PROPOSED MODIFICATIONS
TO DELETE AND ADD SIGNS to N2798 = L2/04-189 Proposal for Cuneiform Encoding.
Expert contribution from Lloyd Anderson, Ecological Linguistics ecoling@aol.com

The additional signs proposed to be added here complete the major portion of common signs needed to satisfy the evidence concerning what are the functional independent units of the script. Choices of units to encode are not purely arbitrary, because they have consequences for complexity vs. simplicity of many aspects of implementation. The evidence for single-sign status comes (a) from actual cuneiform texts, the primary criterion, and (b) from the long scholarly tradition of sign lists. These two sources are in full agreement in almost all cases, primarily because the scholarly tradition of sign lists depends on the experts' implicit knowledge of usage in actual texts, and has had 150 years for rough spots to be ironed out.

What the standard tradition recognizes as single signs are almost without exception written closely together, while what the standard tradition recognizes as sequences of signs are written with considerable space between them when such space is available (in lines or indents with few signs, or more accurately, where the total of sign widths is less than the total space). This regularity persists from Early Dynastic through late Cuneiform. It is true of widely differing types of text, whether on stone or clay, whether "literary" or economic / administrative. I have reported on this fact with extensive examples. There has been no counter-argument. (An assertion was made in the past that there is no such thing as "careful" Cuneiform typography, even that the Gudea statues and Hammurabi's code, both carved on stone instead of clay, were not careful. The assertion was clearly false. Scholars recognize scribal errors vs. careful writing. A statement was made that there will be few or no continuous texts produced in Cuneiform. Yet the Finnish project has now printed three epics of Gilgamesh, Etana, and Anzu in Cuneiform, with very clear, even exaggerated spacing between signs and at least in general none within signs. This constitutes a "legacy" treatment consistent both with the long scholarly tradition and to a great degree also with the ancient practice in spacing Cuneiform texts.

It is known in sociolinguistics that meta-discussion about symbol use is far less consistent, far less reliable, and far less valid, than is unconscious use. While writing is on average less automatic than speaking, it is also more consistent when automatic than when consciously manipulated. Names of signs are much more meta-texts than they are like normal texts, and are accordingly not as good evidence (consider the AL SHESHIG which Steve Tinney notes is found named in a sign list as AL SHE).

With each proposed set of additions (one deletion) appear illustrations of the signs, in Neo-Assyrian font style or in the "classical" font style. At the end there are tables showing the high agreement among the various sign lists from the scholarly tradition.
Failures to provide an encoded character for independent functioning units of the script causes much default behavior of the characters to go awry.

(a) Spacing will be screwy, where it could be straightforward, extra space appearing between full signs, not between components of signs (of course understood: where there is room)

(b) It causes violations of the unification desired across time periods. Components of signs change historically in different ways than do the identical-appearing components functioning as independent signs. It is on the level of signs not components that equations are best drawn across time periods. UMBIN (which has now been accepted for encoding) maintains its identity as a sign across substantial time periods. But it does not maintain the identity of its component structure, that is reanalyzed and changes radically.

(c) Minimal contrasts will be violated if, even in texts with adequate spacing, there is no difference between SIGN U GUD SIGN and SIGN UL SIGN. The sign UL consists of two components U plus GUD, it is not a sequence of two signs. Just as one example. Or the example which Piotr Steinkeller discussed where spacing was used by experts to decide which of two distinct content readings was correct for a given passage (PA.DISH vs. GUR).

(d) Creation of fonts will entail substantial additional work if many of the signs have to be made up of parts with context-sensitive renderings [SIGN joined to SIGN] yielding another sign with a single and often irregular glyph. It is much simpler to encode what are known to be single signs from the outset.

(e) Searches for particular lexical content will be considerably more complex if signs are unnaturally decomposed, in particular if some of the signs proposed for addition here are decomposed, the only way to encode them without these signs. Users will normally not want to request a search for the sequence SIGN-1 SIGN-2 U and have the result contain also all sequences SIGN-1 SIGN-2 UL [where UL looks a bit like, but not exactly like, U-joined to GUD]. There are similar oddities in quite a number of cases. Most of them cannot be remembered by users as such a simple pattern as this one.

(f) There will no doubt be other hacks and kludges and patches necessitated by an encoding which is partly of individual signs, partly of sign components. Only a full and detailed study of implementation might let us know of most such problems. Even then, it is safer to simplify. Just as the working group decided early not to have dynamic composition of parts for larger numbers of signs of the type Container x Infixed components, so the compositions required by the current proposal are just as disadvantageous. Let's clean this up from the start.

Following are a relatively small number of single signs not yet provided for, whose addition will clean up most of the common instances which would cause problems.
Proposal 1.
12115 GISH TUG2 PI SHIR TENU SILA3 :: remove
after 122D7 CUNEIFORM SIGN SHIR :: add CUNEIFORM SIGN SHIR TENU

Comments. (a) Steve Tinney Agrees. He notes that <<Borger lists sequences SHIR-tenu SILA3 and GISH TUG PI SHIR2 TENU which support the necessity of this change.>>
(b) Such long sequences are almost certainly never single signs, even if we do not yet know the immediate constituency (how structurally to break them into parts which are functional constituents of their total sequence). They will therefore almost certainly end up deprecated, and it is better to eliminate them before they get into any standard.

***

Proposal 2.

Before 12058 CUNEIFORM SIGN DAG KISIM5 TIMES A PLUS MASH ::
Add CUNEIFORM SIGN DAG KISIM5
Before 12267 CUNEIFORM SIGN NUN LAGAR TIMES GAR ::
Add CUNEIFORM SIGN NUN LAGAR

<table>
<thead>
<tr>
<th>UTUA2</th>
<th>DAG KISIM5</th>
<th>B439</th>
<th>U00000</th>
</tr>
</thead>
<tbody>
<tr>
<td>TUR3</td>
<td>NUN LAGAR</td>
<td>B145</td>
<td>U00000</td>
</tr>
</tbody>
</table>

Comments: (a) Since DAG KISIM5 (also called UTUA2) exists as a unit into which other components can be infixed, as in 12058ff, and since NUN LAGAR (= TUR3) exists as a unit into which other components can be infixed, as in 12267ff, as standard linguistic deduction on the constituent structure of symbol strings indeed demonstrates, therefore these two combinations of components are presumptively functioning units even when there is no infix. This is not inconsistent with general principle 3.4. (c) Although the signs of the group (DAG KISIM5) x infixes are in part artificial, in vocabularies only, the sign TUR3 exists all the way through the time periods starting from arcaic Uruk. Just as the sign UMBIN discussed in a paper for the previous UTC, the early form of TUR3 clearly has the two components superfixed, not in sequence. Its early form is NUN x LAGAR, later NUN LAGAR (components adjacent, even touching, not overlapping), but the identity as single sign persists throughout.

Here and in what follows, single-morpheme names (UTUA2 and TUR3) are given alongside the names which refer to components of signs, rather than sequences of signs. (Signs proposed here for addition could also be ordered differently, or not, if the single-morpheme names are used.) At least most of the DIRI list is quite likely a naming of sign components, of single signs with particular readings, not a naming of sign sequences which have special idiomatic readings. Dr. Wolfgang Heimpel has confirmed that this interpretation is quite plausible.
Proposal 3.

After 122DC CUNEIFORM SIGN SHU2 :: add the following sequence of signs (illustrations and sign numbers from Borger MZL, further sign lists at end of proposal). Although the use of "TIMES" implying infix or overlap may seem surprising to many cuneiformists here, at least two early Uruk forms clearly have the vault of the night sky (origin of SHU2) surrounding the AN or the E2. The "roof" radical of Chinese characters also behaves structurally just like other radicals which surround enclosed parts. These could also be named SHU2 AN (etc.) or EN2 (etc.)

CUNEIFORM SIGN SHU2 TIMES AN
CUNEIFORM SIGN SHU2 TIMES AN THREE TIMES
CUNEIFORM SIGN SHU2 TIMES ASH2
CUNEIFORM SIGN SHU2 TIMES DUN2
CUNEIFORM SIGN SHU2 TIMES ESH
CUNEIFORM SIGN SHU2 TIMES NE
CUNEIFORM SIGN SHU2 TIMES UR SHESHIG

<table>
<thead>
<tr>
<th>EN2</th>
<th>SHU2 x AN</th>
<th>B870</th>
<th>U00000</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHUHUB, KUNGA</td>
<td>SHU2 x MUL</td>
<td>B872</td>
<td>U00000</td>
</tr>
<tr>
<td>GIBL2, KIBIR</td>
<td>SHU2 + ASH2</td>
<td>B875</td>
<td>U00000</td>
</tr>
<tr>
<td>SHUDUN</td>
<td>SHU2 x DUN4</td>
<td>B876</td>
<td>U00000</td>
</tr>
<tr>
<td>LIL3</td>
<td>SHU2 x ESH</td>
<td>B879</td>
<td>U00000</td>
</tr>
<tr>
<td>LIL5</td>
<td>SHU2 x NE</td>
<td>B874</td>
<td>U00000</td>
</tr>
<tr>
<td></td>
<td>SHU2 x UR-sheshig</td>
<td>B880</td>
<td>U00000</td>
</tr>
</tbody>
</table>

CUNEIFORM SIGN SHU4 TIMES BURU14
CUNEIFORM SIGN SHU4 TIMES DIM
CUNEIFORM SIGN SHU4 TIMES DIM TIMES KUR
CUNEIFORM SIGN SHU4 TIMES E2
CUNEIFORM SIGN SHU4 TIMES GA
CUNEIFORM SIGN SHU4 TIMES GAN
CUNEIFORM SIGN SHU4 TIMES GAR
CUNEIFORM SIGN SHU4 TIMES ITI
CUNEIFORM SIGN SHU4 TIMES MU
CUNEIFORM SIGN SHU4 TIMES SI GUNU
CUNEIFORM SIGN SHU4 TIMES URUDU
Comments: (a) These are all assigned sign numbers by Borger MZL and for about half of them also by most of the other sign lists across all time periods where the sign occurs in the lists. The verdict of the tradition is thus unanimous, so far as I am aware. Borger almost always distinguishes very carefully between sequences of signs (not assigned numbers. For IGI see MZL pp.187-8) and single signs (assigned numbers, MZL pp.189-190), showing different spacing in the two cases. Images of Borger's entries on MZL pp.188-189 are at the end of this paper. Though there may be some influence from prior tradition, the prior tradition is itself under the long-term persistent influence of actual text usage.

(b) Where attested in actual texts (the probative type of evidence), the portion SHU2 or SHU4 is never separated from the remaining wedges of the sign, not across "indents" (lines) within a "line" (frame), and not when additional space is used in the cuneiform form of justification or expanded text. (Not probative of single-sign status is meta-discussion of signs such as the naming of sign components. As noted earlier, in sociolinguistics it is known that normal usage is consistent, reliable, and valid while meta-discussions, on a more conscious level, are not. There is absolutely no reason to believe that this difference does not hold for cuneiform. Even if writing is not on average as fully unconscious as speaking, it can often be automatic.)
Proposal 4.
After 12254 \textsc{Cuneiform} \textsc{sign} \textsc{nim} \times \textsc{gar} \textsc{plus} \textsc{gan2} \textsc{tenu} ::
  Add \textsc{Cuneiform} \textsc{sign} \textsc{nin}
Or else name and order as \textsc{Cuneiform} \textsc{sign} \textsc{sal} \textsc{tug2} shown below.

After 122ad \textsc{Cuneiform} \textsc{sign} \textsc{sal} ::
  Add the following signs
\textsc{Cuneiform} \textsc{sign} \textsc{sal} \textsc{kur}

After 122ae \textsc{Cuneiform} \textsc{sign} \textsc{sal} \textsc{lagab} \times \textsc{ash2} ::
  Add the following signs
\textsc{Cuneiform} \textsc{sign} \textsc{sal} \textsc{lagar} (see comment (b) below)
\textsc{Cuneiform} \textsc{sign} \textsc{sal} \textsc{she3}
\textsc{Cuneiform} \textsc{sign} \textsc{sal} \textsc{tug2}
\textsc{Cuneiform} \textsc{sign} \textsc{sal} \textsc{ur} (see comment (b) below)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>\textsc{nin}</th>
<th>\textsc{sal} x \textsc{tug2} or \textsc{sal} x \textsc{kur}</th>
<th>\textbf{b886}, \textbf{b887}</th>
<th>\textbf{u00000}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\includegraphics[width=2cm]{image1.png}</td>
<td>\includegraphics[width=2cm]{image2.png}</td>
<td>\includegraphics[width=2cm]{image3.png}</td>
<td>\textbf{b890}</td>
<td>\textbf{u00000}</td>
<td></td>
</tr>
<tr>
<td>\includegraphics[width=2cm]{image4.png}</td>
<td>\includegraphics[width=2cm]{image5.png}</td>
<td>\includegraphics[width=2cm]{image6.png}</td>
<td>\textbf{b895}</td>
<td>\textbf{u00000}</td>
<td></td>
</tr>
<tr>
<td>\includegraphics[width=2cm]{image7.png}</td>
<td>\includegraphics[width=2cm]{image8.png}</td>
<td>\includegraphics[width=2cm]{image9.png}</td>
<td>\textbf{b897}</td>
<td>\textbf{u00000}</td>
<td></td>
</tr>
<tr>
<td>\includegraphics[width=2cm]{image10.png}</td>
<td>\includegraphics[width=2cm]{image11.png}</td>
<td>\includegraphics[width=2cm]{image12.png}</td>
<td>\textbf{b898}</td>
<td>\textbf{u00000}</td>
<td></td>
</tr>
</tbody>
</table>

Comments: (a) Borger MZL and sign lists for Fara, sometimes Ur III and Rosengarten, assign a single number to these signs, with the exception (b) that two signs are not attested in earlier stages. (c) Text spacing shows the sign \textsc{nin} is a single sign already in archaic Uruk (zero exceptions found in extensive searches). Plate No.1 in Biggs \textit{Abu Salabikh} (the time period of Fara) contrasts single-sign forms in the frames 1 and 9 below, either stretching the sign or leaving blank space, completely different spacing from the free combinations of separate signs in frames 2 to 8. (d) Using the sign name \textsc{nin} avoids having to determine its second component after \textsc{sal}, where there may still be room to discover the true original components (etymology).
Proposal 5.

After 12149 CUNEIFORM SIGN IGI ::

Add the following signs
CUNEIFORM SIGN IGI DUB
CUNEIFORM SIGN IGI E2
CUNEIFORM SIGN IGI ERIM
CUNEIFORM SIGN IGI MIN (see comment (b) below)
CUNEIFORM SIGN IGI NI
CUNEIFORM SIGN IGI RU
CUNEIFORM SIGN IGI SHE3
CUNEIFORM SIGN IGI UR
CUNEIFORM SIGN IGI UR TIMES KASKAL (see comment (b) below)

Comments: (a) Borger MZL and sign lists since Fara assign a single number to these signs, with the exception (b) that IGI MIN and IGI UR TIMES KASKAL are not listed in Fara, and IGI E2 is not listed in Rosenberg (Lagash) or in Schneider (Ur III). At least I have not yet managed to find them there. (c) Text spacing shows that the sign IGI RU (PAD3) is treated as a single sign (extensive text searches, overwhelming practice). (d) Given a Borger MZL reference somewhere to the duals of body parts, presumably like the assumed ligature here with MIN "two", the semantic specialization may mean that this "ligature" turns out to be always obligatory, never written in sequence, and is thus not a ligature but simply an irregular form like our ampersand "&" or like an irregular verb form "went" instead of "goed". A separate sign to be encoded. I don't know the answer to this question yet, I hope it is in other chapters of Borger MZL.
Concordance to Sign Numberings for Signs Proposed Here as Additions

Tables show where I have been able to identify each sign in traditional numbered sign lists. Since the universal practice has been to assign a number to those forms regarded as single signs, not to sign sequences or to mere components (fragments) of signs, these numbered lists can be taken as a default list of distinctive signs. Of course there are differences between different scholars, but those differences mostly concern the rarer signs or sign sequences, whichever they turn out to be. The omission of a sign from one list may also reflect lack of use in a particular scribal tradition, it need not say anything about the scholar's belief whether the form is a single sign or a sequence. Study of actual texts is real evidence which confirms the sign lists. In tables below, the sign names are not regularized, and group labels in larger print can be disregarded.

Borger's MZL list is especially extensive. Since he is also careful to distinguish between sign sequences (those with IGI as first sign see pp.187-188) and single signs (those with IGI as first component see pp.189-190), the MZL list is surely one of the most valuable.

<table>
<thead>
<tr>
<th>Author</th>
<th>List No.</th>
<th>Sign Number</th>
<th>Sign Name</th>
<th>Sign Location</th>
<th>Sign Meaning</th>
<th>Sign Type</th>
<th>Sign Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>In</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Covers; The Vault of the Night Sky

<table>
<thead>
<tr>
<th>List No.</th>
<th>Sign Number</th>
<th>Sign Name</th>
<th>Sign Location</th>
<th>Sign Meaning</th>
<th>Sign Type</th>
<th>Sign Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>In</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

L2/04-223R Anderson - 8 -
Here is how Borger's MZL distinguishes sign sequences in which IGI is the first full sign (p. 188) from complex signs having IGI as leftmost component, where IGI is not a full sign (p. 189). The scholarly tradition is systematic about this. The distinction needs to be made somehow, as it occurs in texts, not just in sign lists. It is not a matter of ligaturing or other formatting (formatting should not be making distinctions between significantly different text content). To merge these is very much like treating "bold" and "loolol" and "boolol" and "loold" as the same sequence of characters, except for formatting. We would not dream of encoding Latin English in such a way.