

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

Doc Type: Working Group Document

Title: Preliminary proposal to encode the Old North Arabian script in the SMP of the UCS

Source: Michael Everson and M. C. A. Macdonald

Status: Individual Contribution

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

Date: 2010-03-30

1. Introduction. The South Semitic script family split from the original consonantal alphabet in the second millennium, and was used exclusively in Arabia and its immediate environs between at least the tenth century BCE and the sixth century CE. There were two branches: the Old North Arabian (Semiticists now tend to use the term Ancient North Arabian or ANA) and the Old South Arabian (also called Ancient South Arabian or ASA), and from the latter the abugida used for Ge'ez and other languages of Ethiopia was derived. First the Old North Arabian, and then the Old South Arabian scripts disappeared in the first half millennium CE, and with the coming of Islam they were replaced by the Arabic script, which was derived from Aramaic via Nabataean.

The South Semitic scripts of Arabia are known only from inscriptions and all represent “dead” languages or dialects. So we cannot know the exact sounds represented by the letters. However, by comparing the etymology of the words in these scripts with related words in other Semitic languages (particularly Arabic) one can work out what is called the “etymological phoneme” expressed by each letter (regardless of how this was actually realized in speech). The etymological phonemic repertoire appears to be the same as that of Classical Arabic, with 28 consonants in Old North Arabian and 29 in Old South Arabian.

Old North Arabian is the collective term for a group of scripts (and the dialects they express) which were used in the western two-thirds of Arabia from Syria to the borders of Yemen. Three were used in oases and form a subgroup “Oasis North Arabian”, while the others were used by nomads, and are known as “Desert North Arabian” (Macdonald 2009 I: 74–96). On the nomenclature used here see Macdonald 2009 III: 32–36. The earliest datable texts (Taymanitic and Thamudic B) are from the mid-sixth century BCE and the Old North Arabian scripts are thought to have fallen out of use after the fourth century CE since they reveal no trace of Christianity. Dadanitic (formerly “Dedanite” and “Lihyanite”) was used in the oasis of Dadan/Dedān (modern al-‘Ulā in northwest Saudi Arabia). There are many formal inscriptions (often carved in relief) on stelae and rock-faces, and hundreds of graffiti. All formal inscriptions and most graffiti run horizontally right-to-left with words separated by a vertical line in the inscriptions. No everyday texts on perishable materials have yet been found, but their existence may be implied by the presence of cursive-style letter shapes which are mixed with monumental-style ones in many public inscriptions. These different forms were once thought to have chronological significance (Caskel 1954: 21–37) but their presence in the same inscriptions shows this is not so (Macdonald *forthcoming a*). There are also many Minaic inscriptions at Dedān and this is the only known occasion on which settled users of an Old South Arabian and an Old North Arabian script lived in close proximity.

2. Processing. Old North Arabian is an alphabetic script consisting only of consonants; vowels are not indicated in the script, though some Dadanitic texts do make limited use of *matres lectionis*. The script has been encoded with right-to-left directionality, which is typical for Dadanitic. (Taymanitic and Thamudic B can also be written from left-to-right as well as right-to-left and boustrophedon. Thamudic C and D are usually written vertically downwards. Safaitic and Hismaic can be written in any conceivable direction including horizontally, vertically (up or down), boustrophedon, coils, zig-zags, etc.) Glyphs may be mirrored in lines when they have left-to-right directionality, though in Safaitic and Hismaic inscriptions this is not always the case; the choice must be a matter for the font for such palaeographic implementations.

3. Ordering. Traditional sorting orders are poorly attested. One abecedary based on “Northwest Semitic” order is known, as is one based on “South Semitic” ordering, similar to that used in Old South Arabian. A number of other “abecedaries” are known which sort the letters by shape; it is assumed that these may be mnemonic systems used to learn the alphabet, which seems to have been passed from person to person rather than imparted via formal schooling.

3.1. Order in the code chart. Because Old North Arabian letters are related to those of Old South Arabian, the order of the characters in the code chart follows the latter. Note that the letter 𐩦 is found only in the Taymanitic branch of Old North Arabian. In the table below, transliteration is given, followed by the Old North Arabian and then the Old South Arabian and Ethiopic:

	h	l	ḥ	m	q	w	s ²	r	b	t	s ¹	k	n	ḥ	ṣ	s ³	f	'	‘	ḍ	g	d	ḡ	ṭ	z	ḏ	y	ṯ	ṣ	
NARB	𐩇	𐩈	𐩉	𐩊	𐩋	𐩌	𐩍	𐩎	𐩏	𐩐	𐩑	𐩒	𐩓	𐩔	𐩕	𐩖	𐩗	𐩘	𐩙	𐩚	𐩛	𐩜	𐩝	𐩞	𐩟	𐩠	𐩡	𐩢	𐩣	𐩤
SARB	𐩥	𐩦	𐩧	𐩨	𐩩	𐩪	𐩫	𐩬	𐩭	𐩮	𐩯	𐩰	𐩱	𐩲	𐩳	𐩴	𐩵	𐩶	𐩷	𐩸	𐩹	𐩺	𐩻	𐩼	𐩽	𐩾	𐩿	𐪀	𐪁	𐪂
ETHI	ሀ	ለ	ሐ	መ	ቀ	ወ	ሠ	ረ	ቤ	ተ	ሰ	ከ	ነ	ኅ	ጸ	—	ፈ	አ	ዐ	ፀ	ገ	ደ	—	ጠ	ዘ	—	የ	—	—	

3.2. Order for sorting. Modern scholars specializing in Old North Arabian prefer an alphabetical order based on the Arabic order; some other Semiticists prefer the Northwest Semitic order. We recommend that the Unicode Collation Algorithm and ISO/IEC 14651 default to the former, but we give both here:

UCA	'	b	t	ṯ	g	ḥ	ḥ	d	ḏ	r	z	s ¹	s ²	s ³	ṣ	ḍ	ṭ	ṣ	'	ḡ	f	q	k	l	m	n	h	w	y
	𐩙	𐩏	𐩐	𐩑	𐩒	𐩓	𐩔	𐩕	𐩖	𐩗	𐩘	𐩙	𐩚	𐩛	𐩜	𐩝	𐩞	𐩟	𐩠	𐩡	𐩢	𐩣	𐩤	𐩥	𐩦	𐩧	𐩨	𐩩	𐩪
	'	b	g	d	ḏ	h	w	z	ḥ	ḥ	ṯ	ṣ	y	k	l	m	n	s ¹	'	ḡ	f	ṣ	ḍ	q	r	s ²	s ³	t	ṯ
	𐩙	𐩏	𐩐	𐩑	𐩒	𐩓	𐩔	𐩕	𐩖	𐩗	𐩘	𐩙	𐩚	𐩛	𐩜	𐩝	𐩞	𐩟	𐩠	𐩡	𐩢	𐩣	𐩤	𐩥	𐩦	𐩧	𐩨	𐩩	𐩪

4. Character names. Although “Ancient North Arabian” has become quite common among Semiticists for this script, the term “Old North Arabian” is still current amongst students of writing systems. Since the UCS tends to use OLD in character names, the script has been called OLD NORTH ARABIAN here, parallel to the already-encoded OLD SOUTH ARABIAN. Some sources involved with writing systems use the term “North Arabic” but this term is misleading; the Old North Arabian dialects and the Old South Arabian languages are not “Arabic” (which is a language not a language family).

Character names have been based on UCS names for Arabic letters, with some alterations to reflect pronunciation. The letter 𐩛 [g] is written GEEM rather than JEEM [dʒ], the letter 𐩛 (which may have been pronounced [ḡ]) is written ES-2, the letter 𐩙 (which may have been pronounced [ʃ]) is written ES-1, and the letter 𐩛 (which may have been pronounced [s]) is written ES-3. This reflects their typical

transliteration: $\aleph s^2$, $\aleph s^1$, and $\aleph s^3$. This last character is found only in the Taymanitic variety of Old North Arabian.

5. Numbers. Three numbers are attested in Old North Arabian: 1 (I, a unit), 10 (\diamond , possibly derived from AYN for $\aleph s^2$ ‘šr ‘10’), and 20 (\aleph , two 10s stacked). The numbers have right-to-left directionality. In the chart below, the third column is displayed in visual order. This table is a reconstruction; the only numbers actually attested are marked with an asterisk (1, 3, 5, 9, 10, 20, 30).

1*	I	1 ←	11	I \diamond	1 + 10 ←
2	II	1 + 1 ←	12	II \diamond	1 + 1 + 10 ←
3*	III	1 + 1 + 1 ←	13	III \diamond	1 + 1 + 1 + 10 ←
4	IIII	1 + 1 + 1 + 1 ←	14	IIII \diamond	1 + 1 + 1 + 1 + 10 ←
5*	IIIII	1 + 1 + 1 + 1 + 1 ←	15	IIIII \diamond	etc
6	IIIIII	1 + 1 + 1 + 1 + 1 + 1 ←	16	IIIIII \diamond	
7	IIIIIII	etc	17	IIIIIII \diamond	
8	IIIIIII		18	IIIIIII \diamond	
9*	IIIIIIII		19	IIIIIIII \diamond	
10*	\diamond	10 ←	20*	\aleph	20 ←
30*	$\diamond\aleph$	10 + 20 ←	40	$\aleph\aleph$	20 + 20 ←
50	$\diamond\aleph\aleph$	10 + 20 + 20 ←	60	$\aleph\aleph\aleph$	20 + 20 + 20 ←
70	$\diamond\aleph\aleph\aleph$	10 + 20 + 20 + 20 ←	80	$\aleph\aleph\aleph\aleph$	20 + 20 + 20 + 20 ←
90	$\diamond\aleph\aleph\aleph\aleph$	10 + 20 + 20 + 20 + 20 ←			

6. Punctuation. In Dadanitic, a vertical word separator is usually used between words, but this is not widely used in the other dialects. In form it is typically identical to OLD NORTH ARABIAN NUMBER ONE. A script-specific word separator has not been proposed here, as the character can do double duty here as in Old South Arabian

7. Glyph shapes. The glyph shapes used above and in the chart below reflect the monumental Dadanitic glyphs. As can be seen in Figure 2, the other alphabets differ considerably, but it is envisaged that these differences are best handled by dialect-specific fonts. This is similar to the model used for Old Italic.

8. Unicode Character Properties

```

10A80;OLD NORTH ARABIAN LETTER HEH;Lo;0;R;;;;;N;;;;;
10A81;OLD NORTH ARABIAN LETTER LAM;Lo;0;R;;;;;N;;;;;
10A82;OLD NORTH ARABIAN LETTER HAH;Lo;0;R;;;;;N;;;;;
10A83;OLD NORTH ARABIAN LETTER MEEM;Lo;0;R;;;;;N;;;;;
10A84;OLD NORTH ARABIAN LETTER QAF;Lo;0;R;;;;;N;;;;;
10A85;OLD NORTH ARABIAN LETTER WAW;Lo;0;R;;;;;N;;;;;
10A86;OLD NORTH ARABIAN LETTER ES-2;Lo;0;R;;;;;N;;;;;
10A87;OLD NORTH ARABIAN LETTER REH;Lo;0;R;;;;;N;;;;;
10A88;OLD NORTH ARABIAN LETTER BEH;Lo;0;R;;;;;N;;;;;
10A89;OLD NORTH ARABIAN LETTER TEH;Lo;0;R;;;;;N;;;;;
10A8A;OLD NORTH ARABIAN LETTER ES-1;Lo;0;R;;;;;N;;;;;
10A8B;OLD NORTH ARABIAN LETTER KAF;Lo;0;R;;;;;N;;;;;
10A8C;OLD NORTH ARABIAN LETTER NOON;Lo;0;R;;;;;N;;;;;
10A8D;OLD NORTH ARABIAN LETTER KHAH;Lo;0;R;;;;;N;;;;;
10A8E;OLD NORTH ARABIAN LETTER SAD;Lo;0;R;;;;;N;;;;;
10A8F;OLD NORTH ARABIAN LETTER ES-3;Lo;0;R;;;;;N;;;;;
10A90;OLD NORTH ARABIAN LETTER FEH;Lo;0;R;;;;;N;;;;;
10A91;OLD NORTH ARABIAN LETTER ALEF;Lo;0;R;;;;;N;;;;;
10A92;OLD NORTH ARABIAN LETTER AIN;Lo;0;R;;;;;N;;;;;
10A93;OLD NORTH ARABIAN LETTER DAD;Lo;0;R;;;;;N;;;;;
10A94;OLD NORTH ARABIAN LETTER GEEM;Lo;0;R;;;;;N;;;;;
10A95;OLD NORTH ARABIAN LETTER DAL;Lo;0;R;;;;;N;;;;;
10A96;OLD NORTH ARABIAN LETTER GHAIN;Lo;0;R;;;;;N;;;;;
10A97;OLD NORTH ARABIAN LETTER TAH;Lo;0;R;;;;;N;;;;;
10A98;OLD NORTH ARABIAN LETTER ZAIN;Lo;0;R;;;;;N;;;;;
10A99;OLD NORTH ARABIAN LETTER THAL;Lo;0;R;;;;;N;;;;;
10A9A;OLD NORTH ARABIAN LETTER YEH;Lo;0;R;;;;;N;;;;;

































```

10A9B;OLD NORTH ARABIAN LETTER THEH;Lo;0;R;;;N;;;;;
10A9C;OLD NORTH ARABIAN LETTER ZAH;Lo;0;R;;;N;;;;;
10A9D;OLD NORTH ARABIAN NUMBER ONE;No;0;R;;;1;N;;;;;
10A9E;OLD NORTH ARABIAN NUMBER TEN;No;0;R;;;10;N;;;;;
10A9F;OLD NORTH ARABIAN NUMBER TWENTY;No;0;R;;;20;N;;;;;

9. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Old North Arabian encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

10. Bibliography

- Caskel, Werner. 1954. *Lihyan und Lihyanisch*. (Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen, Geisteswissenschaften, 4). Cologne: Westdeutscher Verlag.
- Macdonald, Michael C. A. 2004. "Ancient North Arabian", in Roger D. Woodard, ed. *The Cambridge Encyclopedia of the World's Ancient Languages*. Cambridge: Cambridge University Press. ISBN 0-521-56256-2
- Macdonald, Michael C. A. 2009. *Literacy and Identity in Pre-Islamic Arabia*. Farnham: Ashgate.
- Macdonald, Michael C. A. *forthcoming a*. "Towards a re-assessment of the Ancient North Arabian alphabets used in the oasis of al-‘Ulā."
- Macdonald, Michael C. A. *forthcoming b*. "The South Semitic scripts used in Arabia". In Peter T. Daniels (ed.), *The World's Writing Systems*. (2nd edition). Oxford: Oxford University Press.
- O'Connor, M. 1996. "Epigraphic Semitic scripts" in *The World's Writing Systems*, ed. Peter T. Daniels & William Bright. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0

	10A8	10A9
0	 10A80	 10A90
1	 10A81	 10A91
2	 10A82	 10A92
3	 10A83	 10A93
4	 10A84	 10A94
5	 10A85	 10A95
6	 10A86	 10A96
7	 10A87	 10A97
8	 10A88	 10A98
9	 10A89	 10A99
A	 10A8A	 10A9A
B	 10A8B	 10A9B
C	 10A8C	 10A9C
D	 10A8D	 10A9D
E	 10A8E	 10A9E
F	 10A8F	 10A9F

Letters

10A80	𐤇	OLD NORTH ARABIAN LETTER HEH • transliterated as h → 1200 ሀ ethiopic syllable ha → 10A60 𐤇 old south arabian letter ha
10A81	𐤈	OLD NORTH ARABIAN LETTER LAM • transliterated as l → 1208 ለ ethiopic syllable la → 10A61 𐤈 old south arabian letter lamedh
10A82	𐤉	OLD NORTH ARABIAN LETTER HAH • transliterated as 1E25 ሀ → 1210 ሐ ethiopic syllable hha → 10A62 𐤉 old south arabian letter heth
10A83	𐤊	OLD NORTH ARABIAN LETTER MEEM • transliterated as m → 1218 መ ethiopic syllable ma → 10A63 𐤊 old south arabian letter mem
10A84	𐤋	OLD NORTH ARABIAN LETTER QAF • transliterated as q → 1240 ቀ ethiopic syllable qa → 10A64 𐤋 old south arabian letter qoph
10A85	𐤌	OLD NORTH ARABIAN LETTER WAW • transliterated as w → 12C8 ወ ethiopic syllable wa → 10A65 𐤌 old south arabian letter waw
10A86	𐤍	OLD NORTH ARABIAN LETTER ES-2 • transliterated as s2 → 1220 ሠ ethiopic syllable sza → 10A66 𐤍 old south arabian letter shin
10A87	𐤎	OLD NORTH ARABIAN LETTER REH • transliterated as r → 1282 ረ ethiopic syllable ra → 10A67 𐤎 old south arabian letter resh
10A88	𐤏	OLD NORTH ARABIAN LETTER BEH • transliterated as b → 1260 በ ethiopic syllable ba → 10A68 𐤏 old south arabian letter beth
10A89	𐤐	OLD NORTH ARABIAN LETTER TEH • transliterated as t → 1270 ተ ethiopic syllable ta → 10A69 𐤐 old south arabian letter taw
10A8A	𐤑	OLD NORTH ARABIAN LETTER ES-1 • transliterated as s1 → 1230 ሰ ethiopic syllable sa → 10A6A 𐤑 old south arabian letter sat
10A8B	𐤒	OLD NORTH ARABIAN LETTER KAF • transliterated as k → 12A8 ከ ethiopic syllable ka → 10A6B 𐤒 old south arabian letter kaph
10A8C	𐤓	OLD NORTH ARABIAN LETTER NOON • transliterated as n → 1290 ከ ethiopic syllable na → 10A6C ከ old south arabian letter nun
10A8D	𐤔	OLD NORTH ARABIAN LETTER KHAH • transliterated as 1E2B ከ → 1280 ከ ethiopic syllable xa → 10A6D 𐤔 old south arabian letter kheth
10A8E	𐤕	OLD NORTH ARABIAN LETTER SAD • transliterated as 1E63 ሠ → 1338 ሠ ethiopic syllable tsa → 10A6E 𐤕 old south arabian letter sadhe

10A8F	𐤖	OLD NORTH ARABIAN LETTER ES-3 • transliterated as s3 → 10A6F 𐤖 old south arabian letter samekh
10A90	𐤗	OLD NORTH ARABIAN LETTER FEH • transliterated as f → 1348 ፈ ethiopic syllable fa → 10A70 𐤗 old south arabian letter
10A91	𐤘	OLD NORTH ARABIAN LETTER ALEF • transliterated as 02BC → 12A0 አ ethiopic syllable glottal a → 10A71 ሐ old south arabian letter alef
10A92	𐤙	OLD NORTH ARABIAN LETTER AIN • transliterated as 02BD → 12D0 ሐ ethiopic syllable pharyngeal a → 10A72 ሐ old south arabian letter ayn
10A93	𐤚	OLD NORTH ARABIAN LETTER DAD • transliterated as 1E0D ፈ → 1340 ፈ ethiopic syllable tza → 10A73 ፈ old south arabian letter dhadhe
10A94	𐤛	OLD NORTH ARABIAN LETTER GEEM • transliterated as g → 1308 ገ ethiopic syllable ga → 10A74 ገ old south arabian letter gimel
10A95	𐤜	OLD NORTH ARABIAN LETTER DAL • transliterated as d → 12F0 ደ ethiopic syllable da → 10A75 ደ old south arabian letter daleth
10A96	𐤝	OLD NORTH ARABIAN LETTER GHAIN • transliterated as 0121 ገ → 10A76 ገ old south arabian letter ghayn
10A97	𐤞	OLD NORTH ARABIAN LETTER TAH • transliterated as 1E6D ተ → 1320 ጠ ethiopic syllable tha → 10A77 ጠ old south arabian letter teth
10A98	𐤟	OLD NORTH ARABIAN LETTER ZAIN • transliterated as z → 12D8 ረ ethiopic syllable za → 10A78 ረ old south arabian letter zayn
10A99	𐤠	OLD NORTH ARABIAN LETTER THAL • transliterated as 1E0F ፈ → 10A79 ረ old south arabian letter dhalet
10A9A	𐤡	OLD NORTH ARABIAN LETTER YEH • transliterated as y → 12E8 የ ethiopic syllable ya → 10A7A የ old south arabian letter yodh
10A9B	𐤢	OLD NORTH ARABIAN LETTER THEH • transliterated as 1E6F ተ → 10A7B ፈ old south arabian letter thaw
10A9C	𐤣	OLD NORTH ARABIAN LETTER ZAH • transliterated as 1E93 ረ → 10A7C ረ old south arabian letter theth

Numbers

10A9D	𐤤	OLD NORTH ARABIAN NUMBER ONE
10A9E	𐤥	OLD NORTH ARABIAN NUMBER TEN
10A9F	𐤦	OLD NORTH ARABIAN NUMBER TWENTY