Proposal to encode South Arabian Script

Requestors: Sultan Maktari, Kamal Mansour

30 July 2007

Historical Background

There is abundant evidence that South Arabian script was used not only in the southwest corner of the Arabian Peninsula (modern-day Yemen), but actually in the entire Peninsula before the Islamic era. Besides, samples of South Arabian Script have been found as far as Uruk in Mesopotamia, Delos in Greece, as well as Giza in Egypt. Archaeological finds show that the Script was used as far back as the 8th century BCE. It was used in trade, religious writing, and in civil records. Following the spread of Islam in Yemen, the decline of South Arabian Script began in the 6th century CE as it was gradually supplanted by Arabic script.

South Arabian Script was typically known by the name of the then-dominant peoples in the Southern Peninsula. At various times, it was known as Sabaen, Qatabani, Hadramite, among others. Although it was used for a variety of languages, South Arabian Script is most strongly associated with the Hamyiritic language in which it was simply known as "Musnad", i.e. *writing*. Many Peninsular languages borrowed Musnad before introducing their own changes. Prime examples are the Thamudic, Safavadic, Lyhiani scripts which eventually developed into independent scripts.

South Arabian inscriptions typically reveal a right-to-left directionality, although there are also many cases of alternating directions, known as *boustrophedon* writing (see example in figure 1). As with other scripts that alternate direction—such as Old Italic—glyph forms must be mirrored when rendering text that does not have the default directionality.

Musnad inscriptions are found primarily on stone and ceramic material, as well as metallic surfaces. Hundreds of artifacts have been found and, to this day, continue to be discovered. Some of the best examples can be seen in figures 1–4. Figure 1—from the dust jacket of *The Cambridge Encyclopedia of the World's Ancient Languages*—shows a Sabaen inscription. Figures 2–4 depict various artifacts found in the Museum of Yemeni History in Sanaa, Yemen.

South Arabian inscriptions remained obscure for many centuries until two Orientalists, Gesenius and Rödiger, made great strides towards deciphering them in the 19th century.

The westward migration of the Sabaean people into the Horn of Africa introduced the South Arabic consonantal alphabet into the region. The transplanted script formed the roots of the Geez script of Ethiopia which, in time and under presumably external influences, developed into a rich syllabary unlike any other Semitic script in history. Even a cursory examination of the letter forms of Modern Ethiopic writing reveal a striking similarity to South Arabian Script.

Character Repertoire

The character repertoire of Musnad corresponds to the repertoire of Classical Arabic, with the one exception of an additional letter, Samekh, that is also evident in West Semitic alphabets such as Hebrew and Phoenician. With striking clarity, two independent manuscripts—one dated 1452 CE from Arabia, the other 18th century CE from India— show the presumably equivalent letters of Arabic script and Musnad. Nowadays, Semitic scholars are generally in agreement about the South Arabian character repertoire, though no one has yet to discover the original names of the letters. The following extract from Daniels & Bright shows the two mentioned manuscripts that helped Rödiger in deciphering South Arabian script, along with an annotation by Daniels.

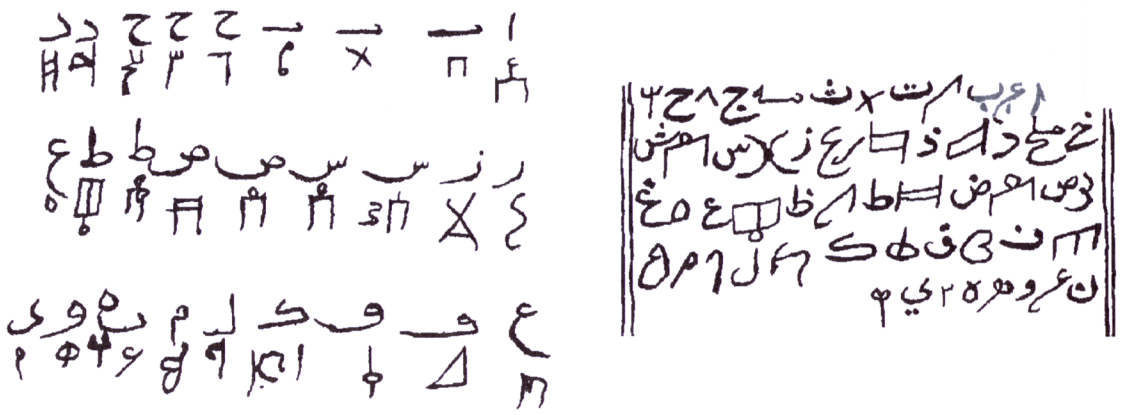


FIGURE 11. Supposed Himyaritic alphabets found in two unrelated manuscripts in the Königliche Bibliothek, Berlin (*left,* ms. 110, Arabic, written in Arabia, dated 856 A.H./1452 C.E.; *right,* ms. 248, Persian, written in India, early 18th century). Arabic key and South Arabian letters (*right to left*):

ي و ه ن م ل ك ق ف غ ع ظ ط ض ص ش س ز ر ذ د خ ح ج ث ت ب ا

(Rödiger 1837, plate).

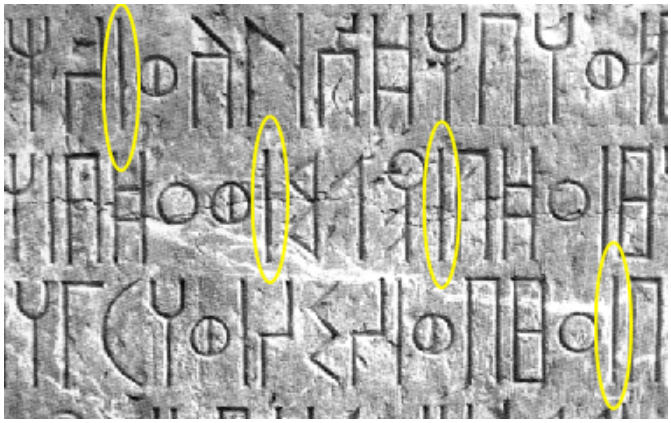
Although the derivation of Musnad is unclear, the difference of seven letters between its repertoire and that of Phoenician makes any direct parentage between the two unlikely. Musnad letters are all consonantal, and unlike other Semitic scripts, it never developed any vocalic notation. There is no evidence of any kind of diacritic marks. Geminate consonants, for instance, are made evident simply by writing the corresponding letter twice. Although inscriptions have been found in varying styles, Musnad letters are always written disjointedly and never developed any connected forms.

Scholars of South Arabian are in agreement about the phonemic repertoire represented by Musnad script. The following extract from Nebes & Stein shows the standard repertoire of 29 consonantal letters matching that of Daniels.

Table 15.1 The Ancient South Arabian consonantal script

Character	Transcription	Character	Transcription
𐩱	h	𐩸	s_3
𐩲	l	𐩹	f
𐩣	ḥ	𐩺	ʾ
𐩥	m	𐩠	ʿ
𐩧	q	𐩵	ḍ
𐩦	w	𐩢	g
𐩬	s_2	𐩴	d
𐩮	r	𐩶	ġ
𐩫	b	𐩷	ṭ
𐩤	t	𐩻	z
𐩪	s_1	𐩽	ḏ
𐩨	k	𐩯	y
𐩭	n	𐩰	ṯ
𐩩	ḫ	𐩼	ẓ
𐩡	ṣ		

While Musnad inscriptions do not separate words with white space, word boundaries are clearly indicated with a vertical bar. In the following illustration, some word separators are enclosed in an oval.



Alphabetic Order

The order of the letters of the Musnad alphabet is not known, though often it is assumed to resemble that of Ethiopic writing since the latter descended directly from it. Similarly, the names of the letters are uncertain. Since the Musnad character repertoire is closest to the Arabic one, it is equally reasonable to use Arabic alphabetic order, for even Ethiopic writing uses Arabic order for some specific types of enumerated lists. This proposal advocates the Arabic order for Musnad, with the Samekh falling between letters Seen and Sheen.

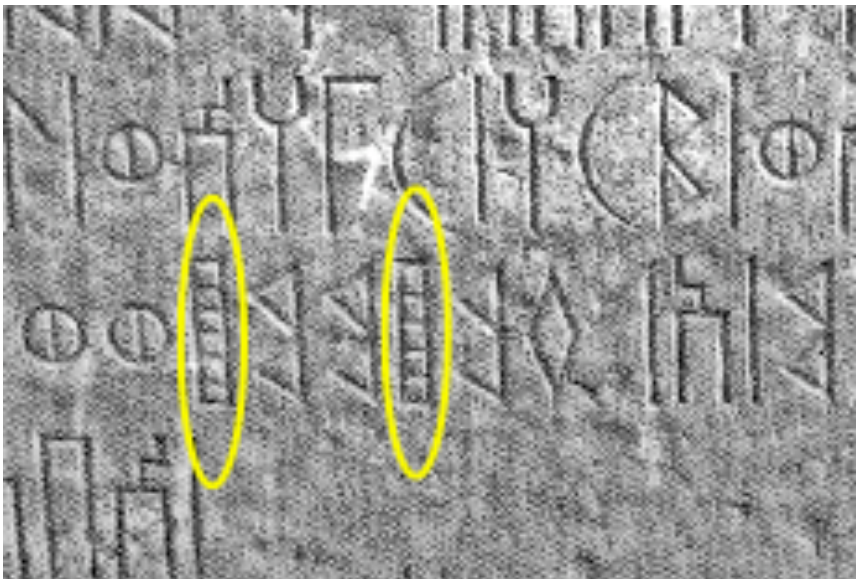
1	Alef	
2	Beh	
3	Teh	
4	Theh	
5	Jeem	
6	Hah	
7	Khah	
8	Dal	
9	Thal	
10	Reh	
11	Zain	
12	Seen	
13	Samekh	
14	Sheen	
15	Sad	
16	Dad	
17	Tah	
18	Zah	
19	Ain	
20	Ghain	
21	Feh	
22	Qaf	
23	Kaf	
24	Lam	
25	Meem	
26	Noon	
27	Heh	
28	Waw	
29	Yeh	

Figures
Using a set of six graphemes, all numbers can be written through juxtaposition in a manner similar to that of Roman numerals. Musnad has no symbol for zero. Following are the essential graphemes alongside their meaning:

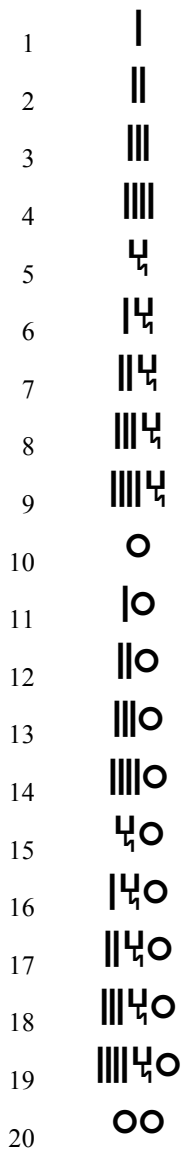
1	𐩱
5	𐩲
10	𐩳
50	𐩴
100	𐩵
1000	𐩶

It is interesting to note that four of the six symbols—5, 10, 100, 1000—are identical to alphabetic letters, the exceptions being 1 and 50. In fact, each of the names of the four numbers starts with the respective letter used to represent them. For instance, *khamsa* (five) starts with the letter Kha, while *'ashara* (ten) begins with Ain. The single stroke used to represent 1 is not alphabetic, but is intuitively clear. The symbol for 50 is thought to have a peculiar derivation. Since the initial letter of *khamseen* (fifty) matches that of *khamsa*, the letter Kha could not be used for it. Instead, by slicing off the lower protrusion of the Meem representing 100, the symbol for 50 intuitively represents half of 100.

Numeric quantities are clearly differentiated from surrounding text with a cross-hatched rectangle. In the following sample, the numeric indicators—highlighted with an oval— surround the number 200. As evident in this inscription, the number of horizontal lines that crisscross the rectangle can vary; in addition, the lines can also be diagonal instead. The numeric indicator and the word separator constitute the only form of punctuation seen in Musnad writing.



Using the above-mentioned graphemes, all numeric values can be created. Unlike Roman numerals, Musnad numbers show an evident right-to-left directionality. The numerals from 1 through 20 are written as follows:



Proposed Characters

1xx00	MUSNAD LETTER ALEF	𐩠
1xx01	MUSNAD LETTER BEH	𐩡
1xx02	MUSNAD LETTER TEH	𐩢
1xx03	MUSNAD LETTER THEH	𐩣
1xx04	MUSNAD LETTER JEEM	𐩤
1xx05	MUSNAD LETTER HAH	𐩥
1xx06	MUSNAD LETTER KHAH	𐩦
1xx07	MUSNAD LETTER DAL	𐩧
1xx08	MUSNAD LETTER THAL	𐩨
1xx09	MUSNAD LETTER REH	𐩩
1xx0A	MUSNAD LETTER ZAIN	𐩪
1xx0B	MUSNAD LETTER SEEN	𐩫
1xx0C	MUSNAD LETTER SAMEKH	𐩬
1xx0D	MUSNAD LETTER SHEEN	𐩭
1xx0E	MUSNAD LETTER SAD	𐩮
1xx0F	MUSNAD LETTER DAD	𐩯
1xx10	MUSNAD LETTER TAH	𐩰
1xx11	MUSNAD LETTER ZAH	𐩱
1xx12	MUSNAD LETTER AIN	𐩲
1xx13	MUSNAD LETTER GHAIN	𐩳
1xx14	MUSNAD LETTER FEH	𐩴
1xx15	MUSNAD LETTER QAF	𐩵
1xx16	MUSNAD LETTER KAF	𐩶
1xx17	MUSNAD LETTER LAM	𐩷
1xx18	MUSNAD LETTER MEEM	𐩸
1xx19	MUSNAD LETTER NOON	𐩹
1xx1A	MUSNAD LETTER HEH	𐩺
1xx1B	MUSNAD LETTER WAW	𐩻
1xx1C	MUSNAD LETTER YEH	𐩼
1xx1D	MUSNAD WORD SEPARATOR	𐩽
1xx1E	MUSNAD NUMERIC INDICATOR	𐩾
1xx1F	MUSNAD NUMBER ONE	𐩿
1xx20	MUSNAD NUMBER FIFTY	𐪀

Although the glyph for word separator (1xx1D) is similar to the One (1xx1F), it is important to encode them both because the first is a punctuation mark while the second is numeric.

While the glyph shapes shown above are typical, there was some degree of variation in the writing of Musnad characters. In figure 5, the column labeled ‘Sabean’ shows the common shape variants for various letters.

Unicode Properties of Proposed Characters

```
1xx00;MUSNAD LETTER ALEF;Lo;0;R;;;;;N;;;;;
1xx01;MUSNAD LETTER BEH;Lo;0;R;;;;;N;;;;;
1xx02;MUSNAD LETTER TEH;Lo;0;R;;;;;N;;;;;
1xx03;MUSNAD LETTER THEH;Lo;0;R;;;;;N;;;;;
1xx04;MUSNAD LETTER JEEM;Lo;0;R;;;;;N;;;;;
1xx05;MUSNAD LETTER HAH;Lo;0;R;;;;;N;;;;;
1xx06;MUSNAD LETTER KHAH;Lo;0;R;;;;;N;;;;;
1xx07;MUSNAD LETTER DAL;Lo;0;R;;;;;N;;;;;
1xx08;MUSNAD LETTER THAL;Lo;0;R;;;;;N;;;;;
1xx09;MUSNAD LETTER REH;Lo;0;R;;;;;N;;;;;
1xx0A;MUSNAD LETTER ZAIN;Lo;0;R;;;;;N;;;;;
1xx0B;MUSNAD LETTER SEEN;Lo;0;R;;;;;N;;;;;
1xx0C;MUSNAD LETTER SAMEKH;Lo;0;R;;;;;N;;;;;
1xx0D;MUSNAD LETTER SHEEN;Lo;0;R;;;;;N;;;;;
1xx0E;MUSNAD LETTER SAD;Lo;0;R;;;;;N;;;;;
1xx0F;MUSNAD LETTER DAD;Lo;0;R;;;;;N;;;;;
1xx10;MUSNAD LETTER TAH;Lo;0;R;;;;;N;;;;;
1xx11;MUSNAD LETTER ZAH;Lo;0;R;;;;;N;;;;;
1xx12;MUSNAD LETTER AIN;Lo;0;R;;;;;N;;;;;
1xx13;MUSNAD LETTER GHAIN;Lo;0;R;;;;;N;;;;;
1xx14;MUSNAD LETTER FEH;Lo;0;R;;;;;N;;;;;
1xx15;MUSNAD LETTER QAF;Lo;0;R;;;;;N;;;;;
1xx16;MUSNAD LETTER KAF;Lo;0;R;;;;;N;;;;;
1xx17;MUSNAD LETTER LAM;Lo;0;R;;;;;N;;;;;
1xx18;MUSNAD LETTER MEEM;Lo;0;R;;;;;N;;;;;
1xx19;MUSNAD LETTER NOON;Lo;0;R;;;;;N;;;;;
1xx1A;MUSNAD LETTER HEH;Lo;0;R;;;;;N;;;;;
1xx1B;MUSNAD LETTER WAW;Lo;0;R;;;;;N;;;;;
1xx1C;MUSNAD LETTER YEH;Lo;0;R;;;;;N;;;;;
1xx1D;MUSNAD WORD SEPARATOR;Po;0;ON;;;;;N;;;;;
1xx1E;MUSNAD NUMERIC INDICATOR;Po;0;ON;;;;;N;;;;;
1xx1F;MUSNAD NUMBER ONE;No;0;R;;;;1;N;;;;;
1xx20;MUSNAD NUMBER FIFTY;No;0;R;;;;50;N;;;;;
```

Line Breaking

Similarly to Arabic script, lines of Musnad text are never broken in the middle of words.

References

د. جواد علي. المفصّل في تاريخ العرب قبل الإسلام. دار العلم للملايين.(١٠ مجلدات). بيروت. ١٩٨٧ م .
Jawwad Ali. Detailed History of the Arabs before Islam. Dâr-il-'Ilm lil-Malayîn. [10 volumes]. (in Arabic). Beirut, 1987.

G. van Beek & Albert Jamme. *An Inscribed South Arabian Clay Stamp From Bethel*. 1958.

A.F.L. Beeston. *A descriptive grammar of epigraphic South Arabian*. London, Luzac, 1962.

A.F.L. Beeston. *Four Sabaean Texts in the Istanbul Archaeological Museum*. 1952.

Florian Coulmas. *The Blackwell Encyclopedia of Writing Systems*. Blackwell Publishers. 1996.

Peter T. Daniels. "Methods of Decipherment" in Peter T. Daniels & William Bright, eds. *The World's Writing Systems.* Oxford University Press, 1996.

Albert Jamme. *Sabaean inscriptions from Ma·hram Bilqîs (Mârib)* [with foreword by Wendell Phillips]. 19??.

Norbert Nebes & Peter Stein. "Ancient South Arabian" in Roger D. Woodard, ed. *The Cambridge Encyclopedia of the World's Ancient Languages*. Cambridge University Press. 2004.

Wendell Philips. *Qataban and Sheba*. Victor Gollancz Ltd, London. 1955.

Museum of Yemeni History, Sanaa, Yemen

Smithsonian Institution. "Written in Stone: Inscriptions from the National Museum of Saudi Arabic". http://www.mnh.si.edu/epigraphy/figs-stones/x-large/color_xl_jpeg/fig02.jpg

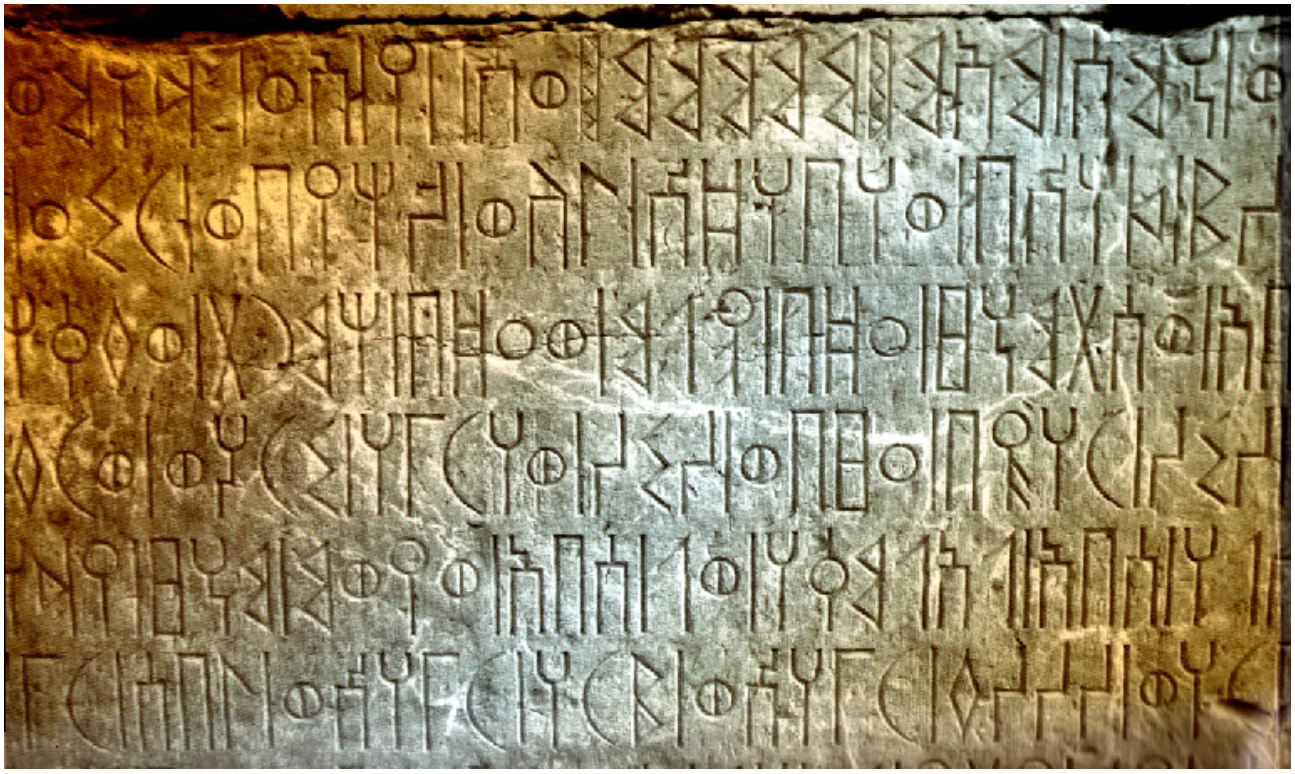


Figure 1

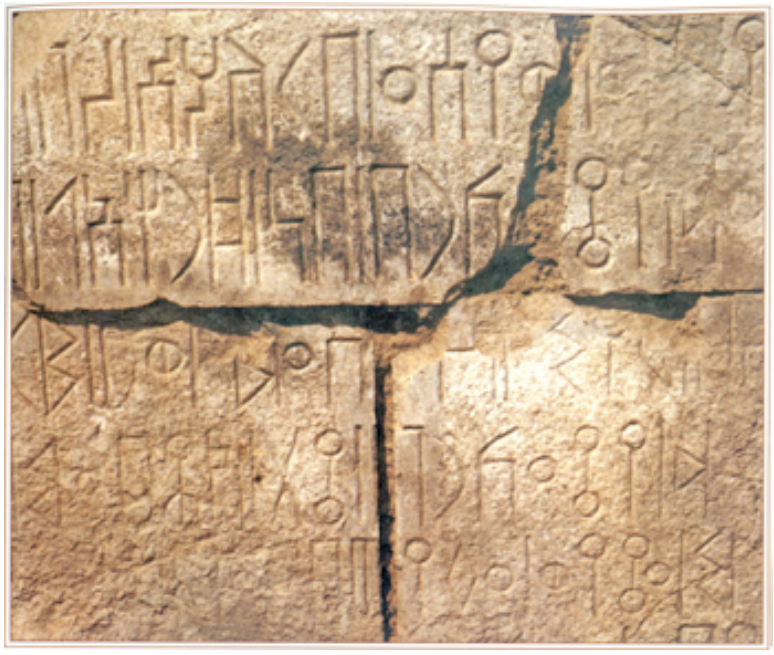


Figure 2

Figure 3

Figure 4

Comparative chart of Ancient Arabian Scripts.

Figure 5