

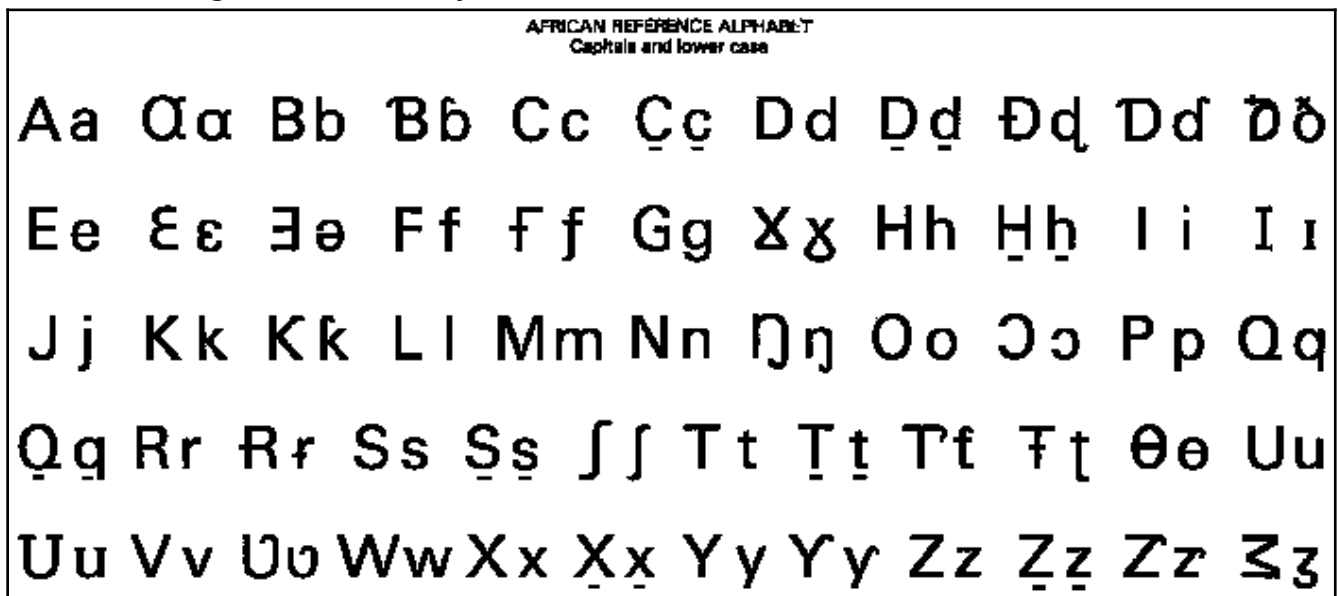
On the 1978 version of the African Reference Alphabet

Eduardo Marín Silva

10/09/2021

Introduction. The African Reference Alphabet (ARA) was a project originating from UNESCO, to create a uniform system, to faithfully transcribe a wide variety of African languages. It went through two major revisions, one in 1978 and another in 1982. The latter is unicameral and is fully supported by Unicode (except for maybe the tilde like letter that could be considered a glyph variant of 0274 ɲ). This document mainly deals with the 1978 version; it is bicameral and has many strange glyphs, that are the subject of our discussion. I give my opinion, on whether or not these characters merit disunification, and if they are better treated as variants, I also suggest a helpful annotation under the base character, so that there's less ambiguity on the nature of the characters.

Here is the original illustration by the creators, that shall be used as the main reference:

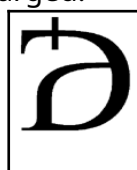


Capital Eth. As it is well known, there are 3 homoglyphs with different looking lowercases. Here we can see the first pair corresponds to "Đđ" which in Unicode correspond to 0189 Đ LATIN CAPITAL LETTER AFRICAN D and 0256 đ LATIN SMALL LETTER D WITH TAIL, so far this isn't problematic at all, since they are in a formal case relation.

However, two positions later, there is a strange letter pair "D̄đ". If the lowercase is interpreted as 00F0 đ LATIN SMALL LETTER ETH, then the uppercase could be interpreted as a glyph variant of 00D0 Đ LATIN CAPITAL LETTER ETH. It seems that the shape was modified to avoid confusability with the previous one; instead of being based on the regular capital D, it seems to be a more "angular" version of the lowercase; therefore being a true hybrid between D and đ.

In my opinion, treating this as a glyph variant of 00D0 Đ is not problematic, but it is important to document it somewhere in the standard, so that people don't confuse it as a distinctive letter.

The glyph is extracted from [this forum post](#) (hosted by SIL) by Peter Constable; discussing glyphic variants of African orthographies. Here it is enlarged:

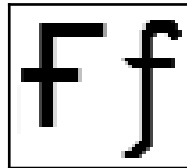


F with longer stroke and f with hook. Following the regular pair of "Ff" there is a pair of letters "Ƒƒ". The uppercase is distinguished by its longer stroke towards the left and the lowercase by the hook at the bottom.

In Unicode, there are two relevant casing pairs: 0191 Ƒ, 0192 ƒ and A798 Ƒ, A799 ƒ with the first being "with hook" and the latter being "with stroke". The uppercase is more similar to A798 Ƒ but without the "hook" below and the lowercase looks just like 0192 ƒ. Notice how for A798 Ƒ instead of having an extra bar like the lowercase, the bar is elongated to the left and a hook is added similarly to 0191 Ƒ. Interestingly, an uppercase letter "F" with an extra bar, is the glyph used for 20A3 ₣ FRENCH FRANC SIGN.

In my opinion, these letters cannot be unified with either of those pairs; not without going against what their names state; and so they merit disunification. The proposed names would be: LATIN CAPITAL/SMALL LETTER OLD AFRICAN F. The names include the term "OLD", because these letters are no longer in common use, but term can be removed so that it now reads: LATIN CAPITAL/SMALL LETTER AFRICAN F. If such a disunification takes place, the bullet note under 0191 Ƒ that reads "African" would need to be removed, and proper cross references should also be added. The new letters can also have a bullet note stating "used in the 1978 version of the 'African Reference Alphabet'".

Here are the enlarged glyphs based on the original illustration:

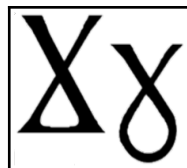


Latin Gamma. After the regular pair of "Gg" there's a pair of letters "Ƴƶ". These letters seem to be based on the shapes for the "Xx" but with the first having a horizontal bar closing the bottom and the latter having a semi-circle appended to the bottom. The evidence points to the lowercase being identified as 0263 ƶ LATIN SMALL LETTER GAMMA, making the uppercase 0194 Ƴ LATIN CAPITAL LETTER GAMMA.

The use of the letters Xx as a base, can be interpreted as the desired appearance, that eventually evolved into different shapes. This makes sense, since if we compare them: Ƴƶ ƶƶ, one can see that the capital and small letters are more distinctive from each other in the earlier design, with the capital having a flat bottom and never going below the baseline. We can even see that mixing the old capital with the new gamma still harmonizes well: Ƴƶ

In my opinion, treating these as glyph variants of the already encoded Latin letters isn't problematic, but it is important to document it somewhere in the standard; particularly the capital form given their pronounced difference (Ƴ vs X). I recommend adding a bullet note under 0194 Ƴ, stating "a glyph variant looks like the capital 'X' but with the bottom closed by a horizontal bar".

Here are the enlarged glyphs based on the original illustration:

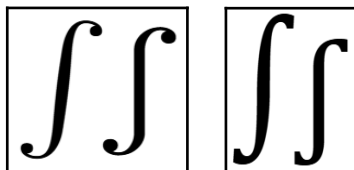


Capital Esh. Following the pair "ƆƆ" there is the following pair "ƆƆ". If the lowercase is identified as 0283 Ɔ LATIN SMALL LETTER ESH, then the uppercase can be interpreted as a glyph variant of 01A9 Σ LATIN CAPITAL LETTER ESH. It seems to be a slightly wider and taller version of the small letter.

In my opinion, treating this as a glyph variant of 01A9 Σ isn't problematic, but it is important to document it somewhere in the standard, given their pronounced difference (Ɔ vs Σ). I recommend

adding a bullet note under 01A9 Σ stating "a glyph variant looks like a slightly taller and wider version of 0283 ʃ".

Here are two enlarged mock-ups of how the proposed casing pair may have looked like with a more detailed font:



Z and z with upper-right hook. Following the pair "Zz" there is a pair "ZZ" that can be identified as the regular letters "Zz" but with a hook attached to the upper right vertex.

In Unicode there is the pair 0224 Z and 0225 z which are named: LATIN CAPITAL/SMALL LETTER Z WITH HOOK. The bullet note under, states that they are used in 'Middle High German' (used for the coronal fricative in that context). On this version of the ARA, this letter was only meant to represent an 'emphatic' version of Z, which is a different sound.

Further complicating things is the pair 2C7F Z and 0240 z named: LATIN CAPITAL/SMALL LETTER Z WITH SWASH TAIL, these letters were invented by linguists in Africa (independently from the ARA) to represent the 'voiced labio-alveolar fricative', present in Shona (it was later adopted by Shona speakers for the 'whistled z' but fell into disuse later). The question then becomes, if the letters in the ARA should be considered variants of these letters instead, since at least they are closer geographically.

In conclusion, while the different placement of the hook isn't enough to disunify, the fact that it was also invented in a different context for a different sound merits disunification. This is further illustrated by the comment on the [Wikipedia article on the ARA](#), claiming that it still can't be represented correctly in Unicode.

In my opinion, the names should be: LATIN CAPITAL/SMALL LETTER Z WITH UPPER-RIGHT HOOK, along with the proper cross references, a bullet note could state "used in the 1978 version of the 'African Reference Alphabet'"

Here are the enlarged glyphs based on the original illustration:



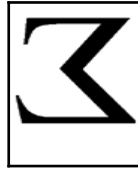
Capital Ezh. Following the Z with upper right hook, we see the following pair of letters "ZZ". If the lowercase is identified as 0292 z LATIN SMALL LETTER EZH, then the uppercase could be interpreted as a glyph variant of 01B7 Z LATIN CAPITAL LETTER EZH.

When comparing the regular glyphs of the ezh (z vs z) it is clear that the uppercase is just a shifted up version of the lowercase. When we do the same, but with the alternate capital (Z vs z), we can see that the uppercase is more distinctive from the lowercase, as it no longer has the round "bowl" at the bottom, and it indeed appears like a reversed version of a certain Greek letter (03A3 Σ GREEK CAPITAL LETTER SIGMA) and in turn, to the capital version of the Latin letter (01A9 Σ LATIN CAPITAL LETTER ESH).

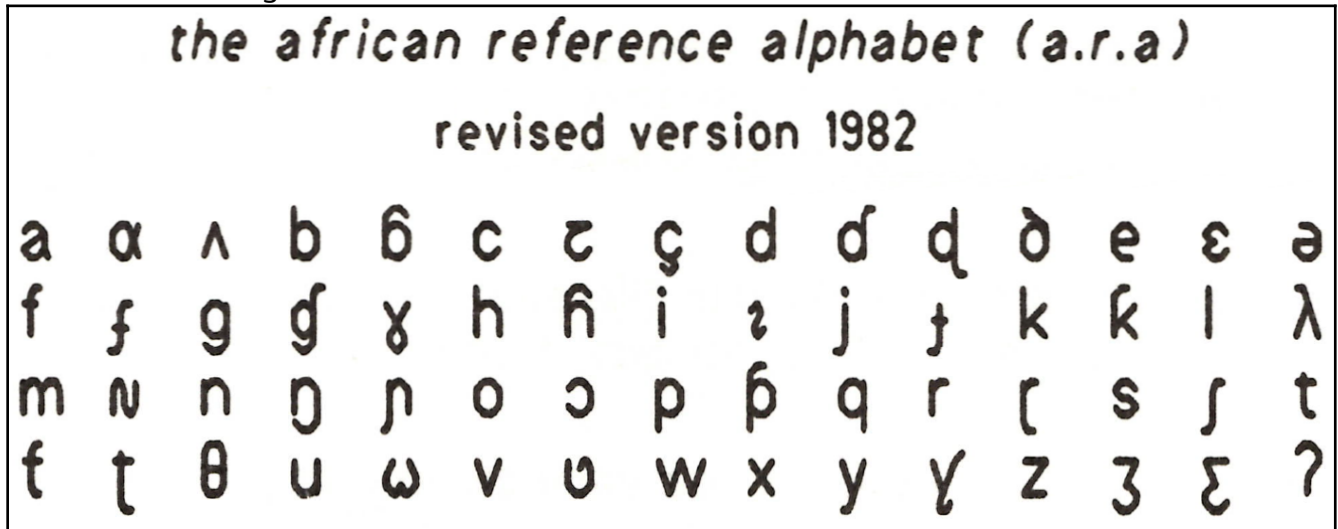
Considering the phonetic relation between the "esh" and the "ezh", this glyph pair seems apt to highlight it (Σ vs Z), which I mention, just to point out that it is plausible people might go for such a glyph, independent of the ARA.

In my opinion, treating this as a glyph variant 01B7 Z isn't problematic, but it's important to document it somewhere in the standard. I propose to add a bullet note under 01B7 Z, stating "a glyph variant, looks like a reversed 'Greek Capital Sigma' (03A3 Σ)".

The glyph was taken from the same [forum post](#) by Peter Constable. Here it is enlarged:



Linearized tilde. To be thorough, we shall also discuss the unusual character from the 1982 version. Here is the original illustration:



After the letter 'm' and before 'n' there is a character that is called "linearized tilde", this name does not imply this is meant for punctuation or just some modifier; it is a proper consonant letter. The position of the letter as well as the glyph invoke the use of the character: 0274 ɱ LATIN LETTER SMALL CAPITAL N.

In my opinion, treating this as a glyph variant of 0274 ɱ isn't problematic, but it is important to document it somewhere in the standard. I suggest adding a bullet note under 0274 ɱ stating "used in the 1982 version of the African Reference Alphabet, where the glyph looks more wavy and it is called the 'linearized tilde'".

Summary. In this document I have proposed annotating 4 characters (00D0 Đ, 0194 Ƴ, 01B7 Ʒ and 0274 ɱ) to better document important glyph variation; although I'm not sure what annotation is fitting for the "eth"; clarifying that the glyph variants in question apply to the 1978 (or 1982 for the last) version of the ARA, may also be added if deemed useful.

I have also proposed the encoding of two casing pairs (for a total of 4 characters), the first consisting of F and f but with strange modifications to their glyphs, as well as the Z and z with an upper-right hook. These two additions would in essence complete the entire alphabet, but out of the two pairs, the latter has definitely the stronger case. The following figures in the following pages illustrate that.

C - T A M A S H E Q

ALPHABET a b d ḍ e ẹ f g gh h i j k kh l m
 n ŋ o p q r ṣ sh t ʈ u w y z ẓ

INVENTAIRE DES SONS REPRESENTES

1. CONSONNES	<u>Labiales</u>	<u>Dentales</u>	<u>Palatales</u>	<u>Vélaires</u>	<u>Glottales</u>
Occlusives sourdes	p (*)	t		k	q
sonores	b	d	j	g	
Emphatiques sourdes		ʈ			
sonores		d			
Constrictives sourdes	f	s	sh	kh	h
sonores		ṣ		gh	
emphatiques sourde		ṣ			
emphatiques sonore		ẓ			
Nasales	m	n	ny	ŋ*	
Latérales		l			
Vibrantes		r			
Semi-voyelles			y	w	

Figure a) Evidence of the original identity of the z with upper-right hook
<http://www.bisharat.net/Documents/Bko66TamasheqTableau.html>

4 Consonnes emphatiques

Les experts se sont penchés sur ce problème particulièrement difficile.

Plusieurs solutions ont été examinées pour la transcription des consonnes emphatiques, comme suit:

(a) majuscules : D T S Z

(b) crossées : ḍ ṭ ṣ ẓ

(c) apostrophe: ‘d ‘t ‘s ‘z

Figure b) Corroborating evidence from another document
<http://www.bisharat.net/Documents/Bamako1966.htm>

The English version proposed an alphabet of 57 letters, given in both upper and lower-case forms. Eight of these are formed from common Latin letters with the addition of an underline mark (̣). Some of the glyphs (two uppercase forms Ḍ and Ẓ, one lowercase form ẓ) cannot be accurately represented in Unicode (as of version 13, March 2020).^[1]

Figure c) Relevant excerpt from the Wikipedia article on the ARA
https://en.wikipedia.org/wiki/African_reference_alphabet

Caractères (1978) [[modifier](#) | [modifier le code](#)]

La version de 1978 comprend 57 caractères dans la version française et 56 dans la version anglaise.

minuscule	a	ɑ	b	ɓ	c	ç	d	ɗ	ɔ	ɔ́	ɔ̃	
majuscule	A	ɑ	B	ɓ	C	Ç	D	ɗ	Ɖ	Ɗ	Ɗ̃	
minuscule	e	ɛ	ə	f	f	g	ɣ	h	ɧ	i	ɪ	
majuscule	E	ɛ	Ǝ	F	Ƒ	G	Ƴ	H	ɥ	I	ɪ	
minuscule	j	k	ƙ	l	m	n	ɲ	o	ɔ	p	q	
majuscule	J	K	Ƙ	L	M	N	Ɲ	O	Ɔ	P	Q	
minuscule	q	r	ɾ	s	ɕ	ʃ	t	ɛ	ɛ́	ə	u	
majuscule	Q	R	Ɛ	S	Ɔ	ƚ	T	Ɛ́	Ɛ́	Ɛ	U	
minuscule	ɔ	v	ɔ	w	x	ɣ	y	z	z	Z	Ʒ	
majuscule	Ɔ	V	Ɔ	W	X	Ƴ	Ƴ	Z	Z	Z	Ʒ	

Remarques :

- les caractères Z majuscule et minuscule n'ont jamais été codés ;

Figure d) Chart on the French Wikipedia forced to use graphics rather than codepoints
https://fr.wikipedia.org/wiki/Alphabet_africain_de_r%C3%A9f%C3%A9rence