ISO/IEC JTC1/SC2/WG2 N3296
L2/07-287
2007-08-29

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

Doc Type: **Working Group Document**
Title: **Towards a proposal to encode the Old South Arabian script in the SMP of the UCS**
Source: **Michael Everson and Daniel Yacob**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Replaces: **N1689**
Date: **2007-08-29**

1. Introduction. The Old South Arabian script was first proposed to WG2 by Michael Everson in N1689 (1998-01-18); more recently Sultan Maktari and Kamal Mansour proposed the script again in L2/07-240 (2007-07-30). This document responds to Maktari and Mansour's contribution and makes recommendations for a character set that could be accepted for encoding in the UCS.

2. Processing. Old South Arabian is an alphabetic script consisting only of consonants; vowels are not indicated in any way in the script. It is typically written right-to-left, but boustrophedon writing is also known. Glyphs are mirrored in lines when they have left-to-right directionality.

3. Ordering and character names. Maktari and Mansour propose character names and ordering based on the Arabic script. While we do recognize that some scholars use Arabic characters for transliteration of Old South Arabian, it is also the case that some scholars use Hebrew and Latin for the same. We consider that, in the context of the UCS, the use of Arabic or Hebrew for character names or ordering would be much more anachronistic than the use of the much more closely-related Ethiopic would be.

Old South Arabian script is the ancestor of Ethiopic, and there is a close relation between the alphabetical order of the two scripts. A. M. Honeyman wrote in 1952 of the Old South Arabian alphabet that "the order of its letters is substantially that of the Ethiopic syllabary and the alphabetic archetype thereof". An inscription discovered at the Oasis of Al-'Ula in 1914 was identified as an abecedary by Honeyman 1952. Ryckmans 1985 gives the order for Old South Arabian as follows (transliteration, Old South Arabian script, Old Gə'əz script, and modern Ethiopic are shown):

h	l	ḥ	m	q	w	š	r	b	t	s	k	n	ḫ	ś	f	’	‘	ḍ	g	d	ġ	ṭ	z	ḏ	y	ṯ	ṣ	ẓ
𐩱	𐩻	𐩧	𐩥	𐩲	𐩵	𐩦	𐩺	𐩢	𐩬	𐩮	𐩫	𐩯	𐩨	𐩩	𐩤	𐩠	𐩡	𐩣	𐩭	𐩪	𐩰	𐩬	𐩹	𐩪	𐩳	𐩬	𐩧	𐩩
ⱎ	ⱂ	ⱘ	ⱋ	ⱔ	ⱏ	ⱐ	ⱄ	ⱂ	ⱍ	ⱑ	ⱆ	ⱀ	–	ⱅ	ⱇ	ⱐ	ⱑ	ⱓ	ⱉ	ⱐ	–	ⱏ	ⱘ	–	ⱏ	–	ⱍ	–
ሀ	ለ	ሐ	መ	ቀ	ወ	ሠ	ረ	በ	ተ	ሰ	ከ	ነ	ኀ	–	ፈ	አ	ዐ	ፀ	ገ	ደ	–	ጠ	ዘ	–	የ	–	ጸ	–

Modern scholars distinguish this "South Semitic" order from the "Levantine" order used for the descendants of Phoenician. The order of the Old South Arabian characters in the UCS code charts should follow the South Semitic alphabetical order because it is attested for Old South Arabian (see Figure 6). For comparison, here is same set in the modern Ethiopic variant of the South Semitic order:

h	l	ḥ	m	š	r	s	q	b	t	ḫ	n	’	k	w	‘	z	y	d	g	ṭ	ṣ	ḍ	f	ś	ġ	ḏ	ṯ	ẓ
𐩣	𐩡	𐩧	𐩨	𐩪	𐩬	𐩮	𐩯	𐩠	𐩱	𐩲	𐩢	𐩤	𐩫	𐩵	𐩦	𐩭	𐩺	𐩥	𐩲	𐩳	𐩮	𐩴	𐩰	𐩷	𐩸	𐩹	𐩻	𐩼
𐩺	𐩡	𐩧	𐩨	𐩪	𐩬	𐩮	𐩯	𐩠	𐩱	𐩲	𐩢	𐩤	𐩫	𐩵	𐩦	𐩭	𐩺	𐩥	𐩲	𐩳	𐩮	𐩴	𐩰	𐩷	𐩸	𐩹	𐩻	𐩼
𐩣	𐩡	𐩧	𐩨	𐩪	𐩬	𐩮	𐩯	𐩠	𐩱	𐩲	𐩢	𐩤	𐩫	𐩵	𐩦	𐩭	𐩺	𐩥	𐩲	𐩳	𐩮	𐩴	𐩰	𐩷	𐩸	𐩹	𐩻	𐩼
𐩣	𐩡	𐩧	𐩨	𐩪	𐩬	𐩮	𐩯	𐩠	𐩱	𐩲	𐩢	𐩤	𐩫	𐩵	𐩦	𐩭	𐩺	𐩥	𐩲	𐩳	𐩮	𐩴	𐩰	𐩷	𐩸	𐩹	𐩻	𐩼

By the same token, Maktari and Mansour’s application of Arabic letter-names to the Old South Arabian script is also anachronistic in the context of the UCS. Although we have inherited no formal letter-names for Old South Arabian, good letter-names can be devised on the basis of Ethiopic, its closest relative. There are two choices. Option A would use the syllable-value of the corresponding Ethiopic base letter, but Option B, which we prefer, would use the older inherited Semitic letter-names. In the table below, we give the Ethiopic first-order syllables, their IPA value in Gə’əz (following Gragg 2004), their transliteration (following Haile 1996), their Gə’əz names, the Old South Arabian letters, their IPA equivalents (following Nebes and Stein 2004), and the two options for UCS character names, re-spelling the consonant values according to the usual UCS conventions.

Ethiopic	*IPA (Gə’əz)*	*Transl.*	*Name*	*OSA*	*IPA (OSA)*	*Option A*	*Option B*
ሀ	[h]	*ha*	*hoy*	𐩣	[h]	HA	HOY
ለ	[l]	*la*	*lawe*	𐩡	[l]	LA	LAWE
ሐ	[ħ]	*ḥa*	*ḥawt*	𐩧	[ħ]	HHA	HAWT
መ	[m]	*ma*	*māy*	𐩨	[m]	MA	MAAY
ቀ	[q]	*qa*	*qāf*	𐩯	[q]	QA	QAAF
ወ	[w]	*wa*	*wawe*	𐩵	[w]	WA	WAWE
ሠ	[ʃ]	*śa*	*śawt*	𐩦	[ɬ]	HLA	HLAWT
ረ	[r]	*ra*	*rə’s*	𐩭	[r]	RA	RES
በ	[b]	*ba*	*bet*	𐩠	[b]	BA	BET
ተ	[t]	*ta*	*tawe*	𐩱	[t]	TA	TAWE
ሰ	[s]	*sa*	*sāt*	𐩮	[s]	SA	SAAT
ከ	[k]	*ka*	*kāf*	𐩫	[k]	KA	KAAF
ነ	[n]	*na*	*nahās*	𐩢	[n]	NA	NAHAAS
ኀ	[x]	*ḫa*	*ḫarm*	𐩤	[x]	KHA	KHARM
—	—	—	—	𐩷	[s̱]	SSA	SAWT
ፈ	[f]	*fa*	*af*	𐩰	[f]	FA	AF
አ	[ʔ]	*’a*	*’alf*	𐩬	[ʔ]	ALF	ALF
ዐ	[ʕ]	*‘a*	*‘ayn*	𐩺	[ʕ]	AYN	AYN
ፀ	[x’]	*ḍa*	*ḍaṗṗā*	𐩴	[ɬ’]	HLLA	HLAFFAA
ገ	[g]	*ga*	*gaml*	𐩲	[g]	GA	GAML
ደ	[d]	*da*	*dant*	𐩥	[d]	DA	DANT
—	—	—	—	𐩳	[ɣ]	GHA	GHAAF
ጠ	[t’]	*ṭa*	*ṭayt*	𐩲	[t’]	TTA	TAYT
ዘ	[z]	*za*	*zay*	𐩹	[z]	ZA	ZAY
—	—	—	—	𐩸	[ð]	DHA	DHAY
የ	[j]	*ya*	*yaman*	𐩾	[j]	YA	YAMAN
—	—	—	—	𐩻	[θ]	THA	THAY
ጸ	[s’]	*ṣa*	*ṣaday*	𐩸	[tˢ’]	TSSA	TSADAY
—	—	—	—	𐩼	[θ’]	THHA	THADAY

Maktari and Mansour also suggest that the script be named “Musnad”, a term by which the script is known in the Arabic language. Since the normal name for the script in English is “Old South Arabian”, that term ought to be preferred. (Compare Höfner’s “Altsüdarabisch”.)

5. Numbers. Old South Arabian uses four of its letters with a numeric value, supplementing these with two explicit numbers. It builds up numbers out of 1 (𐩱, a unit), 5 (KHARM for 𐩧 *ḫms* ‘5’), 10 (AYN for 𐩲 *ʿšr* ‘10’), 50 (𐩬, half of 𐩣 *m*), 100 (MAAY for 𐩣𐩺 *m’t* ‘100’), and 1000 (ALF for 𐩠𐩵 *’lf* ‘1000’). Numbers are flanked on either side by the OLD SOUTH ARABIAN NUMERATOR BRACKET. When 10, 50, or 100 occur before 1000 they serve to indicate multiples of 1000. The numbers have right-to-left directionality. In the chart below, the third column is displayed in visual order.

1	𐩡𐩱𐩡	1 ←	11	𐩡𐩱𐩲𐩡	1 + 10 ←
2	𐩡𐩱𐩱𐩡	1 + 1 ←	12	𐩡𐩱𐩱𐩲𐩡	2 + 10 ←
3	𐩡𐩱𐩱𐩱𐩡	1 + 1 + 1 ←	13	𐩡𐩱𐩱𐩱𐩲𐩡	3 + 10 ←
4	𐩡𐩱𐩱𐩱𐩱𐩡	1 + 1 + 1 + 1 ←	14	𐩡𐩱𐩱𐩱𐩱𐩲𐩡	1 + 3 + 10 ←
5	𐩡𐩧𐩡	5 ←	15	𐩡𐩧𐩲𐩡	2 + 3 + 10 ←
6	𐩡𐩱𐩧𐩡	1 + 5 ←	16	𐩡𐩱𐩧𐩲𐩡	3 + 3 + 10 ←
7	𐩡𐩱𐩱𐩧𐩡	1 + 1 + 5 ←	17	𐩡𐩱𐩱𐩧𐩲𐩡	1 + 3 + 3 + 10 ←
8	𐩡𐩱𐩱𐩱𐩧𐩡	1 + 1 + 1 + 5 ←	18	𐩡𐩱𐩱𐩱𐩧𐩲𐩡	2 + 3 + 3 + 10 ←
9	𐩡𐩱𐩱𐩱𐩱𐩧𐩡	1 + 1 + 1 + 1 + 5 ←	19	𐩡𐩱𐩱𐩱𐩱𐩧𐩲𐩡	3 + 3 + 3 + 10 ←
10	𐩡𐩲𐩡	10 ←	100	𐩡𐩣𐩡	100 ←
20	𐩡𐩲𐩲𐩡	10 + 10 ←	200	𐩡𐩣𐩣𐩡	100 + 100 ←
30	𐩡𐩲𐩲𐩲𐩡	10 + 10 + 10 ←	300	𐩡𐩣𐩣𐩣𐩡	100 + 100 + 100 ←
40	𐩡𐩲𐩲𐩲𐩲𐩡	10 + 10 + 10 + 10 ←	400	𐩡𐩣𐩣𐩣𐩣𐩡	100 (x 4) ←
50	𐩡𐩬𐩡	50 ←	500	𐩡𐩣𐩣𐩣𐩣𐩣𐩡	100 (x 5) ←
60	𐩡𐩲𐩬𐩡	10 + 50 ←	600	𐩡𐩣𐩣𐩣𐩣𐩣𐩣𐩡	100 (x 6) ←
70	𐩡𐩲𐩲𐩬𐩡	10 + 10 + 50 ←	700	𐩡𐩣𐩣𐩣𐩣𐩣𐩣𐩣𐩡	100 (x 7) ←
80	𐩡𐩲𐩲𐩲𐩬𐩡	10 + 10 + 10 + 50 ←	800	𐩡𐩣𐩣𐩣𐩣𐩣𐩣𐩣𐩣𐩡	100 (x 8) ←
90	𐩡𐩲𐩲𐩲𐩲𐩬𐩡	10 + 10 + 10 + 10 + 50 ←	900	𐩡𐩣𐩣𐩣𐩣𐩣𐩣𐩣𐩣𐩣𐩡	100 (x 9) ←
1,000	𐩡𐩠𐩡	1,000 ←	3,000	𐩡𐩠𐩠𐩠𐩡	1,000 (x 3) ←
16,000	𐩡𐩠𐩠𐩠𐩠𐩠𐩠𐩲𐩡	1,000 (x 6) + 10 ←	31,000	𐩡𐩠𐩲𐩲𐩲𐩡	1,000 + 10 (x 3) ←
253,000	𐩡𐩠𐩠𐩠𐩬𐩣𐩣𐩡	1,000 (x 6) + 10 ←			

6. Issue: Punctuation. A vertical word separator is usually used between words in Old South Arabian. In form it is typically identical to OLD SOUTH ARABIAN NUMBER ONE, although in some contexts the word separator is represented as part of the text layout (see Figure 8). For this reason (and that of word-breaking) it is probably inappropriate to try to unify the word separator from the NUMBER ONE. (Doing so would also add another column to the Old South Arabian block.) One option would be to encode a generic WORD SEPARATOR WHITE SPACE VERTICAL BAR in the Supplemental Punctuation block at U+2E32 alongside a WORD SEPARATOR WHITE SPACE DOT since ancient scripts tend to use either a dot or bar and a generic character is wanting. On the other hand, it may not be possible to avoid a third column for Old South Arabian (see §7 below).

7. Issue: “Divine symbols”. Beeston 1962 describes some “divine symbols”, saying that these:

> … have a tendency to become formalized into patterns which resemble alphabetic letters. Thus the “lightning” symbol develops into a form resembling the letter *h* [i.e. 𐩭 HAWT], the “Doppelgriffel” into one resembling *d̲* [i.e. ‘double slate-pencil’, 𐩮 DHAY]. These symbolic “letters” are often placed at the beginning and/or end of texts (e.g. C 634), but have no phonetic significance.

C 634 can evidently be found in the *Corpus Incriptionum Semiticarum* IV; the other source cited in Beeston is: “Adolf Grohmann, ‘Göttersymbole und Symboltiere’, *Denkschr. d. Akad. d. Wiss. in Wien* Bd. 58 (1914), Abh. 1.” We have not yet seen either source.

8. Glyph shapes. Old South Arabian was fairly regular and well-standardized in antiquity. One problem is what the representative glyph shapes should be used in the code charts. In addition to the font shown in Maktari and Mansour's contribution, we show three additional fonts in the table below for consideration.

9. Unicode Character Properties

```
10840;OLD SOUTH ARABIAN LETTER HOY;Lo;0;R;;;;;N;;;;;
10841;OLD SOUTH ARABIAN LETTER LAWE;Lo;0;R;;;;;N;;;;;
10842;OLD SOUTH ARABIAN LETTER HAWT;Lo;0;R;;;;;N;;;;;
10843;OLD SOUTH ARABIAN LETTER MAAY;Lo;0;R;;;;;N;;;;;
10844;OLD SOUTH ARABIAN LETTER QAAF;Lo;0;R;;;;;N;;;;;
10845;OLD SOUTH ARABIAN LETTER WAWE;Lo;0;R;;;;;N;;;;;
10846;OLD SOUTH ARABIAN LETTER HLAWT;Lo;0;R;;;;;N;;;;;
10847;OLD SOUTH ARABIAN LETTER RES;Lo;0;R;;;;;N;;;;;
10848;OLD SOUTH ARABIAN LETTER BET;Lo;0;R;;;;;N;;;;;
10849;OLD SOUTH ARABIAN LETTER TAWE;Lo;0;R;;;;;N;;;;;
1084A;OLD SOUTH ARABIAN LETTER SAAT;Lo;0;R;;;;;N;;;;;
1084B;OLD SOUTH ARABIAN LETTER KAAF;Lo;0;R;;;;;N;;;;;
1084C;OLD SOUTH ARABIAN LETTER NAHAAS;Lo;0;R;;;;;N;;;;;
1084D;OLD SOUTH ARABIAN LETTER KHARM;Lo;0;R;;;;;N;;;;;
1084E;OLD SOUTH ARABIAN LETTER SAWT;Lo;0;R;;;;;N;;;;;
1084F;OLD SOUTH ARABIAN LETTER AF;Lo;0;R;;;;;N;;;;;
10850;OLD SOUTH ARABIAN LETTER ALF;Lo;0;R;;;;;N;;;;;
10851;OLD SOUTH ARABIAN LETTER AYN;Lo;0;R;;;;;N;;;;;
10852;OLD SOUTH ARABIAN LETTER HLAFFAA;Lo;0;R;;;;;N;;;;;
10853;OLD SOUTH ARABIAN LETTER GAML;Lo;0;R;;;;;N;;;;;
10854;OLD SOUTH ARABIAN LETTER DANT;Lo;0;R;;;;;N;;;;;
10855;OLD SOUTH ARABIAN LETTER GHAAF;Lo;0;R;;;;;N;;;;;
10856;OLD SOUTH ARABIAN LETTER TAYT;Lo;0;R;;;;;N;;;;;
10857;OLD SOUTH ARABIAN LETTER ZAY;Lo;0;R;;;;;N;;;;;
10858;OLD SOUTH ARABIAN LETTER DHAY;Lo;0;R;;;;;N;;;;;
10859;OLD SOUTH ARABIAN LETTER YAMAN;Lo;0;R;;;;;N;;;;;
1085A;OLD SOUTH ARABIAN LETTER THAY;Lo;0;R;;;;;N;;;;;
1085B;OLD SOUTH ARABIAN LETTER TSADAY;Lo;0;R;;;;;N;;;;;
1085C;OLD SOUTH ARABIAN LETTER THADAY;Lo;0;R;;;;;N;;;;;
1085D;OLD SOUTH ARABIAN NUMERATOR BRACKET;Lo;0;R;;;;;N;;;;;
1085E;OLD SOUTH ARABIAN NUMBER ONE;Lo;0;R;;;;;N;;;;;
1085F;OLD SOUTH ARABIAN NUMBER FIFTY;Lo;0;R;;;;;N;;;;;
```

10. Bibliography

Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8

Dillmann, August. 1907 (reprinted 1974). *Ethiopic grammar: orthography and phonology, morphology and syntax of the Ethiopic language, also in comparison with other Semitic languages with a general introduction and an index of passages*. Second edition, enlarged and improved, edited by Carl Bezold. Translated, with additions, by James A. Crichton. Amsterdam: Philo Press. ISBN 90-6022-271-7

Gragg, Gene. 2004. "Ge'ez (Aksum)", in Roger D. Woodard, ed. *The Cambridge Encyclopedia of the World's Ancient Languages*. Cambridge: Cambridge University Press. ISBN 0-521-56256-2

Haile, Getatchew. 1996. "Ethiopic writing" in *The World's Writing Systems*, ed. Peter T. Daniels & William Bright. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0

Honeyman, A. M. 1952. "The Letter-Order of the Semitic Alphabets in Africa and the Near East", in *Africa: Journal of the International African Institute, London* 22:136-47.

Kōno Rokurō, Chino Eiichi, & Nishida Tatsuo. 2001. *The Sanseido Encyclopaedia of Linguistics. Volume 7: Scripts and Writing Systems of the World* [*Gengogaku dai ziten* (*bekkan*) *sekai mozi ziten*]. Tokyo: Sanseido Press. ISBN 4-385-15177-6

Küster, Marc W. 2006. *Geordnetes Weltbild: Die Tradition des alphabetischen Sortierens von der Keilschrift bis zur EDV. Eine Kulturgeschichte*. Tübingen: Niemeyer. ISBN 3-484-10899-1

Nebes, Norbert, and Peter Stein. 2004. "Ancient South Arabian", in Roger D. Woodard, ed. *The Cambridge Encyclopedia of the World's Ancient Languages*. Cambridge: Cambridge University Press. ISBN 0-521-56256-2

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

Ryckmans, Jacques. 1985. “L’ordre alphabétique sud-sémitique et ses origines”, in Christian Robin, ed. *Mélanges linguistiques offerts à Maxime Rodinson*. Comptes rendus du Groupe Linguistique d’Études Chamito-Sémitiques. Supplément 12. Paris: Geuthner.

Figures

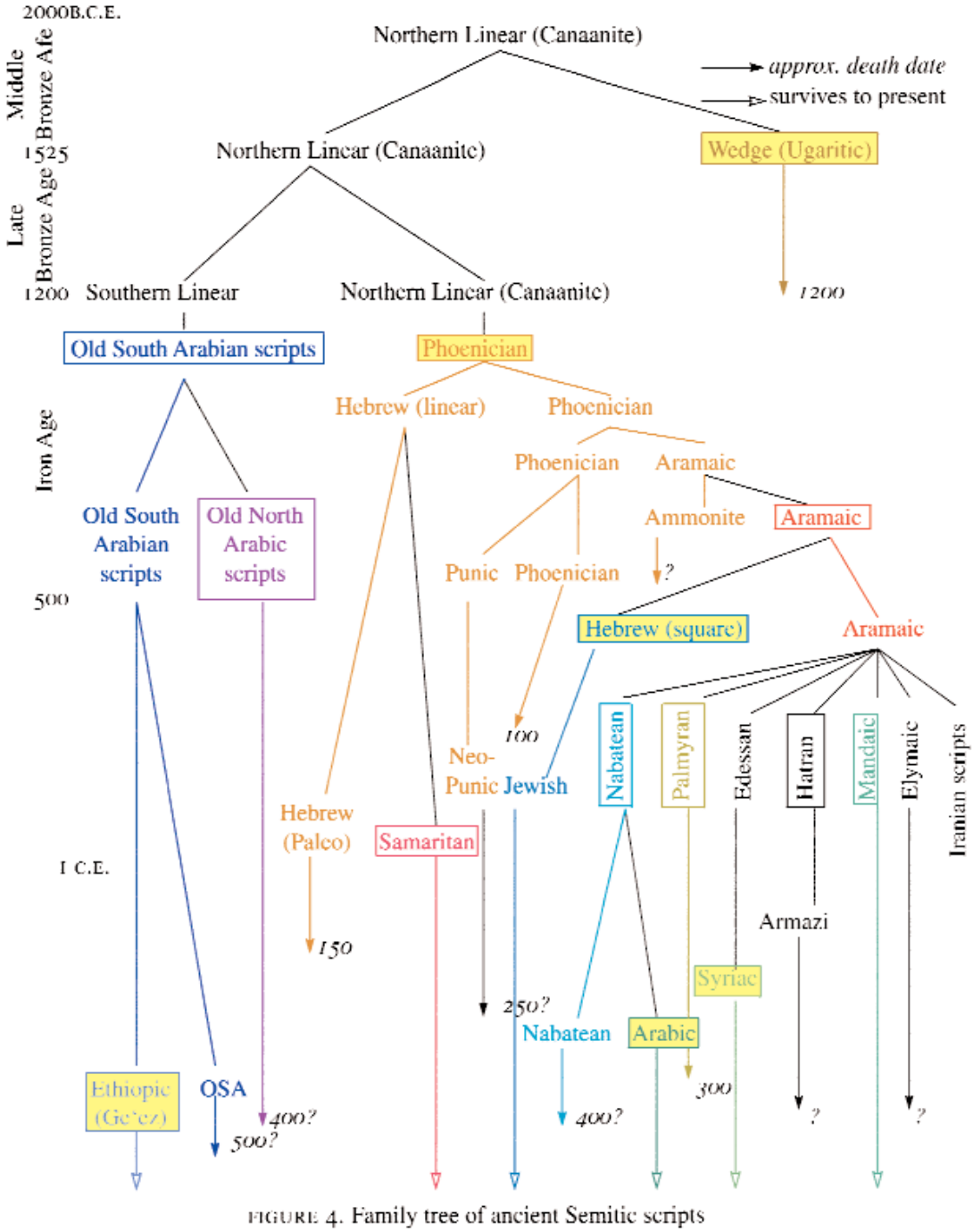


FIGURE 4. Family tree of ancient Semitic scripts

Figure 1. Family tree of ancient Semitic scripts, from O’Connor 1996, to which an “Iranian scripts” node has been added. Boxes have been drawn around nodes which have been or are expected to be encoded. Colour indicates familial unifications. Note the close relationship of Ethiopic and Old South Arabian which use South Semitic ordering, and their distance fom Arabic, Syriac, and Hebrew scripts which use Levantine ordering deriving from Phoenician. While we recognize that scholars who use Arabic script are concerned with Old South Arabian materials, we believe that in the context of the UCS it would be consistent with the practice of our committees to give preference to the genetic relationships when it comes to question of character names and ordering presentation in the code chart.

HIMYARISCH.

Zeichen schwarz	Zeichen relief	Wert	Zeichen schwarz	Zeichen relief	Wert	Ziffern	Wert
		ʾa			l		1
		b			m		2
		dž			n		3
		d			s		4
		ê			ʒ		5
		h			ʔ		6
		w			f, p		10
		z			ṣ		11
		hʿ			ẓ		12
		χ			q		20
		ḍ			r		50
		ṯ			š		100
		y			t		1000
		k			ṯ		?

Die himyarische Schrift wurde im südlichen Arabien gebraucht, bis sie durch die mohammedanische Nesχi verdrängt wurde; sie ist hier unter die afrikanischen Schriften einbezogen worden, weil sie die Wurzel der äthiopischen ist und der Schrift der Berber nahe steht.

Die arabischen Schriftsteller nennen sie *Musnad,* welcher Name zu mancherlei Vermuthungen Anlass gegeben hat, indem er durch *sanad* „stützen" erklärt wird und daher „Säulenschrift" oder auch „gestützte Schrift" bedeuten würde, man hat ihn ferner auf Sindh zurückgeführt, und demnach „indische Schrift" übersetzt; endlich wurde darauf hingewiesen, dass *musnad* auch *spurius* bedeutet, danach wäre sie die „sündhafte Schrift der Ungläubigen". Jedenfalls war sie die Schrift des Sabäismus und ging mit diesem unter.

Himyarisch wurde von rechts nach links, aber auch in wechselnden Furchen (*Bustrophedon*) geschrieben; es kommt nur auf Inschriften vor, theils einfach eingegraben, theils erhaben ausgemeiselt. Die Lautwerte sind nach den Angaben Halevy's (Inschriften von San'a), welcher den Zeichen das neu-arabische Lautsystem zu Grunde legte, beigefügt.

Die Ziffern sind stets eingeklammert, die Zeichen 1—4 werden auch verbunden, z. B. 𐩱 oder 𐩲 2, die Zeichen 5, 10, 100, 1000 sind Sigel, indem die Anfangsbuchstaben (*χamaš, ʒašar, mat, eleph*) die Zahlen vertreten.

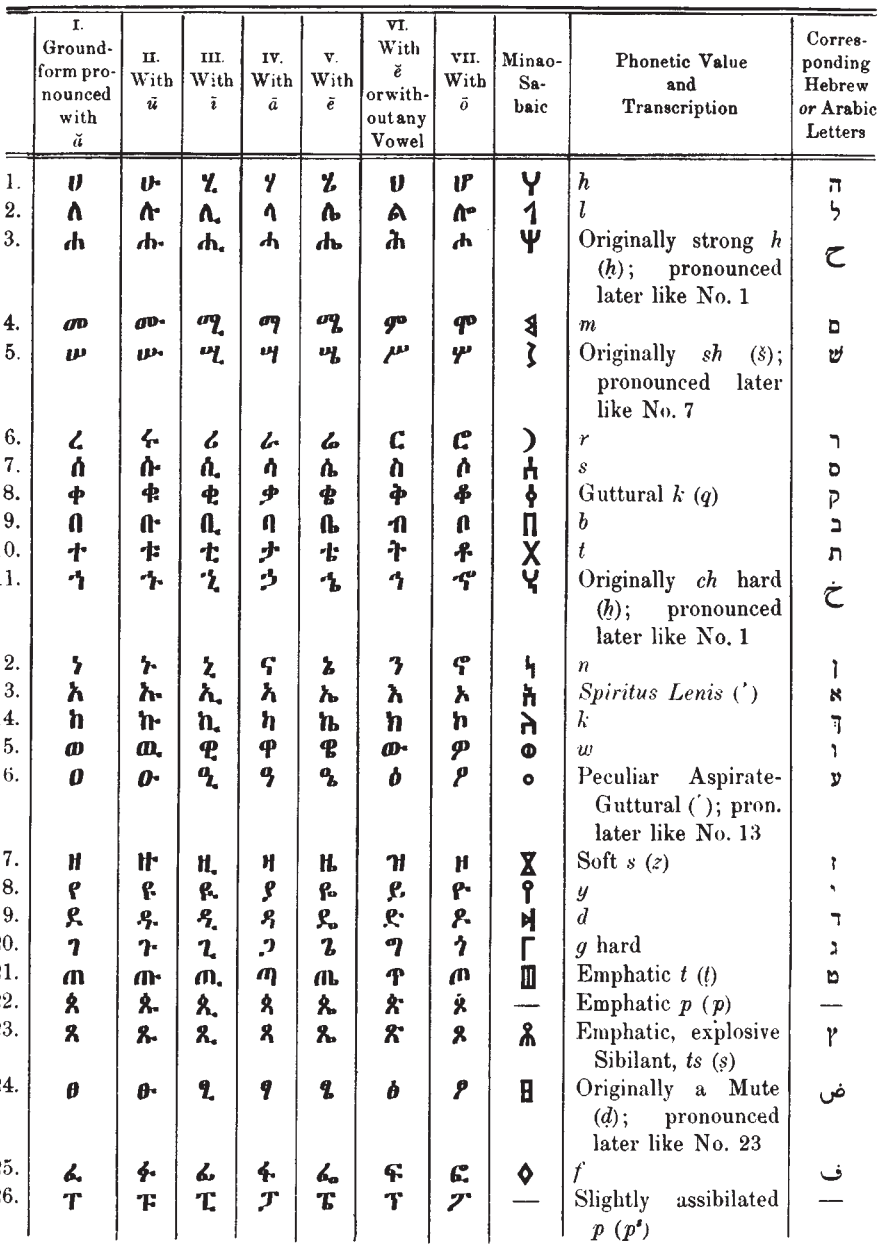
Figure 2. Chart of Old South Arabian, from Faulmann 1990 (1880). The numbers are shown flanked with the OLD SOUTH ARABIAN NUMERATOR BRACKET. Faulmann calls the script "Himyaric" and says: "The Arab writers call it *Musnad*, which name has been the cause of much disputed supposition, in that it is explained by *sanad* 'support' and therefore would also mean 'column writing' or also 'supported writing', furthermore some have attributed it to Sindh, and translated it as 'Indic writing'; finally it was pointed out that *musnad* also meant *spurius*, and accordingly it would be the 'sinful script of the unbelievers."

TABLE 5.7: *Monumental Scripts of Yemen and Ethiopia (Garbini 1979, fig. 10)*[a]

	XXVII	XXVIII	XXIX	XXX
ʾ				
b				
g				
d				
ḏ				
h				
w				
z				
ḥ				
ḫ				
ṭ				
ẓ				
y				
k				
l				
m				
n				
s¹				
ʿ				
ǵ				
p				
ṣ				
ḍ				
q				
r				
s²				
s³				
t				
ṯ				

a. Col. XXVII, Epigraphic South Arabian script (forms in parentheses are of the North Arabic type, found in some older inscriptions); col. XXVIII, Later South Arabian script; col. XXIX, "Thamudic" type of Ethiopic script; col. XXX, Ethiopic consonantal script. The order of letters in this table is artificially based on the North Semitic order; for the ancient South Semitic order, see SECTION 68.

Figure 3. Table of South Arabian scripts, from O'Connor 1996. O'Connor draws specific attention to the fact that the presentation here is in the Levantine order rather than in the South Semitic order.

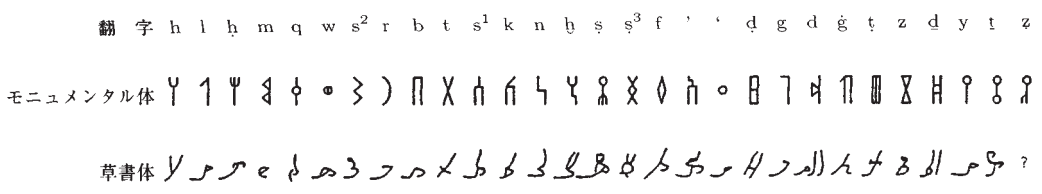


	I. Ground-form pronounced with *ă*	II. With *ŭ*	III. With *ĭ*	IV. With *ā*	V. With *ē*	VI. With *ĕ* or without any Vowel	VII. With *ō*	Minaean-Sabaic	Phonetic Value and Transcription	Corresponding Hebrew *or* Arabic Letters
1.	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ	𐩱	*h*	ה
2.	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ	𐩣	*l*	ל
3.	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ	𐩨	Originally strong *h* (*ḥ*); pronounced later like No. 1	ח
4.	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ	𐩥	*m*	ם
5.	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ	𐩮	Originally *sh* (*ś*); pronounced later like No. 7	ש
6.	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ	𐩧	*r*	ר
7.	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ	𐩯	*s*	ס
8.	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ	𐩲	Guttural *k* (*q*)	ק
9.	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ	𐩢	*b*	ב
10.	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ	𐩫	*t*	ת
11.	ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ	𐩵	Originally *ch* hard (*ḫ*); pronounced later like No. 1	ח
12.	ነ	ኑ	ኒ	ና	ኔ	ን	ኖ	𐩤	*n*	נ
13.	አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ	𐩠	*Spiritus Lenis* (*ʾ*)	א
14.	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ	𐩲	*k*	כ
15.	ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ	𐩺	*w*	ו
16.	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ	𐩴	Peculiar Aspirate-Guttural (*ʿ*); pron. later like No. 13	ע
17.	ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ	𐩬	Soft *s* (*z*)	ז
18.	የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ	𐩹	*y*	י
19.	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ	𐩩	*d*	ד
20.	ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ	𐩡	*g* hard	ג
21.	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ	𐩪	Emphatic *t* (*ṭ*)	ט
22.	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ	—	Emphatic *p* (*p*)	—
23.	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ	𐩰	Emphatic, explosive Sibilant, *ts* (*ṣ*)	צ
24.	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ	𐩦	Originally a Mute (*ḍ*); pronounced later like No. 23	ض
25.	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ	𐩭	*f*	ف
26.	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ	—	Slightly assibilated *p* (*pˢ*)	—

Fig. 42. Characters of the Ethiopic Alphabet

Figure 4. Comparison of Old South Arabian and Ethiopic, from Dillmann 1907 (1974).

〈表 1〉古代南アラビア文字



出典：Robin (1991), p. 130.

Figure 5. Old South Arabian alphabet alongside its cursive form, from Kōno 2001.

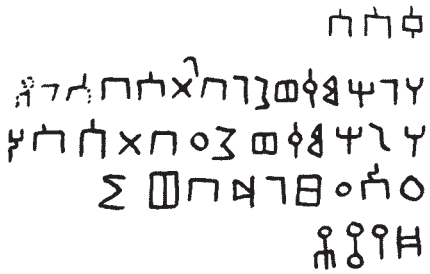


Figure 6. Old South Arabian abecedary in South Semitic order, from Küster 2006. The text reads:

q s s
h l ḥ m q w š r b t s k n ḫ ś/ẓ
h l ḥ m q w š ʿ b t s k n ḫ
f ʾ ʿ ḍ g d ġ ṭ z
d̲ y t̲ ṣ

Figure 7. Fragment from a limestone slab with Sabaean dedicatory inscription, first century CE.

Figure 8. An inscription in limestone from the 4th century CE, taken from the Kunsthistorisches Museum in Vienna, **www.khm.at/system2E.html?/staticE/page1549.html**. On the left and right there are monogram ligatures. Here the vertical bar appears to be part of the text layout, rather than being represented by an individual character.

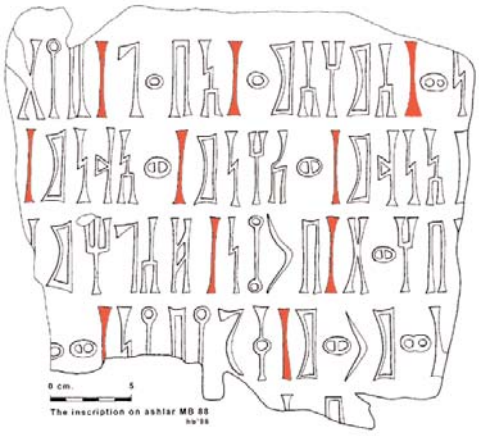


Figure 9. Copy of an Old South Arabian inscription. We have filled in the word separators to highlight them (they are red in the online version of the document). Note the distinctive shapes of the separator.

Figure 10a. A pre-Islamic inscription from the Smithsonian National Museum of Natural History, **www.mnh.si.edu/epigraphy/e_pre-islamic/fig25-musnad2.htm**.

Figure 10b. Transcription of the above. Again, note the distinctive serifs on the word separator.

Figure 11. Votive plaque inscribed with Sabaean dedication, late first millennium BCE, from the Metropolitan Museum of Art, **www.metmuseum.org/toah/ho/04/wap/hob_1993.169.htm**

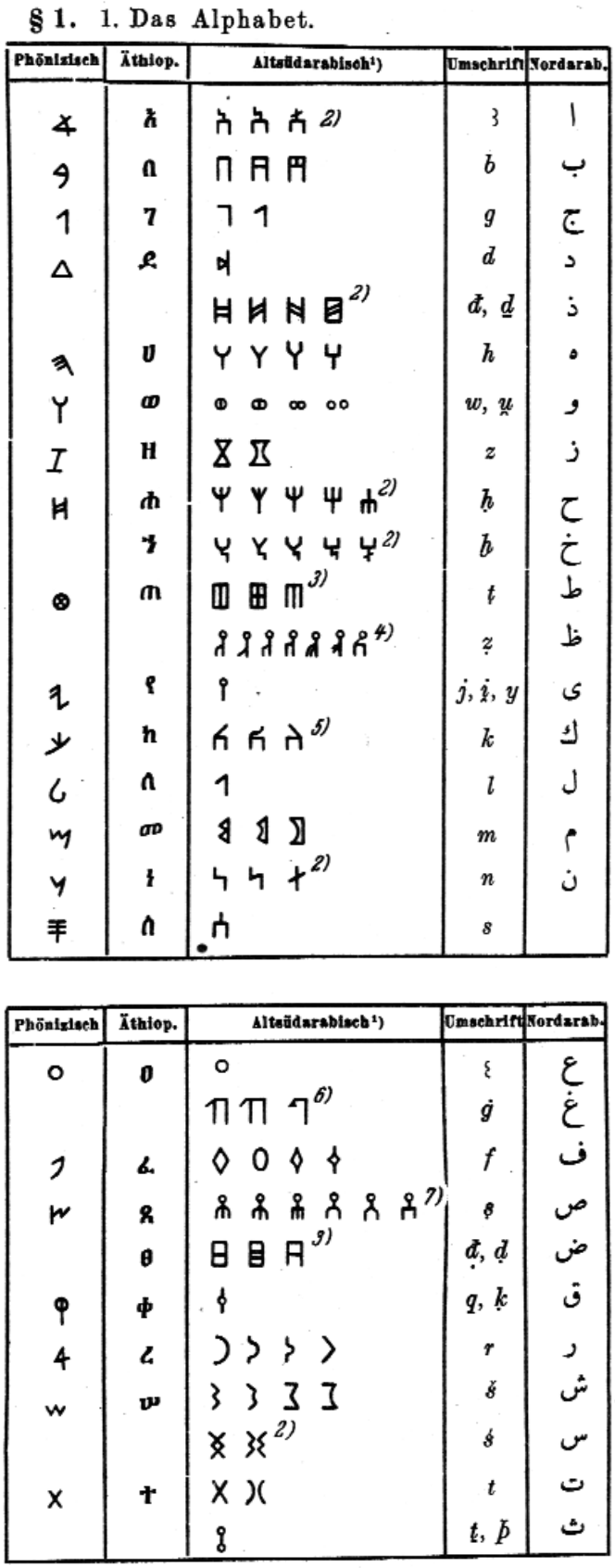


Figure 12. Chart of the Old South Arabian alphabet from Höfner 1943. Phoenician letters are given for comparison. The close glyph relation between Ethiopic and Old South Arabian letterforms is clear. Latin and Arabic transliterations are also given.

§ 8. 8. Zahlzeichen.

1	𐩱	8	𐩻	60	𐩽
2	𐩲	9	𐩼	70	𐩾
3	𐩳	10	𐩿	80	𐩿𐩽
4	𐩴	20	𐩿𐩿	90	𐩿𐩾
5	𐩵	30	𐩿𐩿𐩿	100	𐩷
6	𐩶	40	𐩿𐩿𐩿𐩿	1000	𐩸
7	𐩹	50	𐩽		

Zahlentrenner: 𐩺 𐩺

Figure 13. Chart of Old South Arabian numbers from Höfner 1943.

Beispiele für zusammengesetzte Zahlen:

𐩹𐩽𐩿	17
𐩹𐩽𐩿𐩿𐩿𐩿	47
𐩷𐩷𐩷𐩷𐩷	500
𐩸𐩸𐩸𐩸𐩸𐩸𐩸𐩸	8000
𐩸𐩸𐩸𐩸𐩸𐩸𐩸𐩸𐩸𐩸𐩸𐩸	12000
𐩿𐩿𐩿𐩿𐩿𐩿𐩿𐩿	16000 (in rückläufiger Zeile)
𐩿𐩿𐩿𐩿	31000 (in rückläufiger Zeile)
𐩿𐩿𐩿𐩿	40000
𐩿𐩿𐩿𐩿𐩸𐩸𐩸𐩸𐩸	45000
𐩿𐩿𐩿𐩽𐩸𐩸𐩸	63000
𐩺𐩺	150000 (in rückläufiger Zeile)
𐩷𐩷	200000

Figure 14. Examples of Old South Arabian numbers from Höfner 1943.

Script Table

Pharyngals and Laryngals:	𐩠	𐩡	𐩢	𐩣	
	ʾ	ḥ	ʿ	ḫ	
Velars:	𐩤	𐩥	𐩦		
	ḵ	ḏ	ḳ		
Palatals:	𐩧	𐩨			
	k	g			
Dentals and Interdentals:	𐩩	𐩪	𐩫	𐩬	
	t	d	ṯ	ḏ	
Sibilants:	𐩭	𐩮	𐩯	𐩰	
	s¹	s²	s³	ś	
Emphatics:	𐩱	𐩲	𐩳	𐩴	
	ṭ	ẓ	ṣ	ṣ	
Labials:	𐩵	𐩶			
	f	b			
Liquids and Nasals:	𐩷	𐩸	𐩹	𐩺	
	l	r	m	n	
Semivowels:	𐩻	𐩼			
	w	y			
Numerals:	𐩱	𐩵	𐩿	𐩷	𐩸
	1	5	10	50	100	1000
Numerator bracket:	𐩽					

Figure 15. Chart of the Old South Arabian alphabet and numbers from Beeston 1962.

Row 108: SOUTH ARABIAN DRAFT with three different fonts

hex	Name
60	OLD SOUTH ARABIAN LETTER HOY
61	OLD SOUTH ARABIAN LETTER LAWE
62	OLD SOUTH ARABIAN LETTER HAWT
63	OLD SOUTH ARABIAN LETTER MAAY
64	OLD SOUTH ARABIAN LETTER QAAF
65	OLD SOUTH ARABIAN LETTER WAWE
66	OLD SOUTH ARABIAN LETTER HLAWT
67	OLD SOUTH ARABIAN LETTER RES
68	OLD SOUTH ARABIAN LETTER BET
69	OLD SOUTH ARABIAN LETTER TAWE
6A	OLD SOUTH ARABIAN LETTER SAAT
6B	OLD SOUTH ARABIAN LETTER KAAF
6C	OLD SOUTH ARABIAN LETTER NAHAAS
6D	OLD SOUTH ARABIAN LETTER KHARM
6E	OLD SOUTH ARABIAN LETTER SAWT
6F	OLD SOUTH ARABIAN LETTER AF
70	OLD SOUTH ARABIAN LETTER ALF
71	OLD SOUTH ARABIAN LETTER AYN
72	OLD SOUTH ARABIAN LETTER HLAFFAA
73	OLD SOUTH ARABIAN LETTER GAML
74	OLD SOUTH ARABIAN LETTER DANT
75	OLD SOUTH ARABIAN LETTER GHAAF
76	OLD SOUTH ARABIAN LETTER TAYT
77	OLD SOUTH ARABIAN LETTER ZAY
78	OLD SOUTH ARABIAN LETTER DHAY
79	OLD SOUTH ARABIAN LETTER YAMAN
7A	OLD SOUTH ARABIAN LETTER THAY
7B	OLD SOUTH ARABIAN LETTER TSADAY
7C	OLD SOUTH ARABIAN LETTER THADAY
7D	OLD SOUTH ARABIAN NUMERATOR BRACKET
7E	OLD SOUTH ARABIAN NUMBER ONE
7F	OLD SOUTH ARABIAN NUMBER FIFTY

A. Administrative

1. Title

Towards a proposal to encode the Old South Arabian script in the UCS.

2. Requester's name

Michael Everson and Daniel Yacob

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2007-08-29

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

No.

6b. More information will be provided later

Yes.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Old South Arabian.

1c. The proposal is for addition of character(s) to an existing block

No.

1d. Name of the existing block

2. Number of characters in proposal

32.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category C.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes. See N1689, L2/07-240.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

No.

2b. If YES, with whom?

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Semiticists, Ethiopicists, and other scholars.

4a. The context of use for the proposed characters (type of use; common or rare)
Rare.
4b. Reference
5a. Are the proposed characters in current use by the user community?
Yes.
5b. If YES, where?
Scholarly publications.
6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
No.
6b. If YES, is a rationale provided?
6c. If YES, reference
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
No.
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No.
8b. If YES, is a rationale for its inclusion provided?
8c. If YES, reference
9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
No.
9b. If YES, is a rationale for its inclusion provided?
9c. If YES, reference
10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
No.
10b. If YES, is a rationale for its inclusion provided?
10c. If YES, reference
11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?
No.
11b. If YES, is a rationale for such use provided?
11c. If YES, reference
11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.
11e. If YES, reference
12a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
12b. If YES, describe in detail (include attachment if necessary)
13a. Does the proposal contain any Ideographic compatibility character(s)?
No.
13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?