Proposal to encode ADLAM NASALIZATION MARK for ADLaM script

Abdoulaye Barry, Ibrahima Barry, Peter Constable, Andrew Glass

September 17, 2018

Summary

ADLaM script was added to Unicode in version 9.0 (2016) based on Michael Everson's proposal documents submitted in 2013 and 2014 (L2/13-191 & L2/14-219R). The present document addresses a gap in the ADLaM encoding by proposing a new mark character as an addition to the existing ADLaM block:

Glyph	Code	Character name
,	1E94B	ADLAM NASALIZATION MARK

Usage requirements

ADLaM script uses an apostrophe-like sign called Nyondal to nasalize a succeeding consonant. For example:

ดารกว	/hin.du/	'ancient'
ชาวเรา	/hi.ªdu/	it'

This sign was known at the time the original proposals for ADLaM were submitted, but at that time, the use of European punctuation signs at their standard code points was expected to suffice (Everson 2013: 2, Everson 2014: 2). While implementing and using the ADLaM Unicode encoding, it has become clear that using ' (U+0027) or similar signs to represent ADLaM Nyondal does not work. Specific typographical features of Nyondal make it necessary to encode a distinct sign. These requirements are:

- 1. ADLaM Nyondal is a letter modifier not a punctuation sign and occurs in word internal position
- 2. As a letter, ADLaM Nyondal should have strong right-to-left directionality
- 3. Typographically, ADLaM Nyondal is straight, not hooked
- 4. ADLaM Nyondal is transparent to joining behavior, i.e., adjacent letters connect under it, but it doesn't cause joining by itself.

Comparable signs

Alternative characters already encoded in Unicode have been considered and rejected as follows:

• U+0027 APOSTROPHE

This character is functionally a punctuation character. It is subject to being replaced by a smart quote (U+2018 and U+2019) in many applications resulting in inconsistent spelling across applications. Not all

applications have the option to disable smart quotes and many users are not aware that the feature is enabled. Many applications give the apostrophe special logic in text-selection, etc. When used with joining scripts, it blocks joining behavior between adjacent signs.

• U+02BC MODIFIER LETTER APOSTROPHE

The default form is curved to work with LTR scripts. This character is used in several Latin-based orthography for African languages. As a result, modifying the design for ADLaM specifically will cause inconsistent appearance in mixed, script/language fonts. It blocks joining between adjacent signs. It has a strong-left directionality.

- U+A78B LATIN CAPITAL LETTER SALTILLO
- U+A78C LATIN SMALL LETTER SALTILLO

The default form of these characters are straight; but they have a strong-left directionality and will create the same inconsistent text ordering issues as U+02BC.

U+05F3 HEBREW PUNCTUATION GERESH

The default form is not vertical, and it blocks joining behavior. Though it is strong RTL, it's association with Hebrew script will cause complications with text itemization and font fallback behaviors.

Therefore, the right approach to encoding this sign is to encode as a dedicated character, ADLAM NASALIZATION MARK. The suggestion here is to encode it at the first unassigned code point in the ADLAM block.

Joining behavior

ADLAM Nyondal is spacing and transparent to joining. That is, it has the following characteristics:

Position	With	Without
	Nyondal	Nyondal
Initial, before ADLAM letter	ᆡ'	7
Medial position, between ADLAM letters	קק	Ыa
Final position, after ADLAM letter	ี 'ล	ล
Isolated position, not adjacent to ADLAM letters	I	

Properties

It should have the following properties:

Character properties In the format of UnicodeData.txt:

1E94B;ADLAM NASALIZATION MARK;Lm;0;R;;;;N;;;;

Linebreaking In the format of LineBreak.txt:

1E94B;AL #Lm [1] ADLAM NASALIZATION MARK

Joining categories In the format of ArabicShaping.txt:

1E94B; ADLAM NASALIZATION MARK; T; No_Joining_Group

References

Everson, Michael, 2013 "Preliminary proposal for encoding the Adlam script in the SMP of the UCS". N4488 L2/13-191. http://www.unicode.org/L2/L2013/13191-n4488-adlam.pdf

Everson, Michael, 2014. "Revised proposal for encoding the Adlam script in the SMP of the UCS". N4628R L2/14-219R. http://www.unicode.org/L2/L2014/14219r-n4628-adlam.pdf

	1E90	1E91	1E92	1E93	1E94	1E95
0	b 1E900	2 1E910	1 2	1 E930	g	O 1E950
1	U	5 1E911	<u>U</u>	A 1E931	ട	1 E951
2	<u> </u>	3 1E912	U	? 1E932	Q 1E942	ጎ
3	<u>O</u>	Q 1E913	1 E923	5	<u>면</u>	<u>y</u>
4	H2 1E904	<u>n</u>	1 E924	3	1E944	<u>Ł</u>
5	<u>Ш</u> 1Е905	<u>E</u>	S	Ø 1E935	1E945	£
6	<u>D</u>	<u>5</u>	L e 1E926	M		5
7	A	9 <u>1</u>	W	<u>ජ</u>	℃ ○ 1E947	<u>S</u>
8	<u>U</u>	<u>Y</u>	Q 1E928	<u>5</u>	√ ○ 1E948	X 1E958
9	B	S 1E919	S 1E929	91 1E939	√ / ○ 1E949	P 1E959
Α	1E90A	1E91A	ñ	S	1E94A	
В	U 1E90B	£ 1E91B	G	S 1E93B	1 1E94B	
С	<u>C</u>	A	5	P		
D	1E90D	1E91D	U 1E92D	Æ 1E93D		
Ε	1E90E	<u>u</u>	C) 1E92E	A 1E93E		T
F	Fl. 1E90F	1E91F	d 1E92F	1E93F		1E95F

1E900 Adlam 1E95F

Capital letters 1E900 & ADLAM CAPITAL LETTER ALIF 1E901 ADLAM CAPITAL LETTER DAALI 1E902 ŋ ADLAM CAPITAL LETTER LAAM 1E903 ⊆ ADLAM CAPITAL LETTER MIIM 1E904 12 ADLAM CAPITAL LETTER BA 1E905 W ADLAM CAPITAL LETTER SINNYIIYHE 1E906 D ADLAM CAPITAL LETTER PE 1E907 A ADLAM CAPITAL LETTER BHE 1E909 & ADLAM CAPITAL LETTER E 1E90A 尚 ADLAM CAPITAL LETTER FA 1E90B U ADLAM CAPITAL LETTER I 1E90C C ADLAM CAPITAL LETTER O 1E90D d ADLAM CAPITAL LETTER DHA 1E90E ≥ ADLAM CAPITAL LETTER YHE 1E90F & ADLAM CAPITAL LETTER WAW 1E910 a ADLAM CAPITAL LETTER NUN 1E911 3 ADLAM CAPITAL LETTER KAF 1E912 3 ADLAM CAPITAL LETTER YA 1E913 Ø ADLAM CAPITAL LETTER U 1E914 ဤ ADLAM CAPITAL LETTER JIIM 1E915 원 ADLAM CAPITAL LETTER CHI 1E916 5 ADLAM CAPITAL LETTER HA 1E917 91 ADLAM CAPITAL LETTER QAAF 1E918 5 ADLAM CAPITAL LETTER GA 1E919 8 ADLAM CAPITAL LETTER NYA 1E91A (2 ADLAM CAPITAL LETTER TU 1E91B & ADLAM CAPITAL LETTER NHA Supplementary capital letters 1E91C A ADLAM CAPITAL LETTER VA 1E91D @ ADLAM CAPITAL LETTER KHA 1E91E ♥ ADLAM CAPITAL LETTER GBE 1E91F S ADLAM CAPITAL LETTER ZAL 1E920 1E921 및 ADLAM CAPITAL LETTER SHA Small letters 1E922 & ADLAM SMALL LETTER ALIF 1E923 4 ADLAM SMALL LETTER DAALI 1E924 o ADLAM SMALL LETTER LAAM 1E925 s ADLAM SMALL LETTER MIIM 1E926 ADLAM SMALL LETTER BA 1E927 W ADLAM SMALL LETTER SINNYIIYHE 1E928 @ ADLAM SMALL LETTER PE 1E929 s ADLAM SMALL LETTER BHE ADLAM SMALL LETTER RA 1E92A ⊌ 1E92B @ ADLAM SMALL LETTER E 1E92C 6 ADLAM SMALL LETTER FA 1E92D U ADLAM SMALL LETTER I 1E92E 9 ADLAM SMALL LETTER O 1E92F d ADLAM SMALL LETTER DHA 1E930 C ADLAM SMALL LETTER YHE 1E931 a ADLAM SMALL LETTER WAW 1E932 a ADLAM SMALL LETTER NUN 1E933 # ADLAM SMALL LETTER KAF 1E934 a ADLAM SMALL LETTER YA 1E935 & ADLAM SMALL LETTER U

ADLAM SMALL LETTER JIIM

ADLAM SMALL LETTER CHI

ADLAM SMALL LETTER HA

1E936

1E937

1E938

O

ŏ

כ

```
1E939 of ADLAM SMALL LETTER QAAF
1E93A ʁ
         ADLAM SMALL LETTER GA
1E93B s
         ADLAM SMALL LETTER NYA
1E93C @ ADLAM SMALL LETTER TU
1E93D & ADLAM SMALL LETTER NHA
Supplementary small letters
1E93E a ADLAM SMALL LETTER VA
1E93F
      15
         ADLAM SMALL LETTER KHA
1E940 s
         ADLAM SMALL LETTER GBE
1E941
      В
         ADLAM SMALL LETTER ZAL
1E942
      œ
         ADLAM SMALL LETTER KPO
1E943 # ADLAM SMALL LETTER SHA
Diacritical marks
1E944
         ADLAM ALIF LENGTHENER
1E945
         ADLAM VOWEL LENGTHENER
1E946
         ADLAM GEMINATION MARK
1E947
         ADLAM HAMZA
1E948
         ADLAM CONSONANT MODIFIER
1E949
         ADLAM GEMINATE CONSONANT MODIFIER
1E94A
         ADLAM NUKTA

    called hoortobbhere when drawn above and

          lestobbhere when drawn below the base
          character
1E94B
         ADLAM NASALIZATION MARK
         = nyondal
```

Digits

Digits		
1E950	0	ADLAM DIGIT ZERO
1E951	1	ADLAM DIGIT ONE
1E952	ጎ	ADLAM DIGIT TWO
1E953	ų	ADLAM DIGIT THREE
1E954	Ł	ADLAM DIGIT FOUR
1E955	Ð	ADLAM DIGIT FIVE
1E956	Ŗ	ADLAM DIGIT SIX
1E957	8	ADLAM DIGIT SEVEN
1E958	8	ADLAM DIGIT EIGHT
1E959	٦	ADLAM DIGIT NINE

Punctuation

1E95E	$\boldsymbol{\sigma}$	ADLAM INITIAL EXCLAMATION MARK
1E95F	ล	ADLAM INITIAL QUESTION MARK

₽₽₽₽₽₽₽₽ Es Costlas Vosu Vsient, Hita? SOH TREET FORE PAR COLL n'der حسآج ملاه الحماثات Lancorth The 8 2587 n'galTHE HATE THE SERVER עלעדער **היישטעהיוליור**י B RY WITHUY JULY BU B C F n'gol285 स्प्राया **१८८०५, १८७८ |** Шт. ५, ५, ६०६५७. स्प्र<mark>ायप्र</mark> n'gol Ձ **₹નુમ નુમેત્રમ. ૫೩ મેંગમ** મેંગ ઉત્પાધ મેંગ દેખદેવે ૧ ૧૦૦ વ્યાપ્

Fig. 1. Excerpt from booklet in Fulani discussing democracy, entitled *Demokarāku* (Everson 2014: 14). Highlighted examples show ADLaM Nyondal with correct joining behavior.

حوته الله على - حصا هجس و مسومهم ਭੂਈ * ਉਹਾੜ੍ਹ والمسكم والمستمية ਭੂਸਨ ਫਸਮੀਰ ਹਯਸੀਸੀਣ ਪੂਠਧਾਨ ਫਸਨ ਯਸਹਸਿਨ ਯਰਫ਼ਨ ਟੂ المتاس الله ومع المالية عبي المالية ال ر براك مريم عربية على المعلمة على المعلمة على المعلم المع ਰੂਨ ਨੀ ਸਤੀ ਹਿਣ ਤਾਹੁੰਦੇ ਹਨ। ور ی کردها و الدمال کی همون کردهای لے ک طعبرا حمعیں، قمعہ لاقبلہا کا کہانستاسی ور حمتم علامية للمعلم المتما ع المحالد لم יחהוובש בסקק סאשב שאב פהצייולים שקג שהווזאל. كِياسَةً, سَسَةً, بَهِدِيسَةً; جَهِد عَمَلَم جَهِا عَ طَرَكِين עוווישה הטלישה ניגדסף לצ ਮੁਸ਼ਿਸ਼ਾਨ ਹਨ <mark>ਗ</mark>੍ਰਾਣ ਦੁਸ਼ਦਰਦੂੰ ਤਾਭਾ ਦੀ ਸਾਹਰ ਦੀ ਸੂਬਾਰ ਵਸੰਭਾ ਤਾਲ ਦੀ ਸ਼ਾਹਰ ਵਸੰਭਾ ਤਾਲ र्जेडाउ, र्जाम, ८ उर्जन्य टेजर: उप्त जाह्यवर्जे टावर्जेडर אסעווים א מאסע עין הגהסע אינה التكويسية جهد عملاء فيظها الحالية والمي ميأع فمحدج و 15 معلكم وكير ਪ੍ਰਸੇਰ ਫ਼ਸਪੀਰ ਪਸ ਫੁਣ ਪੁਰਹੁਪ੍ਰ טאליאל מבטן טאסטַטאַ עיַ שאבטן

Fig. 2. Story in Fulani of "the laborer and his children". Examples of ADLaM Nyondal in red.

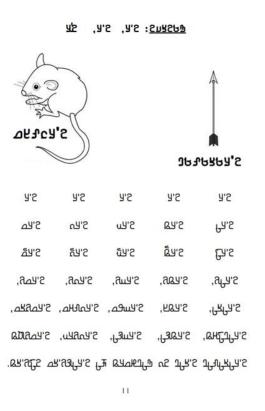


Fig. 3. Sample of ADLaM Nyondal with defective rendering. In this example, the expected joining behavior is not possible due to the use of standard apostrophe that blocks joining.