ISO/IEC JTC1/SC2/WG2 N3291
L2/07-231
2007-07-27

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 add the Samaritan alphabet to the BMP of the UCS
Source: Michael Everson & Mark Shoulson
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2007-07-27

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-familar 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 (*A.B.*) is published in Israel in Samaritan script (along with short articles in Hebrew and Arabic).

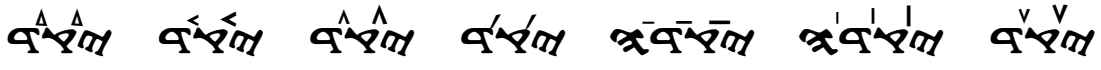
2. Corpus. There are some hundreds of Samaritan manuscripts; one of the largest collections is in the John Rylands University Library at the University of Manchester, including 377 items on parchment and paper. Samaritan MSS 1-27 were acquired in 1901 with the Crawford collection and include what is 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.

	◌ *ê* LONG E	◌ *e* E	◌ *û* LONG U	◌ *u* U
◌ *å:* OVERLONG AA	◌ *å* LONG AA	◌ *å* AA	◌ *î* LONG I	◌ *i* I
◌ *â:* OVERLONG A	◌ *â* LONG A	◌ *a* A		◌ *o* O
		◌ *ă* SHORT A		

These vowel signs are combining characters, each effectively centred between the base letter it follows and the following letter (if any). Examples using the letters YUT, QUF, DALAT, and IY, reading, from right to left *yêqed*, *yå:qåddåh*, *yâ:qâdah*, *yăqăd*, *yûqud*, *yîqid*, *yoqod*:



Other marks are centred over the base letter. Examples are SUKUN, DAGESH, OCCLUSION, and NEQUDAA, reading from right to left *yâqdah*, *yêqqed*, *ḥåbbåh*, *yûq̇ud*:

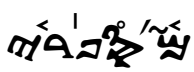
The SUKUN indicates that no vowel follows the consonant; DAGESH indicates consonant gemination; NEQUDAA is an editorial mark which indicates that there is a variant reading of the word. The mark for OCCLUSION "strengthens" the consonant, as here where *w* becomes *b*. Note that in the example

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 ⸗ש̄ב *îšab* 'Esau' contrasts with ⸗שב *'âšu* 'they made'. (Obviously with full pointing the words can be otherwise distinguished.)

The two marks IN and IN-ALAF, which are used to indicate the presence of [ʕ] (Samaritan *'in*, Hebrew *'ayin*) are drawn to the right side of their base letter, reading from right to left *''ḥyk*, *ḥ'ayyåh*:

Issue 1: EPENTHETIC YUT. In some cases, more than one combining mark can be applied to a base letter in Samaritan. The characters affected appear to be vowel signs and the EPENTHETIC YUT. The typical behaviour of the EPENTHETIC YUT is like that of DAGESH: it appears centred atop its consonant (evidently only ⸗ ALAF, ⸗ IY, ⸗ IT, and ⸗ IN (Hebrew *alef, he, ḥet, ayin*) as in the following, reading from right to left, *mi'yyăl*, *mihyyåḥelåk*, *miḥyyowṭ*, *mi'yyåḥûriy*, *bâ'yyår*, *miyyăsfåriy*:

The representation of the last example (*miyyăsfåriy*) is problematic. The EPENTHETIC YUT appears to be spacing here—but why? We have this attested in an edition of the Samaritan Pentateuch, but—since it really does look pretty bad—it could be the result of a shortcoming in the font. We have seen that, typically, *when more than one mark co-occurs with a base consonant, one of them is centred above the base letter and the second takes its place centred between the two letters*. This is easily accomplished when marks like DAGESH centre naturally and the vowels rest to the left. But it may be the case that the font here does not "know" how to centre a vowel sign when followed by an EPETHETIC YUT, and does not "know" to position the EPENTHETIC YUT like a vowel when preceded by another vowel. If the analysis here is correct, the better representation of *miyyăsfåriy* would be as shown below, with VOWEL SIGN I centred over MEM and EPENTHETIC YUT filling the space between the MEM and what follows:



Now, there is still some space added before the ZEN because there are three combining marks here (VOWEL SIGN I + EPENTHETIC YUT + VOWEL SIGN SHORT A), but it certainly makes the wordform less unusual. (It may be the case that the third of these is a syllable-inital spacing MODIFIER LETTER SHORT A; see Issue 2 below for more on that.) But the question of multiple vowels on a base consonant is raised by other examples as well, as in the following, reading from right to left, *'e-umer* 'I will say', *hå-inšem* 'the women':

Again, if the analysis above is correct, these ought to be written thus, with the first vowel sign centred over the base consonant (⸗, ⸗) and the second vowel sign centred between it and the following consonant (⸗, ⸗):

Issue 2: vowels in word-initial and syllable-initial position. Two vowels are known to occur in initial position, before the base character, as in מלפני *ălfåniy* 'before' and מישי *inšiy* 'wives of…'. These are encoded as spacing modifier letters (the exact position to be determined by the font) because combining characters cannot occur in initial position in a word. Users concerned with spoofing possibilities should note the similarity between ˊ MODIFIER LETTER SHORT A and ◌ˊ VOWEL SIGN SHORT A and between ˏ MODIFIER LETTER I and ◌ˏ VOWEL SIGN I.

One suggestion to avoid the spoofing problem would be to use NBSP or NNBSP as a base for these pre-consonantal vowels. This would be unnatural to users of the script. An analysis which would insert NBSP or NNBSP into the set of words at the end of Issue 1 above would not make sense morphologically. אמר *umer* is Hebrew אמר *omer*; the N(N)BSP would interfere with the triliteral root אמ-ר. (It is also improbable that the user community, which is rather small apart from scholars, will require Samaritan script in IDN.)

The argument for a spacing MODIFIER LETTER SHORT A is rather stronger however. As shown above in מיעספרי *miyyăsfåriy* 'from my book' (where *mi-* is 'from'), the syllable initial *ă* has the same function as it would in the word אספרי *ăsfåriy* 'my book'. Moreover, the MODIFIER LETTER SHORT A has an additional function, when used following a letter used numerically, to indicate the thousands, so גˊ = 3000. This is similar to the use of HEBREW PUNCTUATION GERESH for the same thing (׳ג = 3000). It should be noted that GERESH serves a number of functions. It modifies the sound of a letter (ג׳ירפה *ğirafa* 'giraffe'; צ׳יפס *tšips* 'chips'); it marks abbreviations (מס׳, short for מספר *mispar* 'number'); and in transliterations of Samaritan GERESH is used for the syllable initial *ă*, as in נג׳ד for אנגד *ănged*. The MODIFIER LETTER SHORT A—perhaps by another name?—similarly has multiple functions (though the ◌ˀ ABBREVIATION MARK is used with abbreviations in Samaritan).

The suggestion that a single vowel be represented by two different characters is not all that unusual, as similar things occur in other scripts encoded in the UCS. The analogies are not perfect, but מישי *inšiy* 'wives of…' could be written इन्शिय् *inśiy* in Devanagari (with इ and ि) or indeed directly transliterated (in principle) as *ⁱnšy* in Latin, using U+2071 SUPERSCRIPT LATIN SMALL LETTER I and U+0365 COMBINING LATIN SMALL LETTER I. (Macuch uses U+1D49 MODIFIER LETTER SMALL E for schwa when he writes *Qᵉrē*.)

The 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 *ṁ ṅ ḷ ẏ ẇ*. Among the vowels, plain vowels *əm ən əl i u* contrast with two sets of glottalized vowels: *əmṁ ənṅ əlḷ iẏ uẇ* are used when any other vowel follows, and *əm' ən' əl' i' u'* are used word-finally or when an obstruent follows. Compare *ğəm's* 'to lie on the ground' with *ğəmṁìs* 'to lie on the beach'.

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 ב·מ *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 *áfsaq* and *ăfsåq*. 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 *ánged* and *ănged*.)
- The ·ר 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 *ba'uw*.)
- The :ר 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 *'atmâhuw*.)
- The ̈ 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 *šiyy'ålåh*. In the Markeh Shameri font SHIYYAALAA is named "question".)
- The ̇ 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 *zi'yqåh*.)
- 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 Pentatuech. It has many glyph variants. One important variant differs significantly from any of the others; this is the —ר MELODIC QITSA which is used to indicate the end of a sentence "which one should read melodically". Together with : AFSAAQ as —ר: 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 *qiṣṣåh*)
- The =: ZAEF 'outburst' marks expressions of vehemence and anger. (Both Murtonen and the back matter of the Samaritan Pentateuch describe ZAEF explicitly as *zêf* and *z'f*.)
- The ו: TURU 'teaching' marks didactic expressions. (Both Murtonen and the back matter of the Samaritan Pentateuch describe TURU explicitly as *tûru* and *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 *'arkånuw*.)
- 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 *'ännåhuw*. In the Markeh Shameri font ANNAAU is named—in error—"gutteral yut" but it stands next to "yut dagesha" which is the EPENTHETIC YUT.)
- The · WORD SEPARATION POINT is a small dot distinguished from the larger NEQUDAA which is final punctuation like the AFSAAQ. Fossey's example in Figure 5 shows this distinction. We have proposed SAMARITAN WORD SEPARATION POINT but this could be encoded in the Supplemental Punctuation block as a character for general use. The character should have the **Break Opportunity After** property. In the font *Markeh Shameri* used in the weekly newspaper *A.B.*, these two characters are kept distinct; the NEQUDAA glyph is named "semicolon" in this font, and the WORD SEPARATION POINT glyph is named "wordspace". Essentially this character functions exactly as SPACE does, except that it has a dot in it. (This is different from the HYPEHENATION POINT, which functions as a HYPHEN does, except that it has a dot shape and not a horizontal line shape. The HYPEHENATION POINT would not stretch in paragraph justification as SPACE does, and as the WORD SEPARATION POINT—or WORD SPACE WITH DOT?—does.)

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
—ः	←	—ः MELODIC QITSA + ः AFSAAQ
—·ः·ः	←	—· ZIQAA + ः AFSAAQ + · NEQUDAA + ः AFSAAQ
—·ः	←	—· ZIQAA + ः AFSAAQ
—ःः	←	—ः QITSA + ः ATMAAU
—ःः	←	—ः QITSA + ः SHIYYAALAA
ः	←	ः SOF MASHFAAT + ः 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 ः· BAU, ः ATMAAU, and ः 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 MARK EPENTHETIC YUT;Mn;230;NSM;;;;;N;;;;;
081B;SAMARITAN VOWEL SIGN LONG E;Mn;23;NSM;;;;;N;;;;;
```

```
081C;SAMARITAN VOWEL SIGN E;Mn;23;NSM;;;;;N;;;;;
081D;SAMARITAN VOWEL SIGN OVERLONG AA;Mn;23;NSM;;;;;N;;;;;
081E;SAMARITAN VOWEL SIGN LONG AA;Mn;23;NSM;;;;;N;;;;;
081F;SAMARITAN VOWEL SIGN AA;Mn;23;NSM;;;;;N;;;;;
0820;SAMARITAN VOWEL SIGN OVERLONG A;Mn;23;NSM;;;;;N;;;;;
0821;SAMARITAN VOWEL SIGN LONG A;Mn;23;NSM;;;;;N;;;;;
0822;SAMARITAN VOWEL SIGN A;Mn;23;NSM;;;;;N;;;;;
0823;SAMARITAN MODIFIER LETTER SHORT A;Lo;0;R;;;;;N;;;;;
0824;SAMARITAN VOWEL SIGN SHORT A;Mn;23;NSM;;;;;N;;;;;
0825;SAMARITAN VOWEL SIGN U;Mn;23;NSM;;;;;N;;;;;
0826;SAMARITAN VOWEL SIGN LONG U;Mn;23;NSM;;;;;N;;;;;
0827;SAMARITAN MODIFIER LETTER I;Lo;0;R;;;;;N;;;;;
0828;SAMARITAN VOWEL SIGN I;Mn;23;NSM;;;;;N;;;;;
0829;SAMARITAN VOWEL SIGN I;Mn;23;NSM;;;;;N;;;;;
082A;SAMARITAN VOWEL SIGN O;Mn;23;NSM;;;;;N;;;;;
082B;SAMARITAN VOWEL SIGN SUKUN;Mn;23;NSM;;;;;N;;;;;
082C;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;;;;;
083F;SAMARITAN WORD SEPARATOR POINT;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*. Jersualem: 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 e savants*. Nouvelle édition míse à jour à l'occasion du 21e 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 samaritanischan Hebräisch*. Berlin: Walter de Gruyter.

Murtonen, A. 1964. *Materials for a non-Masoretic Hebrew Grammer III: A grammar of the Samatiran 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 Mitwerkung 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). 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		'a	1	*Tet*		ṭ	9	*Pe*		p	80
Beth		b	2	*Yod*		y	10	*Ṣade*		ṣ	90
Gimel		g	3	*Kaph*		k	20	*Quph*		q	100
Daleth		d	4	*Lamed*		l	30	*Reš*		r	200
He		h	5	*Mem*		m	40	*Šin*		š	300
Waw		w	6	*Nun*		n	50	*Tau*		t	400
Zain		z	7	*Samekʼ*		s	60		. :	Pausen	
Kheth		χ	8	*ʾAin*		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		—, ʼ	Lamedh		*l*
Beth		*b* [*b*, *bh*]	Mem		*m*
Gimel		*g* [*g*, *gh*]	Nun		*n*
Daleth		*d* [*d*, *dh*]	Ssamech		*s* [*ß*]
He		*h*	Ajin		ʿ
Waw		*w*	Pe		*p* [*p*, *ph*]
Sajin		*z* [*s* weich]	Ssadhe		*ṣ* [*ß*]
Cheth		*ḥ* [*ch*]	Quph		*q*
Teth		*ṭ*	Resch		*r*
Jodh		*j*	Schin		*š* [*sch*]
Kaph		*k* [*k*, *kh*]	Taw		*t* [*t*, *th*]

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

Allgemeines. Der samaritanischen Schrift bedienten 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:

= *a*, *e*, = *e*, *i*,
= *a*, = *o*, *u*.
= *a*,

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

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		—, '	Lamedh		*l*
Beth		*b, bh*	Mem		*m*
Gimel		*g, gh*	Nun		*n*
Daleth		*d, dh*	Samekh		*s*
Heh		*h*	Ayin		'
Vau		*v, w*	Pe		*p, ph*
Zayin		*z*, soft *s*	Sadhe		*ṣ*, *s* sharp
Cheth		*ḥ, ch*	Koph		*q, k*
Teth		*ṭ*	Resh		*r*
Yod		*j*	Shin		*š, sh*
Caph		*k, kh*	Tav		*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:

= *a, e,* = *e, i,*
= *a,* = *o, u.*
= *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.

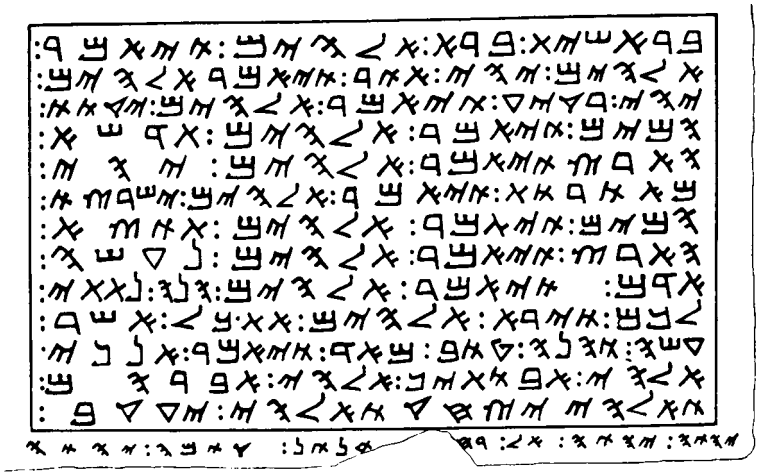


Figure 4. A Samaritan inscription from Naveh 1997. The punctuation marks ˂ AFSAAQ and ^ NEQUDAA are shown.

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	*hwhyl*	*ḥbzm*	*mš*	*tynbw*
.ton Dieu	à Yahwé	un autel	là	et tu bâtiras

bḳ'ay	*tlwḳ*	*hšrwm*	*hšm*	*wnl*	*hwṣ*	*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.

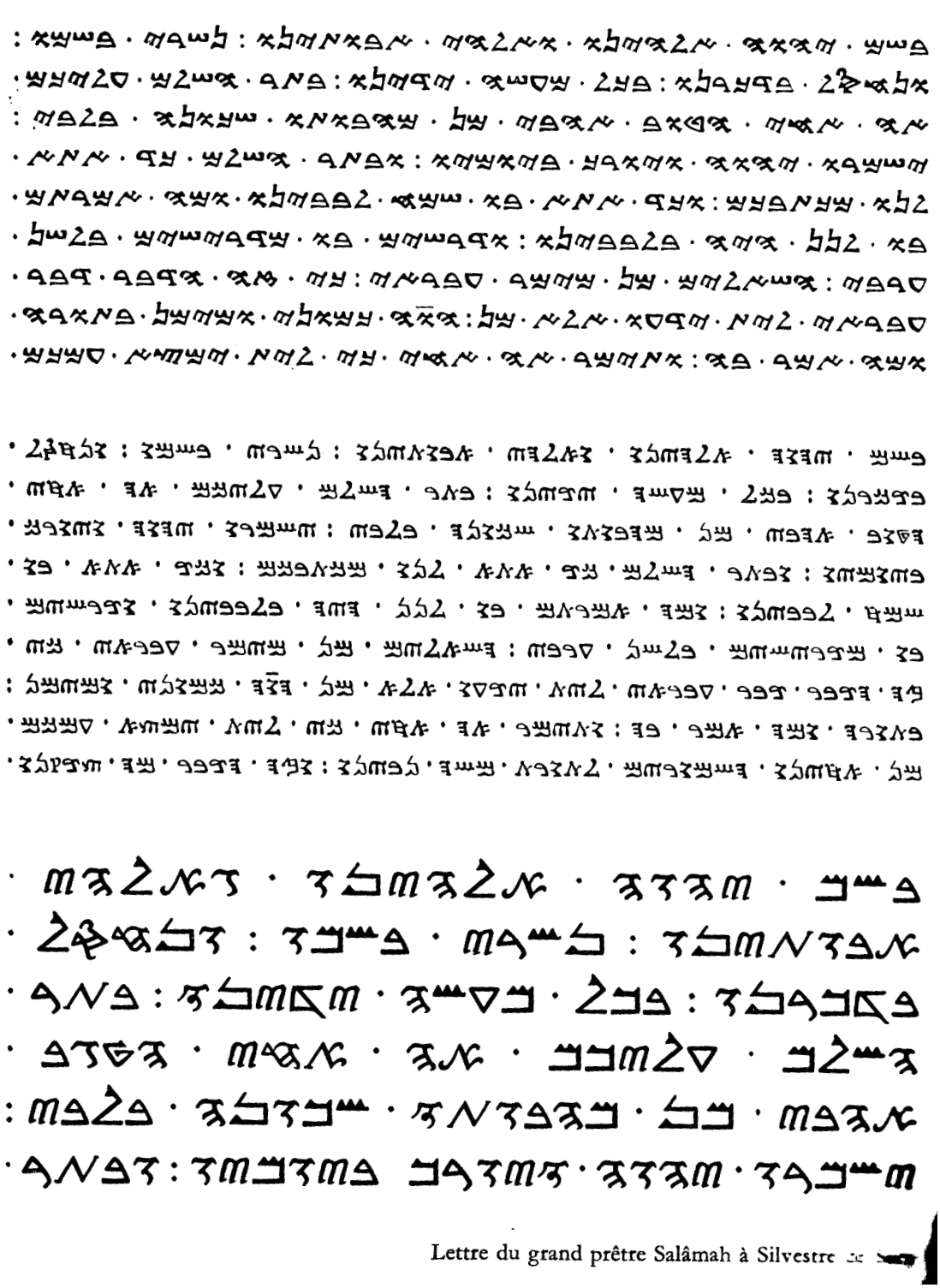


Figure 6. Sample text from the Imprimerie Nationale 1990, showing three styles and two sizes of Samaritan text; ⁝ AFSAAQ and · NEQUDAA are also shown.

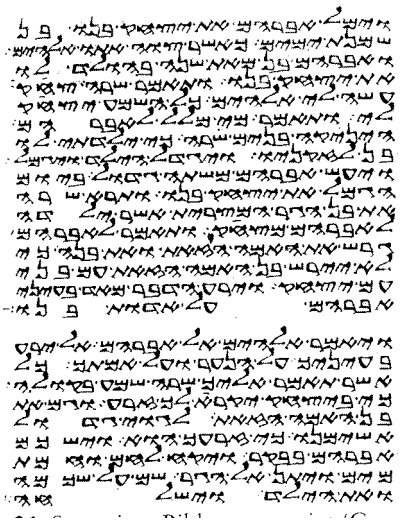
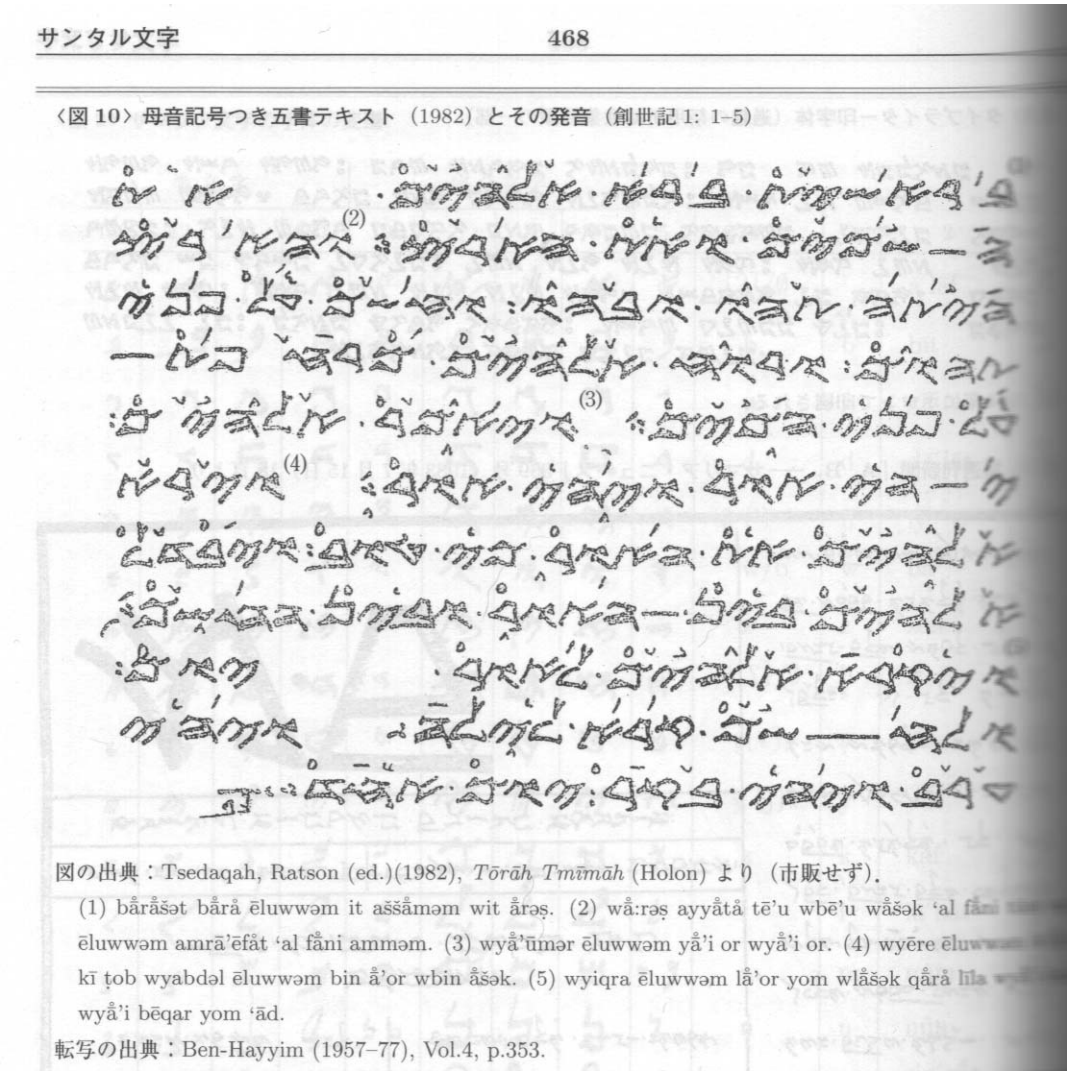


Figure 7. Text from Healey 1990, showing text from a Samaritan Bible (Genesis 21:4–14), in a manuscript dating from the 13th century CE held in the Chester Beatty Library in Dublin (MS 751 27v).



サンタル文字 468

〈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）

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

(1) båråšət bårå ēluwwəm it aššåməm wit åræṣ. (2) wå:ræṣ ayyåtå tē'u wbē'u wåšək 'al fåni
ēluwwəm amrā'ēfåt 'al fåni amməm. (3) wyå'ūmər ēluwwəm yå'i or wyå'i or. (4) wyēre ēluw
kī ṭob wyabdəl ēluwwəm bin å'or wbin åšək. (5) wyiqra ēluwwəm lå'or yom wlåšək qårå lila
wyå'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 *Tōrāh 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
									’ [ʔ]/∅	’	ā´lāf
									b	b	bīt
									g	g	gā´mān
									d	d	dā´lāt
									∅ (’)	h	īy
									w/b	w	bå̊
									z	z	zēn
									’/‘[ʕ]/∅	ḥ	īt
									ṭ[ɬ]	ṭ	ṭīt
									y [j]	y	yūt
									k	k	kå̊f
									l	l	lā´bāt
									m	m	mīm
									n	n	nūn
									s	s	sin´gå̊t, sin´kå̊t
									‘/’/∅	‘	īn
									f/b	f	fī
									ṣ[s]	ṣ	ṣå̊´dīy
									q	q	qūf
									r	r	rīš
									š [ʃ]	š	šān
									t	t	tå̊f

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（モアラド，ラキシュ陶片）　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.

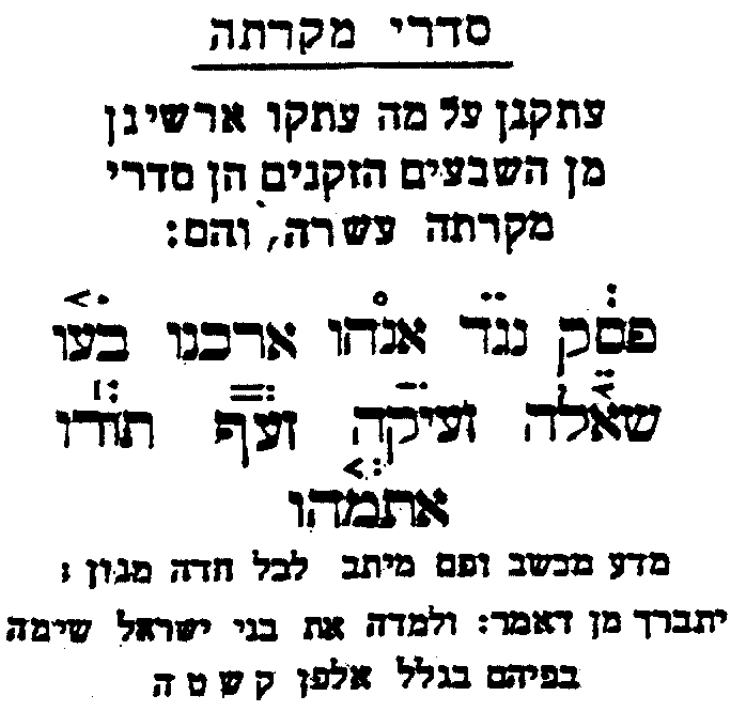
Figure 9. Text from Konô 2001 showing various examples of Samaritan inscriptional and book text, phonetic transcription and names, and Square Hebrew equivalents.

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 (*ánged*, *áfsaq*, and *anâ'u* precede in this order): 4) *arkânu* »submissiveness», consisting of an oblique line, the upper end toward the right /; it marks expressions of meekness and submission; 5) *šîla* »question», consisting of the horizontal colon and an angle with point toward the left below it <̈; it corresponds — in general — to our question-mark; 6) *zîqa* »shout, cry», consisting of a point and of a line after it, the latter normally in a slightly oblique position, falling toward the left /˙; it marks expressions calling attention of human beings; 7) *atmâ'u* »surprise», consisting of the normal colon and an angle (as in no. 5) after it < :; it marks expressions of surprise; 8) *bâ'u* »request, prayer», consisting of one point and the familiar angle <.; it marks humble petitions, above all prayers to God; 9) *zêf* »(burst of) anger», consisting of double no. 6 =:; it marks expressions of vehemence and anger; and 10) *tûru* »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 *ánged*; 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 Abîša). 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

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

פסק נגד אנהו ארכנו בעו
שאלה זיקה זעף תורו
אתמהו

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, ·ר BAU, ̈ר SHIYYAALAA, —· ZIQAA, =: ZAEF, ו: TURU, and ר: ATMAAU.

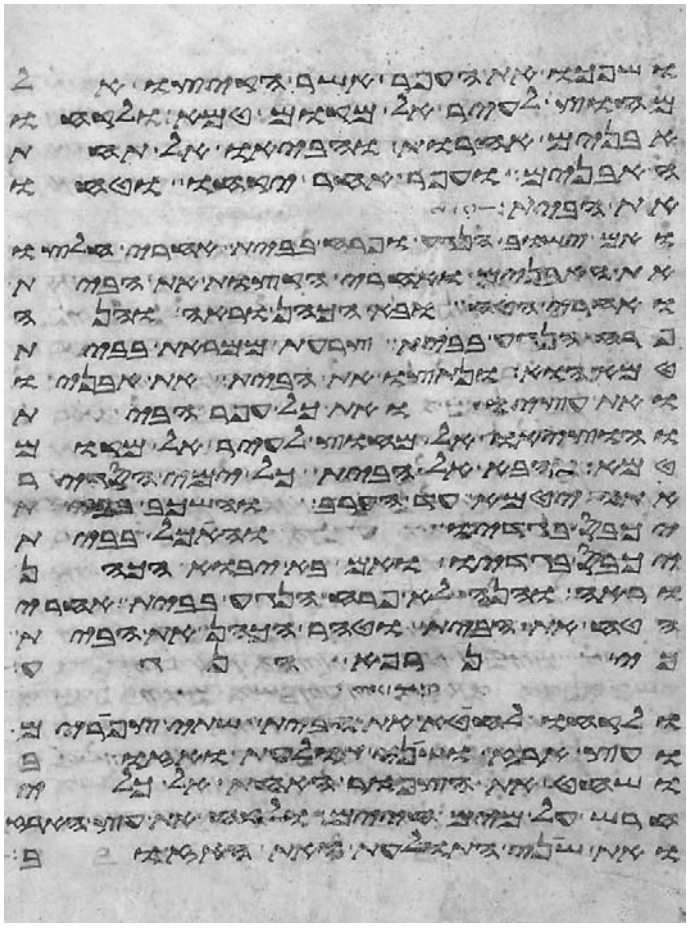


Figure 12. Samaritan manuscript 201 from Ashqelon, Israel, CE 1189. The text shown is Leviticus.

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 ·⁝.

אב"ד · שומרונים · חדשות השומרונים · גליון מס' 879-876

←

מבוא · מלחמת עולם

אחרי · מלחמת · העולם · הראשונה · נמצאו · אנשי · שומרון · בעמדה · קשה · מאוד · בין · הגרמנים · לבין · הבריטים · אנשי · שומרון · נאלצו · לעזוב · את · בתיהם · ולנוע · מצד · לצ

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 MARK EPENTHETIC YUT
1B	SAMARITAN VOWEL SIGN LONG E (fatha al-nida)
1C	SAMARITAN VOWEL SIGN E
1D	SAMARITAN VOWEL SIGN OVERLONG AA (fatha al-ima)
1E	SAMARITAN VOWEL SIGN LONG AA
1F	SAMARITAN VOWEL SIGN AA
20	SAMARITAN VOWEL SIGN OVERLONG A (fatha al-iha)
21	SAMARITAN VOWEL SIGN LONG A
22	SAMARITAN VOWEL SIGN A
23	SAMARITAN MODIFIER LETTER SHORT A
24	SAMARITAN VOWEL SIGN SHORT A (fatha)
25	SAMARITAN VOWEL SIGN LONG U (damma)
26	SAMARITAN VOWEL SIGN U
27	SAMARITAN MODIFIER LETTER I
28	SAMARITAN VOWEL SIGN LONG I (kasra)
29	SAMARITAN VOWEL SIGN I
2A	SAMARITAN VOWEL SIGN O
2B	SAMARITAN VOWEL SIGN SUKUN
2C	SAMARITAN MARK NEQUDAA
2D	(This position shall not be used)
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	SAMARITAN WORD SEPARATION POINT

A. Administrative

1. Title

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

2. Requester's name

Michael Everson & Mark Shoulson

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

Individual contribution.

4. Submission date

2007-07-27

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?