

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

Doc Type: Working Group Document

Title: Proposal to add the Samaritan alphabet to the BMP of the UCS

Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)

Authors: Michael Everson & Mark Shoulson

Status: Individual Contribution

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

Date: 2008-01-25

1. Historical background. Both the Hebrew and the Samaritan scripts ultimately derive from the Phoenician, but by different routes. According to Naveh 1997, by 1050 BCE, the Phoenician script had lost all of the pictographic features which were present in Proto-Canaanite. Phoenician script was adopted by speakers of Aramaic and Hebrew. Hebrew writing began to take on unique features (i.e. those of Palaeo-Hebrew) by the mid-ninth century BCE, and Aramaic writing began to take on its own features by the middle of the eighth century BCE.

The destruction of the First Temple and the exile of educated Hebrew speakers to Babylonia changed things greatly, according to Naveh (p. 78). Later generations returned to Judah, by then a Persian province, where Aramaic was official; many of these people were bilingual in Aramaic and Hebrew, and had given up the Palaeo-Hebrew script which they had used prior to the exile, writing instead in a script derived from Aramaic – having abandoned their original script (pp. 112 ff.). They later developed this script until by the second century CE it had developed into the Jewish script which became the Square Hebrew used today.

The abandonment of one script for another (even if the two scripts are related) is complex, particularly with regard to conservative cultures such as that of the Jews. Naveh suggests that, although Aramaic script was very widespread during the Persian period – indeed being the “international” script *par excellence* – it was not until the official language of the Persian government had become Greek that the by-then-familiar Aramaic came to be modified into the uniquely Jewish script which we know today as the “Hebrew” encoded in the UCS. Apparently some differentiation in function arose between the use of the Aramaic-derived writing (= Square Hebrew) vs. Hebrew-derived writing (Phoenician or Palaeo-Hebrew), with the Pharisees apparently disapproving the Hebrew-derived script. Naveh quotes from the Babylonian Talmud, Sanhedrin 21b:

Originally the Torah was given to Israel in the Hebrew script and in the sacred language; later, in the time of Ezra, the Torah was given in the Assyrian script [i.e. the Aramaic script, introduced by the Assyrians as an official script] and the Aramaic language. They selected for Israel the Assyrian script and the Hebrew language, leaving the Hebrew script and the Aramaic language for the ordinary people.

The Samaritans, who had not gone into exile, did not give up their Palaeo-Hebrew tradition, and continue to use a variety of this script to the present day. According to Naveh, they “believe that they are the true descendants of the sons of Israel”; Rav Hisda explained (in the third century CE) that they are the

“ordinary people” referred to in the Babylonian Talmud cited above. Uniquely Samaritan script features (as distinct from Phoenician/Palaeo-Hebrew) are discernable by the third century CE.

Modern Samaritans continue to make use of this script, and a weekly newspaper $\Delta \cdot \nu$ (A.B.) is published in Israel in Samaritan script (along with short articles in Hebrew and Arabic).

2. Corpus. There are some hundreds of Samaritan manuscripts; one of the largest collections is in the John Rylands University Library at the University of Manchester, including 377 items on parchment and paper. Samaritan MSS 1-27 were acquired in 1901 with the Crawford collection and include what is apparently the earliest dated manuscript (1211 CE) of the whole Samaritan Pentateuch to be found outside Nablus, as well as six other Pentateuchs in whole or in part (two bilingual), three noteworthy theological codices, and interesting liturgical and astronomical texts. Samaritan MSS 28-375 are from the collection of Dr Moses Gaster, acquired by the Library in 1954. Among them are manuscripts of the Pentateuch (including bilingual and trilingual texts), commentaries and treatises, and liturgical, historical, chronological and astronomical codices. There are detailed census lists of the Samaritans and lists of manuscripts in their possession. The Library also holds the substantial, but uncatalogued, correspondence of Dr Gaster with the Samaritan community in Nablus, in Hebrew but written in the Samaritan script. Some other important Samaritan manuscripts are found at the Chester Beatty Library in Dublin (dating to 1211 CE) and at the New York Public Library (dating to 1232 CE).

3. Structure. Samaritan is a right-to-left script. It does not ligate its letters as many right-to-left scripts do, and it does not have explicit final consonants as Hebrew does.

4. Vowels and other marks of pronunciation. Vowel signs are used optionally in Samaritan, as points are used optionally in Hebrew. In modern times, overlong vowels (marked here with circumflex and colon) and long vowels (marked here with circumflex) are distinguished from short vowels by the size of the diacritic. The default vowel sign to be used in transcribing text which does not make the distinction is the smallest one.

	V	v	^	^
	◉ ê LONG E	◉ e E	◉ û LONG U	◉ u U
			<	<
◉ â: OVERLONG AA	◉ â LONG AA	◉ ă AA	◉ î LONG I	◉ i I
—	—	—		^
◉ â: OVERLONG A	◉ â LONG A	◉ a A		◉ o O
		/		
		◉ ă SHORT A		

4.1. The general behaviour of vowel signs. These vowel signs are combining characters, each resting to the left of its base consonant, effectively centred between its base consonant and the following one (if any). Examples using the letters YUT, QUF, DALAT, and IY, reading from right to left: *yêqed*, *yâ:qâdâh*, *yâ:qâdah*, *yăqăd*, *yûqud*, *yîqid*, *yoqod*:



4.2. The behaviour of consonant modifiers. The four marks SUKUN, DAGESH, OCCLUSION, and NEQUDAA are centred over the base letter. Examples are, reading from right to left: *yâqdah*, *yêqqed*, *hăbbâh*, *yûqud*:



These marks modify the consonant and precede the vowel signs where used. The SUKUN indicates that no vowel follows the consonant. The DAGESH indicates consonant gemination; NEQUDAA is an editorial mark which indicates that there is a variant reading of the word. The mark for OCCLUSION “strengthens” the consonant, as here where *w* becomes *b*. Note that in the example *ḥābbāh*, DAGESH stacks atop OCCLUSION *bb*, reflecting the preferred encoding order since consonant quality precedes consonant length. The mark for OCCLUSION also has a secondary use, for instance, to mark personal names to distinguish them from homographs. So *îšab* ‘Esau’ contrasts with *âšu* ‘they made’. (Obviously with full pointing these words can also be distinguished by their vowels.)

The character properties should, if possible, support the priority of these marks over the vowel signs.

4.3. The behaviour of IN and IN-ALAF. The two marks IN and IN-ALAF are used to indicate the presence of [ʔ] (Samaritan ‘in, Hebrew ‘ayin). These are also encoded immediately following their base letter and before any vowel signs. They are drawn to the right side of their base letter. Examples, reading from right to left: *’hyk, ḥ’ayāh*:



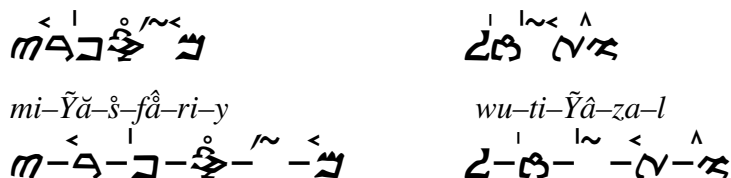
4.4. The behaviour of EPENTHETIC YUT. The EPENTHETIC YUT, transcribed *ÿ* here, represents a kind of glide-vowel which interacts with another vowel sign. It was originally used only with the consonants ALAF, IY, IT, and IN (Hebrew ‘alef, he, het, ‘ayin); those letters used to serve to separate syllables, but lost their sound. The behaviour of the combining epenthetic YUT is the same as that of DAGESH and the other consonant modifiers mentioned above: it is centred above the consonant, with the following vowel sign centred more or less between it and the following letter. Examples, reading from right to left, *bâ’ÿâr, mi’ÿâḥûriy, miḥÿâḥelâk, miḥÿowt, mi’ÿâl*:



To make it clearer what is written here, hyphens might be inserted in the transliterations to show the pattern. A consonant (‘, *h, ḥ, ’*) may be followed by one or two combining marks, as seen here (examples with some of the other combining marks shown above are given for comparison):

<i>bâ-’ÿâ-r</i>	<i>yê-qe-d</i>	<i>yâ-ḳ-da-h</i>
<i>mi-’ÿâ-ḥû-ri-y</i>	<i>yâ:-qâ-dâ-h</i>	<i>yê-ḳe-d (yêqqed)</i>
<i>mi-ḥÿâ-ḥe-lâ-k</i>	<i>yâ:-qâ-da-h</i>	<i>ḥâ-ÿâ-h (ḥābbāh)</i>
<i>mi-ḥÿo-w-t</i>	<i>yă-qă-d</i>	<i>yû-ḳu-d</i>
<i>mi-’ÿă-l</i>	<i>yû-qu-d</i>	<i>’-ḥ-y-k</i>
	<i>yî-qi-d</i>	<i>ḥ’a-yă-h</i>
	<i>yo-ḳo-d</i>	

When epenthetic YUT is not fixed to one of the four consonants listed above, a new behaviour was innovated—not recently—in which the mark for the epenthetic YUT is treated as a spacing character of its own, capable of bearing its own diacritical mark. We transliterate the MODIFIER LETTER EPENTHETIC YUT as *ÿ̇* below. Examples, reading from right to left: *wutiÿ̇âzal, miÿ̇ăsfâriy*:



At some point in the discussions leading to this proposal, the possibility of attaching the combining MARK EPENTHETIC YUT to a NBSP, but this would in effect break single words into two, separating prefixes and marks of conjugation from the root of the word:

mi-Øyǎ-š-fâ-ri-y *wu-ti-Øyâ-za-l*

We believe that the simpler model C((M)V)—where a consonant may be optionally followed by a combining mark (consonant sign or vowel sign) and optionally by a vowel sign—is preferable to a solution which makes use of NBSP. Here are two examples where a single consonant is followed by two vowel signs, reading from left to right, 'eumer 'I will say', *hâ-inšem* 'the women':




hâ-i-n-še-m *'eu-me-r* *yi-šae-l'*

4.5. The behaviour of vowels *ǎ* and *i* in word-initial position. Two vowels are known to occur in initial position, before the base character. These are encoded as spacing modifier letters because combining characters cannot occur in initial position in a word. Users concerned with spoofing possibilities should note the similarity between ' MODIFIER LETTER SHORT A and ' ◌ VOWEL SIGN SHORT A and between ' MODIFIER LETTER I and ' ◌ VOWEL SIGN I. (It is extremely improbable that the user community, which is very small, will require Samaritan script in IDN or similar applications.) Examples, reading from right to left: *ǎlfâniy* 'before' and *inšiy* 'wives of...', using *ǎ* and *i* in the transcription for the modifier letters:



I-n-ši-y *ǎ-l-fâ-ni-y*

The MODIFIER LETTER SHORT A also has an additional function, when used following a letter used numerically, to indicate the thousands, so '𐤀 = 3000. This is similar to the use of HEBREW PUNCTUATION GERESH for the same thing (ג' = 3000). Hebrew GERESH serves a number of functions. It modifies the sound of a letter (ג'ירפָה *gīrafa* 'giraffe'; צ'יפֿס *tšips* 'chips'); it marks abbreviations ('מִס', short for ספֿר *mīspar* 'number'); and in transliterations of Samaritan GERESH is used for the syllable initial *ǎ*, as in נִנְדָ' for אֲנִדָ' *ānged*. The MODIFIER LETTER SHORT A similarly has multiple functions (though the 𐤀 ABBREVIATION MARK is used with abbreviations in Samaritan).

4.5.1 Alternatives previously proposed for this behaviour. It has been suggested that the distinction between MODIFIER LETTER SHORT A and COMBINING SHORT A might cause "difficulty" for users, for instance in searching operations. This suggestion does not seem convincing to us. The word *inšiy* here could be represented, structurally, in Devanagari and Latin, completely isomorphically with the encoding model proposed here (apart from the irrelevant virama); the Latin example uses U+2071 SUPERSCRIPT LATIN SMALL LETTER I and U+0365 COMBINING LATIN SMALL LETTER I:

$\overset{<}{i} + n + \check{s} + \overset{\circ}{i} + y = i\overset{\circ}{n}\check{s}y$
 $\text{इ} + \text{न्} + \text{श्} + \text{ि} + \text{य्} = \text{इन्शिश्य्}$

The concurrent use of combining marks and spacing marks that look very similar is also not unique to Samaritan. In the orthography of Oowekyala, a North Wakashan language spoken in British Columbia, both spacing U+02BC MODIFIER LETTER APOSTROPHE and non-spacing U+0313 COMBINING COMMA ABOVE are used together to indicate glottalization. Among the consonants, plain resonants *m n l y w* contrast with glottalized resonants *ṃ ṇ ḷ ỵ ẉ*. Among the vowels, plain vowels *ə m ə n ə l i u* contrast with two sets of glottalized vowels: *əmṃ ənṇ əlḷ ị ỵ ụ* are used when any other vowel follows, and *ə m' ə n' ə l' i' u'* are used

word-finally or when an obstruent follows. Compare *ḡam*’s ‘to lie on the ground’ with *ḡammîs* ‘to lie on the beach’.

John Hudson suggested that in order to avoid “duplicating” MODIFIER LETTER SHORT A with VOWEL SIGN SHORT A or MODIFIER LETTER I with VOWEL SIGN I, the generally-accepted UCS combining-character model might be abandoned for Samaritan, and all of its many combining diacritical marks might be represented by—effectively—spacing modifier letters that could be rendered correctly by a smart font. While such an encoding model could be made to work, we do not believe it is in the interests of the Samaritan user community itself or of the UCS user community in general to do this, and indeed, we do not believe that MODIFIER LETTER SHORT A with VOWEL SIGN SHORT A are “duplicates” any more than U+02BC MODIFIER LETTER APOSTROPHE and U+0313 COMBINING COMMA ABOVE are. All of the rest of the Semitic scripts follow the standard UCS encoding model. We can imagine no advantage for Samaritan to differ from this model. While word initial *i-* and *a-* are not rare in Samaritan, neither are they a dime a dozen. And this is certainly no reason to abandon a well-understood encoding model for a novel one. Anyone implementing Samaritan will be familiar with Hebrew if not also Arabic. Since spacing MODIFIER LETTER SHORT A is already required in Samaritan as a kind of GERESH to indicate the numeric use of Samaritan letters, the only anomaly to this encoding model is the glyph similarity of MODIFIER LETTER I with VOWEL SIGN I, and that already contrasts with the slightly larger VOWEL SIGN LONG I ... so what’s the benefit in avoiding the current UCS model for encoding of western Semitic scripts? The (I)C((M)V) model for Samaritan is elegant, enables the representation of Samaritan data, and in our view is the optimum encoding model for Samaritan.

5. Punctuation. A large number of punctuation characters is used in Samaritan. These form a coherent and well-defined set, often with a diamond-shape to the dot (in most of the better-designed fonts such as that of the Imprimerie Nationale and the font used in the weekly newspaper *ⲁⲛⲏ A.B.*), and we propose that all of them be encoded as script-specific punctuation. The set as proposed follows the functional description found in Murtonen 1964.

- The · NEQUDAA and : AFSAAQ ‘interruption’ are similar to the Hebrew SOF PASUQ and were used originally to separate sentences, but later to mark lesser breaks within a sentence. The AFSAAQ and the NEQUDAA are the oldest Samaritan punctuation marks. They are sometimes combined together ·: with AFSAAQ preceding NEQUDAA, or vice-versa, :· with NEQUDAA preceding AFSAAQ, or both ·:· as NEQUDAA AFSAAQ NEQUDAA. Both of these characters should have the **Sentence Terminal** property. (Both Murtonen and the back matter of the Samaritan Pentateuch describe AFSAAQ explicitly as *āfsaq* and *ⲁⲫⲥⲁⲗ* *āfsāq*. In the Markeh Shameri font AFSAAQ is named “pause” and NEQUDAA is named “semicolon”.)
- The ¨ ANGED ‘restraint’ indicates a break somewhat less strong than an AFSAAQ. (Both Murtonen and the back matter of the Samaritan Pentateuch describe ANGED explicitly as *anged* and *ⲁⲛⲉⲃ* *anged*.)
- The ⲥⲏ BAU ‘request, prayer’ shows that the preceding is a humble petition, above all prayers to God. (Both Murtonen and the back matter of the Samaritan Pentateuch describe BAU explicitly as *bâ’u* and *ⲃⲏ* *ba’uw*.)
- The ⲥ: ATMAAU ‘surprise’ shows that the preceding is unexpected. (Both Murtonen and the back matter of the Samaritan Pentateuch describe ATMAAU explicitly as *atmâ’u* and *ⲁⲧⲙⲁⲁⲩ* *atmâhuw*.)
- The Ⲏ SHIYYAALAA ‘question’ shows that the preceding is a question. (Both Murtonen and the back matter of the Samaritan Pentateuch describe SHIYYAALAA explicitly as *šîla* and *ⲥⲏⲗⲁ* *šiy’âlâh*. In the Markeh Shameri font SHIYYAALAA is named “question”.)
- The ⲗ ABBREVIATION MARK follows an abbreviation.
- The — ZIQAA ‘shout, cry’ marks expressions calling attention of human beings. (Both Murtonen and the back matter of the Samaritan Pentateuch describe ZIQAA explicitly as *zîqa* and *ⲗⲏⲗⲁ* *zi’yqâh*.)

- The —: QITSA is similar to the ANNAAU (see below) but is used more frequently. The QITSA marks the end of a section, and is may be followed by a blank line to further make the point. It is analogous to the “open” and “closed” sections in the Masoretic Pentateuch. It has many glyph variants. One important variant differs significantly from any of the others; this is the —ϯ MELODIC QITSA which is used to indicate the end of a sentence “which one should read melodically”. Together with : AFSAAQ as —ϯ: it is used to mark the middle part of the Torah (at Leviticus 7:17). (Murtonen describes QITSA explicitly as *qíṣṣa*. In the Markeh Shameri font QITSA is named “final pause”. The Samaritan spelling is $\text{Ⲛⲓⲛⲧⲩ} \text{ qíṣṣáh}$)
- The =: ZAEF ‘outburst’ marks expressions of vehemence and anger. (Both Murtonen and the back matter of the Samaritan Pentateuch describe ZAEF explicitly as *zâf* and $\text{ⲕⲏⲉⲑ} \text{ z'f.}$)
- The ⋮ TURU ‘teaching’ marks didactic expressions. (Both Murtonen and the back matter of the Samaritan Pentateuch describe TURU explicitly as *tûru* and $\text{ⲧⲏⲣⲏ} \text{ tûwruw.}$)
- The / ARKAANU ‘submissiveness’ marks expressions of meekness and submission. (Both Murtonen and the back matter of the Samaritan Pentateuch describe ARKAANU explicitly as *arkânu* and $\text{Ⲡⲕⲁⲛⲏ} \text{ 'arkânuw.}$)
- The ⚭ SOF MASHFAAT is equivalent to the full stop. (In the Markeh Shameri font SOF MASHFAAT is named “full stop”.)
- The ° ANNAAU ‘rest’ indicates that a longer time has passed between actions narrated in the sentences which it separates; it is stronger than the AFSAAQ. (Both Murtonen and the back matter of the Samaritan Pentateuch describe ANNAAU explicitly as *anâ'u* and $\text{Ⲡⲛⲛⲁⲁⲩ} \text{ 'ännâhuw.}$ In the Markeh Shameri font ANNAAU is named—in error—“guttural yut” but it stands next to “yut dagesha” which is the EPENTHETIC YUT.)
- Samaritan distinguishes small dot from the larger NEQUDAA which is final punctuation like the AFSAAQ. Fossey’s example in Figure 5 shows this distinction. The generic U+2E31 WORD SEPARATOR MIDDLE DOT can be used to represent this.

As noted above, the set as proposed follows the functional description found in Murtonen 1964. Reviewers will note that the punctuation as described in secondary sources (Faulmann 1990 (1880), Reichsdruckerei 1924, von Ostermann 1954) some other configurations are also found. These may be conventional or ad-hoc on the part of the writer. The following is not an exhaustive list. The order is right-to-left.

:·	←	: AFSAAQ + · NEQUDAA
·:·	←	· NEQUDAA + : AFSAAQ + · NEQUDAA
—:·	←	—: QITSA + · NEQUDAA
·:	←	· NEQUDAA + : AFSAAQ
—ϯ:	←	—ϯ MELODIC QITSA + : AFSAAQ
—:·:	←	—: ZIQAA + : AFSAAQ + · NEQUDAA + : AFSAAQ
—:·	←	—: ZIQAA + : AFSAAQ
—:ϯ:	←	—: QITSA + ϯ: ATMAAU
—:ë	←	—: QITSA + ë SHIYYAALAA
⚭ ë	←	⚭ SOF MASHFAAT + ë SHIYYAALAA
·⚭	←	· NEQUDAA + ⚭ SOF MASHFAAT

There are other configurations in the MSS which cannot necessarily be composed based on the functional set proposed here. The angle used in ϯ· BAU, ϯ: ATMAAU, and ë SHIYYAALAA for instance has not been encoded uniquely since these elements does not necessarily make sense for Samaritan. The elements alone do not have names or functions—and the functions are given as named entities by Murtonen.

6. Character names. While most of the text samples give Hebrew versions of the names of Samaritan characters in the charts, the Samaritan names as transliterated in Konô *et al.* 2001 (fig. 9) are preferred here.

7. Reference glyphs. The older font charts shown in a number of the Figures below present a normalized 19th-century font style. Modern Samaritan usage prefers fonts which look more like the actual manuscripts. The font used in the chart here was based on a modern font with a certain amount of rectification to enhance a uniform feel.

8. Unicode character properties.

```
0800;SAMARITAN LETTER ALAF;Lo;0;R;;;;N;;;;;
0801;SAMARITAN LETTER BIT;Lo;0;R;;;;N;;;;;
0802;SAMARITAN LETTER GAMAN;Lo;0;R;;;;N;;;;;
0803;SAMARITAN LETTER DALAT;Lo;0;R;;;;N;;;;;
0804;SAMARITAN LETTER IY;Lo;0;R;;;;N;;;;;
0805;SAMARITAN LETTER BAA;Lo;0;R;;;;N;;;;;
0806;SAMARITAN LETTER ZEN;Lo;0;R;;;;N;;;;;
0807;SAMARITAN LETTER IT;Lo;0;R;;;;N;;;;;
0808;SAMARITAN LETTER TIT;Lo;0;R;;;;N;;;;;
0809;SAMARITAN LETTER YUT;Lo;0;R;;;;N;;;;;
080A;SAMARITAN LETTER KAAF;Lo;0;R;;;;N;;;;;
080B;SAMARITAN LETTER LABAT;Lo;0;R;;;;N;;;;;
080C;SAMARITAN LETTER MIM;Lo;0;R;;;;N;;;;;
080D;SAMARITAN LETTER NUN;Lo;0;R;;;;N;;;;;
080E;SAMARITAN LETTER SINGAAT;Lo;0;R;;;;N;;;;;
080F;SAMARITAN LETTER IN;Lo;0;R;;;;N;;;;;
0810;SAMARITAN LETTER FI;Lo;0;R;;;;N;;;;;
0811;SAMARITAN LETTER TSAADIY;Lo;0;R;;;;N;;;;;
0812;SAMARITAN LETTER QUF;Lo;0;R;;;;N;;;;;
0813;SAMARITAN LETTER RISH;Lo;0;R;;;;N;;;;;
0814;SAMARITAN LETTER SHAN;Lo;0;R;;;;N;;;;;
0815;SAMARITAN LETTER TAAF;Lo;0;R;;;;N;;;;;
0816;SAMARITAN MARK IN;Mn;230;NSM;;;;N;;;;;
0817;SAMARITAN MARK IN-ALAF;Mn;230;NSM;;;;N;;;;;
0818;SAMARITAN MARK OCCLUSION;Mn;230;NSM;;;;N;;;;;
0819;SAMARITAN MARK DAGESH;Mn;230;NSM;;;;N;;;;;
081A;SAMARITAN MODIFIER LETTER EPENTHETIC YUT;Lo;0;R;;;;N;;;;;
081B;SAMARITAN MARK EPENTHETIC YUT;Mn;230;NSM;;;;N;;;;;
081C;SAMARITAN VOWEL SIGN LONG E;Mn;23;NSM;;;;N;;;;;
081D;SAMARITAN VOWEL SIGN E;Mn;23;NSM;;;;N;;;;;
081E;SAMARITAN VOWEL SIGN OVERLONG AA;Mn;23;NSM;;;;N;;;;;
081F;SAMARITAN VOWEL SIGN LONG AA;Mn;23;NSM;;;;N;;;;;
0820;SAMARITAN VOWEL SIGN AA;Mn;23;NSM;;;;N;;;;;
0821;SAMARITAN VOWEL SIGN OVERLONG A;Mn;23;NSM;;;;N;;;;;
0822;SAMARITAN VOWEL SIGN LONG A;Mn;23;NSM;;;;N;;;;;
0823;SAMARITAN VOWEL SIGN A;Mn;23;NSM;;;;N;;;;;
0824;SAMARITAN MODIFIER LETTER SHORT A;Lo;0;R;;;;N;;;;;
0825;SAMARITAN VOWEL SIGN SHORT A;Mn;23;NSM;;;;N;;;;;
0826;SAMARITAN VOWEL SIGN U;Mn;23;NSM;;;;N;;;;;
0827;SAMARITAN VOWEL SIGN LONG U;Mn;23;NSM;;;;N;;;;;
0828;SAMARITAN MODIFIER LETTER I;Lo;0;R;;;;N;;;;;
0829;SAMARITAN VOWEL SIGN I;Mn;23;NSM;;;;N;;;;;
082A;SAMARITAN VOWEL SIGN I;Mn;23;NSM;;;;N;;;;;
082B;SAMARITAN VOWEL SIGN O;Mn;23;NSM;;;;N;;;;;
082C;SAMARITAN VOWEL SIGN SUKUN;Mn;23;NSM;;;;N;;;;;
082D;SAMARITAN MARK NEQUDAA;Mn;230;NSM;;;;N;;;;;
0830;SAMARITAN PUNCTUATION NEQUDAA;Po;0;R;;;;N;;;;;
0831;SAMARITAN PUNCTUATION AFSAAQ;Po;0;R;;;;N;;;;;
0832;SAMARITAN PUNCTUATION ANGED;Po;0;R;;;;N;;;;;
0833;SAMARITAN PUNCTUATION BAU;Po;0;R;;;;N;;;;;
0834;SAMARITAN PUNCTUATION ATMAAU;Po;0;R;;;;N;;;;;
0835;SAMARITAN PUNCTUATION SHIYYAALAA;Po;0;R;;;;N;;;;;
0836;SAMARITAN ABBREVIATION MARK;Po;0;R;;;;N;;;;;
0837;SAMARITAN PUNCTUATION MELODIC QITSA;Po;0;R;;;;N;;;;;
0838;SAMARITAN PUNCTUATION ZIQAA;Po;0;R;;;;N;;;;;
0839;SAMARITAN PUNCTUATION QITSA;Po;0;R;;;;N;;;;;
083A;SAMARITAN PUNCTUATION ZAEF;Po;0;R;;;;N;;;;;
083B;SAMARITAN PUNCTUATION TURU;Po;0;R;;;;N;;;;;
083C;SAMARITAN PUNCTUATION ARKAANU;Po;0;R;;;;N;;;;;
083D;SAMARITAN PUNCTUATION SOF MASHFAAT;Po;0;R;;;;N;;;;;
083E;SAMARITAN PUNCTUATION ANNAAU;Po;0;R;;;;N;;;;;
```

8. Bibliography

- Ben-Hayyam, Ze'ev. 2000. *A grammar of Samaritan Hebrew, based on the Recitation of the Law in comparison with the Tiberian and other Jewish traditions*. Jerusalem: Hebrew University Magnes Press. ISBN 1-57506-047-7
- Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8
- Fossey, Charles. 1948. *Notices sur les caractères étrangers anciens et modernes rédigées par un groupe de savants*. Nouvelle édition mise à jour à l'occasion du 21^e Congrès des Orientalistes. Paris: Imprimerie Nationale de France.
- Haarmann, Harald. 1990. *Die Universalgeschichte der Schrift*. Frankfurt: Campus Verlag. ISBN 3-593-34346-0
- Healey, John F. 1990. *The early alphabet*. (Reading the past). London: British Museum. ISBN 0-7141-8073-4
- Hilton, Susanne. 1982. *Oowekeeno oral traditions as told by the late Chief Simon Walkus Sr*. Ottawa: National Museums of Canada.
- Imprimerie Nationale. 1990. *Les caractères de l'Imprimerie Nationale*. Paris: Imprimerie Nationale Éditions. ISBN 2-11-081085-8
- Jensen, Hans. 1969. *Die Schrift in Vergangenheit und Gegenwart*. 3., neubearbeitete und erweiterte Auflage. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Kôno Rokurô, 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
- Macuch, Rudolf. 1969. *Grammatik des samaritanischen Hebräisch*. Berlin: Walter de Gruyter.
- Murtonen, A. 1964. *Materials for a non-Masoretic Hebrew Grammar III: A grammar of the Samaritan dialect of Hebrew*. (Studia Orientalia: 29) Helsinki: Societas Orientalis Fennica.
- Naveh, Joseph. 1997 (1987). *Early history of the alphabet: an introduction to West Semitic epigraphy and palaeography*. Jerusalem: The Magnes Press, The Hebrew University. ISBN 965-223-436-2
- Reichsdruckerei. 1924. *Alphabete und Schriftzeichen des Morgen- und Abendlandes, zum allgemeinen Gebrauch mit besonderer Berücksichtigung des Buchgewerbes. Unter Mitwirkung von Fachgelehrten zusammengestellt in der Reichsdruckerei*. Berlin: Reichsdruckerei.
- von Ostermann, Georg F. 1952. *Manual of foreign languages: for the use of librarians, bibliographers, research workers, editors, translators, and printers*. Fourth edition, revised and enlarged. New York: Central Book Company.

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 Samaritan encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

Figures

SAMARITANISCH.

Name	Zeichen	Wert	Ziffer	Name	Zeichen	Wert	Ziffer	Name	Zeichen	Wert	Ziffer
Aleph	Ⲁ	'a	1	Tet	Ⲅ	t	9	Pe	Ⲣ	p	80
Beth	Ⲃ	b	2	Yod	Ⲇ	y	10	Sade	Ⲉ	s	90
Gimel	Ⲅ	g	3	Kaph	Ⲋ	k	20	Quph	Ⲭ	q	100
Daleth	Ⲇ	d	4	Lamed	Ⲏ	l	30	Resh	Ⲱ	r	200
He	Ⲉ	h	5	Mem	Ⲓ	m	40	Sin	Ⲕ	š	300
Waw	Ⲋ	w	6	Nun	Ⲗ	n	50	Tau	Ⲙ	t	400
Zain	Ⲍ	z	7	Samež	Ⲑ	s	60		· :	Pausen	
Kheth	Ⲏ	χ	8	ʾAin	Ⲕ	ʾa	70		· · :		

Obiges Alphabet zeigt die schöne Form der samaritanischen Schrift in den Büchern, aus derselben entstand später eine cursivere Form, welche jedoch nur unbedeutende Unterschiede von der Druckschrift zeigt. Die samaritanische Schrift war in ältester Zeit

die gemeinsame Schrift der Juden. Der weggeführte Theil derselben änderte die Schrift im Exil und bildete dort die Quadratschrift aus, während die Samaritaner ihre Schrift, von der sie behaupten, sie sei die des Moses, beibehielten.

Figure 1. The Samaritan script, from Faulmann 1990 (1880), with Hebrew names, numeric value, and punctuation.

SAMARITANISCH

Name	Zeichen	Umschrift und Lautwert	Name	Zeichen	Umschrift und Lautwert
Aleph	Ⲁ	-, 'a	Lamedh	Ⲏ	l
Beth	Ⲃ	b [b, bh]	Mem	Ⲓ	m
Gimel	Ⲅ	g [g, gh]	Nun	Ⲗ	n
Daleth	Ⲇ	d [d, dh]	Ssamech	Ⲑ	s [β]
He	Ⲉ	h	Ajin	Ⲕ	'
Waw	Ⲋ	w	Pe	Ⲣ	p [p, ph]
Sajin	Ⲍ	z [s weich]	Ssadhe	Ⲉ	š [β]
Cheth	Ⲏ	h [ch]	Quph	Ⲭ	q
Teth	Ⲑ	t	Resch	Ⲱ	r
Jodh	Ⲓ	j	Schin	Ⲕ	š [sch]
Kaph	Ⲋ	k [k, kh]	Taw	Ⲙ	t [t, th]

Allgemeines. Der samaritanischen Schrift bedienen sich die Juden, bevor sie in die babylonische Gefangenschaft gingen. Dort bildeten sie dieselbe zur Quadratschrift aus. Das Alphabet besteht aus 22 Buchstaben. Die Schrift läuft von rechts nach links. Da dem samaritanischen Alphabet die Vokale fehlen und diese auch nicht durch Punkte, Striche oder Häkchen über oder unter den Buchstaben angedeutet werden (wie im Hebräischen oder Arabischen), nahm man zur Erleichterung des Lesens die folgenden Konsonanten zu Hilfe:

Ⲁ = a, e, Ⲉ = e, i,
 Ⲋ = a, Ⲋ = o, u.
 Ⲏ = a,

Interpunktion. Es erhält der letzte Buchstabe eines Wortes oben einen Punkt. · oder ' oder ·: am Ende eines Satzes, ·· am Ende eines Satztheils, =·: oder —<: am Ende eines Abschnitts, <·: =·: > am Ende eines Kapitels.

Die samaritanischen Typen wurden von der Buchdruckerei G. Kreyssing, Leipzig, zur Verfügung gestellt.

Figure 2. The Samaritan script, from the Reichsdruckerei 1924. It is worth noting that in this book the Hebrew script is given on a different page under a different rubric, showing the Square Script, Rashi, and "Weaver-German" variants, as well as German and Polish handwritten styles.

SAMARITAN

Name	Character	Transliteration and tone value	Name	Character	Transliteration and tone value
Aleph	Ⲁ	—, ’	Lamedh	Ⲍ	l
Beth	Ⲃ	b, bh	Mem	Ⲅ	m
Gimel	Ⲅ	g, gh	Nun	Ⲇ	n
Daleth	Ⲇ	d, dh	Samekh	Ⲉ	s
Heh	Ⲉ	h	Ayin	Ⲋ	.
Vau	Ⲋ	v, w	Pe	Ⲍ	p, ph
Zayin	Ⲍ	z, soft s	Sadhe	Ⲏ	š, s sharp
Cheth	Ⲏ	ḥ, ch	Koph	Ⲑ	q, k
Teth	Ⲑ	t	Resh	Ⲓ	r
Yod	Ⲓ	j	Shin	Ⲕ	š, sh
Caph	Ⲕ	k, kh	Tav	Ⲗ	t, th

This language is a dialect of the Aramaic of Palestine, the best examples of which are found in the literature belonging to the 4th century A.D., in which the alphabet derived from the old Hebrew was used. This had been used by the Jews up to the time of the Babylonian Captivity. The alphabet is still employed for writing Aramaic, Hebrew, and even Arabic. The literature is chiefly of a religious character.

The alphabet consists of 22 characters, and the text reads from right to left. Since there are neither vowels nor diacritical marks above or below the characters, the following consonants are employed as vowel characters:

Ⲁ = a, e,	Ⲏ = e, i,
Ⲉ = a,	Ⲋ = o, u.
Ⲋ = a,	

Punctuation

The last letter of a word is surmounted by a point; : or * or ·: are used at the end of a sentence; .. at the end of a phrase; =·: or —<: at the end of a paragraph; and <·:· = ·:·> at the end of a chapter.

Figure 3. The Samaritan alphabet, from von Ostermann 1954. This book is a handbook for librarians who need to identify and transliterate scripts. The glyphs, vowel samples, and punctuation all appear to have been taken from the Reichsdruckerei materials.

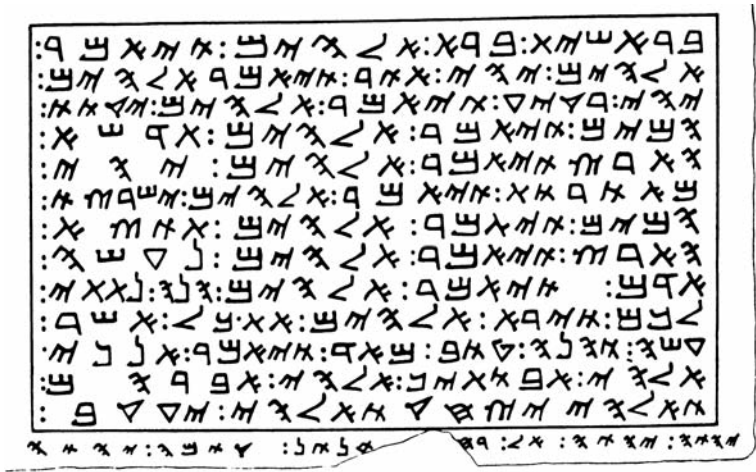


Figure 4. A Samaritan inscription from Naveh 1997. The punctuation marks [^]AFSAAQ and [^]NEQUDAA are shown.

TEXTE SAMARITAIN.

* *ḥṣrt* . *'al* *km'a* . *t'aw* *kyb'a* . *t'a* *dbk*
 .tu tueras Point .ta mère et ton père Honore

: *bngt* . *'al* *f'ant* . *'al*
 tu déroberas Point .tu commettras adultère point

: *kyhl'a* . *hwyl* : *ḥbzm* . *mš* . *tynbw*
 .ton Dieu à Yahwé un autel là et tu bâtiras

* *blk'ay* . *tlwḥ* . *ḥšrw* . *ḥšm* . *wnl* . *hwš* . *hrwt*
 .de Jacob de la communauté en héritage Moïse à nous a donné Une loi

Figure 5. A Samaritan text from Fossey 1948. The small WORD SEPARATION POINT is shown along with the larger punctuation marks : AFSAAQ and · NEQUDAA.

Handwritten Samaritan text in a medieval script, consisting of two columns of text. The characters are dense and feature various vowel signs above and below the letters.

Figure 7. Text from Healey 1990, showing text from a Samaritan Bible (Genesis 21:4–14), in a manuscript dating from the 13th century CE held in the Chester Beatty Library in Dublin (MS 751 27v).

〈図 10〉 母音記号つき五書テキスト (1982) とその発音 (創世記 1: 1-5)

Handwritten Samaritan text with numbered annotations (1) through (5) indicating specific vowel signs. The text is arranged in several lines, with some words separated by hyphens. The annotations are placed above or below the corresponding characters.

図の出典：Tsedaqah, Ratson (ed.)(1982), *Tōrah Tmīmāh* (Holon) より (市販せず)。

- (1) bārāsēt bārā ēluwṣəm it aššāmēm wit āraš. (2) wā:rəṣ ayyātā tē'u wbē'u wāšək 'al fāni
- ēluwṣəm amrā'ēfāt 'al fāni ammēm. (3) wyā'ūmēr ēluwṣəm yā'i or wyā'i or. (4) wyēre ēluwṣəm
- kī ṭob wyabdəl ēluwṣəm bin ā'or wbin āšək. (5) wyiqra ēluwṣəm lā'or yom wlāšək qārā lila
- wyā'i bēqar yom 'ād.

転写の出典：Ben-Hayyim (1957-77), Vol.4, p.353.

Figure 8. Text from Konô 2001 taken from Ratson Tsedaqah's 1982 edition of *Tōrah Tmīmāh*, showing Samaritan vowel signs.

〈表 2〉 サマリア文字の字体の変遷

半文字				完全文字							
13	12	11	10	9	8	7	6	5	4	3	2
𐤀	𐤁	𐤂	𐤃	𐤄	𐤅	𐤆	𐤇	𐤈	ʾ [ʔ]/∅	ʾ	a'laf
𐤉	𐤊	𐤋	𐤌	𐤍	𐤎	𐤏	𐤐	𐤑	b	b	bīt
𐤒	𐤓	𐤔	𐤕	𐤖	𐤗	𐤘	𐤙	𐤚	g	g	gā mān
𐤛	𐤜	𐤝	𐤞	𐤟	𐤠	𐤡	𐤢	𐤣	d	d	dā'lat
𐤤	𐤥	𐤦	𐤧	𐤨	𐤩	𐤪	𐤫	𐤬	∅ (')	h	īy
𐤭	𐤮	𐤯	𐤰	𐤱	𐤲	𐤳	𐤴	𐤵	w/b	w	bā
𐤶	𐤷	𐤸	𐤹	𐤺	𐤻	𐤼	𐤽	𐤾	z	z	zēn
𐤿	𐥀	𐥁	𐥂	𐥃	𐥄	𐥅	𐥆	𐥇	ʿ/[ʔ]/∅	ḥ	īt
𐥈	𐥉	𐥊	𐥋	𐥌	𐥍	𐥎	𐥏	𐥐	t [t]	t	tīt
𐥑	𐥒	𐥓	𐥔	𐥕	𐥖	𐥗	𐥘	𐥙	y [j]	y	yūt
𐥚	𐥛	𐥜	𐥝	𐥞	𐥟	𐥠	𐥡	𐥢	k	k	kāf
𐥣	𐥤	𐥥	𐥦	𐥧	𐥨	𐥩	𐥪	𐥫	l	l	lā'bat
𐥬	𐥭	𐥮	𐥯	𐥰	𐥱	𐥲	𐥳	𐥴	m	m	mīm
𐥵	𐥶	𐥷	𐥸	𐥹	𐥺	𐥻	𐥼	𐥽	n	n	nūn
𐥾	𐥿	𐦀	𐦁	𐦂	𐦃	𐦄	𐦅	𐦆	s	s	sin'gāt, sin'kāt
𐦇	𐦈	𐦉	𐦊	𐦋	𐦌	𐦍	𐦎	𐦏	ʿ/∅	ʿ	īn
𐦐	𐦑	𐦒	𐦓	𐦔	𐦕	𐦖	𐦗	𐦘	f/b	f	fī
𐦙	𐦚	𐦛	𐦜	𐦝	𐦞	𐦟	𐦠	𐦡	ṣ [s]	ṣ	ṣā'diy
𐦢	𐦣	𐦤	𐦥	𐦦	𐦧	𐦨	𐦩	𐦪	q	q	qūf
𐦫	𐦬	𐦭	𐦮	𐦯	𐦰	𐦱	𐦲	𐦳	r	r	rīs
𐦴	𐦵	𐦶	𐦷	𐦸	𐦹	𐦺	𐦻	𐦼	š [ʃ]	š	šān
𐦽	𐦾	𐦿	𐧀	𐧁	𐧂	𐧃	𐧄	𐧅	t	t	tāf

1. 現代ヘブライ文字 2. サマリア文字名 (現代音) 3. 転写 4. 音価 5. 古ヘブライ文字 (アラド, ラキシユ陶片) 6. 五書写本 (1215/6) 7, 11. 17 世紀中葉の写本 8, 12. 現代の写本 (タイプライター印字体) 10. サマリア・ヨシュア記写本 (1513)
出典: Ben-Hayyim (1957-77), Vol. 5, p. 265 を一部修正.

Figure 9. Text from Konô 2001 showing various examples of Samaritan inscriptional and book text, phonetic transcription and names, and Square Hebrew equivalents.

d The rest of the interpunctuation signs have a clear reference to the contents of the expression they mark, often emotionally accented. We enumerate them in the order they usually appear in Samaritan lists (*anged*, *áfsaq*, and *aná'u* precede in this order): 4) *arkánu* »submissiveness», consisting of an oblique line, the upper end toward the right /; it marks expressions of meekness and submission; 5) *šila* »question», consisting of the horizontal colon and an angle with point toward the left below it \angle ; it corresponds — in general — to our question-mark; 6) *zúqa* »shout, cry», consisting of a point and of a line after it, the latter normally in a slightly oblique position, falling toward the left \nearrow ; it marks expressions calling attention of human beings; 7) *atmá'u* »surprise», consisting of the normal colon and an angle (as in no. 5) after it $<$; it marks expressions of surprise; 8) *bá'u* »request, prayer», consisting of one point and the familiar angle $<$.; it marks humble petitions, above all prayers to God; 9) *zéf* »(burst of) anger», consisting of double no. 6 =; it marks expressions of vehemence and anger; and 10) *túru* »teaching», consisting of the familiar colon and of a vertical line after it |; it marks didactic expressions.

e The use of the above marks, however, is anything but consistent. As mentioned, even the appearance of *áfsaq* varies very remarkably between different mss. and — can be added here — even within one and the same ms. The same is the case with *anged*; moreover, among the oldest there are mss. in which it does not appear at all or does rather infrequently (to the latter belongs even the famous »holy» scroll attributed to Abiša). The rest of the signs seem to appear slightly later still, being perhaps invented by one man, at least a part of them. Their appearance remains rather infrequent throughout, except for some certain (different) signs in certain few mss.

f Single words are normally separated by means of a dot on the level of the middle of the normal letters ·, where interpunctuation marks do not appear. In the oldest mss., as well as in a number of less skillfully written younger ones, the latter do not appear at all, and in them the dot usually is put even at the end of lines, where in the other mss. normally only the interpunctuation marks proper appear.

Figure 10b. Discussion of Samaritan punctuation from Murtonen 1964. Murtonen does not have adequate fonts for the punctuation characters.

סדרי מקרתה

עתקנן על מה עתקו ארשינן
 מן השבעים הזקנים הן סדרי
 מקרתה עשרה, והם:

פִּסְק נְגַד אֲנָהּ אֲרַכְנֵנוּ בְּעוֹ
 שְׁאֵלָה וְעִיקָה וְעַף תּוּרוֹ
 אֲתַמְהוֹ

מדע מכשב ופם מיתב לכל חדה מגון :
 יתברך מן דאמר: ולמדה את בני ישראל שימה
 בפיהם בגלל אלפן ק ש ט ה

Figure 11. Discussion in Hebrew of Samaritan punctuation marks. Shown are, from right to left, : AFSAAQ, ° ANGED, ° ANNAAU, [/] ARKAANU, ° BAU, ° SHIYYAALAA, — ZIQAA, =: ZAEF, !: TURU, and √: ATMAAU.

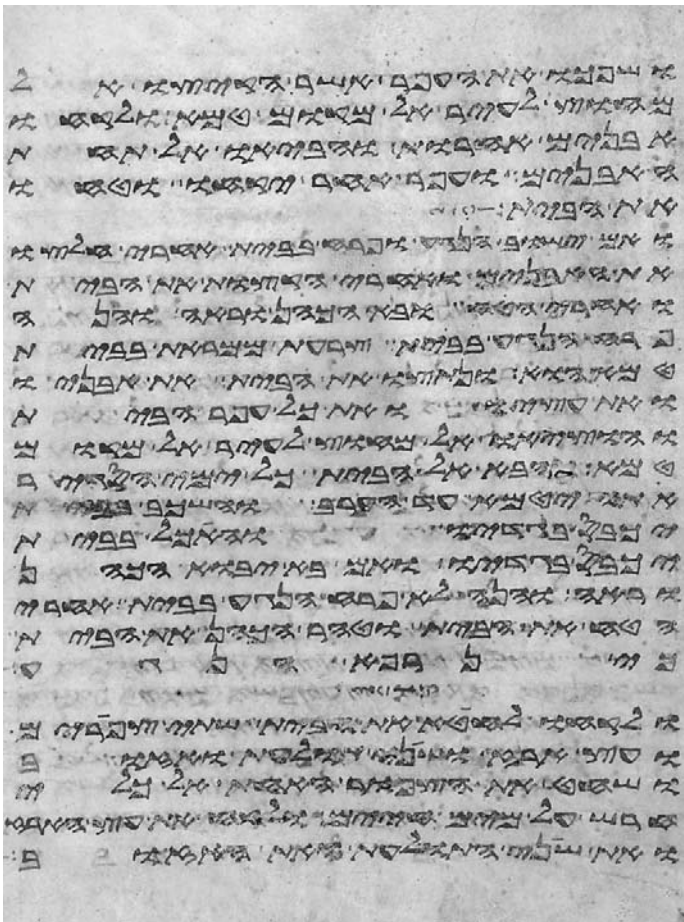


Figure 12. Samaritan manuscript 201 from Ashqelon, Israel, CE 1189. The text shown is Leviticus.



Figure 13. A Samaritan manuscript. Here the WORD SEPARATION POINT is used between words, and NEQUDAA is used at the beginnings of some lines in front of AFSAAQ ∴ and at the end of some lines after AFSAAQ ∴.

אֱלֹהִים יִבְרָא אֶת-הַשָּׁמַיִם וְאֶת-הָאָרֶץ׃ וְהָאָרֶץ חֲדָשָׁה וְרֵקָה׃ וְרוּחַ אֱלֹהִים מְעַבְּרָה עַל-פְּנֵי הַמַּיִם׃ וְיְהוָה אֱלֹהִים בְּרָא אֶת-הָאָדָם׃

וְיְהוָה אֱלֹהִים בְּרָא אֶת-הָאָדָם׃ וְיְהוָה אֱלֹהִים בְּרָא אֶת-הָאָדָם׃ וְיְהוָה אֱלֹהִים בְּרָא אֶת-הָאָדָם׃

וְיְהוָה אֱלֹהִים בְּרָא אֶת-הָאָדָם׃ וְיְהוָה אֱלֹהִים בְּרָא אֶת-הָאָדָם׃ וְיְהוָה אֱלֹהִים בְּרָא אֶת-הָאָדָם׃

Figure 15. A page from the Book of Genesis.

Row 08: SAMARITAN DRAFT

	080	081	082	083
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
A				
B				
C				
D				
E				
F				

hex	Name
00	SAMARITAN LETTER ALAF
01	SAMARITAN LETTER BIT
02	SAMARITAN LETTER GAMAN
03	SAMARITAN LETTER DALAT
04	SAMARITAN LETTER IY
05	SAMARITAN LETTER BAA
06	SAMARITAN LETTER ZEN
07	SAMARITAN LETTER IT
08	SAMARITAN LETTER TIT
09	SAMARITAN LETTER YUT
0A	SAMARITAN LETTER KAAF
0B	SAMARITAN LETTER LABAT
0C	SAMARITAN LETTER MIM
0D	SAMARITAN LETTER NUN
0E	SAMARITAN LETTER SINGAAT
0F	SAMARITAN LETTER IN
10	SAMARITAN LETTER FI
11	SAMARITAN LETTER TSAADIY
12	SAMARITAN LETTER QUF
13	SAMARITAN LETTER RISH
14	SAMARITAN LETTER SHAN
15	SAMARITAN LETTER TAAF
16	SAMARITAN MARK IN
17	SAMARITAN MARK IN-ALAF
18	SAMARITAN MARK OCCLUSION
19	SAMARITAN MARK DAGESH
1A	SAMARITAN MODIFIER LETTER EPENTHETIC YUT
1B	SAMARITAN MARK EPENTHETIC YUT
1C	SAMARITAN VOWEL SIGN LONG E (fatha al-nida)
1D	SAMARITAN VOWEL SIGN E
1E	SAMARITAN VOWEL SIGN OVERLONG AA (fatha al-ima)
1F	SAMARITAN VOWEL SIGN LONG AA
20	SAMARITAN VOWEL SIGN AA
21	SAMARITAN VOWEL SIGN OVERLONG A (fatha al-iha)
22	SAMARITAN VOWEL SIGN LONG A
23	SAMARITAN VOWEL SIGN A
24	SAMARITAN MODIFIER LETTER SHORT A
25	SAMARITAN VOWEL SIGN SHORT A (fatha)
26	SAMARITAN VOWEL SIGN LONG U (damma)
27	SAMARITAN VOWEL SIGN U
28	SAMARITAN MODIFIER LETTER I
29	SAMARITAN VOWEL SIGN LONG I (kasra)
2A	SAMARITAN VOWEL SIGN I
2B	SAMARITAN VOWEL SIGN O
2C	SAMARITAN VOWEL SIGN SUKUN
2D	SAMARITAN MARK NEQUDAA
2E	(This position shall not be used)
2F	(This position shall not be used)
30	SAMARITAN PUNCTUATION NEQUDAA
31	SAMARITAN PUNCTUATION AFSAAQ
32	SAMARITAN PUNCTUATION ANGED
33	SAMARITAN PUNCTUATION BAU
34	SAMARITAN PUNCTUATION ATMAAU
35	SAMARITAN PUNCTUATION SHIYYAALAA
36	SAMARITAN ABBREVIATION MARK
37	SAMARITAN PUNCTUATION MELODIC QITSA
38	SAMARITAN PUNCTUATION ZIQAA
39	SAMARITAN PUNCTUATION QITSA
3A	SAMARITAN PUNCTUATION ZAEF
3B	SAMARITAN PUNCTUATION TURU
3C	SAMARITAN PUNCTUATION ARKAANU
3D	SAMARITAN PUNCTUATION SOF MASHFAAT
3E	SAMARITAN PUNCTUATION ANNAAU
3F	(This position shall not be used)

A. Administrative

1. Title

Proposal to add the Samaritan alphabet to the BMP of the UCS

2. Requester's name

UC Berkeley Script Encoding Initiative (Universal Scripts Project); Authors: Michael Everson and Mark Shoulson

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

Liaison contribution.

4. Submission date

2008-01-25

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

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

Samaritan.

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

1d. Name of the existing block

2. Number of characters in proposal

61.

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 A.

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.

No.

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.)?

Yes.

2b. If YES, with whom?

Alan Crown, Osher Sassoni, Benny Tsedaka

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?

Ecclesiastical and cultural communities.

4a. The context of use for the proposed characters (type of use; common or rare)

Characters are used to write the Samaritan language.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In Israel and the West Bank by Samaritans; also by scholars, ecclesiastical researchers, and librarians.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap; RTL script with modern use.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

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)?

Yes.

11b. If YES, is a rationale for such use provided?

No.

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?