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 ilustrations of the signs, in Neo-Assyrian font style or in the "classical" font style. At the end appears a table showing the high agreeement 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.

(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 some of the signs proposed for addition here. 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 [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.

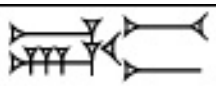
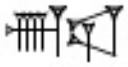
Proposal 1.
12115 GISH TUG2 PI SHIR TENU SILA3 :: remove
after 122D7 CUNEIFORM SIGN SHIR :: add CUNEIFORM SIGN SHIR TENU

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. (illustrations at the end of this set of proposals)

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

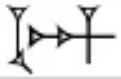
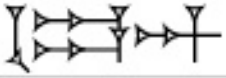
			UTUA2	DAG KISIM5	B439	Uoooooo
			TUR3	NUN LAGAR	B145	Uoooooo

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 archaic 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.

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.) The entire 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.

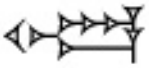
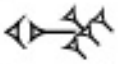
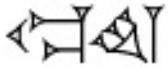
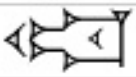
After 122DC CUNEIFORM SIGN SHU2 add the following sequence of signs (illustrations and sign numbers from Borger MZL follow at the end of this proposal). Although the use of "TIMES" implying infix or overlap may seem surprising to many cuneiformists here, at least two early uruk forms clearly have the vault of the night sky (SHU2) surrounding the AN or the E2. The "roof" radical of Chinese characters also behaves structurally just like other radicals which surround enclosed parts. At least two other naming systems could also be used for these.

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

			EN2	SHU2 x AN	B870	Uooooo
			SHUHUB, KUNGA	SHU2 x MUL	B872	Uooooo
			GIBIL2; KIBIR	SHU2 + ASH2	B875	Uooooo
			SHUDUN	SHU2 x DUN4	B876	Uooooo
			LIL3	SHU2 x ESH	B879	Uooooo
			LIL5	SHU2 x NE	B874	Uooooo
				SHU2 x UR-sheshshig	B880	Uooooo

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

After 12055 CUNEIFORM SIGN BUR2
  Add CUNEIFORM SIGN BURU14

				BURU14	B165	Uoooooo
			SHIBIR	SHU4 x BURU14	B666	Uoooooo
			GAKKUL3	SHU4 x DIM	B667	Uoooooo
			GAKKUL	(SHU4 x DIM) x KUR	B668	Uoooooo
			SHITA4	SHU4 x E2 early	B699	Uoooooo
			UTU2	SHU4 x GA	B700	Uoooooo
			SHAGAN	SHU4 x GAN	B684	Uoooooo
			PAD {ŠUG}	SHU4 x GAR	B746	U1227E
				SHU4 x ITI	B664	Uoooooo
			UDUN	SHU4 x MU	B665	Uoooooo
				SHU4 x (SI gunu)	B670	Uoooooo
			GUL	SHU4 x URUDU or DUB ?	B682	Uoooooo

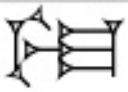
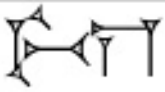
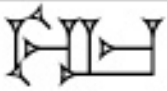
Comments: (a) These are all assigned sign numbers by Borger MZL and for about half of them also by most of the other sign lists across all time periods where the sign occurs in the lists. The verdict of the tradition is thus unanimous, so far as I am aware. Borger almost always distinguishes very carefully 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. 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 CUNEIFORM SIGN NIM TIMES GAR PLUS GAN2 TENU
Add CUNEIFORM SIGN NIN
Or else name and order as CUNEIFORM SIGN SAL TUG2 shown below.

After 122AD CUNEIFORM SIGN SAL
  Add the following signs
CUNEIFORM SIGN SAL KUR

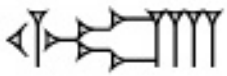
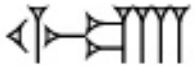
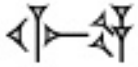
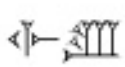
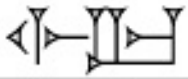
After 122AE CUNEIFORM SIGN SAL LAGAB TIMES ASH2
  Add the following signs
CUNEIFORM SIGN SAL LAGAR (see comment (b) below)
CUNEIFORM SIGN SAL SHE3
CUNEIFORM SIGN SAL TUG2
CUNEIFORM SIGN SAL UR (see comment (b) below)

			NIN	SAL.TUG2 or SAL.KU	B886, B887	Uoooooo
			GEME2	SAL x KUR	B890	Uoooooo
				SAL x LAGAR	B895	Uoooooo
			EGI2	SAL x SHE3 or SAL x NAM2	B897	Uoooooo
			NIG	SAL+UR	B898	Uoooooo

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 NIN is a single sign already in archaic Uruk (zero exceptions found in extensive searches). (d) Choosing the sign name NIN avoids having to name what the second part is after the SAL, which would be specifying at too detailed a level.
***

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)

			AGRIG; GISKIM	IGI x DUB or IGI x UM	B727 (v(2))	Uoooooo
			U6	IGI E2	B728	Uoooooo
			SIG5	IGI ERIM	B729	Uoooooo
				IGI + MIN	B724a	Uoooooo
			IGI.NIG2	IGI x NI	B735, B730	Uoooooo
			PAD3	IGI.RU	B725	Uoooooo
				IGI.SHE3 ?	B732	Uoooooo
			HUL	IGI.UR	B733	Uoooooo
			HUL4	IGI. (UR x KASKAL)	B734	Uoooooo

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, thus a separate sign needing its own encoding. 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

The following tables permit an overview of where each I have been able to identify each sign as included in one of the lists of numbered signs shown. 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 can also reflect a disuse of a sign 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.

Borger's MZL list is especially extensive, and 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.

Hittite Friedrich	Rüster & Neu	von Sod	en	Labat		Borger	Borger MZL	Unicode N2698	Labat readings		Sign Name	Ros.Rép. Lagash	Ur III KWU	Fara (LAK)	Ur Exc. Texts UETII Arch.	PI JN	ATU I	Z. freq.	ZATU
68	34	63	a	87	a		B145	Uooooo	tur3	TUR3	NUN x LAGAR = TUR3	R186	131	77, 78	60	30	239	46x	Z563
	278 in this			281			B439	Uooooo			UTUA2								

Hittite Friedrich	Rüster & Neu	von Sod	en	Labat		Borger	Borger MZL	Unicode N2698	Labat readings		Sign Name	Ros.Rép. Lagash	Ur III KWU	Fara (LAK)	Ur Exc. Texts UETII Arch.	PI JN	ATU I	Z. freq.	ZATU
Covers; The Vault of the Night Sky																			
232	295	273		469		469	B746	U1227E	pad {šug}	PAD {ŠUG}	SHU4 x GAR	R171		398	180		457	18x	Z426
217	275	258		441		441	B698	U1230F	ul	UL	SHU4 x GUD	R012	502	(? 299) 300	221				
210	271	249		429		429	B682	Uooooo	gul	GUL	SHU4 x URUDU or DUB ?	R279	546	389	172	137, 226	454 [455]	1x+atu	Z242
(cf.) 207	266			430		430	B683	Uooooo		GIR4	SHU4 x AD	R015	545	414	181				
				442		442	B699	Uooooo		ŠITA4	SHU4 x E2 early			397	(?230) 170	297		4x	Z535
216	270			428		428	B684	Uooooo		ŠAGAN	SHU4 x GAN	R016	547	407					
				447			B706	Uooooo		NIGIN4	SHU4 x UD		511	401			206	ATU	Z537
				419		419	B671	Uooooo		SAGŠU	SHU4 x SAG		513	316	179				
213	272	243		412		412	B663	Uooooo	muh	UGU	SHU4 x KA	R013	514	413					
206	263			418		418	B670	Uooooo		U.GUN	SHU4 x (SI gunu)		334	179					
				443		443	B700	Uooooo		UTU2	SHU4 x GA			412					
212	274			54		54	B165	Uooooo	—		BURU14			531					
				413	?	413	B666	Uooooo		ŠIBIR	SHU4 x BURU14								
				415	a		B667	Uooooo		GAKKUL3	SHU4 x DIM								
				416		416	B668	Uooooo		GAKKUL	(SHU4 x DIM) x KUR								
				448			B710	Uooooo		KUŠU	ŠU4 x GIR3		504 ??						
				414			B664	Uooooo			SHU4 x ITI								
sub. 205	264			415		415	B665	Uooooo		UDUN	SHU4 x MU								

Hittite Friedrich	Rüster & Neu	von Sod	en	Labat		Borger	Borger MZL	Unicode N2698	Labat readings		Sign Name	Ros.Rép. Lagash	Ur III KWU	Fara (LAK)	Ur Exc. Texts UETII Arch.	PI JN	ATU I	Z. freq.	ZATU
Covers; The Vault of the Night Sky																			
				546		546	B870	Uooooo		EN2	SHU2 x AN		510	356		180		4x+atu	Z138
				547		547	B872	Uooooo		ŠUHUB, KUNGA	SHU2 x AN THREE TIMES								
24				549		549	B876	Uooooo		ŠUDUN	SHU2 x DUN4		503	409					
	88			553	a		B880	Uooooo		ŠU2 x UR-šeššig	SHU2 x UR-sheshshig		570						
				548		548	B875	Uooooo		GIBIL2; KIBIR	SHU2 + ASH2								
				553			B879	Uooooo		LIL3	SHU2 x ESH								
				552			B874	Uooooo		LIL5	SHU2 x NE								

Hittite Friedrich	Rüster & Neu	von Sod	en	Labat			Borger	Borger MZL	Unicode N2698	Labat readings		Sign Name	Ros.Rép. Lagash	Ur III KWU	Fara (LAK)	Ur Exc. Texts UETII Arch.	PI JN	ATU I	Z. freq.	ZATU
				Female SAL = MUNUS																
240	299	300		556			556	B886, B887	Uoooo	nin	NIN	SAL.TUG2 or SAL.KU	R190	795a, 796, 787	522	401	305		11x	Z400
243	305	303		558			558	B890	Uoooo	amat	GEME2	SAL x KUR	R173	793	49	398	303		13x	Z201
239	298	301		557			557	B889	U12070	dam	DAM	DAM	R191	799	523, 46a	291 ?				
241	300	299		555	a		555	B884	U12371	ṣu(m)	ZUM & ZUM2	SAL x "tomb"	R175	788, 791, 790	519					
				555	b			B885	U122AF	zum	ZUM	SAL x (LAGAB x ASH2 (<"tomb"))		791	524					
				555	b			****	Uoooo	zum	ZUM	SAL x			520					
		305		563			563	B898	Uoooo	nig	NIG	SAL+UR			521					
				556	v			B897	Uoooo		EGI2	SAL x SHE3 or SAL x NAM2								
				554	v	1		****	Uoooo			SAL x KAB or SAL x HUB2								
				554	v	2		****	Uoooo			SAL x ME	R154	791a						
				554	v	3		****	Uoooo			SAL x TUK								

Hittite Friedrich	Rüster & Neu	von Sod	en	Labat			Borger	Borger MZL	Unicode N2698	Labat readings		Sign Name	Ros.Rép. Lagash	Ur III KWU	Fara (LAK)	Ur Exc. Texts UETII Arch.	PI JN	ATU I	Z. freq.	ZATU
				IGI																
		261	a	449	n			B724a	Uoooo			IGI + MIN								
208	265	264		455			455	B731	U12149	u3	U3	IGI x DIB	R070	568	432, 433					
226	289	263		451		(n)	451	B726 (v)	U1214A	ar	AR	IGI.RI	R073	557	422					
	292	262		450			450	B725	Uoooo	pa3	PAD3	IGI.RU	R074	558	423	243				
228	290	265		456			456	B733	Uoooo	hul	HUL	IGI.UR	R071	565	428					
227	291			452	(v1)	(n)	452	B727 (v(2))	Uoooo		AGRIG; GISKIM	IGI x DUB or IGI x UM	R072	563	427					
225	293			454			454	B729	Uoooo		SIG5	IGI.ERIM		559-60	424					
				449				B735, B730	Uoooo		IGI.NIG2	IGI x NI	R068	561	435					
				455				B732	Uoooo			IGI.SHE3 ?		? 562, 567	426	244				
												IGI.TUG2			434					
				449		2		B728	Uoooo		U6	IGI + E2			429					
				456	a			B734	Uoooo		HUL4	IGI. (UR x KASKAL)		566						