

**Response to the Proposal
to Encode Phoenician in Unicode**

Dean A. Snyder
8 June 2004

I am a member of the non-teaching, research faculty in the Department of Computer Science, Johns Hopkins University. My educational background includes 12 years of higher education in ancient and modern languages and linguistics, including 3 years of Classics at the University of Oklahoma and 5 years of graduate work in Comparative Semitics at the Oriental Institute of the University of Chicago. My employment history includes 12 years of commercial software engineering, including writing internationalization tools in C++ for a pioneering email application and networking and database integration layers for group-ware applications.

My job at Hopkins has allowed me to apply my computing interests to my research interests in ancient languages. I originated, organized, participated in, and secured funding for the Initiative for Cuneiform Encoding, the group of international cuneiform scholars and Unicode experts responsible for the current Sumero-Akkadian cuneiform encoding proposal to Unicode. I originated and am managing at Johns Hopkins University the Digital Hammurabi Project, which received a 3-year, \$1.65 million research grant from the National Science Foundation to develop both new hardware and new software to capture and render high-resolution 3D images of ancient cuneiform tablets and to complete the encoding of cuneiform.

I therefore write this response against the current proposal (N2746R2) to encode Phoenician in Unicode based on my experience in these two areas - research experience working in the native scripts of several languages and dialects of the ancient Near East and Mediterranean - Hebrew, Aramaic, Ugaritic, Akkadian, Syriac, Arabic, Phoenician, Moabite, Greek, and Latin - and software engineering experience in text processing and internationalization. Basically, I believe the needs to encode texts in the ancient Northwest Semitic scripts are well met by the current Hebrew block in Unicode.

As of right now, I have, in the "Canaanite" folder of my email application, 1480 emails devoted to the topic of encoding Phoenician, collected over the last few months from both the Unicode and Hebrew email lists and from private emails. This has been a very controversial proposal. I have read almost all of these emails more than once; and I'm afraid that I, myself, have been guilty of contributing a significant number of them to this torrent. The debate has become heated and personal at times, but I have striven to limit my involvement to the technical issues at hand. And so, trying not to spill onto you too much froth from this verbal stream, let me give you as succinctly as I can, what I believe

to be the salient arguments for and against a separate Unicode encoding for Phoenician [1].

[Responses against separately encoding Phoenician, most of them my own, but some gathered from others, follow immediately each numbered claim made below in support of a separate encoding.]

1) Phoenician is a separate script.

Phoenician is not a separate script. Phoenician is but one paleographic assemblage in the script continuum known as Northwest Semitic which also includes, for example, Imperial Aramaic and Jewish Hebrew diascripts [2].

These Ancient Northwest Semitic diascripts all:

- 1) have the same 22 (letter) characters,
- 2) in the same order,
- 3) with the same names;
- 4) are written in the same direction, right-to-left;
- 5) are used to write texts in dialects of the same or closely related languages;
- 6) are found in geographically contiguous areas;
- 7) are legible to the same ancient people;
- 8) and have been unified by Semitists for centuries.

The combination of all these traits forms a compelling argument that the Northwest Semitic diascripts form one script, being merely paleographic variants of one another.

2) Phoenician is an historically important node on the West Semitic script tree, with other scripts, most notably Greek, descending from it; Greek does not derive from Hebrew.

There is no doubt that Phoenician is historically important, but there are several lines of evidence supporting the view that the Greeks borrowed the alphabet much earlier than previously thought and borrowed it from some unknown, but non-Phoenician, Canaanite people.

3) Phoenician is illegible to modern Hebrew readers.

Legibility is not a determining factor in script differentiation. Respondents have mentioned, for example cyphers, Fraktur, Suetterlin, etc.

I would add that I can't imagine a more disparate set of glyphs than those of archaic cuneiform compared to, say, Neo-Assyrian cuneiform, and yet I know of no one who would argue for encoding them separately. (Archaic cuneiform is in fact, technically not even cuneiform at all, since the signs were not formed by making wedge-shaped impressions in clay, but rather by drawing the cursive lines of archaic pictograms on the clay.)

But the real problem is that the illegibility argument is based on **modern** Hebrew readers, whereas the encoding is aimed at **ancient** texts, and we have evidence that both Phoenician and Jewish Hebrew diascripts were legible to ancient Jews – for example, they employed both, contemporaneously, in writing Dead Sea scrolls, jar labels, coins, and documents.)

Furthermore there are even examples of the use of Paleo-Hebrew script in modern Israel on coins and public art.

4) Hebrew has a whole slew of "baroque signs" that do not apply to Phoenician.

The same can be said for the accents, minuscules, and symbols of modern Greek compared to ancient Greek, or modern European Latin scripts compared to Roman Imperial Latin. But no one wants to encode these separately even though their descendants are much more complex.

5) There are some modern users of Phoenician who would like to display and process Phoenician in plain text, in particular Indo-Europeanists and script historians. For example, online articles dealing with Greek and Phoenician, should show Phoenician letters, and not Hebrew.

This kind of request is a slippery slope into paleographic quicksand (aka glyph-based encodings). And these are precisely the sorts of things for which text mark-up was designed. Cuneiformists, too, will want to visually differentiate, for example, archaic cuneiform from Neo-Assyrian cuneiform; and **they** will be "forced" to use fonts.

6) If Phoenician is considered a glyphic variation of Hebrew, then it can also be considered a glyphic variation of Greek.

The relationship between Phoenician and Hebrew cannot be compared to the relationship between Phoenician and Greek:

Phoenician and Hebrew have the same 22 consonants; Greek added letters, dropped letters, and re-deployed letters.

Phoenician and Hebrew are written right-to-left; Greek was written right-to-left, left-to-right, and boustrophedon.

7) Jews believe that the Phoenician script is so different that its use in writing a Torah scroll rendered that scroll unfit for ritual use.

But Jews so regard **any** Hebrew script other than square Hebrew. Do we therefore want to separately encode, for example, cursive Hebrew? (Compare also the practise in Amish and Hutterite religious communities where the Bible is still printed in Fraktur instead of

Roman fonts.)

But what is really remarkable about this argument is that it is **proof** that the Phoenician script was legible to literate Jews - the argument was **not** that no one would be able to read it or even that it would be difficult to read, but rather that it was an **undesirable font**.

8) Phoenician lacks Hebrew's final forms.

Ancient Greek lacks word final sigma, along with all the minuscules - nevertheless we do not want to separately encode ancient Greek.

9) It's much harder in software to distinguish scripts that have been unified than to jointly process separated scripts.

The same thing could be said about Fraktur and Latin, or, for that matter, Neo-Assyrian and Old Babylonian cuneiform. The difficulty in separately processing plain text written in these scripts is not incentive enough to justify separately encoding them.

The classification of written materials for bibliographical use is different from the classification of writing systems for encoding. For a reader faced with the choice of locating a Fraktur or Roman edition of a German classic, or a Paleo-Hebrew or Jewish Hebrew edition of the Torah, having that information is clearly valuable and meaningful. On the other hand, for the purposes of encoding, being able to sort, transmit and search both texts the same way is more important and the distinction is properly relegated to fonts (i.e. to rich text) in this case.

10) Semitic scholars who don't want a separately encoded Phoenician in Unicode can simply ignore any text written in it, just as they now ignore multiple competing encodings and transliterations for Northwest Semitic scripts.

But the current practice among Semitists of using multiple encodings and transliterations for Northwest Semitic scripts is a problem all acknowledge. Adding yet another encoding to Unicode will only make matters worse. But what makes the issue more serious is that Unicode is forever, while we can be rid of the current problematic encodings in, say, 10 years. So why would we want to perpetuate problems like ones we are trying to eliminate now?

11) Semitists should not be the only group to decide what should be encoded; even the needs of amateur script enthusiasts should be taken into account.

Absolutely. But in making an informed and fair decision, the convenience that a separate encoding would give to script enthusiasts and even some scholars should be judiciously weighed against the negative impact a separate encoding would have on the scholarly

community that uses the script every day.

Semitists typically study the ancient documents in this script **together**, with a very high degree of cross-pollinating research going on. The languages are similar, the diascripts are similar, and there are pervasive political, cultural, linguistic, and religious interactions over the centuries between the geographically proximate users of this script. That's why they study them **together**, and why there are many degree-granting programs around the world in Northwest Semitics. It's the **combination** of these factors, along with others previously mentioned, that I believe inform a wise decision to keep these diascripts **together** in one encoding and not separate them out.

In the process of hashing out and reviewing these ideas on the Unicode email lists over the last several weeks, I came, at one point, to the conclusion that what was really needed was a separate encoding for Archaic Greek. After all, it is very different from classical Greek, in fact much more different than Phoenician is from Jewish Hebrew. The glyphs are very different; the writing direction was all over the place - right-to-left, left-to-right, boustrophedon; and the character inventories varied from place to place and time to time during the archaic period. Furthermore, encoded Archaic Greek would give to Indo-Europeanists a much more powerful tool than encoded Phoenician. Such a proposal would indeed be exceedingly simple to write, but in the end I rejected the idea for reasons similar to why I think separately encoding Phoenician is not a good idea. Why would I want to search for ΕΡΟΣ in two places when I could find it in one?

[1] Phoenician is an inaccurate name for the paleographic assemblage referenced in this proposal; a better name would be "Old Canaanite". Phoenician is but one of the languages written with this script variant, others being Old Hebrew, Old Aramaic, Ammonite, Moabite, and Edomite. Furthermore, even acknowledging that the ancient Greeks believed, as do some modern scholars, that they borrowed the foundation for their alphabet from the Phoenicians, there is evidence that this may not be correct, that the Greeks really borrowed the script much earlier from some non-Phoenician Canaanite peoples, giving thus another reason to call this paleographic assemblage "Old Canaanite". Nevertheless, for the sake of simplicity and to avoid being overbearing I will use the proposal's term "Phoenician" in my remarks.

[2] Diascript is to script as dialect is to language.