*), 2003, Vol. 100, pp. 198-211. Retrieved in 2009 from <http://www.nytud.hu/nyk/100/amolnar.pdf>

Szepesi Csombor, Márton összes művei (*Complete works of Márton Szepesi Csombor*). Published by Neuman Kht., 2002, Budapest. It is retrieved in 2009 from the Hungarian Electronic Library, <http://mek.niif.hu/05700/05731/html/szepesi0202/szepesi0202.html>

6. Proposal Summary Form

**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:	Proposal for extending the annotation of the character “Y WITH DIAERESIS” in the BMP of the UCS
2. Requester's name:	Dr. Gábor Hosszú and Péter Krauth
3. Requester type (Member body/Liaison/Individual contribution):	Individual Contribution
4. Submission date:	August 12 2009
5. Requester's reference (if applicable):	http://nimrud.eet.bme.hu/hosszu
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:	No	
Name of the existing block:	C1 Controls and Latin-1 Supplement	
2. Number of characters in proposal:	1	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input 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?	Yes	
a. If YES, are the names in accordance with the “character naming guidelines”		

¹. Form number: N3152-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)

in Annex L of P&P document? []

b. Are the character shapes attached in a legible form suitable for review? Yes

5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? Dr. Gábor Hosszú

If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: Dr. Gábor Hosszú, hosszu@eet.bme.hu

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)? Yes

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. – See below.

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.)? Yes
If YES, with whom? Hungarian Standards Institution
If YES, available relevant documents: Yes

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Yes
Reference: Scholars and researchers, medieval texts.

4. The context of use for the proposed characters (type of use; common or rare) Rare
Reference: Historical and contemporary use in traditional family names.

5. Are the proposed characters in current use by the user community? Yes
If YES, where? Reference: Various kind of publications.

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? No
If YES, is a rationale provided? []
If YES, reference: []

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 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?	No
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 character(s)?	No
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	

7. Appendix – Example of the use of character “ÿ” in medieval Hungarian texts

The *Fig. 7-1* presents a part of the texts written by Márton Szepsi Csombor between May 1616 and August 1618 into his notebook.¹ In the text the Hungarian words “irégÿm” (in English: “people who are envious of me”) “menÿre” (in English: “how”) uses the ÿ character.

Az irigy és kedvetlen olvasónak.*

Eörülök rayta ha meg hallok hog ez könyveczkiem utan irégÿm tamattanak, mert akkor veszem esze[m]ben hog nem szinten minden haszontul üres lött irásom. Semmi állapotnak alitom az melynek ninczen irégÿe. Lattam még az Britanniai tengerenis hogy á tüzes tolvajok* az terhes haiok mellet inczelkettenek, az üresseket penigh szabadságosan el boczatottak, á mint amaz köz versis tartya:

Cantabit vacuus coram latrone viator.*

Igy irigy ellenem valo irégÿségedet meg gondolhad menÿre félem, hogy békét hat gye[n]gén folyo kis hireczke[m]nek, mert ezzel énnem semmit, magadnak penigh igen nagyot artasz, ha vétkezni akarsz más vétket végy elő, mert az többinek vagyon valami haszna, gyönyörüsége, csak egyedül az irégÿséghet ismertem minden vétkek közül leg haszontalambnak lenni.

Figure 7-1: Text of Márton Szepsi Csombor from 1616-1618

Fig. 7-2 presents a part of the description of the history of the Dékány family.² In 1588 the ÿ was used in the name of Dékány as *Fig. 7-2* demonstrates it.

¹ Szepsi Csombor, Márton, 1616-1618, <http://mek.niif.hu/05700/05731/html/index.htm>

² Dékány, Csaba, <http://www.dekanycsalad.extra.hu/01.D.doc>

A magyar kultúrtörténetben céhmester, határbíró, hegybíró, tizedes mint városrész előljáró, kurátor, gondnok, esperes, egyházi tisztséget jelent. Az 1300-as évektől Dekan, Decan írásmód fordul elő szerte az akkori Magyarországon, majd a 1500-as évek végén megjelenik Erdélyben a Dekanÿ, és Dekany írásmód. Az ékezetek: Dékán, Dékány, Dékany, De-

Figure 7-2: The Hungarian orthography of the family name “Dékány” in the 16th-17th century

Fig. 7-3/a, b and c) present the parts of the journal article written by Sándor Lázcs about the analysis of medieval texts in Hungarian codices.¹

Az említett döntésekben egyetlen szó sem esik könyvekről. A sümegi várból azonban fennmaradt egy 1597-es leltár, amely a vásárhelyi apácák hátrahagyott értékei között több régi miseruha, oltárterítő és három kép mellett összesen 31 könyvet sorol föl.^[34] Legtöbbjük szerkönyv, de van köztük gyerekeknek szóló tankönyv és egy magyar nyelvű kézirat is „de rebus gestis Apostolorum”.^[35] A leírásokból kitűnik, hogy a *Szegedi Antifonale* utasításainak „lampas ees Laudes megh vagyön yrwan nagy kÿmbe” illetőleg „mas kÿnfbbe keres” minden alapjuk megvolt, ugyanis a leltár egyaránt tartalmaz nagy alakú gradualékat, psalteriumokat, breviariumokat és

Figure 7-3/a) Part 1 of the article of Sándor Lázcs with medieval words that use ÿ

Gábriel Asztrik – a számtalan hiba okán – úgy vélte, hogy az oráció az apáca saját fordítása. Vannak is erre utaló jelek a 37r. levélen: kegyetlenségÿmnek sokasagaer de az te nagy ÿrgalmassagod<erth>nak sokasaga<erth> zereneth ÿrgalmaz.^[56] Úgy tűnik, mintha az író a ragok megválasztásával küzdött volna, de lehetséges, hogy csupán a szakaszban lévő egyik korábbi rag (sokasagaer) zavarta meg. Egyéb fordításra jellemző hiba viszont nincs a szövegben; nincsenek megszokott szóválasztási, egyeztetési problémák, törlései sokkal inkább másolási hibák. A ragtévesztés mindenképpen írható a scriptor figyelmetlenségének, netán eltérő nyelv- érzékének számlájára.

Az imában különben minden javítás másolási hibát korrigál. A gar<bo>losagos (PK 34v:7) tévesztése elképzelhetően a mintapéldányból eredeztethető: az túl közel volt a hanyagul megformált o-hoz. Ugyan ilyen elírás a gondola<ta>tom (PK 35r:4), vagy a kÿ<z>hez

Figure 7-3/b) Part 2 of the article of Sándor Lázcs with medieval words that use ÿ

A *Thewrewk-kódex*ből átvett szövegek közül sorrendben első^[64] a „Hamis tanúk elleni ima”.^[65] A másoló eleinte sokszor tévovázik a könyörgés átírásakor, láthatóan nem tud mit kezdeni a mintapéldány számára ismeretlen helyesírásával. Belekeveredik a ’kewnyorwletews’ szóba: ’kewnywrewletes’ – nem tud mit kezdeni az alul farkas ö-betűvel.^[66] A későbbiekben aztán egyre biztosabban értelmezi a szöveget és saját helyesírása szerint kezd másolni: rendszeresen elhagyja a hosszú mássalhangzók, de ugyanígy a hosszú magánhangzók jelölését is, az alul farkas ö-t többnyire ew-ként, illetőleg e-ként írja át.^[67]

A következő imában, az igen szép Mária-oráció (wagon mondasom az te dedes zÿwedhez oh maria)^[68] másolatában ugyanazokat a helyesírási változtatásokat találjuk, mint az előzőben, de a scriptor már sokkal határozottabban bánik a szöveggel, az o-t többnyire ew-ként, ritkábban e-nek írja át.

Egy másik Mária-ima^[69] esetében a korábbiakhoz hasonlóan változtatja sajátjára mintapéldányának a helyesírását, mint az előző szövegnél, másolatában továbbra is rendszeresen elhagyja a hosszú magánhangzó jelölést. A számára idegen o-ból rendszeresen ew lesz.^[70] Másol – egyre következetesebben – saját nyelvjárása szerint alakítja a szavakat, például mindjárt az ima első sorában; *Thewrewk-kódex*: „Idwezlegÿ te oh zeep ttwyknelkyl walo rosa” megfelelője a *Pozsonyi Kódex*ben: „Idwezlegÿ te oh zep tewÿsknekewl walo rosa”.^[71] Ahol értelemzavaró hibát talál az előtte lévő szövegben, azt javítja: „elewnk my ez zamkywvetesnek wtlgheben lezewnk” (ThK 45v:10–11) a *Pozsonyi Kódex* megfelelő helyén már az új variáns áll: ’mÿgez’ (33r:7).

Figure 7-3/c) Part 3 of the article of Sándor Lázcs with medieval words that use ÿ

Fig. 7-4 presents a part of the scientific article of Ferenc A. Molnár where he presents many Hungarian words with medieval orthography with the character ÿ.²

¹ Lázcs, Sándor, 2006.

² Molnár, A. Ferenc, 2003.

intervokális helyzetben is rövid gy-t jelöl.) Például: PeerK 359: „Maý Eremnek oka : ýmag, | lelký Eden : ýmagh”; CzechK 131: „Zent Istwan kýral. *Imagy*. | Zent Lazlo kýral. *Imagy*”; 129: „Istennek mýnd zent thanetwaaný. *Imagyatok erottem*”; 78: *agyad*; ThewrK: „Kegyelmes zyz maria: *ymagy* | Zepsegews zyz maria: *ymagy*”; 6, 17: *agyad*, 56: *agyad*; stb. A PeerK-ben a kép intervokális helyzetben változatosabb, 21: *agiiadh*, 195: *aggyad*, 279: *aggjad*, 271: *agyon*, 55: *twgya*, 60: *thwgya*, 198: *tuggia* stb. (Az etimológikus elvű helyesírás – pl. 241: *fogadyad* – nem teszi lehetővé a mássalhangzó időtartamának megállapítását.)

Figure 7-4: A part of the article of Ferenc A. Molnár with many examples of the use of the ý