

Proposal to encode Arabic triple dot punctuation mark

Date: May 28, 2003
 Author: Jonathan Kew, SIL International
 Address: Horsleys Green
 High Wycombe
 Bucks HP14 3XL
 England
 Tel: +44 (1494) 682306
 Email: jonathan_kew@sil.org

A. Administrative

1. Title	Proposal to encode Arabic triple dot punctuation mark
2. Requester's name	SIL International (contacts: Jonathan Kew, Peter Constable)
3. Requester type	Expert contribution
4. Submission date	May 28, 2003
5. Requester's reference	
6a. Completion	This is a complete proposal.
6b. More information to be provided?	Only as required for clarification.

B. Technical — General

1a. New script? Name?	No
1b. Addition of characters to existing block? Name?	Yes — Arabic
2. Number of characters in proposal	1
3. Proposed category	A
4. Proposed level of implementation and rationale	1 (simple punctuation character)
5a. Character names included in proposal?	Yes
5b. Character names in accordance with guidelines?	Yes
5c. Character shapes reviewable?	Yes
6a. Who will provide computerized font?	Jonathan Kew, SIL International
6b. Font currently available?	Yes
6c. Font format?	TrueType
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	Yes
8. Does the proposal address other aspects of character data processing?	Yes, suggested character properties are included.

C. Technical – Justification

1.	Has this proposal for addition of character(s) been submitted before?	No
2a.	Has contact been made to members of the user community?	No
2b.	With whom?	N/A
3.	Information on the user community for the proposed characters is included?	Yes
4.	The context of use for the proposed characters	Encoding older Arabic-script documents in African languages.
5.	Are the proposed characters in current use by the user community?	Current use unclear; found frequently in older documents.
6a.	Must the proposed characters be entirely in the BMP?	Preferably
6b.	Rationale?	Closely associated with Arabic script.
7.	Should the proposed characters be kept together in a contiguous range?	N/A
8a.	Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	No
8b.	Rationale for inclusion?	N/A
9a.	Can any of the proposed characters be considered to be similar (in appearance or function) to an existing character?	Appearance somewhat similar to U+2234 THEREFORE
9b.	Rationale for inclusion?	Script-specific punctuation character, not a mathematical operator; properties differ.
10.	Does the proposal include the use of combining characters and/or use of composite sequences?	No
11.	Does the proposal contain characters with any special properties?	No

D. SC2/WG2 Administrative

To be completed by SC2/WG2

1. Relevant SC2/WG2 document numbers
2. Status (list of meeting number and corresponding action or disposition)
3. Additional contact to user communities, liaison organizations, etc.
4. Assigned category and assigned priority/time frame

Other comments

I. Proposal

Traditional orthographic practice when writing African languages such as Hausa, Wolof, Fulani, Mandinka, etc., in Arabic script includes the use of a characteristic punctuation mark as a “stop”. This mark consists of three dots in a triangle, similar to the pattern of dots found on the Arabic letters THEH and SHEEN, but used independently as a punctuation mark.

Although there is some suggestion that this mark is falling out of favor, with present-day writers tending to use more Latin-like punctuation, its widespread use in older texts appears to justify its encoding.

The following character is therefore proposed for addition to the UCS, with a suggested code value of U+0620:

<i>Glyph</i>	<i>USV</i>	<i>Character name</i>	<i>GenCat</i>	<i>CombClass</i>	<i>BidiType</i>
❦	0620	ARABIC TRIPLE DOT PUNCTUATION MARK	Po	0	AL

The character properties proposed here are analogous to those of other Arabic-script punctuation marks such as U+061B and U+06D4.

This character should be treated similarly to the other punctuation marks in collation; it is suggested that it be given a primary weight such that it sorts after U+06D4 ARABIC FULL STOP by default, although this is a rather arbitrary choice. The collation weight can of course be tailored as needed if a different behavior is desired in any particular language.

II. References

- Addis, R. T. 1963. *A study on the writing of Mandinka in Arabic script*. [SOAS library, London: call # LY Mandingo A/574203]
- Piłaszewicz, Stanisław. 1992. *The Zabarma conquest of north-west Ghana and Upper Volta: A Hausa narrative “Histories of Samory and Babatu and others” by Mallam Abu*. Warsaw: PWN—Polish Scientific Publishers.
- Taylor, F. W. 1929. *Fulani-Hausa readings in the native scripts: With transliterations and translations*. Oxford: OUP.

Appendix: Examples

The only punctuation mark that is used is equivalent to a full stop and is composed of three dots thus - ∴. One may occasionally encounter the Arabic interrogation mark ؟, but there is no other common form of punctuation, mainly because any punctuation as we know it can be easily confused with a letter or part of a letter in the Arabic alphabet. N.B. Writers today are tending to use the single dot full stop sign in preference to the three dot sign.

11. Nasals

In Mandinka nasalisation forms an important part of pronunciation. This is reflected throughout the script. Nasalisation is the voicing of a sound as it were through the back of the nose. The easiest examples are those English words ending in 'n', 'm' or 'ng', e.g. 'sin', 'sing', 'come'. You will find if you hold your nose that the last sound cannot be made. It is voiced by a breath through the nose. However, besides these the Mandinka nasalises a number of vowels. This is commonly believed also to be a failing of the English aristocracy who are said to 'speak through their noses!' In 'salāam āleikum', which are Arabic loan words, the marked vowels are nasalised and the 'a' of 'aleikum' is very commonly written to indicate this. A nasalised 'a' is most usually indicated by the use of 'ain' instead of 'alif'. However 'ain' is also used in a peculiarly Mandinka form and is then called 'tinkinteo'. In this form the 'ain' is written in its isolated shape only, is made rather squarer than the true Arabic letter, ع, is written above the line, is never joined on either side and is always followed by 'alif'. It can also take a vowel mark to show what vowel sound is being nasalised. Before further confusing the reader here is an example:-

'Amin' (another Arabic loan word) - عَامِين

This is 'ain' plus 'a' vowel mark, 'alif' alone, 'mim' plus 'i' vowel mark, 'ya' with no mark, 'nun' with no-vowel-sound mark and without its dot. The 'ain' is in its 'tinkinteo' form. The word could also be written

- عَمِين

the 'ain' here appearing in its proper initial form. It would be wrong to write it

- آمِين

with an 'alif' as this would not show the correct pronunciation of the opening sound. Perhaps a better explanation of the use of an initial 'ain' is that it indicates a word commencing with a closed throat or glottal stop. Here is a further example showing the tinkinteo used elsewhere than at the beginning of a word:-

Sango - Rabbit - سِنْعَا

'Sin' with 'a' vowel mark, 'nun' with no-vowel-sound mark, 'tinkinteo' with 'o' vowel mark, 'alif' alone.

It would also be possible to write it - سَانْعَا. 'Sin' with 'ang' vowel mark, alif alone, tinkinteo with 'o' vowel mark, alif alone.

This implies a nasalisation of the 'a' sound and of the 'o' sound.

One might equally well find the tinkinteo dispensed with and the nasal 'a' indicated by a simple vowel mark, thus: سِنُو

Again it would be well understood to write: سِنُو where the nasalisation is indicated merely by the presence of the nun. Despite these definite variations it is fair to say that 'ng' is most commonly expressed by the double vowel mark. The double vowel mark can also be used to show an 'n' or 'm' sound, but here it is more common for the



وغير ان تبارك تدرك في غير شيرك
نقلوه اذ ينون في تنوع واير غير
نحو غير فلهون غير في غير التبارك
نشير نقلوه اذ ينون ويشير في ك
كلور ينشر في غير في غير
كلور في غير في غير في غير
اذ ينون في غير في غير في غير
في غير في غير في غير في غير
ك نقلوه اذ ينون في غير في غير
ذو في غير في غير في غير في غير

Piłaszewicz (1992), page 168. Many examples of the proposed character.

about its real nature as many English phoneticians have: I have seen **puchu** written as 'pupchu' and 'pugchu', whilst Barbot gave 'pukku'.

پُش = puchchu = puchu, horse

عَشَّغْ = achchugo = achugo, to leave

(3) ء. This symbol is always called after its Arabic name, *hamza*, in which language it represents the glottal stop (see para. 3, no. 16), but in Fulani and Hausa it has no sound value. There are various rules for writing it in Arabic, but in purely Fulani and Hausa words it is used with | (often printed with a head, e.g. ٱ or ۞), either above or below according to the pointing, except that with a long 'a' at the commencement of the syllable it is used alone in order to avoid two alifs: this is done in Arabic by the *madda* ~, really an alif written horizontally: ex.

آءِ *āya*, the sign ۞ used as a stop

آيَةٌ 'āyatun', is the same word in Arabic

ٱلْقُرْءَانِ *Alqura'āni*, and قُرْءَانٍ *Kur'āni*, the Koran

ءَدَمَءَوٍ *Adamāwa*, the province.

In the printed character it is separated from the ٱ, but in writing it is shown by a small serif, thus ۞ (sometimes by two serifs, thus ۞ which is really a sort of double hamza written sideways), in which circumstances it is called *alif-hamaja* in Fulani, and *alif da tashi* or *alif da hamza* in Hausa: ex.

printed	أَبُ <i>abu</i> , thing	written	آبُ
„	إِنَا <i>inā</i> , where?	„	آِنَا
„	أَكُ or عَكُ <i>uku</i> or <i>ukku</i> , three	„	آكُ : عَكُ

3073-2

c

1

ڤر ڤول ورتك دغ كاسكم دم عمد نط دا جنفردم عجب
 كبل عيمه جم مع عجم د اد سار فم : باؤم جنفرد
 عم جم ولا كبلع باؤم كو كلمى سى تيز عملر : عجب
 نغ م فم م يمين وشغ غس عم فم خنى ك الله حكيم
 در ما ك فم كسى نطنل نى عجوبغ : عمبودا عطارك
 م پلتم ونط كسى نى : باؤن فم م پتم پشور طوبى
 دولا كسى الله جين بلتم ايسند منغ مع عها اجوتيم
 دا باب جوتيم فم كوتت نوالا جوتير سكا موم
 الله بسند بوبى مع جدرت مع امين هذا والسلام

1*

Ɖerēwol wurtake diga hā suka ma mo ardinɗa hā jangirde
 ma : ujinēje | kofli [sc. kōfli] e yamgo njamu ma e njamu
 Dāda-sāre fuh. Bāwo man jangirde | ummi jam, walā ko fe'i
 bāwo ma kōdume, sei hairu e mo'ere. Andingo : ma fahin,
 mi timmini wittugo ngesa am fuh hande. Ko Allah hokki
 yam | nder māka fuh kabbe chappande nai e jowēgo. Am bō
 ndā ngedāri ko | mi yeɗi ma wonɗo, kabbe tati. Bāwo don
 fahin mi yetti ma yettore dunde | nde walā kempe, Allah juttin
 balde ma, o besda mangu ma, gamā a heutini yam hā bāba
 heutinta biyum fuh, kō bana nō lāmɗo heutinirta suka mūm. |
 Allah besdu baude ma e darja ma, Āmīn. Hāzā wa's-salām.

you three bundles out of the largesse I am making. I am unend-
 ingly grateful to you : may God lengthen your days and grant you
 further honour, for you have helped me on as a father does his son,
 or as a chief advances his servant. May God increase your power
 and glory, Amen. This with weal.

D 2