Proposal to encode Bosnian Arabic characters

Denis Moyogo Jacquerye <moyogo@gmail.com>
2019-10-03

1. Introduction

The Arabic alphabet has been used to write Bosnian from the 15th to the mid 20th century. In the early 20th century Mehmed Džemaludin Čaušević extended the alphabet, adding several letters including two letters that are proposed for encoding:

- ARABIC LETTER NOON WITH INVERTED SMALL V for nj [ɲ],
- ARABIC LETTER HAH WITH INVERTED SMALL V BELOW for ć [tɕ].

Other additional letters, which are already encoded, were used in this orthography:

- 0684 ARABIC LETTER DYEH ڄ for c [ts],
- 0686 ARABIC LETTER TCHEH چ for č [tʃ],
- 062C ARABIC LETTER JEEM ج was used for both dž [dʒ] and đ [dʑ],
- 06B5 ARABIC LETTER LAM WITH SMALL V ڵ for lj [ʎ],
- 06C6 ARABIC LETTER OE ۆ for [u],
- 06C9 ARABIC LETTER KIRGHIZ YU ۉ for [o].

Čaušević's alphabet was used in publications from 1908 until the 1940s. That alphabet is referred to by several names: *arebica*, *harfovica*, *matufovica*, *matufovača*, *mektebica*.

According to Pjanić (2009), several periodicals used arebica: *Tarik* (1908-1911), *Muallim* (1910-1913), *Misbah* (1912-1913) and *Jeni Sabah* (1914); and also the last publication using Arebica was Seid Serdarević *Fikh ʾulʿibādāt* (1941).

Those two letters are mentioned in descriptions of Bosnian Arebica such as Bourgeois (1913), Lehfeldt (2001), Gažáková (2014).

In recent years, there has been interest in Čaušević's orthography. Notably in works by Al-Zubi and Cicak-Al-Zubi, in the *Šegrt Suljica* comics published in the Bosnian children magazine *Elif*. The 062C ARABIC LETTER JEEM ج is used for dž [dʒ] and another letter represents đ [dʑ], for example Al-Zubi and Cicak-Al-Zubi use 0620 ARABIC LETTER AIN WITH THREE DOTS ABOVE غ and a letter ARABIC LETTER HAH WITH SMALL V

BELOW, which could potentially be encoded if in use, can be found in some character charts.

2. Proposal

Encode the following characters:

codepoint	isolated	final	medial	initial	name
U+0889	ڹ	ـڹ	ـڹـ	ڹـ	ARABIC LETTER NOON WITH INVERTED SMALL V
U+088A	ڊ	ـڊ	ـڊـ	ڊـ	ARABIC LETTER HAH WITH INVERTED SMALL V BELOW

3. Unicode character properties

```
0889;ARABIC LETTER NOON WITH INVERTED SMALL V;Lo;0;AL;;;;;N;;;;;
088A;ARABIC LETTER HAH WITH INVERTED SMALL V BELOW;Lo;0;AL;;;;;N;;;;;
```

4.Joining type and group for ArabicShaping.txt

```
0889; DOTLESS NOON WITH INVERTED V ABOVE; D; NOON
088A; HAH WITH INVERTED V BELOW; D; HAH
```

5. Annotations

The following annotations could be updated in NameList.txt.

```
0684 ARABIC LETTER DYEH
    * Sindhi, Bosnian
06B5 ARABIC LETTER LAM WITH SMALL V
    * Kurdish, Bosnian
06C6 ARABIC LETTER OE
    * Uighur, Kurdish, Kazakh, Azerbaijani, Bosnian
06C9 ARABIC LETTER KIRGHIZ YU
    * Kazakh, Kirghiz, Bosnian
```

The following annotations are recommended for NameList.txt.

```
@ Additions for Bosnian:
0889 ARABIC LETTER NOON WITH INVERTED SMALL V
088A ARABIC LETTER HAH WITH INVERTED SMALL V BELOW
```

6. Normalization and confusability issues

ARABIC LETTER NOON WITH INVERTED SMALL V is potentially confusable with 06BA ARABIC LETTER NOON GHUNNA + 065B ARABIC VOWEL SIGN INVERTED SMALL V.

7. Suggested collation

0889 ARABIC LETTER HAH WITH INVERTED SMALL V BELOW can be sorted after 0684 ARABIC LETTER DYEH and before 0686 ARABIC LETTER TCHEH. 088A ARABIC LETTER NOON WITH INVERTED SMALL V can be sorted after 0769 ARABIC LETTER NOON WITH SMALL V and before 0647 ARABIC LETTER HEH.

```
0684 < 0889 < 0686
0769 < 088A < 0647
```

8. Examples

∴

Nous commencerons par donner la suite des caractères serbes avec leurs correspondants arabes.

LETTRES SERBES latines (croates)	VALEUR EN FRANÇAIS	LETTRES ARABES
A	a	ا (آ au commenc. d'un mot).
B	b	ب
C	ts, dans tsar	ج
Ć	ti, dans amitié	ج
Č	tch, dans tchèque	ج
D	d	د
Gj	di, dans adieu	ج
Dž	dj, dans djinn	ج
E	é, è	ء (ا au commencem. d'un mot).
F	f	ف
G	gu, dans gué	غ
H	h (très aspiré)	ح
I	i	ى (اى au commenc. d'un mot).
J	y, dans yeux	ى
K	qu, dans qui	ك
L	l	ل
Lj	ll, dans fille	ل
M	m	م
N	n	ن
Nj	gn, dans vigne	ن

Figure 1 : ARABIC LETTER NOON WITH INVERTED SMALL V (mistakenly with a small v instead of an inverted small v) and ARABIC LETTER HAH WITH INVERTED SMALL V BELOW in Bourgeois 1913, page 298.

Donnons ici quelques exemples des « graphèmes réformés » de Čaušević : le *wăw* surmonté d'un accent circonflexe — و̂ — notait « o », mais avec un accent circonflexe à l'envers — و̌ — il notait par contre « u ». Chacun des phonèmes /c/, /č/ et /ć/ fut noté par un graphème spécifique ; c'est ainsi que /č/, comme en persan et en osmanli, est noté par چ ; pour /ć/ et /c/ ح restait comme signe de base : on y adjoignait un accent circonflexe pour noter /ć/, soit ح̂, et deux points pour noter /c/, soit ح̈. Pour /ń/, on ajoutait au caractère de base ن un accent circonflexe suscrit, soit ن̂, pour /l'/ on munissait le *lăm*, ل, d'un accent circonflexe renversé, soit : ل̌. C'est en particulier à travers l'usage de l'accent circonflexe, droit ou à l'envers, comme signe diacritique, que le caractère de l'écriture arabe fut quelque peu altéré, ce que le précurseur anonyme du XVIe siècle avait précisément évité. Afin de donner au lecteur une idée de la graphie de Čaušević nous lui proposons un extrait de l'introduction au numéro 2 du calendrier *Mekteb* édité par Čaušević lui-même (1326/1908-1909), où celui-ci présente les buts politiques et culturels qu'il poursuit à travers sa réforme :

عربسقو ييسمز، اوتو ييسمو سقوييم يه اى ناش ليبـى،
قر آن عظيم الشان نايسان، از ناشوى دزمو ونى بوچهلو سه به
بيلو سا ازجيبز غزبتى، آلى شكر دراغوم اللهز از بروشلوى
غودنى دوبيسمو چهتيرى ويارسقه قتيفه نا ناشهم يهزهقو
سا عربسقيم حرزفزم، سبومو جز قوييج سادا نا هيئاده
مز سليانا اى مز سليانسقيج أولادا، جهستز غلهدايوجى عربسقو
ييسمز دزلازه دو سوهنى دززنوستى اززويثهنوغ اسلاما.

« Arabsko pismo, ono pismo s kojim je i naš lijepi Qur'ān ʻazīm aš-šān [arabe = « son aspect est splendide »] napisan, u našoj domovini počelo se je bilo sa očiju gubiti, ali šućur [arabe *šukr* = « grâce à »] dragom Allahu u prošloj godini dobismo četiri vjerske knjige na našem jeziku sa arabskom hrufom [arabe *harf*, pl. *hurūf* = « lettre »], s pomoću kojih sada na hiljade muslimana i muslimanskih evlada [arabe *walad*, pluriel *aulād* = « l'enfant »],

Figure 2 : ARABIC LETTER NOON WITH INVERTED SMALL V (mistakenly with a dot in the description text but correctly without a dot in the sample text) and ARABIC LETTER HAH WITH INVERTED SMALL V BELOW (mistakenly with the inverted small v above in the description text but correctly with the inverted small v below in the sample text) in Lehfeldt 2001 showing a sample from the *Mekteb* calendar by Čaušević (1326 H / 1908-1909 CE).

For both *o* and *u* remained *wáw*, however, for *o* there was a circumflex (ۇ) and for *u* an inverted circumflex (ۈ). Finally for *i* Čaušević kept the Arabic *yā'* usually without dots as *alif maqṣūra* (ى). As for consonants, for *h* the Arabic emphatic *ḥā'* (ح) was saved as ه was already reserved for the vowel *e*. Concerning other characters that did not have their equivalent in the Arabic-Ottoman script, all of them received a particular grapheme: *č* remained written like in Persian or Ottoman Turkish (چ); however, this was without any additional functions. For *c* and *ć* the letter ح was preserved as a basic grapheme, although with additional diacritical marks: for *ć* with a circumflex (ح̂) and for *c* with two dots (ح̈). For *nj* the source letter remained the Arabic *nūn,* however, with a circumflex (ن̂) and for *lj* quite logically Arabic *lām* was used, again with a circumflex (ل̌).[49] Of course, the list of *Arebica* reformists was not finished with Čaušević's activities and parallel with him or later other similar and independent efforts appeared.[50]

Figure 3 : ARABIC LETTER NOON WITH INVERTED SMALL V and ARABIC LETTER HAH WITH INVERTED SMALL V BELOW in Gažáková 2014, page 462.
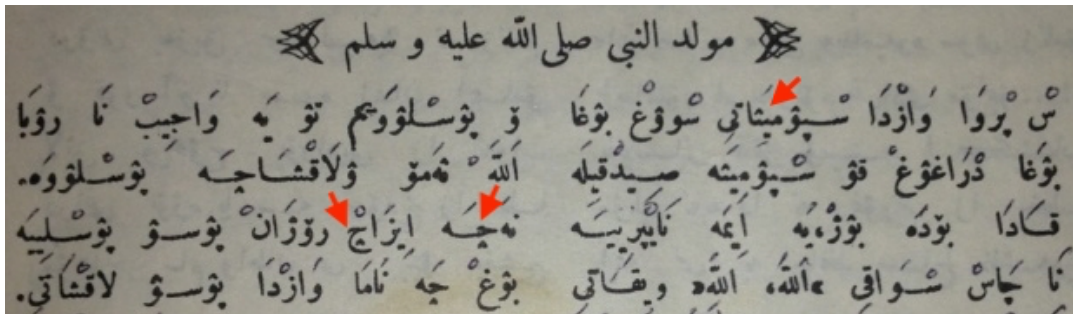

Figure 4 : ARABIC LETTER NOON WITH INVERTED SMALL V and ARABIC LETTER HAH WITH INVERTED SMALL V BELOW in Gašević (1354 H), p. 4.

نا پۆچەتقۆ ژوه (١٩١٦/١٩١٧) شقۆلسقه غۆدینه ایما۔

دۆح پرەد سۆبۆم ازۆیەستان برۆی نۆوع طالیبا، قۆیی چە

ساچیناوانی پرۆی رازرەد مەدرەسە۔ مۆرام س نیما ایزمەجۆ

ۆستالۆغا پۆجەنی ای فیقه۔ شتا ساد؟ نەما نی یەدنه قنیغه ایز

نه زنانۆستی ۆ ناشەم یەزیقۆ۔ ایمادەمۆ نەقۆلیقۆ »علم حالۆوا«

آلی سامۆ زا مەکتەبه۔ طالبی سۆ ایح مەدرەسه وەج ۆچیلی،

زا نیحا سه حۆجه نەشنۆ ویشه۔ نه پرەۆستاده می درۆغۆ،

نەغۆ سۆبۆم پیسانی ۆ ناشەم یەزیقۆ، طالەبی ایسپیسیوانی نا

طابلی دەرس پۆ دەرس، با دا ۆنی پرەپیسۆیۆ ۆ سۆوۆیه

بیلەژنجه۔

Figure 5 : ARABIC LETTER NOON WITH INVERTED SMALL V and ARABIC LETTER HAH WITH INVERTED SMALL V BELOW in Serdarevič 1941.

Već se maša u džepe svione

I vadi njoj knjigu oprošćenja,

Da uzimlje potpuno vjenčanje,

وەچ سه شی اۆجەپه سویۆنه

ای واد بۆی فشیغۆ اۆپرۆشنجه نا

دا اۆز په بۆنبۆنۆ ویبەنجانه

Figure 6 : ARABIC LETTER NOON WITH INVERTED SMALL V in *Takvim 1975*, p. 182–183.

Harfovica

NA LATINICI	NA KRAJU	U SREDINI	NA POČETKU	SAMOSTALNO	NA LATINICI	NA KRAJU	U SREDINI	NA POČETKU	SAMOSTALNO
A a	ـا	ـاـ	آـ	آ	L l	ـل	ـلـ	لـ	ل
B b	ـب	ـبـ	بـ	ب	Lj lj	ـلؐ	ـلؐـ	لؐـ	لؐ
C c	ـج	ـجـ	جـ	ج	M m	ـم	ـمـ	مـ	م
Č č	ـچ	ـچـ	چـ	چ	N n	ـن	ـنـ	نـ	ن
Ć ć	ـج	ـجـ	جـ	ج	Nj nj	ـنؐ	ـنؐـ	نؐـ	نؐ
D d	ـد	ـدـ	دـ	د	O o	ـۇ	ـۇـ	ۇـ	ۇ
Dž dž	ـج	ـجـ	جـ	ج	P p	ـپ	ـپـ	پـ	پ
Đ đ	ـغ	ـغـ	غـ	غ	R r	ـر	ـرـ	رـ	ر

Figure 7 : ARABIC LETTER NOON WITH INVERTED SMALL V and ARABIC LETTER HAH WITH INVERTED SMALL V BELOW in an alphabet table by Al-Zubi, available in Kurgaš (2016).

Figure 8 : ARABIC LETTER NOON WITH INVERTED SMALL V and ARABIC LETTER HAH WITH INVERTED SMALL V BELOW in Al-Zubi and Cicak-Al-Zubi (2013).

9. Bibliography

Amir Al-Zubi, Meliha Cicak-Al-Zubi (2013). "Hadži Šefko i hadži Mefko (TEFERIČ - تفريح ( ) ",
New Muslim Kids, 1 April, 2013. https://newmuslimkids.blogspot.com/2013/04/hadzi-sefko-i-hadzi-mefko-teferic.html

H. Bourgeois (1913). "Transcription arabe du serbe", in *Revue du monde musulman*, Paris,
Ernest Leroux, vol. 23, June 1913, p. 296–304 https://gallica.bnf.fr/ark:/12148/bpt6k103818v/f308.image.langEN

Mirsad Kurgaš (2016). "Renesansa Arebice", *Broj 19 - Number 19*.
http://www.broj19-number19.org/renesansa-arebice/

Salih Gašević (1354 H). *Mevlud*. Available on "Salih Gašević "Mevlud"", *Sandžačka
književnost*, http://sandzackaknjizevnost.com/salih-gasevic-mevlud/

Zuzana Gažáková (2014). "Some Remarks on Aljamiado Literature and the Usage of
Arebica in Bosnia and Heregovina", in Johannes den Heier – Andrea Schmidt – Tamara
Pataridze (eds.), *Script Beyond Borders : a survey of allographic traditions in the Euro-
Mediterranean world*, Leuven : Peeters, p. 453–471. on Academia.edu

Werner Lehfeldt (2001). "L'écriture arabe chez les slaves", in *Slavica occitania*, Alphabets
slaves et interculturalité, Toulouse, no. 12, 2001, p. 267–282. French translation from
German by Roger Comtet. http://w3.slavica-occitania.univ-tlse2.fr/pdf/articles/12/186.pdf

Maksida Pjanić (2009), *Die Arabismen in der Aljamiado Literatur Bosniens*, Universität
Wien. http://othes.univie.ac.at/8081/1/2009-12-16_0048319.pdf

Muhammed Seid Serdarevič (1941). فِقُه العِبادات [Fikh ul-idabat]. Sarajevo. (pictures from
edition of 1968 in "U biblioteci „Behram-beg" Fikhul-ibadat Muhamed Seid Serdarevića
pisan na arebici", *Muftijstvo tuzlansko*, muftijstvotz.ba, https://muftijstvotz.ba/u-
biblioteci-behram-beg-fikhul-ibadat-muhamed-seid-serdarevica-pisan-na-arebici/)

Takvim 1975 [تقويم ١٣٩٥]. Partially available in the "Mektebica حارفؤويجا Harfovica"
Facebook group: https://www.facebook.com/1421893648022473/photos/a.1963238813887951/1963238833887949/

Antun Vujic *et al.* (2013), "Arabica", *Proleksis Enciklopedija*, Leksikografski zavod Miroslav
Krleža. http://proleksis.lzmk.hr/9245/

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646
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*.

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)

A. Administrative

1. **Title:** *Proposal to encode Bosnian Arabic characters*
2. Requester's name: *Denis Moyogo Jacquerye <moyogo@gmail.com>*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *2019-10-03*
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: *NO*

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: *YES*
     - Name of the existing block: *Arabic Extended-B*
2. Number of characters in proposal: *2*
3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary		B.1-Specialized (small collection)	X	B.2-Specialized (large collection)	
C-Major extinct		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? *YES*
   - a. If YES, are the names in accordance with the "character naming guidelines" *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?

     Denis Moyogo Jacquerye
   - b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

     <moyogo@gmail.com>
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 issue

   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 (presentation, sorting)*

8. 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 UAX#44: 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.

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, available relevant documents: **Examples**

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? — **YES**
   Reference: **See proposal**

4. The context of use for the proposed characters type of use; common or rare) — **YES**
   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? — **Yes**
   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: