Elaine Keown
Tucson, Arizona
k_isoetc@yahoo.com

Fri, June 4, 2004

L2/04-181

REBUTTAL
to
"Final proposal for encoding the Phoenician script in the UCS"
ISO/IEC JTC1/SC2/WG2 N2746 L2/04-141 2004-04-26
http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2746.pdf

Note: This rebuttal responds to the above document, available online since April 28, 2004. It also responds in part to a slightly revised document, WG2-N2746R, which is still privately posted (as of Jun. 4, 2004).

When publicly available, N2746R2 will probably be found at: http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2746R2.pdf . On June 15, 2004, N2746R2 will be discussed and possibly voted on during the 'new scripts' session of the UTC / L2 meeting (in Markham, northeast of Toronto).

Contents of Rebuttal:

1. Introduction

In the 1990s, a Unicode Consortium member proclaimed:

" An A is an 𝗔 is an 𝒜 ….."

to express Unicode's **character / glyph model** in the simplest way.

The question for June 15, 2004, is,……..**can we say**:

" An aleph is an aleph is an aleph …." ?

Or do we need to say that some alephs, beths, gimels, etc. need their own separate encoding block?

Currently Unicode encodes Hebrew and Syriac in two separate blocks (there are also blocks for Ugaritic, Ethiopic, and Coptic).

In the 'Hebrew' block, the 'representative glyphs' are 'square Aramaic.' The Syriac block has 'Estrangelo' glyphs. However, Unicode Syriac documentation clearly says that the block is intended to cover other varieties of Syriac script.

The Phoenician proposals imply, among other implications, that

> **the Unicode 'Hebrew' block is inadequate to cover several 22-letter 'Early Linear Canaanite' versions of the Semitic alphabet**.

The Phoenician proposals also suggests that the proposed block cover earlier mostly 'pictographic' writing with 23 or more letters (guesses as to the number of letters varies with the scholarly study).

Unicode's Character / Glyph Model

The Unicode Consortium's text representation model is called the

'**character / glyph model'**.

The model tries to develop the idea of an 'abstract character' which has a pool of glyphs, all of which can be '**unified**' under the one abstraction, the '**character**.'

Hence, " an A is an 𝗔 is an 𝘈 ….." no matter how tall, curly, heavy, short, comical, or 'malformed' its particular glyph might appear. So this huge pool of 'A's will all be represented in the Universal Character Set computer code by one number (= one code point) for the computer to manipulate.

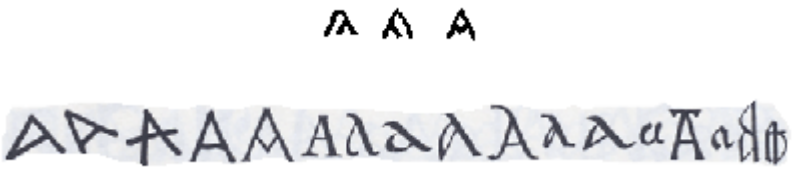


Figure 1 Two sets of 'A' glyphs

The character/glyph model has been applied to Chinese / Japanese / Korean / old Vietnamese characters (abbreviated CJKV), and to many other writing systems since the 1980s.

2. Four Rebuttal Positions

This rebuttal states that

> an aleph is an aleph is an aleph, since about 1200 B.C.E.

Note: Based on Akkadian and Ugaritic evidence, both the name ‘aleph’ and its sound predate aleph’s appearance as a regular ‘linear Canaanite’ letter.

Just after 1200 B.C.E., the 22-letter **very late** Proto-Canaanite writing (e.g., 'El-Khadr arrowheads) loses its pictographic character and becomes more readable. However, the direction of writing is still not stable. The epigrapher F. M. Cross observes that in the 11th century one sees the “fixed horizontal right-to-left habit of Early Linear inscriptions” (Cross 1980).

Four rebuttal positions:

- ‘Phoenician’ (early 22-letter, right-to-left ‘Linear Canaanite’) should not be encoded as a separate block. That is, standard ‘Phoenician’ should be regarded as a set of glyphs with no significant technical differences from ‘Hebrew.’ There are glyph differences, but they can be regarded as the usual variation in glyphs seen with Roman, Greek, or other scripts more familiar to the western eye than ‘Phoenician’ or other early Semitic glyphs.

- Also, Unicode should not encode scripts which are still being deciphered today. Certainly it should not do this under its current encoding practices. The revised ‘Phoenician’ proposal still suggests encoding the earlier Proto-Canaanite script, which is under lengthy study by epigraphers.

- Unicode should consider corpus size and the true needs of epigraphers when encoding new archaic ‘script’ blocks. For small ancient corpora, i.e., Wadi el-Hol, Proto-Sinaitic, Proto-Canaanite, standard encoding practices are actually irrelevant.

  A possible encoding block for these might be a PUA font for every symbol in the tiny corpus. For the Wadi el-Hol corpus--two inscriptions--that might mean a font of 28 units. For Proto-Canaanite—29 inscriptions, plus storage jar handles—a font of 207 items would cover the corpus.

- The so-called 'Phoenician numbers' should not be encoded as such. The numbers are originally from hieratic Egyptian, and are found around the Mediterranean (and possibly down the coast of West Africa, as well) in various better-known or obscure character sets.

  It would be better to designate an area within the Archaic Scripts block for archaic numbers, hopefully with excellent documentation.

3. Other Online Responses to N2746

Starting April 28, 2004, online discussion on Phoenician can be read at:

1) ANE discussion list
Thread: [ANE] Phoenician Unicode Proposal: Expert Feedback Requested
https://listhost.uchicago.edu/pipermail/ane/2004-April/012937.html
https://listhost.uchicago.edu/pipermail/ane/2004-April/012946.html

2) From the main Unicode online archives. To read online, when the dialogue box comes up, you fill in 'unicode-ml' in top line, and then 'unicode' in bottom.

a. April Proposal announcement:
http://www.unicode.org/mail-arch/unicode-ml/y2004-m04/0494.html .
b. Feedback in April/May:
http://www.unicode.org/mail-arch/unicode-ml/y2004-m05/0000.html ,
http://www.unicode.org/mail-arch/unicode-ml/y2004-m05/0003.html or
http://www.unicode.org/mail-arch/unicode-ml/y2004-m05/0154.html .
c. Discussion of proposed Phoenician numbers starts at:
http://www.unicode.org/mail-arch/unicode-ml/y2004-m05/1270.html

4. Context for Rebuttal

For general historical reference, I include a table of Semitic script information for the period just **before** the **revised** proposal (~3100 BCE – 1900 BCE).

SEMITIC SCRIPT OVERVIEW CHART 1
selected scripts, 3100 BCE – 1700 BCE

Script Name, Dates	Number of Items in Script	Script Direction(s)	Corpus Size, Material	When Deciphered
Akkadian cuneiform 3100 BCE-2nd CE / AD	~600	Left-to-right and right-to-left	**Huge**: **hundreds of thousands** of clay tablets (**500,000** in basement of British Museum alone)	1840s-present writing well understood
Tel el-Amarna Canaano-Akkadian cuneiform 1385 BCE-1355 BCE	?	Left-to-right ?	**250** clay tablets; diplomatic correspondence containing odd hybrid Canaano-Akkadian language, including One letter to **Tutankhamun** (!)	1900s-present
Wadi el-Hol 2000-1900 BCE	**23-27** proto-letters	**very variable**	**2** inscriptions (maybe 36 letters; see Altschuler)	still being deciphered--found 1990s
Other 'Pictographic' Writing ('Old Canaanite' or 'Proto-Sinaitic' 1800-1700 BCE	1. **27** ? (Albright) 2. **26** ? (Puech) 3. **23** ? (Colless)	**very variable**	**44** inscriptions in caves, on clay, on base of statues (sphinx)	ongoing since 1905

In a 1990 article in the journal Abr-Nahrain, Prof. Brian E. Colless makes the following observations about **Proto-Sinaitic**:

- only 1/3 of the letters can be deciphered with certainty
- there is "No set direction for the line of writing."

J. Naveh observed: "it would be premature to state that the Proto-Sinaitic inscriptions have been satisfactorily deciphered."

5. Rebuttal of Specific Proposal Passages

5.1 Section C. Technical - Justification, 4a:

> "The context of use for the Proposed characters……..
> Phoenician script is proposed to unify **Proto-Canaanite**, [emphasis mine]
> Punic, Phoenician proper, Late Phoenician Cursive,
> Phoenician papyrus, Siloam Hebrew, Hebrew seals, Ammonite,
> Moabite, Palaeo-Hebrew.

OBJECTIONS 1-2:
During the period between early Ugaritic and about 1100 B.C., most Canaanite languages acquired a simplified sound system. In its heyday, Ugaritic still had **27 consonants** and 27 consonantal letters. However, between the heyday of Ugaritic and the stabilizing of linear alphabetic script, most Canaanite languages lost 4-5 sounds. So the emerging linear alphabets 'shrank' over about 400 years as the sound systems simplified.

So-called 'Phoenician' is a **right-to-left**, **22-letter** script written in horizontal lines, used starting ca 1000 B.C.E. See Chart 2 for a description of some glyphs it is supposed to cover. The earlier stable alphabet period, 1200-1000 B.C. is frequently called '**Byblian**.'

1. Please note that Proto-Canaanite **has a larger repertoire than 'Phoenician'**.
2. In addition, Proto-Canaanite still has **variable direction** (vertical, boustrophedon, left-to-right, right-to-left, or all of these).

SCRIPT OVERVIEW CHART 2:
Covers period from 1700 – 900 BCE;
includes Ugaritic and Hebrew for comparison.

Script Name	Number of Letters in Script	Script Direction(s)	Corpus Size, Material	When Deciphered
Ugaritic	27 consonants 3 vowels	left-to-right; sometimes, right-to-left	**Large**; **hundreds** of clay tablets	ongoing since 1929
Proto-Canaanite	1. **23** letters 2. **>23** letters	**variable**	**29** inscriptions (total 183 letters) + Gezer storage jars (24 letters)	ongoing
'Phoenician'	**22** letters	right-to-left	a few longer inscriptions many short ones	mostly deciphered
Earliest Hebrew	**22** letters	right-to-left	some longer many short	mostly deciphered

OBJECTION 3:
Proto-Canaanite is not yet a deciphered script. As Prof. B. E. Colless (1991) observes:

"The same difficulties will confront us in Canaan as in Sinai: illegibility through damage or scribal incompetence; inconsistency in depicting the object represented by a particular pictograph; variation in the direction of writing; ambiguity caused by the lack of vowels and the dearth of punctuation ……(although word dividers are sometimes in evidence in early Canaanite inscriptions."

5.2 Section D. Proposal

Quotes from first and second paragraphs:
"The Phoenician alphabet and its successors ….

"The Phoenician alphabet is a forerunner of the Etruscan, Latin, Greek, Arabic, **Hebrew** [emphasis mine] , and Syriac scripts among others,"

"Phoenician is quintessentially illustrative of the historical problem of where to draw lines in an evolutionary tree of continuously changing scripts in use over thousands of years……The historical cut that has been made here considers the line from Phoenician to Punic to represent a single continuous branch of script evolution."

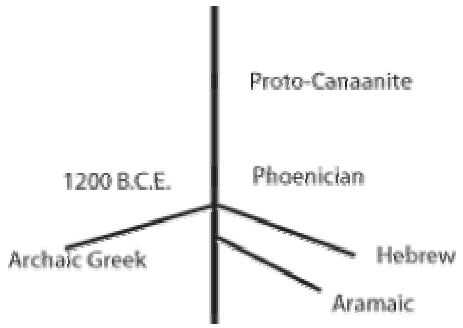
OBJECTION 4:
Recently Mr. Everson made it very clear that he is using the **underlying model of a tree** to represent Semitic alphabetic evolution. The tree model has been used since the 1800s. Mr. Everson is arguing above that

'**Phoenician' must be encoded because it is an important 'node' on the evolutionary tree of script development**.

Trees also come with an implied chronology; Mr. Everson is using a 'middle-of-the-road' chronology, where 'Phoenician' starts just after 1200 B.C. However, the primary evidence for early Semitic writing is the epigraphs themselves, viz:

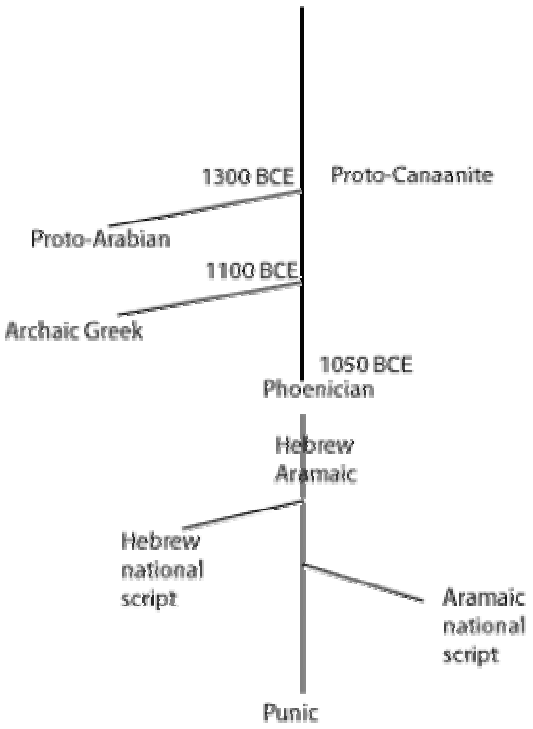
Ugaritic	Wadi el-Hol		Proto-Sinaitic		
Early Proto-Canaanite			Later Proto-Canaanite		
Very Late Proto-Canaanite			Byblian	Phoenician	
Ya'udic	Ammonite	Hebrew	Moabite	Aramaic	Deir 'Alla

From this primary evidence, Mr. Everson produces the derived script tree below, where so-called 'Phoenician' is a major tree node:



But in the broader literature, interpretations of Semitic script history vary greatly, viz:

Naveh (1987): "…..Proto-Arabian, which evolved from the Proto-Canaanite script about 1300 B.C.E., and the archaic Greek—about 1100 B.C.E. The Phoenician script is the direct offshoot (…) of Proto-Canaanite…. …independent Hebrew script….middle of ninth century….Aramaic branched off a century later….."



Cross (1980): "the issue in question is when a Hebrew national script broke away from the Old Canaanite or Early Linear Phoenician script, as the case may be."

Hooker (1990): "Thus three main West Semitic scripts emerged from the earlier **Byblian** [emphasis mine] linear alphabet. The primary one was the Phoenician, from which the Aramaic and Hebrew scripts are usually thought to be derived."

Millard (1976): "Now it is beyond all doubt that the Byblian, Palestinian, Phoenician, and Aramaic scripts are all related, the Byblian being the most archaic"

"The alphabet of 1000 B.C. was not an isolated phenomenon, it was one system of writing in a region that had known many, some for almost two millennia."

Colless (1988): "The Greek alphabet itself was borrowed from the original Semitic alphabet used by Phoenicians, Canaanites, Israelites, and Aramaeans,"

5.3 Section D. Proposal -- Processing

"Typical fonts for the Phoenician and especially Punic have very exaggerated descenders. These descenders help distinguish the main line of Phoenician evolution toward Punic from the other (*e.g.* Hebrew) branches of the script, where the descenders instead grew shorter over time."

OBJECTION 5:
This paragraph displays a serious misunderstanding of the development of Hebrew scripts. The letter forms which we now call 'finals' are actually the original earlier long letterforms used, for example, in writing all letters, no matter where they fall in the word, in the Persian period.

What happened after the Persian period is that the original very long kafs, mems, pehs and tsadis developed into **2** forms:
1) the original older long form remained long and became the 'final' forms of the letters used at the ends of Hebrew words today, encoded in Unicode as 4 distinct codepoints.

2) newer, shorter letterforms slowly developed, where the long tails did disappear and eventually became the 'medial' letter forms still used today (also encoded in Unicode).

5.4 Figures

"Figure 1….Table of Phoenician…"

The Shipitbaal inscription is actually 'Byblian', a Canaanite dialect from Byblos, north of Phoenicia proper. See Millard.

"Figure 2 ….(note inside says 'Phoenician inscription of Ahiram..')"
Ahiram inscription is also 'Byblian.'

> "Figure 13……..The Tetragrammaton in Phoenician script is
> Indicated with the large black arrow;…."

For all other publications, this figure would state that the Tetragrammaton is in Neo-Palaeo-Hebrew (S. Birnbaum).

> "Figure 14……the Tetragrammaton in Phoenician script
> alongside Greek text…."

Same as above---in all other publications, this would say Neo-Palaeo-Hebrew.

Letterforms (Figure 1)

Firmage, Richard. The Alphabet Abecedarium. David R. Godine, Publishers: Boston, 1993, p. 27.
Vine, Brent. Studies in Archaic Latin Inscriptions. Innsbrucker Beiträge zur Sprachwissenschaft: Innsbruck, 1993.

Bibliography and Links

Altschuler, Eric L. "Gloss of One of the Wadi el-Hol Inscriptions." ANES 39 (2002), pp. 201-204.
Birnbaum, S. The Hebrew Scripts.
Colless, Brian E. "Recent Discoveries Illuminating the Origin of the Alphabet." Abr-Nahrain vol. XXVI (1988). E. J. Brill: Leiden, pp. 30-67.
Colless, Brian E. "The Proto-Alphabetic Inscriptions of Canaan." Abr-Nahrain vol. XXIX (1991). E. J. Brill: Leiden, pp. 18-66.
Colless, Brian E. "The Proto-Sinaitic Inscriptions of Sinai." Abr-Nahrain vol. XXVIII (1990), pp. 1-52.
Cross, F. M. "Early Alphabetic Scripts." Symposia Celebrating the Seventy-Fifth Anniversary of the American Schools of Oriental Research (1900-1975). ASOR: Cambridge, 1979, pp. 97-111.
Cross, F. M. "Newly found inscriptions in Old Canaanite and Early Phoenician script." BASOR 238 Spring 1980, pp. 1-20.
Darnell, John Coleman. 2002. Theban Desert Road Survey in the Egyptian Western Desert 1, Gebel Tjauti Rock-Inscriptions 1-45 and Wadi el-Hol Rock Inscriptions 1-45, OIP 119. Chicago: Oriental Institute Press.
Hetzron, Robert. "Semitic Languages." International Encyclopedia of Linguistics, Vol. 3. Oxford University Press: Oxford, 1992, pp. 412-417.
Hooker, J. T. "Ancient Writing from Cuneiform to the Alphabet." Reading the Past. University of California Press: Berkeley, 1990, p. 222.
Logan, Robert K. The alphabet effect. William Morrow & Co.: New York, 1986.
Millard, Alan R. "The Canaanite Linear Alphabet and Its Passage to the Greeks." Kadmos 15 1976, p. 130-144.
Moran, W. A. (1992) The Amarna Letters. Baltimore: John Hopkins University.

Naveh, J. "Proto-Canaanite, Archaic Greek, and the script of the Aramaic Text on the Tell Fakhariyah Statue." Ancient Israelite Religion: Essays in Honor of Frank Moore Cross. Fortress Press: Philadelphia, 1987.
Naveh, J. Early History of the Alphabet: An Introduction to West Semitic Epigraphy and Palaeography. The Magnes Press: Jerusalem, 1982.
Sass, Benjamin. Studia Alphabetica: On the Origin and Early History of the Northwest Semitic, South Semitic, and Greek Alphabets. Universit
Vines, Brent. Studies in Archaic Latin Inscriptions. Innsbrücker Beiträge zur Sprachwissenschaft: Innsbruck, 1993.
Zaborski, A. "Afro-Asiatic Languages." International Encyclopedia of Linguistics, Vol. 1. Oxford University Press: Oxford, 1992, pp. 36-37.

Links

"Discovery of Egyptian Inscriptions Indicates an Earlier Date for Origin of the Alphabet" 1999, NYTimes on the Web:
http://www.library.cornell.edu/colldev/mideast/alphorg.htm

Kogan, L. and A. Militarev. Semitic etymological dictionary:
http://starling.rinet.ru/cgi-bin/query.cgi?flags=eygtnnl&basename=%5Cdata%5Csemham%5Csemet
(Part of SFI (Santa Fe Institute's) EHL "Evolution of Human Language" project, also funding from MacArthur Foundation, head: Sergei Starostin)

Wadi el-Hol inscription, on West Semitic Research Project site, USC
http://www.usc.edu/dept/LAS/wsrp/information/wadi_el_hol/inscr1.jpg